

Faserbasierte Simulation von Humanhaar



- Dissertation -

zur Erlangung des Doktorgrades (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen
Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Dipl.-Ing. Gerrit Sobottka
Bonn, Dezember 2008

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn.

1. Referent: Prof. Dr. Andreas Weber (Universität Bonn)
2. Referent: Prof. Dr. Gabriel Zachmann (TU Clausthal)

Tag der Promotion: 20.03.2009

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert.

Erscheinungsjahr: 2009

Zusammenfassung

Gegenstand dieser Arbeit ist die Entwicklung eines faserbasierten parametrischen Haarmodells zur numerischen Simulation komplexer Faseranordnungen. Parametrisch bedeutet, dass neben Geometrieinformationen grundlegende, die Faserinteraktion beeinflussende Faktoren, wie Reibung, Temperatur, Luftfeuchte, Steifigkeit, Ladung, etc. mit in den Modellierungsprozess einfließen. Das Modell setzt sich aus zwei Hauptkomponenten zusammen: a) einem mechanischen Modell zur Beschreibung des mechanischen Verhaltens bzw. der Deformation einer einzelnen Faser unter externen Belastungen und b) einem Interaktionsmechanismus, der Kontaktpunkte oder Durchdringungen zwischen den Fasern detektiert und durch Rückprojektion der Geschwindigkeiten in den Raum der zulässigen Systemzustände in ein lokal physikalisch korrektes Kontaktverhalten, unter Beachtung von Reibung, transformiert. Die korrekte Handhabung der Faser-Faser-Interaktionen ist der Schlüssel zur Volumenerhaltung in komplexen Faseranordnungen. Die Effizienz des Modells wird an einer Reihe komplexer numerischer Beispiele aufgezeigt. Darüber hinaus werden Algorithmen zur optischen Darstellung von komplexen Faseranordnungen, sowie zur Rekonstruktion von Frisuren aus groben Oberflächeninformationen, wie sie bspw. nach der Abtastung von Skulpturen vorliegen, entwickelt.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Warum faserbasiert ?	2
1.2	Beiträge	4
2	Eigenschaften von Humanhaar	7
2.1	Einleitung	7
2.2	Aufbau der Keratinfaser	7
2.3	Geometrische Eigenschaften	9
2.3.1	Haarlänge	9
2.3.2	Querschnitt	10
2.3.3	Welligkeit	11
2.3.4	Haarkörper	12
2.3.5	Haardichte	13
2.4	Mechanische Eigenschaften	14
2.4.1	Elastisches Verhalten	14
2.4.1.1	Linearelastischer Bereich	15
2.4.1.2	Fließbereich	16
2.4.2	Rheologische Eigenschaften	17
2.4.3	Hysterese mechanischer Parameter	26
2.4.4	Einfluss von Feuchtigkeit	27
2.4.5	Einfluss von Fluiden	30
2.4.6	Einfluss der Temperatur	31
2.4.7	Biegung	33
2.4.8	Torsion	35
2.4.9	Reibung	38
2.4.10	Elektrostatische Effekte	42
2.4.11	Kosmetische Behandlung	42
2.4.11.1	Kämmen und Bürsten	42
2.4.11.2	Glätten der Haare	44
2.4.11.3	Haarsprays	44
3	Mechanische Modelle für Fasern	47
3.1	Einleitung	47
3.1.1	Diskrete und kontinuierliche Modelle	47
3.1.2	Herleitung von Bewegungsgleichungen	48
3.2	Spezielle Theorie der COSSERAT-Stäbe	49

3.2.1	Gleichgewichtsbedingungen	51
3.2.2	Konstituierende Gleichungen	53
3.2.3	Darstellung der lokalen Basis	56
3.2.4	Zwangsbedingungen	58
3.2.5	Theorie von KIRCHHOFF	59
3.2.6	Bewegungsgleichungen der Faser	59
3.2.7	Allgemeine Theorien	61
3.2.8	Verwandte Ansätze	62
3.3	Lösung der statischen COSSERAT-Gleichungen	65
3.3.1	Symbolische Lösung	66
3.3.1.1	Symbolische Integration	67
3.3.1.2	Randwertproblemlöser	73
3.3.1.3	Numerische Beispiele	75
3.3.2	Numerischer Lösungsansatz	79
3.3.2.1	Energie der Faser	82
3.3.2.2	Diskretisierung	84
3.3.2.3	Numerische Beispiele	85
3.4	Lösung der dynamischen COSSERAT-Gleichungen	86
3.4.1	Generalisiertes Alpha-Verfahren	88
3.4.2	Stabilitätsanalyse	88
3.4.3	Diskretisierung der Bewegungsgleichungen	90
3.4.4	Randbedingungen	94
3.4.5	Integration der Kinematikgleichungen	95
3.4.6	Rekonstruktion der Mittellinie	95
3.4.7	Relaxationsverfahren	96
3.4.8	Alternative Verfahren	99
3.4.9	Numerische Beispiele	100
3.4.9.1	Stab unter Gravitation	101
3.4.9.2	Helix mit Terminallast	101
3.4.9.3	Gerader Stab mit Drehmoment	101
3.4.9.4	Vorgespannte Spiralfeder	101
3.4.9.5	Buckling	102
3.4.9.6	Faseranordnung unter Zugbelastung	103
3.4.9.7	Faseranordnung unter Gravitation	104
3.4.9.8	Faseranordnung unter Bewegung des Lagers	104
3.4.10	Diskussion	105
4	Elektrostatik auf Haaren	111
4.1	Einleitung	111
4.2	Berechnung elektrostatischer Einflüsse	112
4.3	Fernfeldapproximation durch Multipolentwicklung	113
4.3.1	Abtastung des elektrostatischen Feldes	115
4.4	Schnelles Multipolverfahren (FMM)	116
4.4.1	Interaktionslisten	117

4.4.2	Interaktion zwischen Partikeln und Oberfläche	119
4.4.3	Randbedingungen	119
4.5	Numerische Beispiele	120
5	Rekonstruktion von Frisuren aus groben Oberflächeninformationen	127
5.1	Einleitung	127
5.2	Verwandte Ansätze	127
5.3	Der Algorithmus	130
5.3.1	Repräsentation von Randbedingungen	131
5.3.2	Parametrisierung der Wachstumsregion	131
5.3.3	Leithaargenerierung	133
5.3.4	Einsatz lokaler Dichteinformationen	135
5.3.5	Finaler Schnitt und Strähneninterpolation	136
5.4	Numerische Beispiele	137
5.5	Diskussion	138
6	Photorealistische Darstellung von Faseransammlungen	141
6.1	Einleitung	141
6.2	Streuphänomene bei Haarfasern	142
6.3	Fernfeldmodell für Haare	144
6.3.1	Die Streufunktion	144
6.3.2	Normalstreuung - Streuung am dielektrischen Kreis	146
6.3.3	Einfluss der Cuticula - Longitudinale Streuung	148
6.4	Entwurf eines Nahfeldmodells	149
6.4.1	Einfluss indirekter Beleuchtung	151
6.5	Diskussion	152
7	Kollisionserkennung für 1D-Strukturen	155
7.1	Einleitung	155
7.2	Detektion von Faser-Umgebungs-Kollisionen	155
7.2.1	Distanzfelder	156
7.2.2	Charakteristikenmethode	157
7.3	Detektion von Faser-Faser-Kollisionen	159
7.3.1	Räumlich und topologisch angepasste Hierarchien	160
7.3.2	Subquadratische Kollisionserkennung	162
7.3.3	Hierarchieaufbau	162
7.3.4	Kombinatorische Beschreibung von Hüllkörpern	165
7.3.5	Hierarchiebasierte Kollisionserkennung	165
7.3.6	Aufrechterhaltung einer Hierarchie	167
7.3.7	Hüllkörper	168
7.3.7.1	Achsenparallele Quader (AABBs)	168
7.3.7.2	Diskret orientierte Polyeder (k-DOPs)	169
7.3.7.3	Orientierte Quader (OBBs)	172
7.3.7.4	Kugeln	174

7.4	Performancevergleich	175
7.4.1	Statischer Faserbündeltest	177
7.4.2	Dynamischer Faseranordnungstest	177
7.4.3	Statischer Faseranordnungstest	178
7.4.4	Parametrischer DOP-Test	182
7.5	Diskussion	182
7.6	Kontinuierliche Kollisionserkennung	183
8	Kollisionsantwort	187
8.1	Verfahren zur Kollisionsbehandlung	187
8.1.1	Strafkräfte und Anordnungsoptimierung	187
8.1.2	Analytische Methoden	189
8.1.3	Impulsbasierte Techniken	190
8.1.4	Projektive Verfahren für komplexe Kontaktsituationen	190
8.1.4.1	Zeitschrittverfahren	191
8.1.4.2	Phase maximaler Kompression	193
8.2	Systeme mit unilateralen Zwangsbedingungen	194
8.2.1	Stöße und Reaktionen in LAGRANGE-Systemen	197
8.2.2	Stoßgesetze	200
8.2.3	Differentialinklusionen	201
8.2.4	Diskretisierung	203
8.3	Ein anisotropes Reibungsmodell für Keratinfasern	204
8.3.1	Der Reibungskegel	205
8.3.2	Maximalleistungsungleichung	206
8.3.3	Tangentialreaktion	209
8.4	Das Zeitschrittverfahren	210
8.4.1	Generalisierung auf $SE(3)$	211
8.5	Kollisionsantwort bei kinematischen Ketten	214
8.5.1	Verfahren I: Impulskorrektur	215
8.5.1.1	Gelenkbedingung	216
8.5.1.2	Integration	217
8.5.1.3	Gleichungskopplung	218
8.5.2	Verfahren II: Linearisierter Ansatz	220
8.5.3	Integration elastischen Verhaltens	222
8.5.4	Algorithmus	223
8.5.5	Integration mit COSSERAT-Modellen	224
8.6	Numerische Beispiele	224
8.6.1	Horizontale Anordnung auf Kugel	225
8.6.2	Vertikale Anordnung auf Kugel	226
8.6.3	Vertikale Anordnung auf Terrain	226
8.6.4	Horizontale Anordnung auf schiefer Ebene	229
8.6.5	Diskussion	229
8.6.5.1	Inelastische Stöße	233
8.6.5.2	Reibungsanisotropie	239

8.6.5.3	Harmonisierung der Zwangsbedingungen	240
8.7	Ausblicke	240
A	Anhang	243
A.1	Notation	243
A.2	Der LAGRANGE-Formalismus	243
A.3	Berechnung der Wurzeln	247
A.4	Analytische JACOBI-Matrizen für die α -Methode	248
A.5	Elliptische Integrale und JACOBIsche elliptische Funktionen	251
A.6	Definition der Räume	252
A.6.1	Spezielle Orthogonale Gruppe, $\mathcal{SO}(3)$	252
A.6.2	Spezielle Euklidische Gruppe, $\mathcal{SE}(3)$	252

Inhaltsverzeichnis

1 Einleitung

Die physikalisch basierte Simulation menschlicher Haare ist seit Beginn der 90er Jahre fester Gegenstand der Forschungsbemühungen im Bereich der Computergrafik. So wurden in dieser Zeit in über 90 Publikationen weltweit, überwiegend aus dem universitären Bereich, eine Vielzahl von Ansätzen zur physikalisch basierten Generierung, Simulation und photorealistischen Darstellung von Haaren oder Faseransammlungen¹ im Allgemeinen vorgestellt. Der Bedarf an qualitativ hochwertigen Modellen ist dabei enorm. Mit der Intensivierung der Produktion von computergenerierten Animationsfilmen werden die virtuellen Charaktere zunehmend mit komplexen, beweglichen Frisuren ausgestattet (Shrek, DreamWorks 2001; Final Fantasy, Hironobu Sakaguchi 2001; Madagascar, DreamWorks Animation 2005). Die Komplexität solcher Faseransammlungen stellt aber nach wie vor eine große Herausforderung dar, die unter teilweise erheblichem manuellem Aufwand, bspw. der Verwendung von Keyframetechniken, bewältigt werden muss. Physikalisch basierte Ansätze können hier helfen, die Kosten des Produktionsprozesses zu senken. Die Beratung erfolgt dabei oftmals durch aus dem universitären Bereich eingekaufte Spezialisten. Auch die kosmetische Industrie hat mittlerweile den Zeitgeist erkannt und durch die Initiation von kooperativen Forschungsprojekten mit Universitäten Interesse signalisiert. Die Unterschiede in den Ansprüchen von Film- oder Computerspiele-Industrie einerseits und der kosmetischen Industrie andererseits liegt dabei primär in der Genauigkeit bzw. der Prognosefähigkeit solcher Modelle. Akkurate Vorhersagen sind in der Regel nur dann möglich, wenn das Modell auf einer soliden mathematisch-physikalischen Theorie basiert. Diese Forderung disqualifiziert viele der bis dato vorgestellten ad-hoc Ansätze, die den perzeptorischen Erfahrungen nicht widersprechen mögen und sich damit für den Einsatz bei der Film- oder Spieleproduktion durchaus eignen, aber andererseits keine Vorhersagen erlauben. Hochwertige Modelle können darüber hinaus wesentlich zur Überwindung des wahrnehmungspsychologischen „Uncanny Valleys“ virtueller Charaktere beitragen.

Bei der Simulation von Faseranordnungen kann man drei Hauptaspekte ausmachen:

1. Die Modellierung komplexer Faseranordnungen oder Frisuren: Die nach wie vor weit verbreitete manuelle Modellierung komplexer Frisuren mit entsprechenden Modellierungswerkzeugen (CINEMA 4D, LIGHTWAVE, MAYA, etc.) bedarf, neben einiger künstlerischer Fingerfertigkeit, auch einer Menge an Kopier- und Einfügeoperationen. Alternative Verfahren unterstützen den Designprozess durch Verwendung von 3D-Eingabe-Peripherie („virtueller Friseur“). Daneben existiert mittlerweile auch eine Reihe von Ansätzen, die komplexe Frisuren aus einer Serie von Photos oder Oberflächenscans von Skulpturen zu rekonstruieren vermögen. Die Ergebnisse eines solchen Prozesses dienen u.a. als Anfangswerte für eine Dynamiksimulation.

¹Wir ziehen den Begriff der Faseransammlung im Folgenden dem der Haarsträhne oder Frisur vor.

2. Das Gebiet der Dynamiksimulation befasst sich mit der Frage, wie sich die Fasern einer solchen Anordnung bzw. die Anordnung als Ganzes unter Einwirkung physikalischer Einflüsse (Gravitation, Ladung, Material, Temperatur, Luftfeuchte, etc.) mechanisch verhalten. Mit einer der ersten Publikation auf diesem Gebiet durch Rosenblum et al. [Rosenblum et al., 1991], in dem die Autoren ein primitives Masse-Feder-Modell zur Simulation einzelner Haarfasern bzw. Haarsträhnen vorschlagen, werden Partikelsysteme als erstes Simulationsparadigma für deformierbare Fasern etabliert. Neuere Ansätze, wie der hier vorgestellte, machen Gebrauch von quasi-kontinuumsmechanischen Modellen, wie bspw. der speziellen Theorie der COSSERAT-Stäbe. Solche Ansätze stellen recht hohe Anforderungen an potentielle numerische Lösungsverfahren, bieten dafür aber universellere Beschreibungsmöglichkeiten des Materialverhaltens.
3. Will man mit den ersten beiden Aspekten über den Status eines abstrakten Zahlenfriedhofs hinaus gelangen, bleibt als dritter Aspekt die visuelle Darstellung einer Faseranordnung auf einem Bildschirm oder auf dem Drucker. Das geschieht primär unter dem Aspekt des Photorealismus. Geeignete Modelle beschreiben, wie das Licht, das von definierten Lichtquellen ausgeht, an den Fasern gestreut wird. Ein Meilenstein auf diesem Gebiet ist der Beitrag von Marschner et al. [Marschner et al., 2003], der die Theorie der Lichtstreuung an dielektrischen Zylindern auf Haare überträgt. Wir werden diesen Ansatz auf die Darstellung blonder Haare erweitern.

1.1 Warum faserbasiert ?

Warum eignen sich bestehende Modelle nicht zur Vorhersage des Verhaltens von Haaren unter definierten Bedingungen? Der überwiegende Teil der bis dato vorgestellten Modelle arbeitet auf der Basis von Haarsträhnen. Im Prinzip wird dabei nur die Mittelliniengeometrie in Form von sparsam über die Kopfoberfläche verteilten Leithaaren berechnet und fehlendes Material in einem anschließenden Interpolationsprozess generiert. Solange man den faserbasierten Level einfach überspringt und den Vorgängen, die zwischen den einzelnen Fasern einer solchen Anordnung ablaufen, keine Beachtung schenkt, müssen die dort liegenden Ursachen für das globale Haarverhalten zwangsläufig unverstanden bleiben.

Im Gegensatz dazu verfolgt die vorliegende Arbeit erstmalig das Paradigma einer faserbasierten Haarsimulation. D.h., dass wir Modelle entwickeln, die das global mechanische Verhalten von Ansammlungen auf der Basis einzelner Fasern beschreiben. Die Haarsträhne ist dann eben nicht die Konsequenz des Interpolationsprozesses über einer deformierten Leitfaser, sondern das Ergebnis einer physikalisch korrekten Bearbeitung der Faser-Faser-Interaktionen innerhalb dieser Haarsträhne. Für die Demonstration der Vorzüge eines faserbasierten Simulationsparadigmas ist es nicht notwendig, einen ganzen Kopf mit 130.000 oder mehr einzelnen Fasern zu berechnen. Exemplarische Strähnen mit 200 bis 300 Fasern sind völlig ausreichend und sollen deshalb auch im Mittelpunkt unseres Interesses stehen.

Ein Modell, das solches leistet, setzt sich aus zwei Hauptkomponenten zusammen.

1. Dem mechanischen Modell für die Beschreibung des mechanischen Verhaltens bzw. der Deformation einer einzelnen Faser unter externen Belastungen. Die Faser darf dabei belie-

bige Materialeigenschaften haben. Eine Beschränkung auf Haare oder allgemeiner, Keratin als Werkstoff, ist dabei nicht notwendig.

2. Der Weg von der Einzelfaser zur Ansammlung führt schließlich durch das unwegsame Tal der Kollisionsbehandlung. Der Grund dafür liegt in dem komplexen Wechselspiel zwischen den Fasern solcher Verbände. Kontaktpunkte oder Durchdringungen zwischen den Fasern sind zu detektieren und durch Applikation geeigneter Maßnahmen in ein lokal-physikalisch korrektes Kontaktverhalten, bspw. unter Beachtung von Reibung, zu transformieren. Die korrekte Handhabung der Faser-Faser-Interaktionen ist der Schlüssel zur Volumenerhaltung bei komplexen Faseranordnungen.

Das Modell für die Einzelfaser in Kombination mit den Interaktionsmechanismen zwischen den Fasern liefert ein parametrisches Haarmodell. Parametrisch bedeutet, dass grundlegende, die Faserinteraktion beeinflussende Faktoren, wie Reibung, Temperatur, Luftfeuchte, Steifigkeit, etc. neben den Geometrieinformationen mit in den Modellierungsprozess einfließen.

Während die erste Komponente von theoretischer Seite durch die weiter oben schon erwähnte spezielle Theorie der COSSERAT-Stäbe abgedeckt ist, umfasst der Begriff des Interaktionsmechanismus neben theoretischen Aspekten der Stoßmodellierung auf der Basis von Maßdifferentialinklusionen auch das Problem der Bestimmung nächster Nachbarn als potentielle Kollisionskandidaten, sowie die paarweise Abstandsbestimmung der Fasern, bspw. für die Berechnung der COULOMB-Interaktionen.

Die Vorwärtssimulation einer Faseranordnung auf der Basis des parametrischen Haarmodells läuft dann wie folgt ab. Gegeben seien Anfangswerte in Form einer beliebigen Ausgangsgeometrie und Metainformationen in Form von Materialeigenschaften, Umweltbedingungen, Schrittweiten, externen Belastungen, etc. Diese Werte entsprechen den Eingabespezifikationen des parametrischen Haarmodells. Fassen wir dieses zunächst als Blackboxprozess auf, dann erhalten wir als Ausgabe eine neue Geometrie, die den zeitlich fortgeschrittenen Zustand der Ansammlung repräsentiert. Diese Rohdaten dienen zur erneuten Eingabe in das parametrische Haarmodell oder als Ausgangsdaten für einen „Lichtsimulator“, der daraus eine photorealistische Darstellung erzeugt, vgl. Abb. (1.1).

Worin unterscheiden sich Anfangswerte von den Parametern? Anfangswerte umfassen die Startwerte der unbekanntenen Größen, deren zeitliche Entwicklung beschrieben werden soll (i.d.R. sind das Positionen und Geschwindigkeiten). Parameter sind dagegen bekannte Größen, die den Systemzustand genauer spezifizieren. Sie können durchaus von der Zeit und vom Ort abhängen. Anfangswerte sollten nicht mit den Parametern verwechselt werden, auch wenn sie in jedem Zeitschritt neu an das Modell übergeben werden. Sie folgen immer aus dem zeitlich vorhergehenden Zustand.

Die Vorgänge innerhalb der Blackbox lassen sich relativ grob wie folgt skizzieren. Im ersten Schritt müssen Durchdringungen der Fasern untereinander und mit der Peripherie detektiert werden. In einem weiteren Schritt sind Maßnahmen zur Behebung etwaiger Fehlzustände zu ergreifen. Das kann durch die Korrektur von Positionen, Geschwindigkeiten oder Kräften geschehen. Im dritten Schritt erfolgt die eigentliche numerische Integration der Differentialgleichungen, welche die Deformation der Fasern beschreiben, auf der Basis der korrigierten Systemgrößen und unter Auswertung aller expliziten Kraftfelder, wie Gravitation oder Elektrostatik. Die neuen Konfigurationen der einzelnen Fasern der Anordnung sind das Resultat dieses Prozesses. Hier gilt es

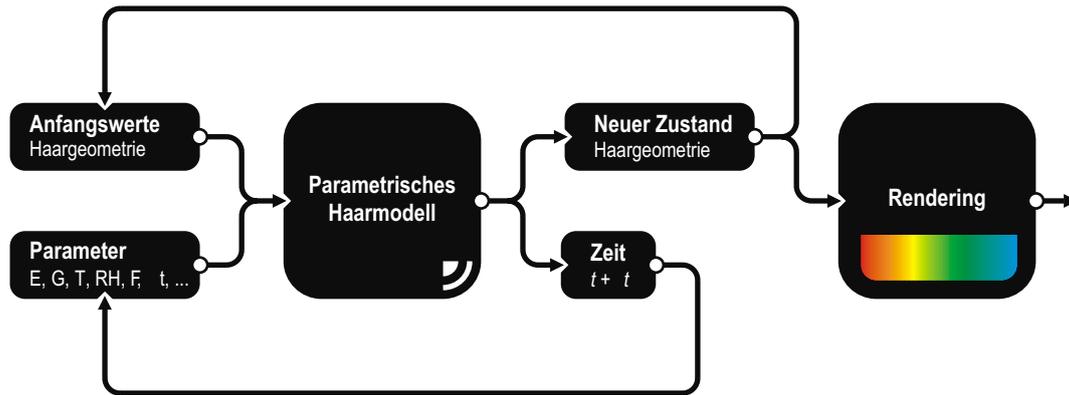


Abbildung 1.1: Einsatz eines parametrischen Haarmodells zur physikalisch basierten Simulation von Faseranordnungen.

zu beachten, dass sich die neuen Systemgrößen u.U. schon direkt aus der Korrektur der Fehlzustände ergeben können. Man bezeichnet das Korrekturprinzip dann als Zeitschrittverfahren. Eine anschließende Integration zeitlich bereits fortgeschrittener Systemgrößen ist hier nicht mehr notwendig.

Die einzelnen Systemkomponenten werden in den folgenden Kapiteln genau beschrieben.

1.2 Beiträge

Gegenstand dieser Arbeit ist die Entwicklung eines faserbasierten parametrischen Haarmodells zur numerischen Simulation komplexer Faseranordnungen. Die computergestützte Darstellung komplexer Faseranordnungen muss unter dem Gesichtspunkt der numerischen Stabilität und unter dem Aspekt des Berechnungsaufwandes als äußerst anspruchsvoll gelten. Da sich ein solches Modell notwendigerweise aus einer Reihe von Teilaspekten zusammensetzt, ist die Arbeit modular aufgebaut. Die folgende Liste gibt eine Übersicht über die Art und den Umfang der einzelnen wissenschaftlichen Beiträge.

1. Kap. (2) gibt eine detaillierte Übersicht über die geometrischen und physikalischen Eigenschaften von Humanhaar. Als Quelle dienten dabei über 70 Fachbeiträge und Bücher aus dem Bereich der kosmetischen und Textilindustrie.
2. Kap. (3) gibt eine Einführung in die spezielle Theorie der COSSERAT-Stäbe, wobei das System partieller Differentialgleichungen, welches das dynamische Gleichgewicht einer Faser unter externen Belastungen beschreibt, Schritt für Schritt hergeleitet wird. Der Schwerpunkt liegt dann auf der Entwicklung numerischer Lösungsverfahren für die im Faserkontext auftauchenden Randwertprobleme auf der Basis der statischen und dynamischen COSSERAT-Gleichungen. Wir zeigen, dass der statische Spezialfall einer nicht scherbaren und nicht dehnbaren Faser auf analytische Lösungsausdrücke mit JACOBISCHEN elliptischen Funktionen führt. Daraus entwickeln wir einen Randwertproblemlöser und zeigen die Grenzen der Verwertbarkeit der analytischen Lösung bei der Fasersimulation auf.

Im Gegensatz dazu führt der dynamische Sonderfall einer nicht scherbaren und nicht dehnbaren Faser auf ein steifes System nicht-linearer gekoppelter partieller Differentialgleichungen. Auf der Basis der generalisierten Alpha-Methode, einem speziellen numerischen Integrationsverfahren, das dem Bereich der Strukturmechanik entstammt und darüber hinaus über „schöne“ Eigenschaften verfügt, gelingt die Entwicklung eines Relaxationsverfahrens für die dynamischen COSSERAT-Gleichungen. Im Gegensatz zu existierenden Ansätzen zur Simulation von Fasern aus dem Bereich der Computergrafik, die primär auf der LAGRANGESchen Mechanik beruhen, zeigen wir damit, dass die unter numerischen Gesichtspunkten als äußerst bedenklich einzustufende Nichtdehnbarkeitsbedingung weder eliminiert noch aufgeweicht werden muss, sondern die durch sie induzierten Probleme mit geeigneten Verfahren direkt angegangen werden können. Unser Ansatz ist damit auch der erste, der das zugrunde liegende Randwertproblem direkt löst.

3. In Kap. (4) diskutieren wir das Problem der Berechnung des elektrostatischen Feldes auf menschlichen Haaren und entwickeln ein effizientes Verfahren zur Simulation elektrostatischer Effekte. Der Ansatz stellt im Wesentlichen eine Kombination aus einem statischen COSSERAT-Modell und der bekannten schnellen Multipol Methode (FFM) zur schnellen Auswertung der paarweisen COULOMB-Interaktionen dar. Wir zeigen zudem, wie Umweltbedingungen (Temperatur, Luftfeuchte) die Auswirkungen elektrischer Ladungen auf der Haaroberfläche beeinflussen.
4. Kap. (5) befasst sich mit der (Re)-Konstruktion von realistischen Frisuren auf der Basis von Rand- oder Oberflächeninformationen, die man bspw. bei der Oberflächenabtastung von Skulpturen erhält. Der Ansatz basiert auf dem statischen COSSERAT-Modell. Der Einpassungsvorgang einzelner Fasern in das Haarkörpervolumen unter gleichzeitiger Beachtung vorgegebener Randbedingungen erfolgt durch Lösen eines Optimierungsproblems.
5. Kap. (6) befasst sich mit dem wichtigen Thema der photorealistischen Darstellung von Faseransammlungen. Wir diskutieren dabei den Ansatz von Marschner et al. [Marschner et al., 2003], der eine Analogie zwischen der Lichtstreuung an dielektrischen Zylindern und Haarfasern herstellt, und erweitern diesen im Hinblick auf die Darstellbarkeit blonder Haare. Insbesondere zeigen wir zum ersten mal den Einfluss indirekter Beleuchtung auf die realistische Darstellung heller Haartypen auf.
6. In Kap. (7) diskutieren wir das Problem der Kollisions- bzw. Selbstkollisionserkennung in eindimensionalen Strukturen. Wir vergleichen dabei räumlich adaptive mit topologiebasierten Verfahren und demonstrieren, dass topologiebasierte Verfahren unter bestimmten Bedingungen ein besseres Laufzeitverhalten zeigen. Ein Schwerpunkt liegt dabei auf der Untersuchung der Effizienz parametrischer DOP-Bäume mit variabler Anzahl an Richtungen. Der Richtungsgenerator selbst basiert auf der Theorie der GRASSMANN-Räume.
7. Kap. (8) gibt zunächst einen Überblick über die theoretischen Grundlagen der Behandlung von Stößen in Systemen mit Trockenreibung auf der Basis der Theorie der Differentialinklusionen. Dieses Konzept wurde ursprünglich von dem französischen Mathematiker Jean-Jaques Moreau entwickelt und fand Niederschlag in einem so genannten Zeitschrittverfahren für die Granulat- oder Geschiebesimulation, dass sich bei der Kollisionsbehand-

lung in komplexen Faseranordnungen als äußerst robust erweist. Kollisionen werden auf der Basis der Geschwindigkeiten behandelt. Wir bauen hier auf einem Beitrag von Kaufman et al. [Kaufman et al., 2005] auf, der Moreaus Zeitschrittverfahren (marginal) erweitert. Wir entwickeln ausgehend von der Theorie der Limitflächen (vgl. [Goyal et al., 1991]) ein anisotropes Reibungsmodell, das in der Lage ist, den durch die Morphologie der Keratinstruktur begründeten und weithin bekannten richtungsabhängigen Reibungseffekt von Keratinfasern, darzustellen. Wir demonstrieren die Effizienz des Ansatzes an exemplarischen Faseranordnungen mit mehreren hundert einzelnen Fasern und zeigen dabei, dass der akkuraten Kollisionbehandlung bei der faserbasierten Simulation komplexer Faseranordnungen eine existentielle Bedeutung hinsichtlich der Volumenerhaltung zukommt.

Im Rahmen dieser Arbeit ist die Fasersimulationssoftware Rapunzel entstanden, die alle vorgestellten Algorithmen integriert. Rapunzel ist im weitesten Sinne ein Starrköpersimulator, bei dem zwischen Objekten und Faseranordnungen Assoziationen errichtet werden können. Die Software kann über eine XML-basierte Konfigurationsdatei angesteuert werden und verfügt über einen interaktiven OpenGL-basierten Szenenvisualisierer, vgl. Abb. (1.2). Die Konfigurationsdatei er-



Abbildung 1.2: Faserbasierte Dynamiksimulation mit der Fasersimulationssoftware Rapunzel. Die Eingabespezifikationen umfassen neben den geometrischen und physikalischen Informationen über die Ausgangsgeometrie auch Angaben über Simulationszeitdauern, Schrittweiten, Ausgabeformate etc. Die Simulationsergebnisse werden nach jedem Schritt im spezifizierten Format exportiert und stehen dann als Rohdaten für den Hochleistungsfaserrenderer LightSimulator zur Verfügung.

laubt die Spezifikation von Faser- und Objektgeometrien, Materialeigenschaften, Ausgabeformaten, Integrationsschrittweiten, Simulationsdauer, etc. Es besteht zudem eine Dateischnittstelle zu dem Hochleistungsfaserrenderer LightSimulator (Universität Bonn).

2 Eigenschaften von Humanhaar

2.1 Einleitung

Im Folgenden soll ein kurzer Überblick über die physikalischen Eigenschaften von menschlichen Haaren gegeben werden. Als Informationsquellen dienen die Literatur aus dem Bereich der Kosmetikbranche bzw. der Medizin, sowie diverse Zeitschriftenbeiträge unterschiedlicher Fachrichtungen, die in den letzten Jahrzehnten zu diesem Thema veröffentlicht worden sind.

Die Aufklärung der Haarstruktur bzw. die Verifikation bestehender Modelle sowie die Aufdeckung von Zusammenhängen zwischen Krankheitsbildern und pathologischen Veränderungen der Haareigenschaften stehen im Mittelpunkt des Interesses der nicht kosmetisch motivierten Haarforschung. Infolgedessen werden Untersuchungen oftmals unter Bedingungen durchgeführt, die in dieser Form im „Haar-Alltag“ nicht auftreten. Dennoch scheint die Angabe qualitativer Zusammenhänge unter unnatürlichen Bedingungen sinnvoller, als bestimmte Themenbereiche gänzlich unbeleuchtet zu lassen.

Haare gehören zur Gruppe der *Keratine* und sind morphologisch und von ihren physikalischen Eigenschaften her eng verwandt mit der Schafswolle, die sehr viel eingehender untersucht worden ist. So sind zum Beispiel die Isothermen der Feuchtigkeitsaufnahme bis zum 95 %-Level praktisch identisch [Menkart et al., 1966]. In der Literatur wird der Bezug zur Wolle oft hergestellt, wenn empirische Befunde für Haare nicht vorliegen. Dieser Vorgehensweise wollen wir uns im Folgenden anschließen.

2.2 Aufbau der Keratinfaser

In der Literatur findet man häufig den Hinweis, dass die mechanischen Eigenschaften von Haaren primär über die Struktur verstanden werden können. Deshalb wird im Folgenden kurz auf den mikro-strukturellen bzw. molekularen Aufbau der Haarfaser eingegangen.

Haare gehören zur Gruppe der Keratinfasern und bestehen im wesentlichen aus einer Ansammlung dicht gepackter, keratinhaltiger Filamente. *Keratine* selbst sind Faserproteine mit α -helikaler Struktur, die durch Disulfidbindungen stabilisiert werden [Nelson & Cox, 2000]. Der Begriff des Keratins wird in der Literatur oftmals synonym für biologische Materialien verwendet, die solche Proteine enthalten. Als typische Beispiele können hier neben Haaren auch Wolle, Horn, Nägel, Hufe, Stacheln etc. angeführt werden. Haare sind morphologisch und von den physikalischen Eigenschaften her eng verwandt mit der Schafswolle.

Die so genannten α -Keratine¹ bestehen aus Polypeptidketten, die durch Kondensation von Aminosäuren entstehen [Orfanos & Happle, 1990]. Keratine sind reich an Cystin, einer Amino-

¹Das α rührt vom typischen Röntgen-Beugungsmuster her, dem α -Pattern, das für diese Art von Proteinen charakteristisch ist.

2 Eigenschaften von Humanhaar

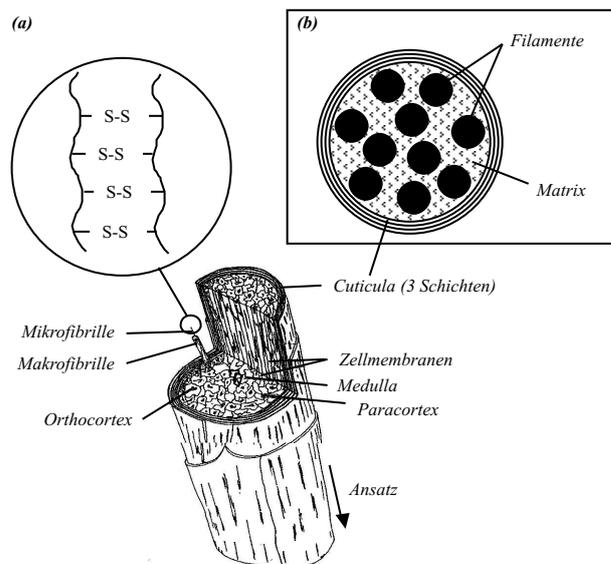


Abbildung 2.1: Struktur des menschlichen Haars. (a) Räumliche Ansicht, [Jollès et al., 1997]; (b) Schematischer Querschnitt mit mechanisch relevanten Komponenten, [Robbins, 2002].

säure, die sich durch ihren Schwefelgehalt auszeichnet. Die Thiolgruppen zweier Cystinreste können durch Ausbildung einer kovalenten Bindung mehrere Polypeptidketten miteinander verbinden.

Die physikalischen Eigenschaften von Humanhaar werden maßgeblich durch die Wasseraufnahmefähigkeit seiner Zellkomponenten bestimmt [Zahn, 1998]. Diese hängt von der chemischen Zusammensetzung ab, insbesondere dem Verhältnis von hydrophilen und hydrophoben Aminosäureresten und dem Vernetzungsgrad der Proteine. Die charakteristische Unlöslichkeit und Resistenz gegen enzymatischen Abbau begründen sich durch den hohen Gehalt an der Aminosäure Cystin bzw. den damit verbundenen hohen Schwefelgehalt.

Morphologisch können beim Haar drei Hauptkomponenten unterschieden werden: (i) die Cuticula oder äußere Hülle des Haars, (ii) der Cortex, der von der Cuticula umschlossen wird (iii) die Medulla, eine poröse Röhre im Zentrum des Cortex, vgl. Abb. (2.1).

Die Aufgabe der Cuticula besteht im Schutz der Haare vor chemischen und physikalischen Belastungen durch die Außenwelt [Orfanos & Ruska, 1968a]. Die Cuticula ist von röhrenförmiger Struktur und bildet den Mantel des Haars, bestehend aus verbundenen dachziegelförmig übereinander angeordneten Cuticulazellen von 35-40 μm Dicke, den *Schuppen*. Die Cuticulazellen setzen sich ihrerseits aus drei Schichten, der *Epi-*, *Endo-* und *Exocuticula* zusammen.

O'Connor et al. [O'Connor et al., 1995] haben die Kinetik des Hydratationsvorganges und die damit einhergehenden morphologischen Veränderungen der Haaroberfläche durch *Atomic Force Microscopy* erstmals direkt beobachten können. Die *Epicuticula* ist bedingt durch den hohen Grad der Vernetzung sehr hart und neigt weniger zum Quellen. Sie fungiert damit als Diffusionsbarriere. Die hydrophile *Endocuticula* bildet bedingt durch ihre Wasseraufnahmefähigkeit einen wichtigen Diffusionsweg. Ihre Quellfähigkeit und leichte enzymatische Abbaubarkeit sind die Folge hoher Konzentrationen an hydrophilen Aminosäureresten.

Der Cortex setzt sich aus in axialer Richtung verlaufenden spindelförmigen Zellen in zylindrischer Anordnung zusammen. Die Rindenzellen sind durch eine interzelluläre Kittsubstanz mit-

Tabelle 2.1: Unterschiedliche Typen der Kopfbehaarung, durchschnittliche maximale Länge und Radius in Abhängigkeit vom Alter, vgl. [Robbins, 2002].

Haartyp	Alter [a]	l_{\max} [cm]	r_{\max} [μm]
Lanugo	<1	15,0	20
Kind	1 - 12	60,0	60
Erwachsene	>13	100,0	100
Vellus	>30	0,1	4

einander verbunden [Orfanos & Ruska, 1968b]. Das Cortexkeratin besteht aus feinsten Filamenten von 6-8 nm Länge, den *Mikrofibrillen*. Mehrere solcher Mikrofibrillen sind zu *Makrofibrillen* zusammengelagert, die in den Cortexzellen liegen. Sie enthalten das α -Keratin. Die Cortexzellen sind von unterschiedlicher chemischer Zusammensetzung. Das Haar des Afrikaners besteht im Bogeninneren der Kräuselung aus cystinreichen Cortexzellen (Paracortex), auf der Außenseite aus cystinärmeren Zellen (Orthocortex). Bei glatten Haaren sind beide Zelltypen konzentrisch verteilt.

Im Zentrum des Cortex kann als weitere Zellkomponente gelegentlich die Medulla auftreten, die von porösem Charakter ist. Ihre Anwesenheit hat destabilisierenden Einfluss auf die Zugfestigkeit.

Generell sind die Zellen der Cuticula, des Cortex und der Medulla fest miteinander verkittet. Das kann auf dem Wege der chemischen Bindung über Kittsubstanzen geschehen, andererseits treten auch rein mechanische, druckknopfartige Verbindungen auf.

Die Haarfarbe wird durch den Pigmentfarbstoff *Melanin* verursacht, der in den Cortexzellen eingelagert ist. Er kommt in den beiden Modifikationen *Phäomelanin* und *Eumelanin* vor [Swift, 1997], deren Verhältnis die natürliche Haarfarbe bestimmt. Das Eumelanin kann alle Farbtöne von braun bis schwarz annehmen, Phäomelanin überwiegt bei blonden und roten Haaren. Der Verlust der Pigmentierung mit fortschreitendem Alter (Ergrauen der Haare) geht in der Regel nicht mit Änderungen im physikalischen Verhalten der Haare einher [Hollfelder et al., 1995], das fortschreitende Alter an sich dagegen schon. Unpigmentierte Haare sind etwas gröber und welliger als pigmentierte.

2.3 Geometrische Eigenschaften

2.3.1 Haarlänge

Die Haarlänge hängt vom jeweiligen Haartypus ab. Die pränatalen Haare oder Lanugo fallen in der Regel schon vor der Geburt oder kurz danach aus und erreichen Längen von kaum mehr als 15 cm. Sie werden direkt vom präpubertären Haartypus, dem primären Terminalhaar abgelöst. Seine Struktur ist deutlich gröber als das Lanugo und erreicht Längen von bis zu 60 cm. Mit dem Einsetzen der Pubertät und den damit einhergehenden hormonellen Veränderungen lässt sich eine gröbere Haarstruktur mit Längen von bis zu 100 cm feststellen, das so genannte sekundäre Terminalhaar. Zwischen dem zwanzigsten und dreißigsten Lebensjahr führen dann erneute hormonelle Umstellungen zum Wachstum von feineren und oftmals kürzeren Haarfasern. Dieser Effekt kann bis zum Übergang des Terminalhaares zum Vellus führen, das bei einem sehr kleinen

Durchmesser von ca. $4\ \mu\text{m}$ Längen von kaum mehr als $0,1\ \text{cm}$ erreicht. Dieser feine Haartypus ist mit bloßem Auge kaum wahrnehmbar und lässt die betroffenen Regionen kahl erscheinen. Tab. (2.1) fasst die Daten zusammen. Die hier genannten Längen stellen Durchschnittswerte dar. Auf „Langhaar- Wettbewerben“ können des öfteren Haarlängen von über $150\ \text{cm}$ beobachtet werden. Den Rekord hält derzeit die Chinesin Xie Qiuping aus der Provinz Guangxi mit mehr als $560\ \text{cm}$. Die Wachstumsrate von Haaren liegt bei ca. $15\ \text{cm/a}$.

2.3.2 Querschnitt

Der Haardurchmesser schwankt in der Regel mit dem Alter und nimmt seinen größten Wert mit der Pubertät an. Die Welligkeit und die Querschnittsgeometrie des Haares sind genetisch bedingt und von Rasse zu Rasse unterschiedlich, so dass eine Untergliederung der Eigenschaften in Abhängigkeit von der jeweiligen Rasse sinnvoll erscheint. Robbins [Robbins, 2002] unterscheidet hier *kaukasisch*, *mongolisch* und *äthiopisch*. Dabei stellt der kaukasische mit einem Durchmesser von $50\text{-}90\ \mu\text{m}$ den feinsten Haartyp dar. Interessanterweise besitzen Haare keinen runden, sondern einen eher elliptischen Querschnitt. Modifikationen mit nieren-/ birnenförmigem oder gar dreieckigem Querschnitt treten vereinzelt auf, primär im Zusammenhang mit pathologischen Veränderungen. Dabei ist das Verhältnis der beiden Haupttradien charakteristisch für die jeweilige Rasse. Der Quotient übt wesentlichen Einfluss auf das Biegeverhalten aus. Die Form des Querschnitts unterliegt innerhalb einer Rasse und innerhalb der Kopfbehaarung einer Person u.U. erheblichen Schwankungen. Als Faustregel gilt, dass der Durchmesser der Haare bei einer Person um mehr als den Faktor zwei variiert [Robbins, 2002]. Genau genommen ändert sich der Faserquerschnitt in axialer Richtung. Die Änderungen sind jedoch klein und können deshalb vernachlässigt werden. Querschnittswerte sind deshalb immer als Mittelwerte anzusehen.

Vernall [Vernall, 1961] hat die Haare von jungen, männlichen Personen unterschiedlicher Nationalität im Alter zwischen 20 und 30 Jahren hinsichtlich ihrer Querschnittsgeometrie untersucht². Die Variabilität bezüglich der Geometrie zwischen den Rassen ist dabei signifikant größer als zwischen den Individuen einer Rasse und hier wiederum signifikant größer, als beim Individuum selbst.

Die Werte der Hauptdurchmesser bewegen sich zwischen $60\ \mu\text{m}$ beim West-Europäer und $125\ \mu\text{m}$ beim Chinesen (kleinster bzw. größter über alle Proben beobachteter Wert). Die Haare des Afrikaners sind in Bezug auf den Hauptdurchmesser durchschnittlich am größten, die eines West-Europäers am kleinsten (Chinesen und Inder liegen dazwischen).

Betrachtet man den Wert des kleinsten Durchmessers, so liegen die Werte zwischen $43\ \mu\text{m}$ beim West-Europäer und $100\ \mu\text{m}$ beim Chinesen. Chinesen verfügen im Durchschnitt über die dicksten, Europäer dagegen über die dünnsten Haare. Der intra-personelle Unterschied bewegt sich zwischen $25\ \mu\text{m}$ beim Individuum afrikanischer Abstammung und $313\ \mu\text{m}$ beim Inder.

In Bezug auf die Querschnittsfläche verfügen West-Europäer mit $2.017\ \mu\text{m}^2$ über den kleinsten und die Chinesen mit $10.105\ \mu\text{m}^2$ über den größten Wert.

Als Index für den Grad der Elliptizität wird der Quotient der beiden Haupttradien herangezogen. Die Indices variieren hier zwischen 1,25 und 1,68. Das Haar des Afrikaners zeichnet sich

²Es wird eine etwas andere Aufteilung bezüglich der Rassen verwendet als in [Robbins, 2002]. Die etwas unglücklichen Bezeichnungen *Negroider* und *negroid* werden im folgenden durch die moderneren Begriffe *Afrikaner* bzw. *afrikanisch* ersetzt.

Tabelle 2.2: Größter bzw. kleinster mittlerer Durchmesser, sowie Querschnittsfläche bei Haarfasern von Personen unterschiedlicher ethnischer Zugehörigkeiten [Swift, 1995, Robbins, 2002].

Zugehörigkeit	d_{max} [μm]	d_{min} [μm]	d_{max}/d_{min} [-]	A [μm^2]
West-Europäer	81,94	56,74	1,44	3,786
Chinesen	94,28	76,79	1,23	5,817
Inder	92,94	66,49	1,40	4,994
Afrikaner	98,23	58,52	1,68	4,648
Kaukasier	63,93	47,28	1,35	2,411
Mongolen	79,53	61,96	1,28	4,085
Athiopier	90,62	51,70	1,75	4,006

durch seine Flachheit und folglich durch den größten Index aus. Bei den Chinesen findet man den kleinsten Index und damit die rundeste Form. Die Geometrie bei West-Europäern und Indern liegt dazwischen. In Tab. (2.2) sind die Werte zusammengestellt. Besonders das afrikanische Haar fällt neben seiner ausgeprägten Elliptizität auch durch seine extreme axiale Torsion auf [Menkart et al., 1966].

Swift [Swift, 1995] führt, basierend auf den Erkenntnissen zur Elliptizität, weitergehende Überlegungen zur Bevorzugung bestimmter Deformationsrichtungen von Haarfasern durch, auf die im Zusammenhang mit dem Biegeverhalten von Haaren weiter unten kurz eingegangen werden soll.

2.3.3 Welligkeit

Die longitudinale Struktur oder Welligkeit des Haares stellt eine der wichtigsten Größen für das Haarstyling dar und variiert insbesondere mit der Rasse. Die extreme Lockigkeit des Afrikanerhaares findet man sonst nur bei den Populationen Papua-Neuguineas [Ogle & Fox, 1999]. Im asiatischen Raum überwiegen Haare mit glatter Konfiguration, wohingegen im europäischen Raum alle Arten vertreten sind. Die Haargeometrien sind schematisch in Abb. (2.2) dargestellt. Dauerwelle oder die Applikation von Haarrelaxern bewirken eine Änderung der Welligkeit. Je nachdem, ob das Haar eine mehr glatte oder eine welligere Struktur aufweist, fällt der Einfluss physikalischer Effekte wie Reibung und Steifigkeit größer oder kleiner aus. Robbins und Reich [Robbins & Reich, 1986] stellen eine Methode zur Bestimmung der Welligkeit vor. Dabei wird die Länge l_c des einzelnen Haares im Normalzustand und im gestreckten Zustand (l_t) durch Anhängen eines Gewichtes von 1 g gemessen. Sieht man die Konfiguration des einzelnen Haares vereinfacht als Sinuskurve an, kann durch Abzählen der Maxima n die Welligkeit quantifiziert werden zu $C = n \cdot l_t/l_c$ ³. In der Regel weisen Haarfasern aber diesbezüglich große Unregelmäßigkeiten auf. Nach dem Durchnässen der Haare oder der Applikation von Ölen können die Kohäsionskräfte des Fluides in den kapillaren Zwischenräumen der Fasern Werte erreichen, bei denen die Wellen ganz verschwinden. Reibung und Gewicht des Haftwassers verstärken den Effekt, der sich besonders deutlich bei kaukasischem und orientalischem Haar zeigt. Krauses Haar neigt dagegen deutlich weniger zum Nässe bedingten Formverlust. Die Werte für C in Tab. (2.3)

³Leider fehlen in dieser Untersuchung Angaben für n oder l_t , so dass sich die Werte lediglich für den Vergleich untereinander eignen.

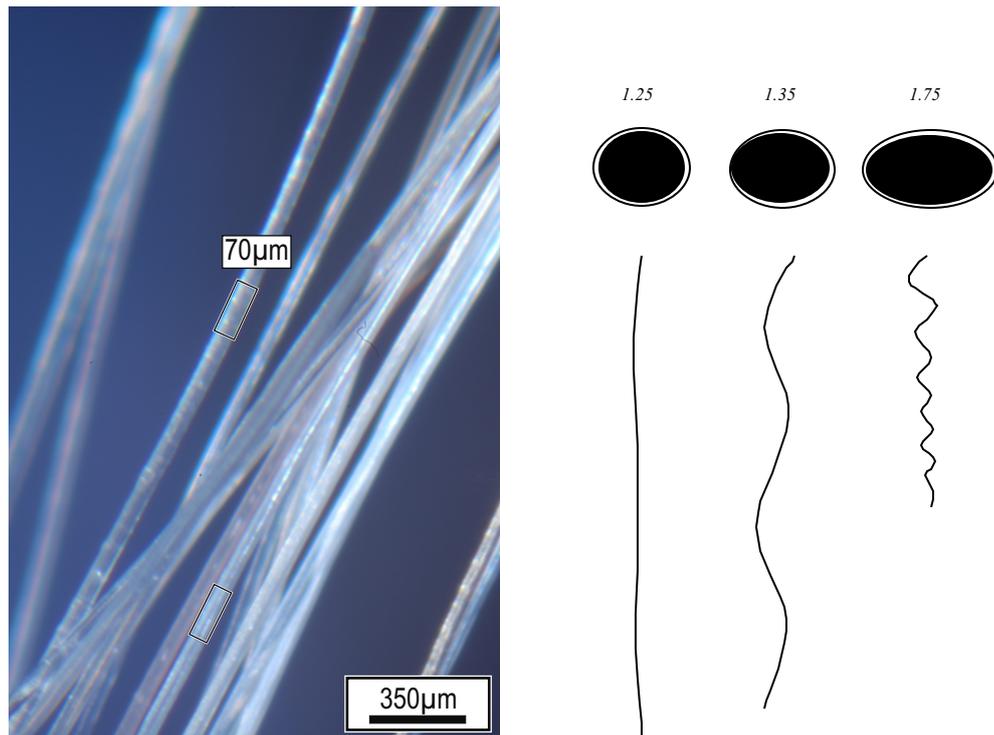


Abbildung 2.2: Links: Lichtmikroskopische Aufnahme einer Haaranordnung (©Universität Bonn). Rechts: Welligkeit und Elliptizität bei Haarfasern von Personen unterschiedlicher ethnischer Zugehörigkeit: asiatisch, kaukasisch, afrikanisch, [Robbins, 2002].

können nur als grobe Anhaltswerte dienen.

Die Haarfasern ein und desselben Haarschopfes zeigen mit zunehmender Dicke eine vergrößerte Welligkeit [Yin et al., 1976]. Über den Grund kann man nur spekulieren. Eine mögliche Erklärung könnte sein, dass dünnes Haar ausschließlich in glatter Konfiguration gebildet wird. Andererseits setzen dünne Fasern Glättungseffekten wie dem Druck durch benachbarte Fasern, Eigengewicht, Kämmen und Bürsten weniger Widerstand entgegen, so dass sie infolge der alltäglichen Haarpflege möglicherweise ihre Form verlieren.

2.3.4 Haarkörper

Der Haarkörper ist das sichtbare Volumen der gesamten Haarmasse. Die Ausprägung des Haarkörpers wird im wesentlichen bestimmt durch die Steifigkeit, die Querschnittsgeometrie und den Durchmesser der Faser, die longitudinale Form sowie durch Faserinteraktionen wie Reibung und Adhäsion [Robbins, 2002]. Bei der Längsdehnung von Haaren geht der Durchmesser quadratisch, bei Biegung und Torsion dagegen mit der vierten Potenz ein, so dass die Längsdehnungen infolge Eigengewicht vernachlässigbar sind. Schon kleine Änderungen des Durchmessers sollten sich auf das Aussehen des Haarkörpers auswirken. Tolgyesi et al. [Yin et al., 1976] haben den Einfluss des Faserdurchmessers auf den Haarkörper von erwachsenen amerikanischen Frauen un-

Tabelle 2.3: „Welligkeit“ C gemessen an $l_c = 25,4$ cm langen Haarsträhnen nach diversen kosmetischen Behandlungen im nassen und trockenen Zustand (P = Pomade, SAC = Stearalkoniumchlorid, SLS = Natriumlaurylsulfat, BL = Bleichen), [Robbins & Reich, 1986].

Haartyp	Behandlung	Welligkeit	
		trocken	nass
Orientalisch	P	0,0	-
	SAC, SLS, BL	1,6	0,0
Kaukasisch	P	0,0	-
	SAC, SLS, BL	2,6	0,0
Kraus I	P	18,0	-
	SAC, SLS, BL	18,0	14,0
Kraus II	P, SAC, SLS, BL	16,0	-

ter kosmetischen Gesichtspunkten untersucht. Die Haarsteifigkeit steht hierbei im Vordergrund, Reibung und Kohäsion bleiben dagegen unerwähnt.

Die Zunahme der Welligkeit führt zu einem deutlich fülligeren Haarkörper. Die Kraft, die beim Bürsten der Haare aufgewendet werden muss, nimmt mit der Faserdicke zu.

Der Vorgang der Dauerwelle bewirkt eine Zunahme der Welligkeit und der Reibung; durch Bleichen wird die Reibung erhöht. Deshalb führt diese Art der kosmetischen Behandlung zu einer Vergrößerung des Haarkörpervolumens.

Umgekehrt wirken Haare nach der Behandlung mit einer Pflegespülung schlaff und der Haarkörper verliert deutlich an Fülle, wahrscheinlich durch Herabsetzung des Reibungsbeiwertes [Robbins, 2002]. Die Behandlung mit hochreinigenden Shampoos führt einerseits zum Rückgang der Faseradhäsion, andererseits zu einer erhöhten Reibung, da die Entfernung des Talgs von der Haaroberfläche die natürlichen Unebenheiten freilegt (cuticuläre Schuppen).

2.3.5 Haardichte

Die Haardichte übt maßgeblichen Einfluss bei der Konstitution des Haarkörpers aus. Für Männer und Frauen kaukasischer Abstammung kann ein Durchschnittswert von 340 Haare/cm² (± 76) angenommen werden, sofern keine kosmetischen Behandlungen vorgenommen wurden [Orfanos & L.Brückmann, 1982]. Andere Untersuchungen ergaben Werte von 192 Haare/cm² (± 12) [Pecoraro & Astore, 1990].

Während sich in jungen Jahren keine geschlechterspezifischen Unterschiede feststellen lassen, findet man bei Männern eine hormonell bedingte Abnahme der Haardichte in der frontoparieta- len Region. Kosmetische Behandlungen (Dauerwelle, Färben, Bleichen), die vorrangig von Frauen vorgenommen werden, können u.U. im Alter zu einer leichten Abnahme der Haardichte führen. Durchschnittswerte für die angesprochene Populationsgruppe findet man in Tab. (2.4).

Tabelle 2.4: Haardichte bei Frauen und Männern unterschiedlichen Alters, [Orfanos & L.Brückmann, 1982].

Alter [a]	weiblich [n/cm ²]	männlich [n/cm ²]
16 - 29	332	346
30 - 55	352	329

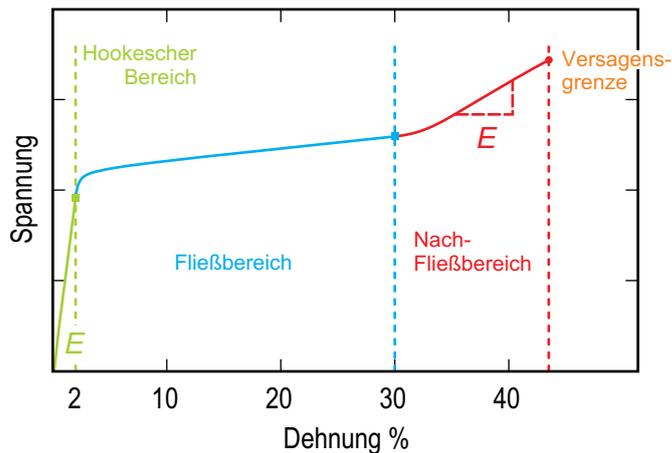


Abbildung 2.3: Verlauf der Spannungs-Dehnungskurve bei Haaren. Das Haar wird bei konstanter Rate gedehnt und die materialinternen Spannungen, die sich als Antwort auf die Dehnung aufbauen, gemessen. Die Kurve zeigt drei charakteristische Bereiche: Linearelastischer Bereich, Fließbereich und Nachfließbereich. Bei Belastungen über diese Grenze hinaus tritt Materialversagen ein, [Jollès et al., 1997].

2.4 Mechanische Eigenschaften

Die am häufigsten auftretenden Belastungen sind Dehnung, Stauchung, Scherung, Biegung und Torsion, wobei die letzten zwei als Kombinationen von Dehnung, Stauchung und Scherung aufgefasst werden können. Biegung und Torsion spielen bei der Deformation von Haaren eine wichtige Rolle. Trotzdem ist das Dehnungsverhalten am meisten untersucht worden⁴. Gerade im Bereich der Medizin, wird durch intensive Erforschung des mechanischen Verhaltens versucht, Rückschlüsse über das Vorliegen bestimmter Erkrankungen, das Alter und sogar das Geschlecht einer Person zu ziehen. Letztere Informationen sind insbesondere im Bereich der forensischen Medizin von großem Interesse, weil Haare in der Regel an jedem Tatort gefunden werden können. Ihre geringe Dicke macht sie fast unsichtbar für das ungeübte Auge und infolgedessen unterbleiben Beseitigungsmaßnahmen durch Kriminelle. Haare haben zudem die charakteristische Eigenschaft, dass sie resistent gegenüber so ziemlich allen Umwelteinflüssen sind und noch lange nach dem Tode einer Person Rückschlüsse über bspw. deren Ernährungsgewohnheiten zulassen.

2.4.1 Elastisches Verhalten

Das Dehnungsverhalten von Fasern wird im klassischen Zugversuch bestimmt. Das einzelne Haar wird bei gegebener Temperatur und Luftfeuchte mit einer konstanten Rate⁵ gedehnt und die Widerstandskraft des Materials gemessen. Trägt man die Spannungen σ (Verhältnis von Verlängerung zur Gesamtlänge) gegen die Dehnungen ε auf, ergibt sich eine Kurve mit drei charakteristischen Bereichen: (i) HOOKEScher oder linearelastischer Bereich (pre-yield), (ii) Fließbereich (yield) und (iii) Nachfließbereich (post-yield), Abb. (2.3).

Andere wichtige Parameter der Spannungs-Dehnungskurve sind der Modul der Elastizität E_α , die HOOKESche Grenze, die Bruchdehnung und die Bruchspannung, bei der das Material versagt, der Modul des Nachfließbereiches E_β und die Dehnungsarbeit, die der Fläche unter der Kurve entspricht.

⁴Interessanterweise gibt es trotz der Häufigkeit der Versuche keine Standards.

⁵Üblich sind hier 25 mm/min.

Tabelle 2.5: Schallgeschwindigkeit und E-Modul bei unterschiedlicher rel. LF, $T=23\text{ }^{\circ}\text{C}$, [Goldsmith & Baden, 1970].

rel. LF [%]	Schallgeschwindigkeit [km/s]	E-Modul [10^5 N/cm^2]
6	3,0	11,7
53	2,6	8,8

2.4.1.1 Linearelastischer Bereich

Im HOOKEschen oder linearelastischen Bereich sind die Dehnungen proportional zu den Spannungen, d.h., die Kurve weist eine in etwa konstante Steigung auf, die sich als materialabhängige Kenngröße interpretieren lässt. Die Steigung entspricht dem Elastizitätsmodul E (oder auch E-Modul oder YOUNGSche Zahl).

Biologische Materialien folgen aber strenggenommen keinem elastischen Verhalten, sondern zeigen vielmehr ein zeitabhängiges Verformungsverhalten [Fung, 1993]. Deshalb ist der Name HOOKEscher Bereich nicht ganz angemessen [Feughelman, 1982]. Der E-Modul setzt sich aus einem zeitunabhängigen elastischen und einem zeitabhängigen viskosen Anteil zusammen. Leichte Schwankungen des E-Moduls in Abhängigkeit von der Dehnungsrate sind deshalb nicht auszuschließen. Auf die Zusammenhänge wird weiter unten kurz eingegangen.

Robbins [Robbins, 2002] gibt für den E-Modul einen Wert von $3,89 \cdot 10^5\text{ N/cm}^2$ (60 % rel. LF) an⁶. Goldsmith und Baden [Goldsmith & Baden, 1970, Goldsmith & Baden, 1971] stellen ein alternatives Verfahren zur Bestimmung des E-Moduls vor. Die vom Querschnitt des Materials unabhängige Schallgeschwindigkeit und der E-Modul stehen über $v_s = \sqrt{E/\rho}$ in Beziehung. Daraus kann bei bekannter Dichte⁷ der E-Modul durch Messung der Schallgeschwindigkeit für unterschiedliche Dehnungen bestimmt werden. Die Schallgeschwindigkeit und der E-Modul für unterschiedliche rel. Luftfeuchten sind in Tab. (2.5) aufgeführt.

Weiterhin findet man eine Altersabhängigkeit des Durchmessers, der Zugfestigkeit und der Dehnungsarbeit bis zum Materialversagen. Abb. (2.4) zeigt die Zusammenhänge, die bei japanischen Männer im Alter von zwei bis 91 Jahren festgestellt werden konnten [Naruse & Fujita, 1971]. Für den E-Modul wird hier ein Wertebereich von $1,5\text{--}4,6 \cdot 10^5\text{ N/cm}^2$ angegeben (20 °C, 65 % rel. LF). Swanbeck und Nyren [Swanbeck & Nyren, 1970] haben die Änderung der mechanischen Eigenschaften bei Patienten, Männer und Frauen kaukasischer Abstammung im Alter zwischen 20 und 56 Jahren, mit unterschiedlichen Formen von Haarerkrankungen untersucht. Signifikante Änderungen konnten nicht festgestellt werden. Für den E-Modul liegen die Werte hier zwischen $1,64\text{--}2,80 \cdot 10^5\text{ N/cm}^2$ (20 °C, 100 % rel. LF).

Die Frage, ob eine Haarprobe von einer männlichen oder einer weiblichen Person stammt, oder gar das Alter einer Person, lässt sich nicht anhand der Zugfestigkeit beantworten⁸, wie Johri und Jatar konstatieren [Johri & Jatar, 1982]. Allerdings liefert ihre Studie Durchschnittswerte für den E-Modul von $1,23\text{--}4,10 \cdot 10^5\text{ N/cm}^2$ (30 °C, 30 % rel. LF) für männliche und weibliche Personen (vermutlich indischer Abstammung) im Alter zwischen 15 und 35 Jahren.

⁶Wegen der Feuchtigkeits- bzw. Temperaturabhängigkeit werden die Umgebungsbedingungen immer mit angegeben, falls bekannt.

⁷[Goldsmith & Baden, 1970] geben für die Dichte des menschlichen Haares einen Wert von $\rho = 1,3\text{ g/cm}^3$ an.

⁸Obwohl ein Zusammenhang zwischen Alter und Zugfestigkeit durchaus besteht, wie weiter oben angesprochen wurde.

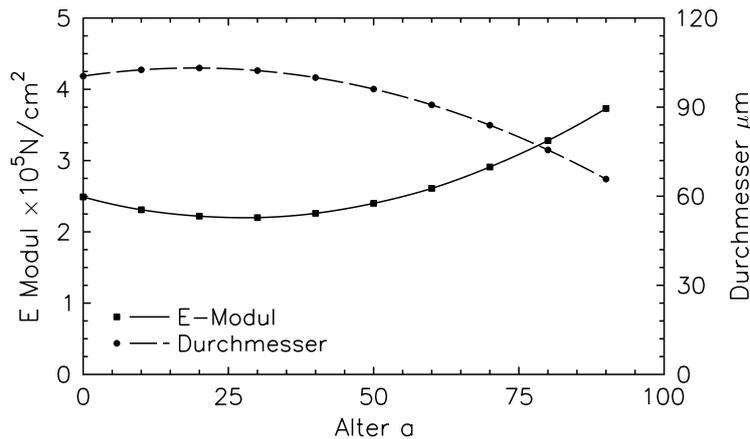


Abbildung 2.4: Statistisch ermittelte Zusammenhänge zwischen Alter und E-Modul bzw. Haardurchmesser, [Naruse & Fujita, 1971].

Sakai et al. [Sakai et al., 2000] geben einen Wert von $0,38 \cdot 10^5$ bzw. $1,03 \cdot 10^5 \text{ N/cm}^2$ bei 40°C bzw. 90°C (in Wasser) an. Diese Werte sind für Dehnungsraten von $0,5\text{--}5,0 \text{ %/min}$ quasi konstant. Interessanterweise kann man nicht davon ausgehen, dass dicke Haarfasern eine größere Steifigkeit aufweisen als dünne [Nikiforidis et al., 1993]. Die Werte für den E-Modul sind bei dünnen Haarfasern ($< 70 \mu\text{m}$) signifikant größer als bei dicken ($> 80 \mu\text{m}$) Haaren derselben Person. Der Grund könnte in der Ausprägung der porösen Medulla liegen, die bei dicken Haarfasern einen größeren Querschnitt im Verhältnis zur Gesamtfläche aufweist, als bei dünnen. Das führt zu einer Destabilisierung des Haarcortex gegenüber mechanischen Beanspruchungen. Die Entnahmeregion der Haarfasern (occipital oder frontoparietal) hat dagegen keinen Einfluss auf den E-Modul. In Tab. (2.6) sind die Werte für dünne und dicke Fasern dargestellt.

2.4.1.2 Fließbereich

Beim Überschreiten der HOOKEschen Grenze bei $\epsilon \approx 2 \text{ %}$ führen bereits kleine Spannungszunahmen zu überproportional großen Längenänderungen (ca. 30 %), oder anders ausgedrückt: das Material reagiert bei weiterer Dehnung nicht mehr durch die antagonistische Erhöhung der materialinternen Spannungen, sondern beginnt sich plastisch zu Verformen. Es ist lediglich noch ein geringer Spannungsanstieg zu verzeichnen, das Material fängt an zu fließen. Aus diesem Grund wird dieser Bereich auch als Fließbereich bezeichnet.

Auf molekularer Ebene vollzieht sich der Übergang des Keratins vom α -helikalen in den β -gefalteten Zustand [Feughelman, 1973]. Die Dehnungen sind vollständig reversibel, wenn man Haarfasern unter Wasser hält oder ihnen einen längeren Erholungszeitraum einräumt. Ruetsch und Weigmann [Ruetsch & Weigmann, 1996] konnten zeigen, dass solche extremen Belastungen trotz der Wiederherstellung der mechanischen Eigenschaften zu irreversibler Schädigung der Cuticula führen und damit die These bestätigen, dass die Cuticula keinen Beitrag zur Zugfestigkeit von Haaren leistet. Im Bereich großer Dehnungen führen Scherkräfte in den äußeren Haarschichten zum Abheben der cuticulären Schuppen. Es ist zu vermuten, dass dadurch der Reibungskoeffizient beeinflusst wird. Im nassen Zustand ist der E-Modul des Cortex etwa 20 mal größer als der Modul der Cuticula. Bei der Torsion dagegen hat die Cuticula einen nicht zu vernachlässigenden Einfluss.

Tabelle 2.6: Durchschnittswerte für E_α bei dünnen und dicken Haaren, 23 °C, 45 % rel. LF; cf. [Nikiforidis et al., 1993].

Probe	<70 μm [10^5 N/cm^2]	>80 μm [10^5 N/cm^2]
1	3,44	1,70
2	4,89	2,00
3	4,22	1,70
4	4,44	1,70
5	4,22	2,55
6	6,44	2,44
7	4,00	2,22
8	5,40	2,33
9	4,67	2,55
10	5,55	2,89
11	4,44	1,70
12	4,20	1,80
13	5,55	2,50
14	3,89	2,55
15	3,55	2,44
\bar{x}	4,59	2,20

Durch Tests konnte gezeigt werden, dass die Arbeit, die zum Ausreißen eines Haares aufgebracht werden muss, näherungsweise der HOOKEschen Grenze entspricht [Robbins, 2002]. Die Konsequenz von zu großen Zugkräften ist das Versagen der Haarwurzel, noch lange bevor Materialermüdung eintritt.

Über die obere Grenze (Fließpunkt) des Fließbereiches hinaus ändert sich die Länge des Haares wieder langsamer. Die Spannungen sind hier wieder proportional zu den Dehnungen. Das Verhalten ist hier, wie im HOOKEschen Bereich, in etwa linearelastisch. Die Steigung der Kurve wird durch den E-Modul E_β des Nachfließbereiches beschrieben. Das Verhältnis der Steigungen der drei Bereiche beträgt etwa 100:1:10 [Feughelman, 1982]. Beanspruchungen über die Bruchdehnung hinaus führen zu einem schlagartigen Materialversagen.

2.4.2 Rheologische Eigenschaften

Die Rheologie beschäftigt sich traditionell mit dem Verformungsverhalten von Materie. Gerade in dieser Hinsicht ist das Verhalten vieler biologischer Materialien irgendwo zwischen denen eines elastischen Festkörpers und denen einer viskosen Flüssigkeit anzusiedeln. Ersterer speichert bei Deformation die Energie, die durch die Arbeit äußerer Kräfte verrichtet wird und stellt bei Entlastung seinen Ausgangszustand, seine ursprüngliche Form, exakt wieder her. Eine viskose Flüssigkeit auf der anderen Seite verliert ihre Form unter Belastung unwiederbringlich. Die durch eine bestimmte Dehnung in einer Faser verursachten Spannungen bzw. die durch eine bestimmte Spannung verursachte Dehnungen hängen auch von der Dauer der Belastung und der mechanischen Historie, also allen vorausgegangenen Belastungen ab. D.h., erfolgt die Beanspruchung der Faser über eine bestimmte Dauer, bspw. mit einer konstanten Dehnung, verformt sich das Mate-

rial plastisch, es fängt an zu kriechen. Diese Verformungen sind in der Regel reversibel. Entlastet man die Faser, findet man einen instantanen Abfall der Spannungen bzw. Dehnungen auf einen bestimmten Wert, d.h., eine teilweise Wiederherstellung des Ausgangszustandes. Darüber hinaus erfolgt die Relaxation des Materials als Funktion von der Zeit. Dieser Vorgang wird durch die so genannte Relaxationsfunktion $G(t)$ beschrieben, die sich als materialabhängige Kenngröße darstellt. Kriechen und Relaxation sind damit zwei genau entgegengesetzte Vorgänge.

Basis für die Beschreibung dieser Art von Deformationen bildet die lineare Theorie der Viskoelastizität, die im Wesentlichen in BOLTZMANN'S Superpositionsprinzip ihren Niederschlag findet.

Die im Kriechexperiment bspw. durch Vorgabe einer Rechteckspannung ($\sigma(t) = \sigma_0, t_1 < t < t_2, \sigma(t) = 0$ sonst) in der Probe auftretenden Dehnungen lassen sich im Wesentlichen in drei Anteile $\varepsilon_1, \varepsilon_2$ und ε_3 zergliedern: $\varepsilon_1, \varepsilon_2$ sind die unmittelbare elastische bzw. die verzögerte Deformation; ε_3 hingegen ist der NEWTONSche Fluss, also genau der Anteil der Deformation, der dem NEWTONSchen Viskositätsgesetz für viskose Flüssigkeiten folgt. Dabei sind alle drei Anteile exakt proportional zur aufgebrauchten Spannung. Das Belastungsprogramm definiert hier eine Kriechkomplianz $J(t)$, die nur von der Zeit abhängt: $\varepsilon(t)/\sigma = J(t) = J_1 + J_2 + J_3$, wobei die J_i den ε_i entsprechen. Die unmittelbare elastische Deformation ist zeitunabhängig, während der verzögerte Anteil von der Zeit abhängt. Beide Anteile sind vollständig reversibel, z.B. wenn die Probe entlastet wird. Der rein viskose Anteil hingegen ist nicht reversibel und macht das Verhalten hochgradig nicht linear. In der Regel tritt der NEWTONSche Fluss von starren Polymeren nicht bei gewöhnlichen Temperaturen auf, weil ihre Viskosität zu groß ist. Selbst in Fällen, in denen geringe viskose Anteile auftreten können, bspw. bei linear amorphen Polymeren oberhalb der Glassübergangstemperatur, werden sie in der Regel von den elastischen Anteilen dominiert. Kristalline Strukturen hingegen haben keine viskosen Komponenten. Im Interesse der Einfachheit der Theorie werden also in der linearen Viskoelastizität alle viskosen Terme vernachlässigt.

Im Relaxationsexperiment wird die Entwicklung der Spannung gemessen, während man die Dehnungen der Probe konstant hält. Die Relaxationsfunktion ist $G(t) = \sigma/\varepsilon$. Wenn keine viskosen Verformungen (Fluss) auftreten, strebt der Relaxationsmodul gegen einen endlichen Wert, wohingegen er bei Auftreten viskoser Anteile in hinreichend langen Zeiträumen ganz auf null abfallen kann. Für den ersten Fall erhält man dann den Gleichgewichtsmodul G_∞ nach unendlicher Zeitdauer.

Die genannten Zusammenhänge lassen sich auch in formaler Form darstellen. Das Superpositionsprinzip von BOLTZMANN ist eine mathematische Ausdrucksform der linearen Viskoelastizität. BOLTZMANN hat vorgeschlagen, dass (i) das Kriechverhalten eine Funktion der gesamten mechanischen Vorgeschichte der Probe ist, dass (ii) jeder Belastungsschritt einen unabhängigen Beitrag zur endgültigen Deformation leistet und dass (iii) diese endgültige Deformation durch Aufaddieren aller Anteile erhalten wird.

Gegeben sei ein einfaches Belastungsprogramm mit den inkrementellen Spannungen $\Delta\sigma_1, \Delta\sigma_2, \dots, \Delta\sigma_n$, die zu den Zeiten $\tau_1, \tau_2, \dots, \tau_n$ wirken, Abb. (2.5). Nach dem Superpositionsprinzip nimmt die Kriechfunktion dann zum Zeitpunkt t den Wert

$$\varepsilon(t) = \Delta\sigma_1 J(t - \tau_1) + \Delta\sigma_2 J(t - \tau_2) + \dots + \Delta\sigma_n J(t - \tau_n) \quad (2.1)$$

an, wobei man $J(t - \tau)$ als Kriechkomplianzfunktion bezeichnet. Der Beitrag jedes einzelnen Belastungsschrittes besteht also aus dem Produkt eines Spannungsinkrementes und einer zeitabhängigen Funktion, der Kriechkomplianzfunktion, die wiederum nur von dem Zeitintervall

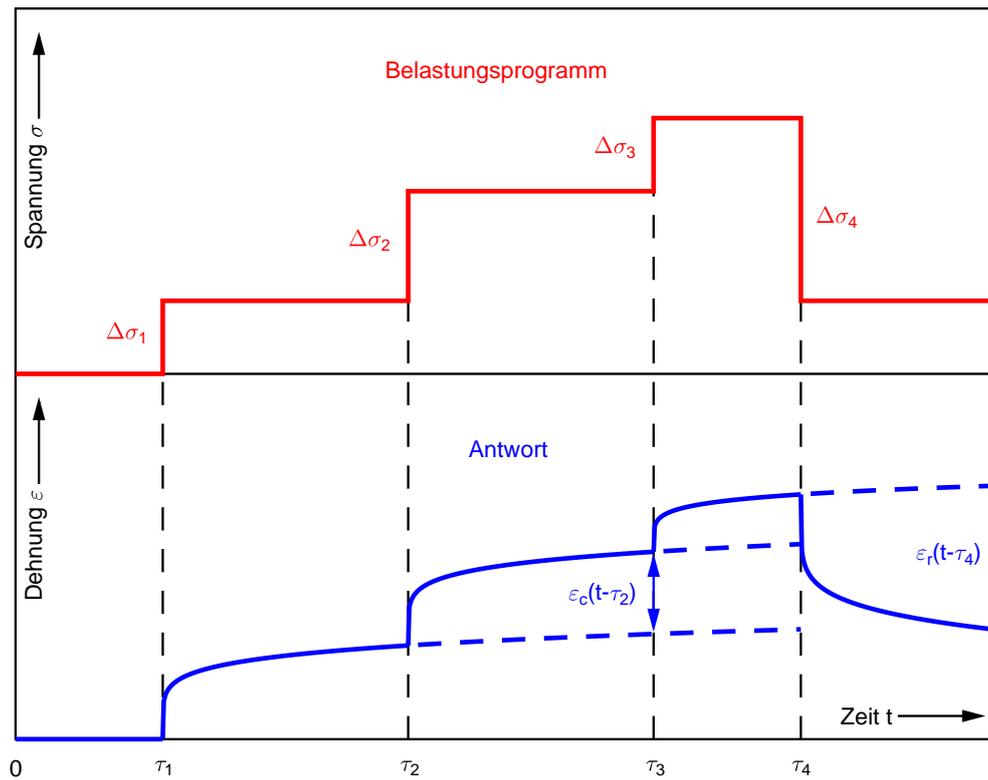


Abbildung 2.5: Qualitatives Kriechverhalten von linear viskoelastischen Materialien.

zwischen Einsetzen des Spannungszinkrementes und dem Augenblick, zu dem das Kriechen gemessen wird, abhängt. Das lässt sich generalisieren,

$$\varepsilon(t) = \int_{-\infty}^t J(t - \tau) d\sigma(\tau). \quad (2.2)$$

Das Superpositionsprinzip nach BOLTZMANN definiert also eine Integralgleichung und wird infolgedessen auch als *Integralrepräsentation* der linearen Viskoelastizität bezeichnet.

Den unmittelbaren elastischen Anteil zieht man vor das Integral, während der verbleibende Term unter dem Integral in mathematisch korrekter Form geschrieben wird, was auf das DUHAMEL-Integral führt,

$$\varepsilon(t) = \left[\frac{\sigma}{G_u} \right] + \int_{-\infty}^t J(t - \tau) \frac{d\sigma(\tau)}{d\tau} d\tau. \quad (2.3)$$

Die untere Integrationsgrenze ist $-\infty$, weil nach dem BOLTZMANN-Prinzip alle vorangehende Elemente der Belastungsgeschichte Berücksichtigung finden müssen. Im einfachsten Fall wird das Integral einfach durch die diskrete Summe der unterschiedlichen Antwortterme aufgefasst. Betrachten wir bspw. die ersten zwei Schritte unseres Programms in Abb. (2.5). Zum Zeitpunkt $t = \tau_1$ erfolgt die Aufgabe eines Belastungszinkrementes der Größe $\Delta\sigma_1$, zum Zeitpunkt $t = \tau_2$ ein

weiteres Inkrement der Größe $\Delta\sigma_2$. Die dadurch initiierten Dehnungen sind $\varepsilon_1 = \Delta\sigma_1 J(t - \tau_1)$ und $\varepsilon_2 = \Delta\sigma_2 J(t - \tau_2)$ bzw. die Gesamtdehnung $\varepsilon(t)$ die Summe aus beiden, $\varepsilon(t) = \varepsilon_1 + \varepsilon_2$. Die zusätzliche durch den zweiten Belastungsschritt verursachte Dehnung $\varepsilon_c(t - \tau_2)$ ist damit genauso groß, als wenn die Zuladung ohne Vorbelastung zum selben Zeitpunkt τ_2 erfolgt wäre. Das ist eine direkte Konsequenz des Superpositionsprinzips.

Betrachten wir weiterhin die Spannungsinkremente $\Delta\sigma_2$ und $\Delta\sigma_3$, die zum Zeitpunkt τ_2 bzw. τ_3 einsetzen und die sich anschließende Spannungsreduktion um $\Delta\sigma_4$ zum Zeitpunkt τ_4 . Der Verlauf der Dehnungsfunktion, wenn t größer als τ_4 ist, ergibt sich durch Addition der Einzelanteile $\varepsilon_2 = \Delta\sigma_2 J(t - \tau_2)$, $\varepsilon_3 = \Delta\sigma_3 J(t - \tau_3)$ und $\varepsilon_4 = -\Delta\sigma_4 J(t - \tau_4)$. Die Erholungsfunktion $\varepsilon_r(t - \tau_4)$ (der Index r steht hier für *recovery*) ist die Differenz zwischen dem durch die Anfangsspannungen $\Delta\sigma_2$ und $\Delta\sigma_3$ ausgelösten Kriechen ohne Berücksichtigung des Spannungsabfalls σ_4 (gestrichelt fortgesetzte Linie im Diagramm) und der eigentlichen gemessenen Antwort, $\varepsilon_r(t - \tau_4) = \Delta\sigma_2 J(t - \tau_2) + \Delta\sigma_3 J(t - \tau_3) - [\Delta\sigma_2 J(t - \tau_2) + \Delta\sigma_3 J(t - \tau_3) - \Delta\sigma_4 J(t - \tau_4)] = \Delta\sigma_4 J(t - \tau_4)$. Hieraus wird deutlich, dass sowohl die Kriechfunktion als auch die Erholungsfunktion die gleichen Werte annehmen. Das ist eine weitere direkte Konsequenz des Superpositionsprinzips.

Das Entspannungsverhalten kann man analog zum Kriechverhalten durch das Superpositionsprinzip ausdrücken. Bspw. hätte man für die Serie von inkrementell wirkenden Dehnungen $\Delta\varepsilon_1, \Delta\varepsilon_2, \dots, \Delta\varepsilon_n$, die zu den Zeitpunkten $\tau_1, \tau_2, \dots, \tau_n$ einsetzen, eine resultierende Gesamtspannung im Material, von

$$\sigma(t) = \Delta\varepsilon_1 G(t - \tau_1) + \Delta\varepsilon_2 G(t - \tau_2) + \dots + \Delta\varepsilon_n G(t - \tau_n) = \sum_{i=1}^n \Delta\varepsilon_i G(t - \tau_i). \quad (2.4)$$

$G(t - \tau)$ ist die schon bekannte Relaxationsfunktion. In integraler Form, nach Extraktion des unmittelbaren Anteils an der Spannungsfunktion und Einführung des Gleichgewichtmoduls G_r ,

$$\sigma(\varepsilon, t) = [G_r \varepsilon] \int_{-\infty}^t G(t - \tau) \frac{d\varepsilon(\tau)}{d\tau} d\tau. \quad (2.5)$$

Der Kriechvorgang und die Entspannung wirken also genau entgegengesetzt zueinander und deshalb lässt sich diese Beziehung auch formalisieren zu

$$\int_0^t J(t - \tau) \frac{dG(\tau)}{d\tau} d\tau = 1. \quad (2.6)$$

Das Integral über das Produkt aus Kriechkomplianzfunktion und der zeitlichen Änderung der Relaxationsfunktion muss also konstant bleiben oder nach einer Normierung der Definitionen von $J(\tau)$ und $G(\tau)$ gleich dem Einselement sein. Die Größe der Anfangsbelastung entscheidet, wie schnell die Relaxationsfunktion $G(t)$ ihren Gleichgewichtswert erreicht. Bei sehr hohen Anfangsbelastungen kann sich die Dauer der Abnahme über viele Stunden erstrecken. Das hängt u.a. von der Art des Gewebes bzw. seiner Dehnbarkeit ab.

Chapman [Chapman, 1973] hat das Entspannungsverhalten von Wollfasern bei kleinen Biegedeformationen (0,5-4,0 %) hinsichtlich der Fragestellung untersucht, inwieweit sich das Entspannungsverhalten unter Anwendung der Theorie der linearen Viskoelastizität beschreiben lässt. Den rheologischen Eigenschaften während der Biegung kommt im Bereich der Textilindustrie eine besondere Bedeutung zu, da sie das Trageverhalten von Kleidung (Faltenwurf, Flexibilität etc.) maßgeblich bestimmen. Über die Frage, ob das Biegeverhalten aufgrund von Daten aus Dehnungsversuchen ausreichend beschrieben werden kann, herrscht in der Fachwelt eine geteilte Meinung. Die

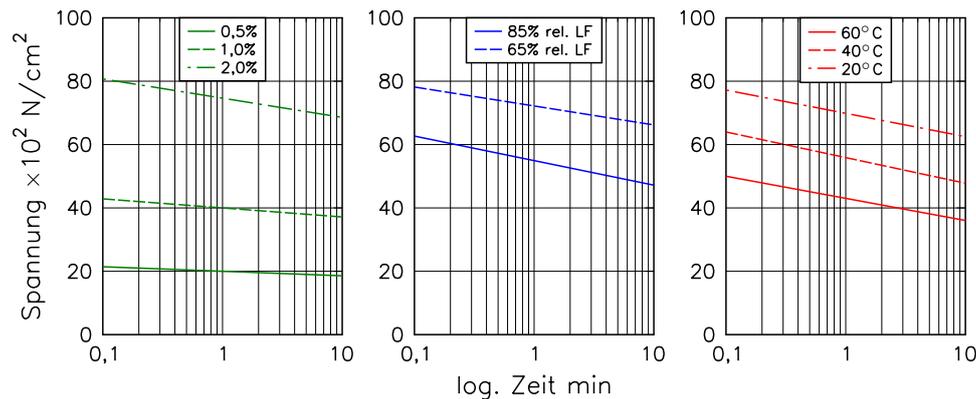


Abbildung 2.6: Relaxationskurven von Merino-Wolle. Als Standardsetting wurde eine Anfangsdehnung von 2 % und eine Relaxationsdauer von 10 min bei 20 °C und 65 % rel. LF festgelegt. Die Kurven zeigen die Entwicklung des Relaxationsmoduls bei Variation der einzelnen physikalischen Qualitäten a) Anfangsdehnung, b) rel. LF und c) Temperatur, wobei die anderen Größen konstant gehalten werden, [Chapman, 1973].

während der Biegung auftretenden Dehnungen sind inhomogen. Die Spannungen nehmen bei Wollfasern für Dehnungen < 4 % linear mit der Zeit ab, wenn sie gegen die log. Zeit aufgetragen werden. Bei größeren Dehnungen weisen die Kurven eine Krümmung auf. Der Relaxationsmodul⁹ bei 1 % Dehnung und 1 min Relaxationszeit liegt bei $4,0 \cdot 10^5$ N/cm². Die lineare Theorie der Viskoelastizität setzt voraus, dass die mechanischen Eigenschaften wie die Relaxationsfunktion $G(t)$ invariant bezüglich der Zeit sind. Dies trifft bei Wolle für kleine Dehnungen (< 1 %) zu. Die Erholungskurven (Abb. (2.6)), also der zeitliche Rückgang der Dehnungen nach dem Absetzen der Spannungen, sind im Bereich bis 1 % in etwa identisch. Mit zunehmender Dehnung kommt es zu bleibenden Verformungen (viskose Anteile).

Trägt man die nach Absetzen der Dehnung verbleibende Deformation für unterschiedliche Relaxationszeiten gegen die log. Zeit auf, erhält man eine Kurvenschar, deren einzelne Kurven sich durch verschieben entlang der Zeitachse zu einer Mastererholungskurve komponieren lassen, vgl. Abb. (2.7). Daraus kann gefolgert werden, dass die Relaxationszeiten von Wolle innerhalb eines Zeitintervalls von 0,1-1000 min eine fast konstante Verteilungsfunktion haben.

Die mechanischen Eigenschaften von Keratin ändern sich allerdings im Verlaufe der Zeit, sie unterliegen einem Alterungsprozess. Dieses Phänomen ist von Textilien her bekannt: sie zeigen ein deutlich verbessertes Erholungsverhalten in Bezug auf Faltenwurf, wenn sie vor entsprechenden Tests über einen längeren Zeitraum unter konstanten Umgebungsbedingungen gelagert werden. Das entspricht einem kontrollierten Alterungsprozess. Je länger die Textilien so gelagert werden, desto besser sind auch ihre Erholungseigenschaften. Durchnässen, gekoppelt mit einer anschließenden Trocknung, also ein „Entaltern“ des Materials, hat den temporären Verlust der verbesserten Eigenschaften zur Folge, die sich jedoch im Verlaufe der Zeit wieder einstellen. Während des Alterungsprozesses nimmt die Steifigkeit von Wollfasern zu, die Fähigkeit zur Stressrelaxation hingegen ab, Chapman [Chapman, 1975]. Diese Änderungen sind besonders markant bei

⁹Spannung geteilt durch die Dehnung, bei der die Probe gehalten wird, ausgedrückt als Funktion der Zeit.

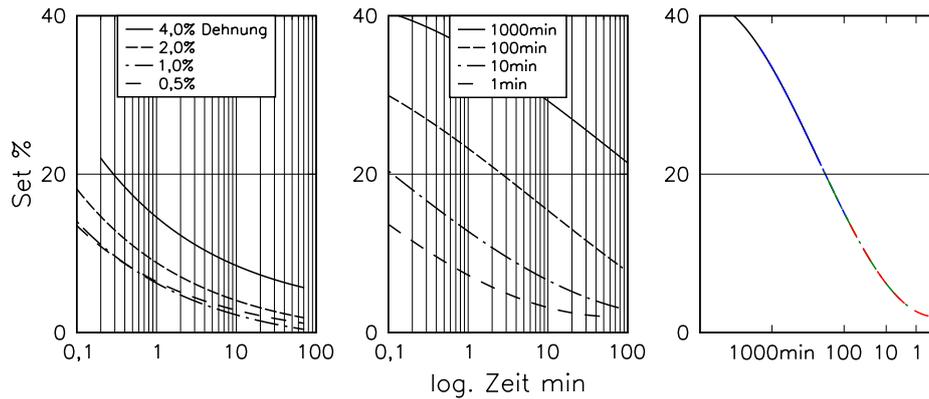


Abbildung 2.7: Erholungskurven für Merino-Wolle. a) Verbleibende Deformation in Abhängigkeit von der Zeit für unterschiedliche Biegedehnungen; b) Verbleibende Deformation in Abhängigkeit von der Zeit für unterschiedliche Relaxationszeiten; c) Mastererholungskurve, durch Verschiebung der einzelnen Teilkurven von b); nach [Chapman, 1973].

Torsion. Entsprechende Änderungen wurden auch in Bezug auf Biegedehformationen beobachtet.

Chapman [Chapman, 1975] hat die Änderung des rheologischen Verhaltens von Keratin während eines mehrere Jahre andauernden kontrollierten Alterungsprozesses untersucht. Die Alterung beobachtet man bei Raumtemperatur bevorzugt über die Änderung des E-Moduls, der sich linear mit dem Logarithmus der Alterungszeit ändert. Bei höheren Temperaturen läuft der Prozess schneller ab. Der Alterung liegt eine molekulare Entspannung des Keratinsystems zugrunde, die bei konstanten Umgebungsbedingungen so lange läuft, bis die mechanischen Parameter einen konstanten Wert annehmen. Werden die Umgebungsbedingungen geändert, strebt der Entspannungsprozess einen Gleichgewichtszustand an, der mit den neuen Bedingungen kompatibel ist.

Zur Erhebung quantitativer Ergebnisse werden dann folgende Versuchsreihen durchgeführt. Die hoch gealterten Fasern beansprucht man nach einer einstündigen Konditionierung durch Biegung bis zur einer 1 %-Dehnung und lässt den sofort einsetzenden Relaxationsvorgang für 1 (10, 100, 1000) Minuten gewähren, um die Faser anschließend aus dieser Deformation zu entlassen. Man wiederholt diese Versuchsreihe, allerdings mit dem Unterschied, dass die Faser jetzt vor jedem Relaxationsvorgang durch Behandlung mit Wasser zunächst „entaltert“ und anschließend 10 (100, 1000) Minuten bei kontrollierten Bedingungen konditioniert („gealtert“) wird.

Der Relaxationmodul verläuft beim Auftragen gegen die logarithmische Zeit in etwa linear bis zu einer Relaxationsdauer von 100 min. Man findet zudem eine fast konstante Verschiebung der Relaxationsfunktion um etwa $0,3 \text{ N/cm}^2$ nach oben pro Alterungsklasse (10, 100, 1000 min bzw. hochgealterte Fasern). Damit ist $G(t)$ nicht nur von der Relaxationszeit, sondern auch von der Alterungszeit des Keratinsystems selbst abhängig und folglich eine der Grundvoraussetzungen für linear viskoelastisches Verhalten, nämlich die der zeitlichen Invarianz der mechanischen Eigenschaften, nicht mehr erfüllt.

Die Erholung von einer vorausgegangenen Deformation wird ausgedrückt durch den verbleibenden prozentualen Anteil an der ursprünglichen Dehnung (1 %). Da der Relaxationsmodul, wie oben beschrieben, mit der Alterungszeit variiert, verhält sich Keratin nicht linear viskoelas-

tisch. Dennoch stimmen im Falle hochgealterter Fasern die empirischen Werte sehr gut mit der Vorhersage durch das BOLTZMANNsche Superpositionsprinzip überein, und zwar für alle vier Relaxationszeiten. Signifikante Abweichungen treten besonders dann auf, wenn die Relaxationszeit mindestens von derselben Ordnung ist, wie die Alterungsdauer. Für die Fälle, in denen die Relaxationszeit die Alterungszeit überschreitet, schlägt Chapman [Chapman, 1975] eine modifizierte Superposition der Form

$$\sigma(t) = \sum_{i=0}^n \Delta \varepsilon_i G(t_A, t - \tau_i)$$

vor. $G(t_A, t - \tau_i)$ hängt neben der Relaxationszeit t auch von der Alterungszeit t_A (oder absoluten Zeit) ab.

Die absolute Zeit fungiert hier als Selektor, der immer die zur Alterungsdauer passende Relaxationsfunktion auswählt. In [Chapman, 1975] sind Kurven für 10, 100, 1000 Minuten und hochgealterte Fasern angegeben, vgl. Abb. (2.7). Für andere Zeiten benutzt man zweckmässigerweise ein geeignetes logarithmisches Inter-Extrapolationsverfahren. Analog dazu kann auch eine Komplianzfunktion $J(t_A, t - \tau_i)$ angegeben werden.

Der Anstieg der Umgebungstemperatur kann den Prozess der Alterung wesentlich beschleunigen. Werte über 50 °C führen nach wenigen Stunden bereits zu Alterungserscheinungen, die sonst erst nach Jahren erreicht werden [Rigby et al., 1974].

Im Zusammenhang mit der Angabe von Werten für die verschiedenen Moduli findet man in der Literatur häufig die folgenden Ausdrücke für den Wassergehalt in einer Probe, die einer exakten Definition folgen.

- Moisture regain (Feuchtigkeitsaufnahme) oder einfach nur Regain R : Das ist das Verhältnis des Anteils an absorbiertem Wasser in der Probe zur Masse der Trockenprobe.

$$R = \frac{\text{Masse des absorbierten Wassers}}{\text{Masse der Trockenprobe}} \times 100. \quad (2.7)$$

- Moisture content (Feuchtigkeitsgehalt) M : Das ist das Verhältnis des Anteils an absorbiertem Wasser in der Probe zur Masse der ungetrockneten Probe.

$$M = \frac{\text{Masse des absorbierten Wassers}}{\text{Masse der ungetrockneten Probe}} \times 100. \quad (2.8)$$

Trockenprobe bedeutet, dass das gesamte physikalisch gebundene Wasser durch entsprechende Maßnahmen, bspw. Trocknung im Ofen oder unter Verwendung eines Trocknungsmittels (P_2O_5), aus der Probe entfernt wurde. Zwischen der Feuchtigkeitsaufnahme und dem Feuchtigkeitsgehalt besteht die einfache Beziehung

$$M = \frac{R}{1 + \frac{R}{100}}.$$

Tao und Postle [Tao & Postle, 1989a] haben ein Modell zur Abschätzung des viskoelastischen Verhaltens von Keratinkompositen entwickelt. Solche Komposite setzen sich aus zwei oder mehreren unterschiedlichen Phasen zusammen, die jede für sich genommen als Kontinuum angesehen werden kann und die über ihre Oberfläche miteinander verbunden sind, vgl.

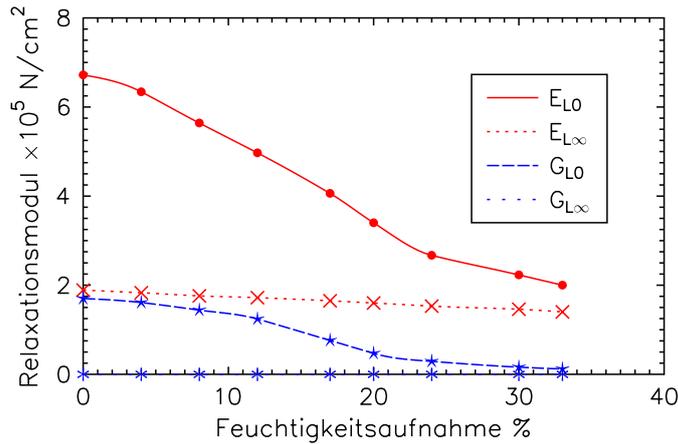


Abbildung 2.8: Effektiver longitudinaler Zug- bzw. Schubrelaxationsmodul in Abhängigkeit von der Feuchtigkeitsaufnahme, [Tao & Postle, 1989a].

[Feughelman, 1959]. Keratin kann als natürliches, faserverstärktes Komposit mit kristallinen, wasserundurchlässigen Mikrofibrillen angesehen werden, die in eine amorphe, wasserdurchlässige Matrix eingebettet sind. Wenn die Anordnung der Filamente in der Matrix einer zufälligen Verteilung folgt oder ein hexagonales Gitter bildet, ist das Komposit transversal isotrop.

Auf der Basis des statischen Korrespondenzprinzips wurden die longitudinalen und transversalen mechanischen Eigenschaften für Keratinkomposite mit viskoelastischen Filamenten und isotrop viskoelastischer Matrix analysiert und für unterschiedliche Feuchtigkeitsaufnahmen berechnet. Die Ergebnisse decken sich sehr gut mit experimentellen Messungen an Wollfasern. Die Anfangswerte für den effektiven, also auf das Gesamtkomposit bezogenen, longitudinalen Zug- bzw. Schubrelaxationsmodul (E_{L0} bzw. G_{L0}) bewegen sich zwischen $6,7 \cdot 10^5 \text{ N/cm}^2$ in der Trockenprobe und $2,0 \cdot 10^5 \text{ N/cm}^2$ bei maximaler Feuchtigkeitsaufnahme. Für den Torsionsrelaxationsmodul betragen sie $1,70 - 0,12 \cdot 10^5 \text{ N/cm}^2$. Die Zunahme des Wasseranteils bewirkt ein Absinken der Anfangswerte E_{L0} und G_{L0} auf $2,0$ bzw. $0,12 \cdot 10^5 \text{ N/cm}^2$. Die finalen Werte $E_{L\infty}$, gemessen zu dem Zeitpunkt, an dem die zeitlichen Änderungen des Relaxationsmoduls vernachlässigbar klein werden, können als konstant angesehen werden. Der Wert des Schubrelaxationsmoduls $G_{L\infty}$ beträgt unabhängig vom Wassergehalt immer 0 . Das Verhältnis E/G beschreibt den Grad der Anisotropie. Zu Belastungsbeginn ist die Anisotropie am kleinsten, Abb. (2.8).

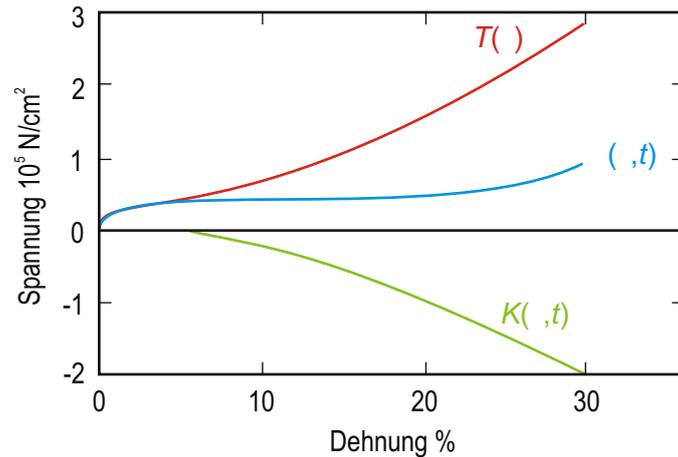
Basierend auf denselben Prinzipien haben Tao und Postle [Tao & Postle, 1989b] ein Modell für die thermale bzw. hygrale Viskoelastizität der Keratinfaser formuliert. Das Quellen der Matrix infolge Feuchtigkeitsaufnahme bzw. die thermische Ausdehnung werden durch die mechanischen und themalen/hygralen Eigenschaften der Kompositkonstituenten ausgedrückt. Der thermale Ausdehnungskoeffizient liegt bei $5,18 \cdot 10^{-4} / ^\circ\text{C}$.

Nikiforidis et al. [Nikiforidis et al., 1992, Nikiforidis et al., 1993] haben die Zugeigenschaften von Haaren unter Berücksichtigung des viskoelastischen Verhaltens mit einer speziell konstruierten Maschine bestimmt. Die Spannungen hängen von der Dehnungshistorie ab und können bei linearem Verlauf in einen zeitabhängigen und einen zeitunabhängigen Anteil untergliedert werden

$$\sigma(\varepsilon, t) = T(\varepsilon) + \int_0^t T[\varepsilon(t - \tau)] \frac{\partial G(\tau)}{\partial \tau} d\tau.$$

$T(\varepsilon)$ ist die Spannungsantwort des Materials infolge eines Einheitsprungs in der Dehnung, $G(\tau)$

Abbildung 2.9: Verlauf des elastischen und des viskosen Anteils der Spannungsantwort bei Haaren, [Nikiforidis et al., 1992].



ist die normalisierte Relaxationsfunktion mit $G(0) = 1$. Im Bereich großer Dehnungen zeigen biologische Materialien aber in der Regel kein lineares Verhalten mehr. In diesem Fall kann die Spannung experimentell durch eine schrittweise Erhöhung der Dehnung um jeweils 1 % mit einer anschließenden hinreichend langen Relaxationsphase approximiert werden. Die Dehnungen haben quasi instantan zu erfolgen. Es gilt

$$\sigma(\varepsilon, t) = \sum_{i=1}^n \left\{ T_i(\varepsilon) + \int_{t_{i-1}}^{t_i} T_i[\varepsilon(t_i - \tau)] \frac{\partial G_i(\tau)}{\partial \tau} d\tau \right\},$$

mit $t_0 = 0$. $G_i(\tau)$ ist die Relaxationsfunktion in jedem Schritt und kann vereinfacht durch eine Exponentialfunktion $G(\tau) = A_i e^{(-k_i \tau)}$ dargestellt werden. Die Summe der Integrale ist der zeitabhängige Anteil $K(\varepsilon, t)$ und $T(\varepsilon)$ der zeitunabhängige Anteil der Spannungen. Der Verlauf der beiden Anteile ist in Abb. (2.9) abgebildet.

Ein alternativer Ansatz zur Messung des Kriech- bzw. Relaxationsverhaltens ist das Aufbringen wechselnder Dehnungen bspw. mittels entsprechender Vibrationsvorrichtungen, bei gleichzeitiger Messung der resultierenden Spannungen. Sobald ein dynamisches Gleichgewicht erreicht ist, ändern sich Spannungen und Dehnungen sinusförmig. Die Dehnungen folgen den Spannungen mit einer bestimmten zeitlichen Verzögerung. Diese Phasenverschiebung δ kann genau angegeben werden, wenn man die Entwicklung der Spannungen bzw. der Dehnungen in Abhängigkeit von der Kreisfrequenz ω und der Zeit t schreibt, $\sigma = \sigma_0 \sin \omega t + \delta$, $\varepsilon = \varepsilon_0 \sin \omega t$; mittels Additionstheorem hat man $\sigma = \sigma_0 \sin \omega t \cos \delta + \sigma_0 \cos \omega t \sin \delta$. Aus der letzten Gleichung geht hervor, dass sich die Spannung aus zwei Anteilen zusammensetzt: einen in Phase verlaufenden Beitrag der Größe $\sigma_0 \cos \delta$ und einen gegenphasig zur Dehnung verlaufenden der Größe $\sigma_0 \sin \delta$. Die Spannungs-Dehnungsbeziehung kann besser durch die Gleichung $\sigma = \varepsilon_0 G_1 \sin \omega t + \varepsilon_0 G_2 \cos \omega t$ beschrieben werden, mit den beiden Moduln $G_1 = \sigma_0 / \varepsilon_0 \cos \delta$ und $G_2 = \sigma_0 / \varepsilon_0 \sin \delta$. Es ist deshalb nahe liegend, die komplexe Darstellung des Moduls $G^* = G_1 + iG_2$ bzw. $\tan \delta = G_2 / G_1$ zu wählen. Der in Phase verlaufenden Realteil G_1 wird als Speichermodul bezeichnet, weil er die Energie angibt, die in der Faser gespeichert wird. Der gegenphasig verlaufende Imaginärteil G_2 hingegen wird als Verlustmodul bezeichnet, weil er den Grad der bei Dehnung dissipierten Energie angibt, der die interne Viskosität der Haarfasern beschreibt. Den Energieverlust pro Spannungs-

Dehnungszyklus erhält man aus folgender Beziehung,

$$\Delta E = \omega \varepsilon_0 \int_0^{2\pi/\omega} (G_1 \sin \omega t \cos \omega t + G_2 \cos^2 \omega t) dt = \pi G_2 \varepsilon_0^2. \quad (2.9)$$

In den meisten Fällen ist der Verlustmodul G_2 klein im Vergleich zu G_1 , so dass $G_1 \approx G^*$ ist. Üblicherweise definiert man das dynamische mechanische Verhalten einer Faser durch den Modul $G \equiv G^*$ und den Phasenwinkel $\tan \delta = G_2/G_1$. Für hinreichend kleine G_2 gilt die Näherung $\delta \approx \tan \delta$. Die Phasenverschiebung selbst ist ein Maß für die interne Dämpfung oder Viskosität der Faser.

Huck und Baddiel [Huck & Baddiel, 1971] haben den G-Modul mittels der Methode des oszillierenden Stabes für kleine Dehnungen (im Bereich von 1–2 %) bestimmt. Für unbehandelte Haarfasern liegen die Werte für G_1 und G_2 bei $4,10 \cdot 10^5 \text{ N/cm}^2$ ($\pm 0,66$) bzw. $0,177 \cdot 10^5 \text{ N/cm}^2$ (25 °C, 75 % rel. LF). Der Unterschied beträgt hier gerade einmal eine Größenordnung. Generell korreliert der Verlustmodul G_2 mit dem Schwefelgehalt der Haarkeratine, G_1 dagegen nicht. Ein Einfluss von Substanzen für das Haarsetting auf G_1 bzw. G_2 ist statistisch nicht nachweisbar (vgl. [Huck & Baddiel, 1971], Tab. II).

Dynamische Tests, bspw. unter Einsatz von Vibrationsapparaturen haben gegenüber den klassischen (statischen) Zugversuchen den Vorteil, dass sie den *in vivo* auftretenden Belastung menschlicher Haare, bspw. durch Kopfbewegungen, Kämmen oder Tragen von Kopfbedeckungen, wesentlich näher kommen. Denn die zeitlichen Skalen, in denen man sich hier bewegt, liegen in der Größenordnung von einer Sekunde oder weniger. Auch die auftretenden Dehnungen betragen maximal 1-2 % und bleiben damit im linearelastischen Bereich.

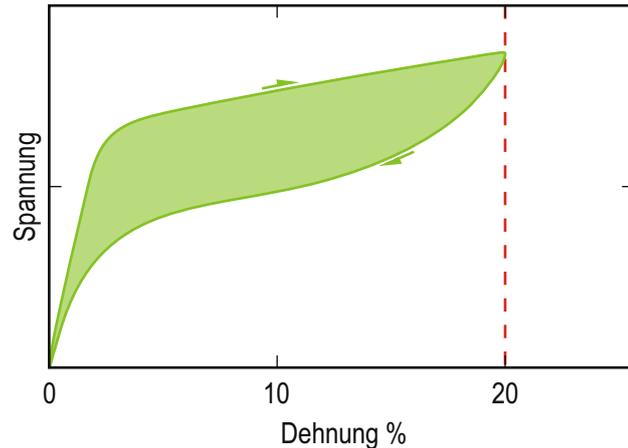
Da Relaxation und Kriechen zwei genau entgegengesetzte Prozesse sind, wie weiter oben bereits angedeutet wurde, kann man in Analogie zum komplexen Modul G^* auch eine komplexe Komplianz $J^* = J_1 - iJ_2$ einführen. Beide Größen stehen über $G^* = 1/J^*$ in Beziehung.

2.4.3 Hysterese mechanischer Parameter

Der klassische Zugversuch ist eine quasi-statische Testmethode, die das Verhalten von Haaren eindimensional beschreibt. Natürliche Belastungssituationen, die bspw. beim Kämmen oder Bürsten von Haaren auftreten, lassen sich besser in dynamischen Tests durch eine Sequenz von Lastspielen simulieren, bei denen in vorgegebenen zeitlichen Abständen genau definierte Be- und Entlastungszyklen aufeinander folgen. Die Pfade für Be- und Entlastung im Diagramm sind nicht identisch, da ein Teil der Energie infolge interner Reibung (interne Viskosität der Keratin-Matrix) dissipiert. Sie liegen vielmehr auf den beiden Seiten einer frei verschieblichen Geraden, die man in einem bestimmten Winkel gegen die Abszissenachse in das Diagramm einträgt und beschreiben damit eine Hysteresiskurve. Während die Hysteresisschleifen zu Beginn des Lastspielprogramms nicht identisch sind, stabilisieren sie sich bei zyklischer Wiederholung des Vorganges und konvergieren gegen ihre Endform. Das Material ist dann konditioniert. Der Einfluss der Dehnungsrate auf die Form der Hysteresisschleifen ist vernachlässigbar.

Untersuchungen der oben genannten Art liegen nur für Woll- bzw. Haarfasern in aquatischen Systemen vor. Eine Besonderheit aquatischer Systeme besteht darin, dass selbst große Dehnungen der Faser von 30 % und mehr vollständig reversibel sind. D.h., hier laufen die Hysteresiskurven in den Nullpunkt zurück, Abb. (2.10).

Abbildung 2.10: Typischer Verlauf der Hysteresiskurve von Woll- und Haarfasern in Wasser, [Deem & Rieger, 1968].



Unter Wasser zeigen Wollfasern bei Dehnungsraten von 100 %/min einen geringfügig größeren Widerstand als bei 0,1 %/min [Feughelman, 1973]. Das bedeutet, dass der Einfluss der Dehnungsrate auf die Hysteresiseigenschaften gering ist. Dies deckt sich mit der Beobachtung [Fung, 1993], dass eine Variation der Belastungsrate zumindest für lebendes Gewebe ohne größeren Einfluss auf die Hysteresisschleife bleibt. Eine Steigerung der Rate um das 1000-fache führt lediglich zu einer 1-2 %-igen Zunahme der Belastungsantwort.

Häufiges Wiederholung des Belastungs/Entlastungsvorganges führt zu einer fortschreitenden Abnahme der Spannungen [Feughelman, 1973]. Ein vollständiger Belastungszyklus (20 % Dehnung) wird bei 0,1 %/min in mehr als 6 h durchlaufen.

Deem und Rieger [Deem & Rieger, 1968] haben solche Versuche an chemisch modifizierten Haarfasern für große Dehnungen ($\epsilon = 30\%$) durchgeführt (in Wasser). Die Form des Entlastungsastes der Hysteresiskurve lässt sich sowohl durch die Temperatur als auch durch die Art der chemischen Modifikation wesentlich beeinflussen. Der Belastungsast zeigt dagegen lediglich geringe Formänderungstendenzen bei Variation der besagten Parameter.

Das Hysterisisverhältnis (Dehnungs- zu Wiederherstellungsarbeit bei 20 oder 30 % Dehnung) ist ein charakteristischer Parameter der Keratinfaser. Chemische Modifikationen, wie sie z.B. beim Bleichen oder bei der Dauerwelle auftreten, führen zu einem veränderten Hysterisisverhältnis. Das Verhältnis ist innerhalb eines bestimmten Bereiches proportional zur Temperatur. Der lineare Verlauf der Hysterisisverhältnis-Temperatur-Kurve zeigt einen Knick bei etwa 45 °C (Änderung der Steigung). Die Steigung des Bereiches vor dem Übergangspunkt ändert sich bei Modifikation von Disulfidbindungen, die des Bereiches danach unterliegt der Beeinflussung durch andere Bindungsarten.

2.4.4 Einfluss von Feuchtigkeit

Haare sind extrem hygroskopisch, sie enthalten ca. 15 Gew.-% Eigenfeuchte. In feuchtem Klima lassen sie sich um bis zu 100 % ihrer Ursprungslänge dehnen, unter Normalbedingungen immerhin noch um etwa 50 %. Ohne äußere Krafteinwirkung ändert sich ihre Länge bei 100 % rel. Luftfeuchte um ca. 2 % [Kassenbeck, 1999]. Der Wassergehalt bei Sättigung beträgt ca. 30 %, vgl. [Menkart et al., 1966]. Die Feuchtigkeitsaufnahme führt zu einem Anschwellen des Durchmes-

2 Eigenschaften von Humanhaar

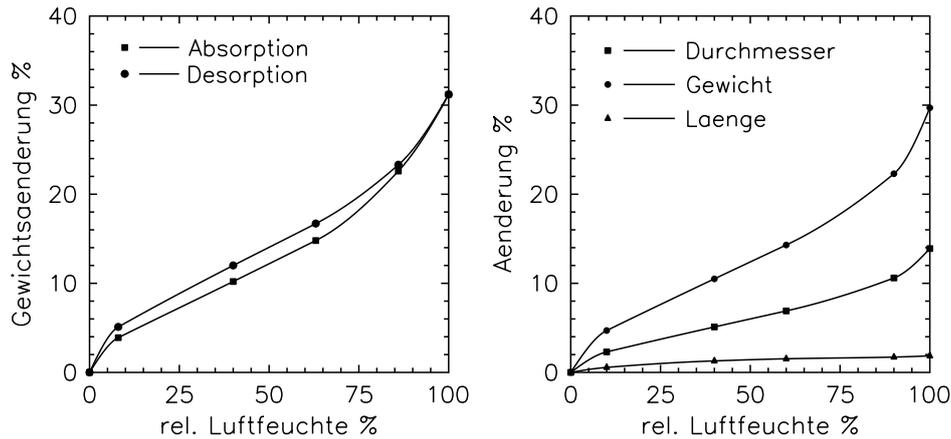


Abbildung 2.11: Links: Feuchtigkeitshysterese; die Änderung des Gewichts folgt bei der Absorption/Desorption unterschiedlichen Pfaden, [Robbins, 2002]. Rechts: Änderung des Gewichts, Durchmessers und der Länge einer Keratinfaser in Abhängigkeit von der rel. Luftfeuchte (Isothermen, 25 °C), [Robbins, 2002]

sers um bis zu 16 %. Der Feuchtigkeitsgehalt von Haaren ist abhängig von der rel. Luftfeuchte.

Seit langem bekannt ist das Phänomen der Feuchtigkeitshysterese bei Haaren, vgl. Abb. (2.11), links. Der Wassergehalt im Gleichgewichtszustand ist bei abnehmender relativer Luftfeuchte immer etwas größer als bei Zunahme derselben [Feughelman, 1997]. In Abb. (2.11) (rechts) sind die Isothermen der Feuchtigkeitsänderung dargestellt. Bezüglich Absorption und Desorption bestehen geringe rassistische Unterschiede, die sich im Bereich von ca. 1 % bewegen [Menkart et al., 1966].

Über den Zusammenhang von E-Modul und Luftfeuchte liegen lediglich Untersuchungen an Wollfasern vor (Merino). Ein ähnlicher Zusammenhang für Haare kann postuliert werden, da (i) die Dehnbarkeit von Haaren ebenfalls mit zunehmender rel. Luftfeuchte steigt und der E-Modul dagegen abnimmt, (ii) die quantitative Wasserbindefähigkeit von Wolle und Haaren in Abhängigkeit von der rel. Luftfeuchte identisch sind [Robbins, 2002].

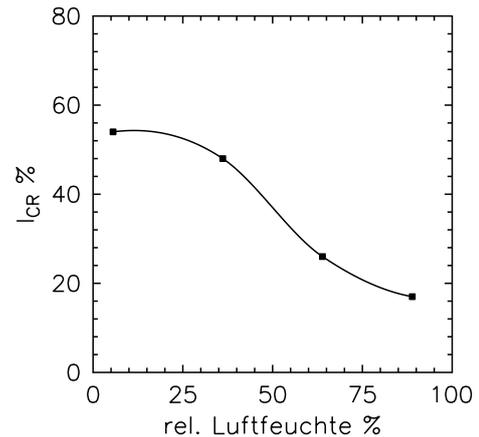
Luftfeuchte beeinflusst die mechanischen Eigenschaften von Humanhaar in erheblichem Maße. Locken, die mit Wasser gesetzt werden, so genannte Wasserwelocken, zeigen je nach Umgebungsbedingungen unterschiedliche Haltbarkeit, vgl. z.B. [Feughelman, 1997]. Der Einfluss der Luftfeuchte wird in der Regel über den Haltbarkeitsindex quantifiziert. Der Index I_{CR} (engl.: curl retention) ist das Verhältnis der Längendifferenz der ungewickelten Haarlocke und der gewickelten nach einer bestimmten Inkubationszeit (3 h) bei bekannter Luftfeuchte zur Länge im ungewickelten Zustand,

$$I_{CR} = \frac{L_{\infty} - L_3}{L_{\infty}}.$$

Zwischen I_{CR} und dem E-Modul, gemessen bei unterschiedlichen rel. Luftfeuchten, besteht ein linearer Zusammenhang [Breuer, 1972]. Dasselbe gilt für die Beziehung zwischen I_{CR} und der Kraft, die zum Herbeiführen einer 10 %-igen Dehnung aufgebracht werden muss (bei unterschiedlicher rel. Luftfeuchte, vgl. Abb. (2.12)).

Die Haltbarkeit von Wasserwelocken kann durch entsprechende Pflegeprodukte wesentlich

Abbildung 2.12: Entwicklung des Haltbarkeitsindex von wasser-gesetzten, unbehandelten Haarlocken in Abhängigkeit von der rel. Luftfeuchte, [Breuer, 1972].



erhöht werden.

Der Effekt der Wasseraufnahme kann über molekulare Modelle des Keratins verstanden werden. Das gängigste Modell ist das des *Polyelektrolyt-Gels*. Es fasst Haar als Netzwerk von Polymeren mit α -helikaler kristalliner oder amorpher Struktur auf. Uniaxiale Belastungen überführen die Moleküle vom kristallinen in den ungeordneten Zustand. Sofern keine Überdehnung vorliegt, nimmt das Netzwerk bei Entlastung wieder eine stabile Form durch Reformation der kristallinen Bereiche an. Die Polymerketten sind über Disulfidbrücken miteinander verbunden und nehmen damit feste Positionen ein. Nach dem Aufbrechen dieser Bindungen lassen sich die Moleküle frei gegeneinander verschieben. Bei der Aufnahme von Wasser – es wird chemisch gebunden – verschiebt sich das Gleichgewicht zwischen den kristallinen und den amorphen Regionen. Alle Prozesse, die mit dem Übergang vom kristallinen in den amorphen Zustand gekoppelt sind (z.B. die Dehnung), werden demnach durch die Aufnahme von Wasser bzw. durch die Luftfeuchte beeinflusst. Breuer [Breuer, 1972] hat eine Gleichung basierend auf der thermodynamischen Theorie des Schmelzverhaltens von Polymeren vorgeschlagen, über die sich die Reaktionskraft in einer gedehnten Haarfaser im Gleichgewichtszustand bestimmen lässt,

$$f = \frac{\Delta H^\circ}{\Delta L} \left(1 - \frac{T}{T_S^\circ} \right) + \frac{RT}{\Delta L} \log \left(\frac{1 + K_1 a}{1 + K_2 a} \right), \quad (2.10)$$

mit $T_S^\circ = \Delta H^\circ / \Delta S^\circ$, wobei ΔH° , ΔS° und ΔL die Differenzen der partiellen molaren Wärmen ($\approx 20,0$ J/g), der Entropien und der Längen ($\approx 0,7$ cm/g) der Aminosäurereste in der kristallinen und der amorphen wasserfreien Konformation sind. T_S und T_S° sind die Schwundtemperaturen des Haares bei gegebener Luftfeuchte und bei Abwesenheit von Wasser, K_1 ($0,34 \text{ mol}^{-1}$) sowie K_2 ($0,29 \text{ mol}^{-1}$) sind die Reaktionskonstanten des Wassers bei Bindung an den Aminosäurerest im kristallinen bzw. amorphen Zustand; a ist die Aktivität des Wassers. Dieser Ausdruck erklärt die auftretenden Kräfte gewissermaßen auf molekularer Ebene.

Bis zu einer Dehnung von ca. 1-2 % ist das mechanische Verhalten der Keratinfaser in etwa linear viskoelastisch. Der Anstieg der Steifigkeit bei Rückgang des Wassergehaltes ist zeitabhängig. Das Verhalten der Faser kann idealisiert durch ein System aus parallel geschalteter Feder und Feder-Dämpfer-Element beschrieben werden, vgl. [Feughelman, 1982]. Die Feder leistet dabei einen Beitrag von $1,4 \cdot 10^5 \text{ N/cm}^2$ (in Wasser) zur Steifigkeit. Die Viskosität des Dämpfungs-

elementes hängt von der Feuchtigkeit ab. Feughelman [Feughelman, 1959] zieht zur Erklärung ein Zwei-Phasen Modell heran. Es setzt sich aus einer wasserundurchlässigen Phase aus zylindrischen Stäben und der umschließenden wasserdurchlässigen Matrix-Phase zusammen. Nach Schätzungen leisten Matrix und Mikrofibrillen im Keratin je einen Beitrag von $0,2 \cdot 10^5$ N/cm² bzw. $1,2 \cdot 10^5$ N/cm² zur Gesamtsteifigkeit. Curiskis und Feughelman [Feugelmann & Curiskis, 1983] haben das Konzept des Zwei-Phasen Modells mittels der *Finite-Elemente-Methode* für unterschiedliche Volumenverhältnisse von Filament und Matrix untersucht. Die Keratin-Matrix zeigt mit steigendem Wassergehalt zunehmend anisotropes Verhalten. Exakte Berechnungen fordern die zusätzliche Berücksichtigung der transversalen Material-Eigenschaften. Für die POISSON-Zahl scheint 0,49 ein realistischer Wert.

2.4.5 Einfluss von Fluiden

Es ist eine bekannte Tatsache, dass sich Haare nach zu schneller Entspannung von einer vorausgehenden Belastung, z.B. nach dem Hairstyling, schlaff anfühlen. Darüber hinaus kann auch durch Öle und aquatische Systeme Einfluss auf die mechanischen Eigenschaften von Haarfasern genommen werden. Die Vorgänge sind im Sinne einer mechanischen Konditionierung aufzufassen und im Bereich der Haarpflege durchaus erwünscht. Bei kleinen Dehnungen (≤ 1 %), bei denen die Haarfasern viskoelastisches Verhalten zeigen, findet man Änderungen im Dehnungs- bzw. Relaxationverhalten nach Immersion der Fasern in bzw. Entnahme aus bestimmten Fluiden.

Das Relaxationsverhalten nach Immersion in Ölen bspw., ist abhängig von der Konditionierungsvorgeschichte des Öls bzw. der Haarfaser. Wurde das Öl bei einer geringeren Luftfeuchte als die Haarfaser konditioniert, zeigen sich Längenkontraktionen und Zunahmen der Kraft (*overshoot*), die zum Halten einer 1 %-igen Dehnung aufgebracht werden muss ($F_{1\%}$). Diese erreicht relativ schnell ihr Maximum und fällt anschließend langsam auf das Ausgangsniveau ab [Gamez-Garcia, 1993]. Wurde das Haar dagegen bei einer geringeren Luftfeuchte als das Öl konditioniert, findet man Längenzunahmen und einen Rückgang von $F_{1\%}$ (*undershoot*), vgl. Abb. (2.13). Dieses Phänomen lässt sich auch in unterschiedlichen Salzlösungen beobachten, mit dem Unterschied, dass (i) ein Maximum deutlich schneller erreicht wird und (ii) $F_{1\%}$ nicht mit der Zeit abfällt, solange die Faser in der Lösung gehalten wird. Nach der Entnahme fällt $F_{1\%}$ innerhalb weniger Minuten um etwa 85 % ab. In einer zweiten, lang andauernden Phase, die sich über mehrere Stunden erstrecken kann und die von der Luftfeuchte beeinflusst wird, vollzieht sich der Rückgang auf das Ausgangsniveau. Der Übergangspunkt von der schnellen zur langsamen Phase der Wiederherstellung lässt sich physikalisch als Übergang von der FICKSchen zur *viskoelastischen Diffusion* deuten. Dieser *undershoot*-Effekt ist aus der Wollforschung wohlbekannt, vgl. [Rigby et al., 1974].

Von Interesse ist in diesem Zusammenhang der Index der Relaxation, also die Änderung der Kraft, die zum Halten einer konstanten Dehnung (1 %) nach einer bestimmte Zeitspanne noch aufgebracht werden muss. Der Index erreicht seinen Gleichgewichtszustand deutlich später als $F_{1\%}$. Diese zeitliche Differenz ist die Ursache für die gute Frisierbarkeit von Haaren nach ein bis zwei Stunden Trockenzeit.

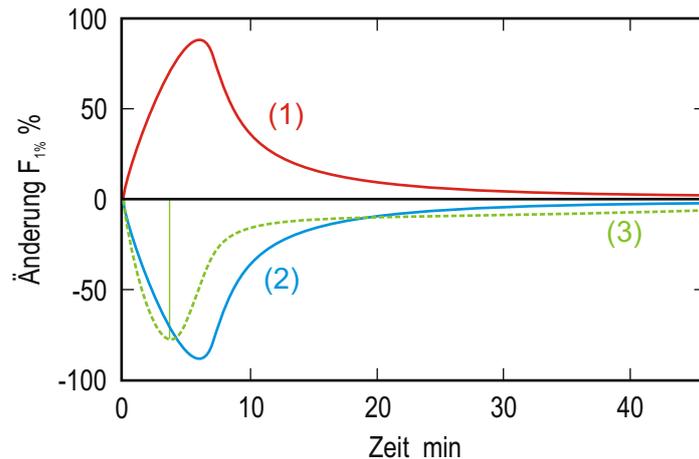


Abbildung 2.13: Overshoot/Undershoot-Effekt bei Immersion von Haarfasern nach Relaxation in unterschiedlichen Fluiden. (1) In Öl, bei geringerer rel. LF konditioniert als die Faser; (2) Öl, bei höherer rel. LF konditioniert als die Faser; (3) Nach Immersion und Entnahme aus der Salzlösung (senkrechter Strich), [Gamez-Garcia, 1993].

2.4.6 Einfluss der Temperatur

Die Ergebnisse des klassischen Zugversuchs zeigen eine deutliche Abhängigkeit von der Temperatur. Der Verlauf der Spannungs-Dehnungskurve wird mit steigender Temperatur flacher. Trägt man die Steigungen des HOOKESchen Bereichs für unterschiedliche Messungen gegen die Temperatur auf, zeigt die Kurve bei 0 °C und 75 °C einen Knick, vgl. Abb. (2.14) und [Sakai et al., 2000].

Die Übergänge vom HOOKESchen zum Fließbereich, sowie vom Fließbereich zum Nach-Fließbereich sind ebenfalls temperaturabhängig und weisen Steigungsänderungen bei 0 °C bzw. 80 °C auf.

Nagase et al. [Nagase et al., 2000] haben ein Modell zur Erklärung der wichtigsten physikalischen Eigenschaften von Haaren entwickelt, das aus zwei Komponenten mit je zwei Zuständen besteht. Der Übergang vom einen zum anderen Zustand ist gekennzeichnet durch eine charakteristische, stark vom Wassergehalt abhängige Transitionstemperatur. Sie beträgt in einem wässrigen Medium für die eine Komponente etwa 70 °C (hohe T_c^h) und für die andere etwa 0 °C (tiefe T_c^l). Die Werte verschieben sich bei zunehmendem Feuchtigkeitsgehalt nach oben.

Oberhalb dieser Temperaturen sind beide Komponenten plastisch, unterhalb fest und elastisch. Zu den strukturellen Bestandteilen der T_c^h -Komponente zählen Makrofibrillen und Exocuticula, zu denen der T_c^l -Komponente der Zellmembrankomplex, die Endocuticula und die Matrix. Die hohe Übergangstemperatur lässt sich mit der großen Anzahl an Disulfidbindungen erklären. Die Matrix, wie auch die Exocuticula, bestehen dagegen aus nicht keratinartigen Proteinen und bestimmten Lipiden mit deutlich weniger S-S Bindungen.

Auf physikalischer Ebene scheint ein Wechsel vom Gel zum Sol als Art des Übergangs wahrscheinlich. Im T_c^h -Bereich liegen überwiegend Disulfid-Bindungen, im Bereich der T_c^l vorwiegend Wasserstoffbrücken- bzw. ionische Bindungen vor. Mit zunehmender Temperatur werden einige der Bindungen aufgebrochen und das Polymernetz geht in einen leicht verformbaren Zu-

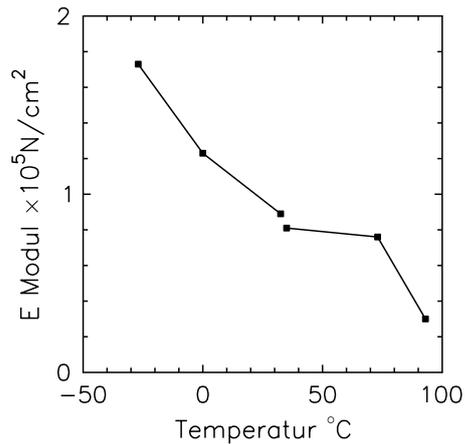


Abbildung 2.14: Abhängigkeit des E-Moduls von der Temperatur, [Sakai et al., 2000].

stand über. In diesem Zustand ist das Eindiffundieren von chemischen Substraten stark erleichtert. Der Übergang ist reversibel.

Das Modell für die mechanischen Eigenschaften von Haaren kann vereinfachend als System aus parallel geschalteten Federn und Dämpferelementen aufgefasst werden. Die T_c^l -Komponente wird durch ein einzelnes Dämpferelement abgebildet, da keine elastischen Anteile beobachtet wurden. Federn repräsentieren T_c^h -Komponenten, die α -Helices in den Mikrofibrillen bilden das MAXWELL-Element, vgl. Abb. (2.15).

Bei Temperaturen unterhalb T_c^l sind im HOOKEschen Bereich alle drei Elemente an der Verformung beteiligt. Bei Überschreitung dieses Übergangspunktes fällt der Einzeldämpfer weg, da die T_c^l -Komponente in den viskosen Sol-Zustand übergeht. Für $T > T_c^h$ fällt auch die Feder weg und das weitere Verhalten wird durch das MAXWELL-Element (α -Helices) bestimmt. Dieses Modell bezieht, im Gegensatz zum Ansatz von Feughelman [Feughelman, 1959], den mechanischen Beitrag der Substanz zwischen den Makrofibrillen und der Endocuticula mit in die Betrachtung ein.

Das Settingverhältnis bei der Herstellung von Wasserwellocken (tiefe Relaxationstemperatur) ist stark von der Behandlungstemperatur abhängig. Liegt diese oberhalb von T_c^h , ist infolge der Plastizität beider Komponenten eine bleibende Verformung möglich. Bei Werten unterhalb von T_c^l ist der Zustand elastisch, eine bleibende Verformung wird verhindert. Im Bereich zwischen den beiden Übergangstemperaturen liegen lediglich die strukturellen Elemente der T_c^l im plastischen Zustand vor, die der T_c^h bleiben elastisch. Die Stabilität ist mittelmäßig und bleibt bei Werten $< T_c^h$ ungefähr konstant. Der Einfluss der Relaxationstemperatur besteht darin, dass die Deformation gleichsam eingefroren wird. Liegt sie über T_c^l , ist der Zustand plastisch und nur die T_c^h -Komponente ist für die Formgebung verantwortlich.

Relaxation von Wasserwellocken in unterschiedlichen Lösungsmitteln zeigen, dass der Wert der Dielektrizitätskonstante ϵ über die Haltbarkeit entscheidet. $\epsilon < 20$ führt zu besserer Haltbarkeit infolge einer Anhebung von T_c^l .

Der Wert von T_c^l hängt stark vom Feuchtigkeitsgehalt der Haare ab. Bei nassen Haaren sinkt er in der Regel unter Raumtemperatur, so dass ein Setting bequem möglich ist. Nach dem Trocknen steigt er wieder und die Form bleibt erhalten. Aufgrund dieses Effektes verlieren Lockenwicklerfrisuren bei steigender Luftfeuchte schnell ihre Form.

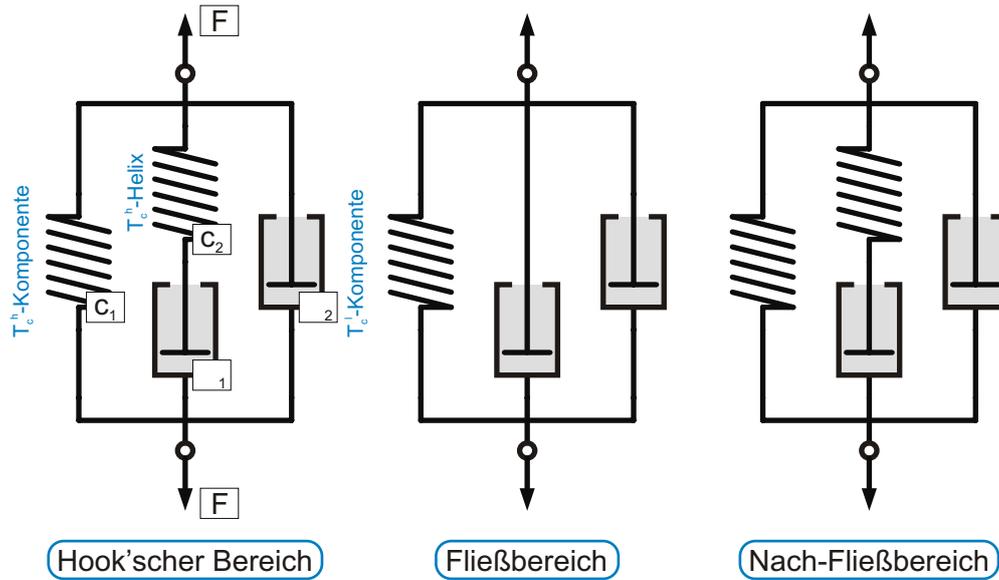


Abbildung 2.15: Strukturelles Modell für die mechanischen Eigenschaften von Haaren, [Nagase et al., 2000]. Die Feder repräsentiert die T_c^h -Komponenten, das Dämpferelement die T_c^l -Komponenten, das kombinierte Feder-Dämpfer-Element die α -Helices in den Mikrofibrillen. Die Aktivierung der drei Elemente ist von der Temperatur abhängig.

Chemische Substanzen wie Färbemittel können mit steigender Temperatur einfacher in die Haarstrukturen diffundieren. Der Diffusionskoeffizient steigt mit dem Erreichen von T_c^h .

2.4.7 Biegung

Die Biegung der Haarfaser stellt eine wichtige Deformationsgröße dar. Dabei entsteht im äußeren Bereich der Krümmung eine Zone maximaler Dehnung, die zur Krümmungsinne Seite hin über die neutrale Faser in die Kompressionszone übergeht. Die neutrale Faser verläuft durch die Schwerpunkte der Querschnitte. Der vom Haar der Biegung entgegengebrachte Widerstand ist die Biegesteifigkeit $E_B I$, wobei I das Flächenträgheitsmoment darstellt ($I_y = \int_A x^2 dA$, x ist der Abstand von der neutralen Faser). Vom Standpunkt der Elastizitätstheorie her lassen sich Dehnung und Kontraktion beim ideal elastischen Material durch den E-Modul beschreiben. Die Biegung kann im Bereich kleiner Verformungen in Analogie zur Biegung von Stäben vereinfacht durch eine Differentialgleichung vierter Ordnung beschrieben werden,

$$EI_2 \frac{d^4 x}{dz^4} - Q_x = 0, \quad EI_1 \frac{d^4 y}{dz^4} - Q_y = 0,$$

wobei

$$Q = -\frac{dF}{dl},$$

also die Änderung der Kraft entlang der Stabachse im deformierten und z die Stabachse im undeformierten Zustand ist. I_1 bzw. I_2 sind die Hauptflächenträgheitsmomente, die Richtungen von

Zugehörigkeit	I_2/I_1	I_1/I_1^w
West-Europäer (w)	2,09	1,00
Chinesen	1,51	2,85
Inder	1,95	1,83
Afrikaner	2,82	1,32

Tabelle 2.7: Steifigkeit der Haarfasern von Personen unterschiedlicher ethnischer Zugehörigkeiten, [Swift, 1995].

x bzw. y fallen mit den Hauptträgheitsachsen zusammen. Kleine Verformung heißt in diesem Zusammenhang, dass die Ableitung der Tangente $d\mathbf{T}/dl$ entlang des Stabes klein ist.

Haarfasern neigen zu viskoser Verformung und verfügen keineswegs über einen homogenen Querschnitt, so dass der Modul der Biegung in der Regel vom E-Modul, der aus Zugversuchen bekannt ist, abweicht.

Aufgrund der elliptischen Querschnittsgeometrie zeigt das Haar in Richtung der beiden Hauptachsen unterschiedliche Biegesteifigkeiten. Die Flächenträgheitsmomente berechnen sich dabei zu

$$I_1 = \frac{1}{4} \cdot \pi \cdot a \cdot b^3, \quad I_2 = \frac{1}{4} \cdot \pi \cdot a^3 \cdot b,$$

wobei a und b der große bzw. der kleine Halbmesser der Ellipse sind. Die Biegung findet bevorzugt in Richtung des kleinen Halbmessers statt. Eine Vergrößerung des großen Halbmessers um 25 % hätte einen Anstieg der Biegesteifigkeit um 25 %, eine Zunahme des kleinen Halbmessers um denselben Wert fast eine Verdopplung der Steifigkeit zur Folge.

Mit den in Tab. (2.2) gegebenen Haardurchmessern unterschiedlicher ethnischer Zugehörigkeiten errechnet man als Verhältnis I_2/I_1 Werte zwischen 1,5 und 2,8, vgl. Tab. (2.7). Bei afrikanischen Haarfasern kann die Biegesteifigkeit in der Richtung des großen Halbmessers folglich bis zu dreimal so groß sein, wie senkrecht dazu. Ob sich dieser theoretische Aspekt in irgendeiner Weise auf das makroskopische Erscheinungsbild auswirkt, lässt sich nicht beurteilen. Allerdings scheinen sich Parameter wie Geometrie und Biegesteifigkeit durchaus in der Haarmechanik bemerkbar zu machen. Von Swift [Swift, 1995] wird in diesem Zusammenhang auf das *Syndrom nicht kämmbarer Haare* verwiesen, bei dem man eine dreieckige Querschnittsgeometrie als Ursache annimmt. Das Verhältnis der Biegesteifigkeiten von Dreieck zu Ellipse beträgt etwa $1,20a/b$. Das Phänomen kann aber nicht alleine aus dieser Tatsache heraus erklärt werden.

Interessanterweise wird der Cuticula im Gegensatz zum Zugverhalten ein Einfluss auf das Biegeverhalten zugesprochen, der sich anteilmäßig zwischen 40 % und 50 % bewegen kann. Es ist davon auszugehen, dass Agenzien, welche die cuticuläre Struktur zu beeinflussen vermögen, damit indirekt auch das Biegeverhalten beeinflussen.

Scott und Robbins [Scott & Robbins, 1978] haben eine spezielle Methode zur Messung des Biegeverhaltens von Haaren entwickelt. An den Enden des einzelnen Haares werden Gewichte von 0,1 g fixiert und das Haar anschließend über einem Haken ausbalanciert. Über den sich einstellenden Abstand d der beiden Schenkel kann der E-Modul zu

$$E_B = \frac{\pi \cdot T \cdot d^2}{2 \cdot A^2}$$

ermittelt werden, wobei T die an den Enden ausgeübte Kraft ist. Diese Methode liefert näherungsweise die gleichen Werte wie der klassische Zugversuch (vgl. Tab. (2.8)), solange ein kreis-runder

Tabelle 2.8: E-Modul aus Biege- und Zugversuchen unter Annahme eines kreisrunden Querschnitts, 60 % rel. LF, 75 °F. K = Süd-Koreanisches Haar; H, L = weibl., kaukasisch, 15 bzw. 12 Jahre, [Scott & Robbins, 1978].

Probe	lin. Dichte [$\mu\text{g}/\text{cm}$]	E_B [$10^5 \text{ N}/\text{cm}^2$]	E_S [$10^5 \text{ N}/\text{cm}^2$]	E_B/E_S [-]
K	99,50	4,23	3,68	1,15
L	94,90	3,54	3,82	0,93
K	89,20	4,29	3,43	1,25
L	71,80	4,25	3,83	1,11
H	69,20	4,11	3,75	1,10
L	67,70	3,35	3,96	0,85
L	54,60	3,60	4,12	0,88
L	52,90	4,69	3,98	1,18
H	52,60	3,74	4,03	0,93
H	42,30	3,23	4,21	0,77
L	34,40	2,89	4,33	0,67
H	31,30	3,58	3,59	1,00
\bar{x}	63,40	3,79	3,89	0,97
Δx		13,90	6,70	

Querschnitt angenommen wird. Der Abstand d der Schenkel ist direkt proportional zur Steigung des HOOKEschen Bereiches. Auf der Basis der Werte aus [Scott & Robbins, 1978] (Abb. 5) wurde E_B für drei Fasern von unterschiedlicher linearer Dichte berechnet. Der Verlauf von E_B bei steigender Luftfeuchte ist in Abb. (2.16) dargestellt.

Die Biegesteifigkeit nimmt mit steigender Luftfeuchte ab. Dickere Fasern zeigen eine größere Sensitivität gegenüber Änderungen der Luftfeuchte.

2.4.8 Torsion

Haare erfahren bei natürlicher Beanspruchung durch Wind, Kämmen, Bürsten etc. neben Biegung, auch eine Verdrehung oder Torsion des Faserquerschnitts. Dauerwelle oder der Einsatz von Lockenwicklern verleiht Haarsträhnen eine Spiralförmigkeit. Die Beanspruchung von Haarlocken durch externe Kräfte, bspw. Gravitation, führt zu einer Verformung, bei der neben Biegung dann auch eine Verdrehung des Faserquerschnitts auftritt. Der Widerstand, den der Querschnitt der Verdrehung entgegensetzt, wird als Torsionssteifigkeit bezeichnet. Die Starrheit hingegen ist das Drehmoment, das erforderlich ist, um eine Umdrehung pro cm Länge herbeizuführen. Diese Größe spielt eine wichtige Rolle in der Kosmetikindustrie. Messverfahren zur Bestimmung des Torsionsverhaltens von textilen Fasern basieren auf dem Prinzip des Torsionspendels. Eine Last am Ende der Faser oszilliert mit der Periode T und definiert den Torsions- oder Schubmodul durch

$$G = \frac{128 \cdot \pi \cdot I \cdot l}{T^2 \cdot d^4}. \quad (2.11)$$

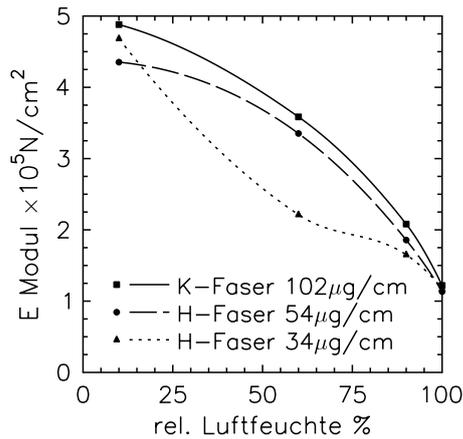


Abbildung 2.16: Änderung des E-Moduls der Biegung E_B bei Fasern unterschiedlicher linearer Dichte, berechnet auf der Basis der Werte aus [Scott & Robbins, 1978].

Dabei ist l die Länge, d der Durchmesser der Faser, I das Trägheitsmoment des Pendels. Daraus ermittelt man den Widerstand gegen Verdrehen, die so genannte Starrheit Γ , als

$$\Gamma = \frac{8 \cdot \pi^3 \cdot I \cdot l}{T^2} \quad (2.12)$$

Starrheit und Torsionsmodul stehen über den Querschnitt A der Faser durch

$$G = \frac{\Gamma}{A^2} \quad (2.13)$$

in Beziehung. Da die Amplitude der Torsionsschwingung mit der Zeit abnimmt, führt man das *logarithmische Dekrement*

$$\delta = \frac{1}{n} \ln \frac{a_1}{a_n} \quad (2.14)$$

als weitere charakteristische Größe ein. Dabei ist n die Anzahl der Oszillationen, a_1 und a_n die Amplitude der ersten bzw. letzten Oszillation. Diese Größe beschreibt die Torsionselastizität der Faser. Je größer die Dämpfung, desto weniger elastisch ist die Faser bzw. desto größer die interne Viskosität. Für $\delta = 0$ liegt ein ideal elastisches Verhalten vor. Man kann in Analogie zur Biegung den Modul G' einführen, der den Torsionsverlust eines Systems beschreibt. Er steht mit dem logarithmischen Dekrement über $G' = \delta G / \pi$ in Beziehung. In Tab. (2.9) sind Standardwerte für den E-Modul und den Torsionsmodul aufgeführt.

Feuchtigkeit beeinflusst die Torsionseigenschaften in stärkerem Maße als die Dehnungseigenschaften. Während der E-Modul der Dehnung lediglich um den Faktor 2 abnimmt, verzeichnet man beim Torsionsmodul bereits eine Abnahme um den Faktor 4, wenn die rel. Luftfeuchte von

Modul	[10 ⁵ N/cm ²]	E _{65%} /E _{100%}
E_S	3,89	2,62
E_B	3,79	2,40
G	0,89	4,10

Tabelle 2.9: Dehnungs-, Bieigungs- und Torsionsmodul von Haaren bei 60-65 % rel. Luftfeuchte, in 10⁵ N/cm² [Robbins, 2002].

Tabelle 2.10: Werte des Torsionsmoduls von Haaren nach unterschiedlicher mechanischer bzw. chemischer Behandlung bei 65 % rel. Luftfeuchte, in 10^5 N/cm^2 [Wolfram & Albrecht, 1985].

Behandlung	G [10^5 N/cm^2]	$\Gamma_{65\%}/\Gamma_{100\%}$ [-]	δ [-]
Unbehandelt	1,02	0,26	0,40
Gebleicht	1,05	0,14	0,44
Gefärbt (blond)	1,07	0,23	0,42
Gefärbt (schwarz)	1,08	0,23	0,45
Dauerwelle	1,01	0,09	0,56
Relaxed	0,91	0,06	0,65

65 % auf 100 % ansteigt, vgl. [Bogaty, 1967]. Das logarithmische Dekrement zeigt eine deutliche Zunahme bei steigender Luftfeuchte. Das macht es verständlich, dass Haare sich im nassen Zustand leichter frisieren lassen und ihren Zustand nach dem Trocknen beibehalten. Feuchtigkeit macht Haare plastisch formbar, so dass ihnen mit Lockenwicklern eine neue Form verliehen werden kann. Im trockenen Zustand nimmt die Steifigkeit der Haarfaser ihren ursprünglichen Wert an. Das Setting ohne Verwendung chemischer Substanzen führt zu Frisuren, die ihre Form bei steigender Luftfeuchte schnell wieder verlieren. Einige Werte nach diversen Behandlungen sind in Tab. (2.10) angegeben.

Bei dauergewelltem oder chemisch reduziertem Haar findet man bei Luftfeuchten unterhalb von 90 % eine geringfügig erhöhte, oberhalb davon eine etwas kleinere Steifigkeit, als bei nicht modifiziertem Haar, vgl. Abb. (2.17), links. Zudem ist die plastische Formbarkeit etwas größer, als vor einer solchen Behandlung.

Phillips [Phillips, 1987] hat den Einfluss des Alterungsprozesses auf die Torsionseigenschaften von Wollfasern untersucht. Unter dem Alterungsprozess versteht man die zeitliche Entwicklung der mechanischen Eigenschaften von Keratinfasern nach dem Entfernen aus einer wassergesättigten Umgebung oder einem Wasserbad. Die anschließende Konditionierung bei definierter Luftfeuchte ist ein dynamischer Prozess, bei dem sich allmählich der Gleichgewichtswassergehalt einstellt. Das ist dann der Fall, wenn die Verdunstungsrate mit der Absorptionsrate im Gleichgewicht steht. Der Begriff der Alterung rührt daher, dass die mechanischen Eigenschaften von Keratinfasern, die unter konstanten Bedingungen gehalten werden, über lange Zeiträume (im Bereich von Monaten) einer stetigen Änderung unterliegen. Ein Durchnässen der Faser hat die Wiederherstellung ihrer ursprünglichen Eigenschaften zur Folge. Mit dem Alterungsprozess ist u.a. eine Abnahme der Dämpfung gegenüber dynamischer Beanspruchung auf Torsion bzw. eine Zunahme der Starrheit verbunden. Dieser Prozess kann sich unter ungestörten Bedingungen über Jahre erstrecken, allerdings mit immer kleiner werdender Abnahmerate, vgl. Abb. (2.17), rechts.

Die Plausibilität der angegebenen Werte lässt sich über den Zusammenhang zwischen Torsionsmodul und E-Modul $G = E/[2(1 + \nu)]$ überprüfen. Dabei ist ν die Poisson-Zahl [Fung, 1993]. Mit den Werten aus Tab. (2.9) findet man kaum Übereinstimmungen. Feughelman [Feughelman, 1959] wendet diese Beziehung mit $\nu = 0,35-0,5$ auf Wolle an. Der Torsionsmodul sollte folglich etwa $G = E/3$ betragen, was bei Wolle zumindest für trockene Fasern mit einer Toleranz von $\approx 10\%$ zutrifft.

Anders als bei der Zugbelastung, leisten die äußeren Schichten der Cuticula einen wesentlichen Beitrag zur Torsionssteifigkeit. Über die Veränderung der Torsionseigenschaften

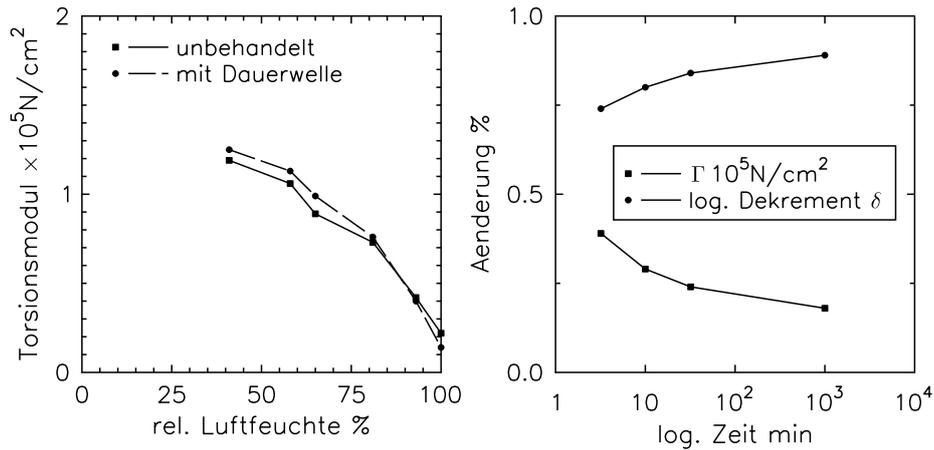


Abbildung 2.17: Links: Änderung des Torsionsmoduls in Abhängigkeit von der rel. Luftfeuchte bei Haaren mit und ohne Dauerwelle, [Bogaty, 1967]. Rechts: Änderung der Torsionseigenschaften von Wolle in Abhängigkeit von der Zeit. Starrheit Γ und logarithmisches Dekrement δ . Bei 30 °C, 65 % rel. Luftfeuchte, [Phillips, 1987].

kann auf etwaige Schädigungen der Cuticula geschlossen werden. Wolfram und Albrecht [Wolfram & Albrecht, 1985] geben für den Torsionsmodul im nassen Zustand einen Wert von $0,18 \cdot 10^5 \text{ N/cm}^2$ an. Der Torsionsmodul für die Cuticula (nass) kann nach folgender Formel abgeschätzt werden,

$$G_{\text{Cut}} = \frac{G \cdot d_2^4 - G_{\text{Cort}} \cdot d_1^4}{d_2^4 - d_1^4}.$$

d_1 ist der Durchmesser des Faserkerns, ohne Cuticula und d_2 der Gesamtdurchmesser. G_{Cort} ist der Torsionsmodul des Cortex ($0,24 \cdot 10^5 \text{ N/cm}^2$). Daraus berechnet sich der Wert G_{Cut} zu $0,012 \cdot 10^5 \text{ N/cm}^2$.

2.4.9 Reibung

An den Berührflächen von Körpern treten Kräfte dissipativen Charakters auf, die Reibungskräfte. Während die mikroskopischen Mechanismen an den Interaktionsflächen zweier Körper nach wie vor weitgehend unverstanden sind – man postuliert hier mit lokal plastischen Verformungen einhergehende Ausbildung von chemischen Bindungen, zu deren Überwindung dann Scherkräfte notwendig sind – beschreiben phänomenologische Gesetze, wie jenes von COULOMB, das durchschnittliche mechanische Verhalten an der Kontaktfläche rein approximativ.

Die COULOMBSche Reibungskraft, die der Bewegung eines Körpers auf einer festen Unterlage ohne Schmiermittel (Trockenreibung) entgegenwirkt, ist gegeben durch

$$F_R = \mu_K \cdot F_N,$$

wobei F_N die Kraft normal zur Kontaktfläche darstellt. Der Körper haftet hingegen, solange die Kraft die folgende Bedingung erfüllt,

$$F_R \leq \mu_S \cdot F_N.$$

Dabei ist μ_S der Koeffizient der Haftreibung und μ_K der Gleitreibungskoeffizient. Diese Beiwerte setzen die zum Einleiten bzw. Aufrechterhalten der Bewegung erforderliche Kraft mit der Anpresskraft in Beziehung. Es gilt: $\mu_S \geq \mu_K$. Das heißt, dass die Kraft, die den Körper in Bewegung setzt, größer ist, als jene die sich seiner Bewegung, wenn sie einmal in Gang gekommen ist, entgegensetzt. Das führt zu den bekannten „stotternden“ Bewegungsabläufen von Körpern, die bspw. an einer Feder über eine raue Unterlage gezogen werden.

Bei Anwesenheit von Schmiermitteln gilt das Gesetz der COULOMB-Reibung nicht. Ausnahmen bilden die Fälle, in denen die Substanzen gleichmäßig über die Reibungsflächen verteilt sind. Morton und Hearle weisen darauf hin [Morton & Hearle, 1997], dass Schmiermittel nur einen geringen Einfluss auf die Reibung zwischen Fasern haben und den Reibungskoeffizienten selten um mehr als 0,2 reduzieren. Feuchtigkeit bspw. infolge Luftfeuchte, sowie die Anpresskraft zwischen zwei Berührflächen üben deutlichen Einfluss auf das Reibungsverhalten aus. Der Theorie nach sollte der Reibungskoeffizient unabhängig von der Größe der Belastung normal zur Reibungsfläche sein. In der Praxis zeigen sich aber gerade im Bereich kleiner Belastungen bei der Faserreibung deutliche Zunahmen von μ_K . Ein weiterer Punkt ist die oft vereinfachend angenommene Unabhängigkeit des Reibungsbeiwertes von der Gleitgeschwindigkeit. Morton und Hearle weisen darauf hin [Morton & Hearle, 1997], dass diese Annahme genau genommen nur für kleine Geschwindigkeiten im Bereich von 2-90 cm/min zulässig ist. Bei höheren Geschwindigkeiten muss mit einer dramatischen Zunahme des Reibungskoeffizienten gerechnet werden. Wir werden später im Zusammenhang mit der Modellierung von Reibung COULOMBS Gesetz noch genauer untersuchen.

Bei der Berührung zweier gekreuzter Fasern bildet sich an der Kontaktstelle eine kreisförmige Abplattung aus, deren Durchmesser d_k mit der Anpresskraft über den empirischen Ausdruck

$$F_N \propto d_k^{2,7}$$

in Beziehung steht. Daraus lässt sich für den Reibungskoeffizienten die Gleichung

$$\mu_S = F_R/F_N = kF_N^{-0,26} d^{0,52}$$

ableiten, wobei k eine Konstante ist, vgl. [Morton & Hearle, 1997].

Die experimentelle Ermittlung von Reibungskräften bei Berührung von Haaren mit Materialien gestaltet sich einfacher als zwischen Haaren untereinander. Der Koeffizient der kinetischen Reibung kann nach [Robbins, 2002] oder [Morton & Hearle, 1997] durch Spannungsmessungen an einer Haarfaser, die man über einen Dorn mit Kreisquerschnitt und definierter Dicke laufen lässt und an deren Ende eine kleine Zugkraft T_2 ausgeübt wird, näherungsweise durch Anwendung von AMONTONS Gesetz bestimmt werden zu

$$\mu_K = \frac{1}{\theta} \ln \frac{T_2}{T_1}.$$

θ ist der Winkel des berührenden Bogensegmentes in Radiant und T_1 bzw. T_2 sind die Zugspannungen, die man vor und nach dem Passieren des Dorns misst. Bei kleinem Reibungskoeffizienten ändert sich das Verhältnis T_2/T_1 bei einem Kontaktwinkel von $\pi/2$ bis 4π (zwei volle Umwicklungen) nur unwesentlich. Die Exponentialfunktion macht sich jedoch mit steigendem Reibungsbeiwert nachteilig bemerkbar. Die abgehende Zugspannung T_2 steigt exponentiell, was bspw. bei

Material Behandlung	Richtung #	μ_S		μ_K	
		trocken	nass	trocken	nass
Haar/Haar	1			0,15	0,340
<i>unbehandelt</i>	1				0,250
<i>Dauerwelle 2×</i>	1				0,310
<i>+Bleichen</i>	1				0,490
<i>ABS gewaschen</i>	1		0,30		0,350
<i>+Pflegespülung</i>	1		0,20		0,260
<i>+ABS gewaschen 1×</i>	1		0,16		0,190
<i>+ABS gewaschen 3×</i>	1		0,29		0,360
Haar/Hartgummi	2			0,19	0,380
<i>Shampoo</i>	1				0,425
	2				0,546
<i>Pflegespülung</i>	1				0,293
	2				0,475
Haar/Nylon	1			0,14	0,220
Haar/Aluminium	1			0,12	0,180
Haar/Polyethylen	1			0,22	0,290
Haar/Glas	1				1,400
Wolle/Wolle	1	0,13		0,11	
	2	0,61		0,38	
	3	0,21		0,15	
Wolle/Rayon	1	0,11		0,09	
	2	0,39		0,35	
Wolle auf Nylon	1	0,26		0,21	
	2	0,43		0,35	

Tabelle 2.11: Statischer μ_S und dynamischer Reibungskoeffizient μ_K für die Interaktion von trockenen und nassen Haaren mit unterschiedlichen Materialien. Dabei haben die Richtungswerte die folgende Bedeutung: 1 → Bewegung mit den Schuppen, 2 → Bewegung gegen die Schuppen und 3 → gleiche Faserrichtung (bei Faser-Faser Reibung), vgl. [Schwartz & Knowles, 1963, Morton & Hearle, 1997, Robbins, 2002].

Garnen, die in Textilverarbeitungsmaschinen über Spulen abgewickelt werden, zu Materialversagen führen kann.

In Tab. (2.11) sind einige Reibungsbeiwerte für Keratinfasern bei der Interaktion mit unterschiedlichen Materialien aufgeführt.

Die Ermittlung des statischen Reibungskoeffizienten μ_S kann über den Gleitwinkel einer kleinen Haarschleife bestimmt werden, die auf zwei parallelen Haarfasern aufgesetzt und durch eine Druckkraft belastet wird. Hier gilt, $F = A \cdot S$, mit der Kontaktfläche A und der Scherfestigkeit S der berührenden Materialien.

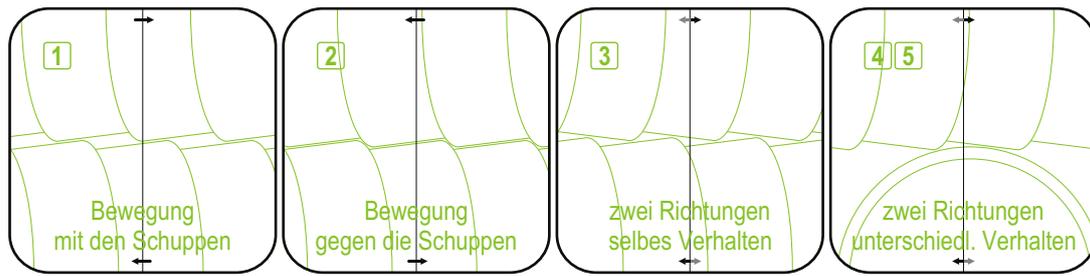


Abbildung 2.18: Der *Differential Friction Effect* beschreibt das Phänomen der Richtungsabhängigkeit der Reibung bei Haaren. Hier können prinzipiell fünf Fälle unterschieden werden, [Schwartz & Knowles, 1963].

Die Reibung zwischen den Haarfäsern ist bedingt durch die morphologische Struktur der Haaroberfläche anisotrop, also von der Richtung abhängig. Dieses Phänomen wird als *Differential Friction Effect* bezeichnet und ist seit langem bekannt, vgl. [Adams et al., 1990]. Schwartz und Knowles [Schwartz & Knowles, 1963] unterscheiden bezüglich der relativen Reibungsrichtung von Haarfäsern untereinander fünf Fälle, vgl. Abb. (2.18). Fall (1) zeigt die Reibung zweier paralleler, aber gegensätzlich gerichteter Haarfäsern. Die Relativbewegung verlaufe antiparallel zur Wachstumsrichtung, also mit den Schuppen. Fall (2) zeigt die gleiche Ausgangssituation, wie im ersten Fall. Allerdings verlaufe die Bewegungsrichtung jetzt parallel zur Wachstumsrichtung, also gegen die Schuppen. Diese blockieren, so dass die Reibung offensichtlich größer sein muss, als im ersten Fall. Fall (3) zeigt zwei parallele, jedoch gleich gerichtete Fasern. Die Relativbewegung ist hier unerheblich, denn es reiben nur die Schuppen jeweils einer Haarfaser. Die Reibung liegt irgendwo zwischen denen aus Fall (1) und Fall (2). Fall (4) und (5) zeigen die zwei möglichen Bewegungsrichtungen zweier senkrecht zueinander stehender Haarfäsern. Für beide Fälle setzt man unterschiedliche Reibungsbeiwerte an, weil die Bewegung einmal mit und einmal gegen die Schuppen erfolgen kann. Alle Reibungsfälle lassen sich auf diese fünf Diagramme zurückführen. Jede dazwischenliegende Bewegung kann in die Hauptrichtungen zerlegt werden.

Dass die Reibung in Fall (2) am größten ist, folgt nicht notwendigerweise aus einer oberflächlichen Betrachtung der Skizzen, sondern ist vielmehr die Konsequenz einer genauen Analyse der lokalen Kontaktverhältnisse unter Beachtung von Kontaktwinkelhäufigkeiten und der Größe der Kontaktflächen. Eine detaillierte Analyse findet man in Morton und Hearle [Morton & Hearle, 1997].

Die Richtungsabhängigkeit der Reibung kann durch Shampoos und Pflegespülungen beeinflusst werden, vgl. siehe Tab. (2.11). Der Effekt von Shampoos ist oftmals kumulativ, so dass mehrfaches Waschen zu einer schrittweisen Beeinflussung der Reibungsbeiwerte führt. Den Effekt einer solchen sukzessiven Behandlung kann man in Tab. (2.11) sehen. Die Bildung natürlicher Fette setzt die Reibung herab. Talg füllt die Zwischenräume benachbarter Schuppen der Cuticula aus, so dass die Rauigkeit abnimmt [Corcuff et al., 1993]. Generell wird die Reibung mit zunehmendem Faserdurchmesser größer. Änderungen der rel. Luftfeuchte, der Temperatur sowie das Bleichen oder die Dauerwelle beeinflussen das Reibungsverhalten von Haaren mehr oder weniger stark.

2.4.10 Elektrostatische Effekte

Mechanische Belastungen, die mit dem Zusammenpressen von Haarfasern einhergehen, führen infolge unterschiedlicher Elektronenaffinitäten zu Ladungstrennungen und folglich zur Ausbildung von Reibungs- bzw. Piezoelektrizität. Wie schnell die Ladungen abgebaut werden, hängt vom elektrischen Widerstand ab. Haare begünstigen infolge ihres hohen elektrischen Widerstandes die Ausbildung elektrostatischer Ladungen. Die Konsequenz ist das so genannte *Flyaway Hair*, das man besonders gut bei frisch gewaschenem Haar beobachten kann. Hierbei spielt der Haartyp eine wichtige Rolle. Je glatter das Haar, desto größer ist die Tendenz zum Abstehen der Fasern. Mit zunehmender Welligkeit hindern sich die Fasern durch den höheren Grad der Verflechtung gegenseitig am Austreten aus den Haarschichten [Robbins, 2002].

Die Richtungsabhängigkeit der Reibung hat einen interessanten Effekt zur Folge. Zieht man aus einem Bündel gleich orientierter Haare eine einzelne Faser heraus, so bildet sich auf der Faser entweder eine positive oder eine negative Ladung aus, je nachdem, ob die Faser am unteren Ende oder an der Spitze herausgezogen wird. Ist die Faser den anderen entgegengerichtet, bildet sich keine Ladung aus.

Die Zunahme der Luftfeuchte geht mit der Abnahme des elektrischen Widerstandes bzw. einer Zunahme der Leitfähigkeit einher. Der Widerstand von Haaren beträgt bei 85 % rel. Luftfeuchte etwa $R=1,0-1,7 \cdot 10^{13} \Omega$. Feuchtes Haar zeigt eine deutlich geringere Tendenz zur Ladungsbildung. Da das Haarkeratin stark hygroskopisch ist, kann der Widerstand durch die Elektrolytkonzentration beeinflusst werden. Shampoos oder Pflegespülungen verringern zum einen den elektrischen Widerstand, zum anderen die Reibung, so dass weniger Kämmarbeit geleistet werden muss und infolge dessen weniger Ladung aufgebaut wird. Bei Widerständen unterhalb von $R=10^8 \Omega$ verteilt sich die Ladung zudem besser über die Haaroberfläche. Temperaturänderungen zeigen ähnliche Wirkungen, ein Anstieg der Temperatur um 10°C führt zu einer Abnahme des Widerstandes um das Fünffache.

2.4.11 Kosmetische Behandlung

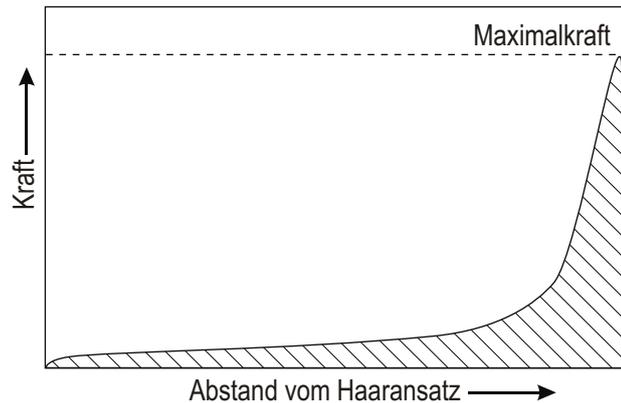
Typische kosmetische Behandlungen der Haare umfassen das Kämmen und Bürsten, das Bleichen, Färben, die Dauerwelle, das Waschen mit Shampoos und Pflegespülungen, sowie die Applikation von Haarsprays, Haarlacken, Haargelen oder anderen Pflegeprodukten. Der Einfluss auf die physikalischen Eigenschaften, insbesondere die mechanischen, ist unterschiedlich stark ausgeprägt. Die Erforschung der Wirkung von Kosmetika auf Haut und Haar stehen im Mittelpunkt des Interesses der kosmetischen Industrie und sind von daher eingehend untersucht worden.

2.4.11.1 Kämmen und Bürsten

Robbins und Reich [Robbins & Reich, 1986] haben den Einfluss globaler Haareigenschaften auf die Kämmbarkeit von Haaren untersucht. Diese kann als Funktion von statischer und dynamischer Reibung (F^s , F^k), Steifigkeit (S), Welligkeit (C), Durchmesser (d) und elektrischer Ladung (Q) aufgefasst werden,

$$K = f(F^s, F^k, S, C, d, Q).$$

Abbildung 2.19: Verlauf der Kämmkraft mit Maximum im Bereich der Haarspitzen, [Yin et al., 1976].



Adhäsive Effekte, die über gewöhnliche Reibung hinausgehen, bleiben hier unberücksichtigt. Das Kämmverhalten kann durch die zu überwindende Maximalkraft bzw. die zu verrichtende Arbeit charakterisiert werden. Die Kämmbarkeit bei Haaren kaukasischen oder orientalischen Ursprungs ist im nassen Zustand größer als im trockenen. Bei krausen Haaren verhält es sich umgekehrt. Afrikanisches Haar ist infolge seines in axialer Richtung ständig wechselnden Durchmessers und seiner starken Torsion etwa fünfmal schwerer zu kämmen, als kaukasisches Haar [Kassenbeck, 1999]. Haupteinflussfaktoren sind dabei die Faserkrümmung und die Reibung. Größere Steifigkeit führt zu besserer Kämmbarkeit. Der Durchmesser hat hingegen keinen nennenswerten Einfluss auf Maximalkraft und Arbeit. Der Einfluss einer Behandlung auf die Kämmbarkeit lässt sich durch einen linearen Ausdruck abschätzen [Robbins & Reich, 1986],

$$K = -N_1 \cdot F^k - N_2 \cdot F^s - N_3 \cdot Q - N_4 \cdot C + n_1 \cdot S + n_2 \cdot d.$$

Dabei sind die N_i Konstanten. Der Beitrag der Größen mit kleinem n_i zur Kämmbarkeit ist gering. Beim Kämmen oder Bürsten der Haare ist der größte Widerstand im Bereich der Spitzen zu überwinden, vgl. Abb. (2.19).

Bedingt durch Witterungseinflüsse weisen die Fasern im Bereich der Spitzen die größte Rauigkeit auf. In solchen Bereichen stehen mitunter bis zu 30 einzelne Fasern in Kontakt [Swift, 1997]. Es kommt zur Ausbildung von *Haarfilzen*. Bei langsam ansteigenden Kämmkräften gleiten die Haare übereinander hinweg. Bei großer ruckartiger Belastung dagegen stellen sich die Fasern teilweise senkrecht und werden übereinander hinweg gebogen. In dieser Konfiguration erfährt die Faser eine Scherbeanspruchung in axialer Richtung, die sich parabolisch über den kleinen Halbmesser verteilt. An der äußeren Enden des Halbmessers verschwinden die Scherspannungen, in der Ebene des großen Halbmessers werden sie dagegen maximal. Ihr Wert wird größer, wenn der kleine Halbmesser oder die Krümmung zunehmen. Bei andauernder Belastung wird die Ausbildung von Haarspliß begünstigt, der sich in Richtung Haarspitze fortpflanzt. Das Herabsetzen des Reibungsbeiwertes der Haare durch Applikation von schmierenden Substanzen oder die Erhöhung der Plastizität durch Feuchtigkeit können die Auswirkungen solcher Scherbeanspruchungen vermindern.

Die Kraft, die beim Bürsten der Haare aufgewendet werden muss, nimmt offensichtlich mit der Faserdicke zu [Yin et al., 1976]. Diese Beobachtung widerspricht der Theorie. Demnach sollten infolge der zunehmenden Biegesteifigkeit die von den Zähnen des Kammes oder der Bürste aus-

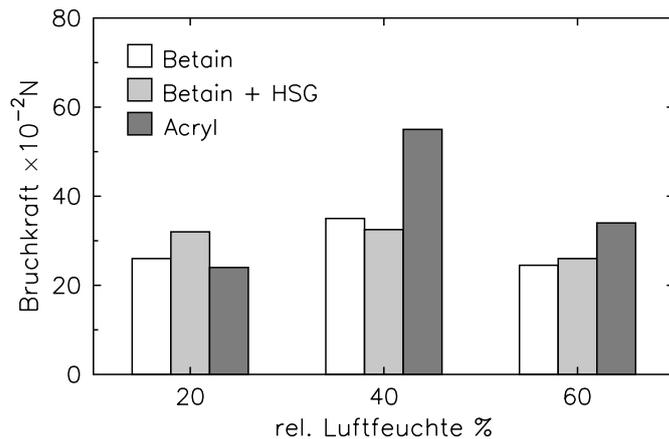


Abbildung 2.20: Einfluss der rel. Luftfeuchte aus Bruchkraft von Haarsprayverbindungen basierend auf unterschiedlichen Polymeren, [Wickett et al., 1992].

geübten Belastungen zu einer Spreizung oder zu einem Auseinanderdrücken der Fasern führen, welche der Ursache vorwegläuft. Die aufzubringende Kraft sollte also mit zunehmender Dicke der Haare kleiner werden. Kleine Unebenheiten mit einer Amplitude in der Größenordnung des Faserdurchmessers könnten hier eine Rolle spielen. Außerdem nimmt mit der Faserdicke auch die Reibung zu, wie weiter oben bereits angedeutet wurde.

2.4.11.2 Glätten der Haare

Glätten der Haare kann als Gegenteil des Dauerwellprozesses angesehen werden. Das Ziel ist die Entfernung der Wellen aus dem Haar.

Generell kann man dauerhaftes Glätten nur durch eine *Superkontraktion* um mindestens 5 % der Ursprungslänge erreichen. Dafür eignen sich Agenzien, die eine solche Superkontraktion zu induzieren in der Lage sind. Sie sind in der Regel stark alkalisch ($\text{pH} > 13$) oder enthalten LiCl und entfalten dadurch eine starke Quellwirkung [Wong et al., 1994]. Legt man z.B. lockige Haare in 1N NaOH ein, strecken sie sich ohne äußere Krafteinwirkung. Das radiale Quellen der Keratinfasern im alkalischen Milieu führt zu einer wesentlichen Vergrößerung der Kurvenradien und damit zu einer temporären Glättung. Folgt eine Superkontraktion um ca. 5 % der Ausgangslänge, so bleibt die Glättung dauerhaft. Eine dauerhafte Glättung kann auch durch eine Dehnung der Fasern in kochendem Wasser um ca. 20 % ihrer Ausgangslänge erreicht werden. Glättungseffekte infolge Hitzeeinwirkung, bspw. durch Lockenwickler, sind hingegen nur von temporärer Dauer, wenn sie nicht mit einer Superkontraktion einhergehen. Diese findet ihre Ursache auf molekularer Ebene in Änderungen der Sekundärstruktur des Keratins. Diese Änderungen sind nicht reversibel. D.h., dass dauerhaft geglättete Fasern nach Einlegen in Wasser nicht mehr ihre ursprüngliche Form annehmen.

2.4.11.3 Haarsprays

Unter Verwendung von Haarsprays bzw. Festigern kann die Struktur des Haarkörpers fixiert werden. Die Haltbarkeit hängt u.a. ab von der Partikelgröße, der Fließfähigkeit der Klebstoffe und von der Art der Bindung, die sich zwischen den Fasern ausbildet. Die Glasübergangstemperatur

Tabelle 2.12: Haltbarkeit von Haarsprayverbindungen unterschiedlicher Klassen, gerundet, gerundet nach [Wickett et al., 1992]. UHH = *ultra-high hold*, HH = *hard-to-hold*, RH = *regular-hold* und LH = *light-hold*

Klasse	Maximalkraft [N]
UHH	0,35 - 0,40
HH	0,25 - 0,30
RH	0,20 - 0,25
LH	0,10

des Harzes und seine Viskosität sind wichtige Einflussgrößen zur Bestimmung der Adhäsionskräfte. Experimentell lässt sich die Stärke solcher Bindungen in Zugversuchen ermitteln. Man durchschneidet eine an der Berührungsstelle verklebte Haarschlaufe (ca. 2 cm) und belastet anschließend bis zum Versagen der Klebestelle auf Zug [Wickett et al., 1992]. In der Regel tritt ein adhäsives Versagen eher ein, als ein kohäsives. D.h., dass eine größere Tendenz zur Ablösung der Faseroberfläche vom Klebstoff besteht, als zum Bruch des Klebstoffkörpers selber. Die Größe der Oberfläche hat Einfluss auf die Festigkeit der Bindung. Handelsübliche Produkte lassen sich nach der Haltbarkeit klassifizieren: *ultra-high hold (UHH)*, *hard-to-hold (HH)*, *regular-hold (RH)* und *light-hold (LH)*. In der Regel korrespondieren die Bezeichnungen mit der Stärke der von den Produkten ausgebildeten Verbindungen. Die Stärke hängt stark von den Inhaltsstoffen der Produkte ab. Acrylpolymerer weisen die größte Festigkeiten bei 40 % rel. Luftfeuchte auf, vgl. Abb. (2.20). Einige typische Durchschnittswerte sind in Tab. (2.12) zu finden.

Die meisten Haarsprays werden klebrig, sobald das Lösungsmittel verdampft ist. Die Zunahme der Steifigkeit ist proportional zur Viskosität der Polymerlösung und zur aufgetragenen Gesamtmenge und nimmt direkt nach der Applikation im Vergleich zum unbehandelten Haar um 30-50% zu. Dieser Initialeffekt kann auch eine Folge der Kapillarkräfte sein, die zwischen den Fasern herrschen, solange das Lösungsmittel noch flüssig ist.

Je höher das Molekulargewicht des Polymers, desto größer die Steifigkeit, die sich einstellt [Jachowicz & Yao, 1996]. Aber auch dieser Zusammenhang ist von der rel. Luftfeuchte abhängig.

2 *Eigenschaften von Humanhaar*

3 Mechanische Modelle für Fasern

3.1 Einleitung

Die Modellierung schlanker, insbesondere torsionssteifer Strukturen ist in den letzten Jahren in das Blickfeld des computergraphischen Interesses gerückt. Kabel, Schläuche, Katheter oder Haare bleiben den etablierten Masse-Feder-Systemen weitestgehend verschlossen, sollen sie mechanisch korrekt modelliert werden. Kontinuumsmechanische Ansätze, welche die Deformation solcher Strukturen unter Einwirkung äußerer Kräfte beschreiben, wurden bereits von KIRCHHOFF [Kirchhoff, 1859] und später durch die Brüder COSSERAT vorgestellt, vgl. [Cosserat & Cosserat, 1907b, Cosserat & Cosserat, 1907a, Cosserat & Cosserat, 1908]. Die Grundidee ist dabei die Ausstattung der Struktur mit einem lokalen Koordinatensystem. Dieser lokale „Direktor-Ansatz“ erlaubt eine akkurate Modellierung lokaler Deformationen wie Scherung, Biegung und Verdrehung. Materialeigenschaften lassen sich durch die in der Kontinuumsmechanik typischen konstituierenden Gleichungen elegant in den theoretischen Rahmen einbetten. Linear- oder viskoelastisches Verhalten sind dann nur mehr eine Frage des Materialgesetzes.

3.1.1 Diskrete und kontinuierliche Modelle

Bei der Modellierung von Fasern gibt es traditionell zwei Sichtweisen: die der diskreten kinematischen Kette und die kontinuierliche. Die Bewegung der einzelnen Glieder folgt den NEWTON-EULER-Gleichungen (bzw. dem zweiten NEWTONschen Gesetz¹ im Falle von Punktmassen). Gelenkbedingungen zwischen den Kettengliedern diskreter Modelle werden üblicherweise durch die Addition von Korrekturkräften und Momenten erzwungen. Richtung und Betrag folgen dabei aus der Annahme abstrakter Federn („schwache Bedingung“) zwischen den Gliedern oder durch Applikation des LAGRANGE-Multiplikatoransatzes („starke Bedingung“). Bei reduzierten Koordinatendarstellungen kann die Nullbeschleunigung, die den Zusammenhalt der Körper gewährleistet, auf der Basis des aus der Robotik stammenden Featherstone-Algorithmus direkt berechnet werden, vgl. [Mirtich, 1996].

Torsionsdeformationen und die Abbildung von Materialeigenschaften bleiben bei Masse-Feder-Systemen ein ad-hoc-Ansatz.

Bei der kontinuierlichen Sichtweise hingegen wird die Deformation der Faser durch ein System von Differentialgleichungen beschrieben. Beispiele hierfür sind die COSSERAT-Theorie und die Theorie der Gelenkträger.

Primitive Masse-Feder-Systeme fanden bereits 1988 zur Simulation von Würmern und Schlangen durch Miller Verwendung, vgl. [Miller, 1988]. In einer der ersten Publikation auf dem Gebiet der Haarsimulation durch Rosenblum et al. [Rosenblum et al., 1991], schlagen die Autoren ein

¹*Lex secunda: Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur.*

Masse-Feder-Modell zur Darstellung einzelner Haarfasern bzw. Haarsträhnen vor. Die Verwendung solcher Modelle finden wir auch in [Plante et al., 2001, Plante et al., 2002]. In ihrem Beitrag ersetzen Byoungwon et al. [Choe et al., 2005] die Punktmassen durch starre Körper, womit zusätzlich Orientierungen für die Messungen lokaler Deformationen zur Verfügung stehen. In dem Beitrag von Hadap und Thalmann [Hadap & Magnenat-Thalmann, 2001] bzw. Chang et al. [Chang et al., 2002] werden Fasern auf der Basis des Featherstone-Algorithmus berechnet. Anjyo et al. [Anjyo et al., 1992] fassen die Faser als Kontinuum auf und beschreiben ihre Deformation auf der Basis des Gelenkträgermodells, das durch zwei gewöhnliche Differentialgleichungen vierter Ordnung definiert ist. Dieser Ansatz wurde von Daldegan et al. [Daldegan et al., 1993], Kurihara et al. [Kurihara et al., 1994] und Ward et al. [Ward et al., 2002] recycelt. Die Einführung des Begriffes der COSSERAT-Theorie in den Bereich der Computergraphik erfolgte mit dem Beitrag von Pai [Pai, 2002]. Allerdings beschränken sich seine Betrachtungen nur auf den statischen Fall. Die ersten dynamischen Ansätze sind das Superhelixmodell [Bertails et al., 2006], die geometrisch exakten dynamischen Splines [Theetten et al., 2007] und die COSSERAT-Rod-Elemente [Spillmann & Teschner, 2007], welche die Theorie COSSERAT-Stäbe auf der Basis des LAGRANGE-Formalismus phänomenologisch nachbilden.

3.1.2 Herleitung von Bewegungsgleichungen

Eine Frage, die sich an dieser Stelle unmittelbar stellt, ist jene nach der Gewinnung der Bewegungsgleichungen für den kontinuierlichen Fall.

Mechanische Probleme lassen sich oftmals einfacher durch die Anwendung des Energie- oder Impulsbegriffs ausdrücken, als durch Kräfte. Dadurch gelangt man zur LAGRANGESchen Mechanik, die in diesem Sinne eine Neuformulierung der klassischen Mechanik darstellt. Hier untersucht man den Konfigurationsraum eines Systems, indem man die Zustände bzw. ihre Energien durch eine minimale Menge von unabhängigen Systemparametern, den reduzierten Koordinaten, bzw. deren Zeitableitungen oder Geschwindigkeiten, ausdrückt. Die reduzierten Koordinaten müssen nicht notwendig mit den kartesischen Koordinaten zusammenfallen, sie müssen das System nur eindeutig beschreiben.

Verwandt mit dem LAGRANGE-Ansatz ist die HAMILTONSche Mechanik, die wir ausgehend vom LAGRANGE-Formalismus auf dem Wege der Variablentransformation (LEGENDRE-Transformation) erreichen. Hierbei wird durch Substitution der generalisierten Geschwindigkeiten durch generalisierte Impulse der Phasenraum des Systems betrachtet, der oft tiefe Einblicke in die mathematische Struktur des zugrunde liegenden Problems zulässt. Beide Ansätze, der nach LAGRANGE und der nach HAMILTON erlauben die systematische Herleitung der Bewegungsdifferentialgleichungen komplexer Systeme, die der Anschauung oftmals nur schwer zugänglich sind. Ersterer Fall liefert n Gleichungen zweiter Ordnung, letzterer hingegen $2n$ Gleichungen erster Ordnung, die kanonischen Gleichungen.

Der LAGRANGE-Ansatz hat sich mittlerweile als robustes Handwerkszeug im Bereich der Computergrafik etabliert. Gelingt es, kinetische und potentielle Energie für das System in Form einer Funktion aufzustellen, folgen die Bewegungsgleichungen direkt durch die festen Differentiationsregeln des LAGRANGE-Formalismus. Vor diesem Hintergrund ist es verständlich, dass die beiden „vollständigsten“ bisher im Bereich der Computergrafik eingeführten Modelle zur Simulation von Fasern, diesem Ansatz folgen. Wir werden dem Superhelixmodell [Bertails et al., 2006], so-

wie den COSSERAT-Rod-Elementen [Spillmann & Teschner, 2007] in den nächsten Abschnitten einige Aufmerksamkeit schenken. Der LAGRANGE-Formalismus ist im Anhang Abs. (A.2) beschrieben.

Eine Alternative zum LAGRANGESchen Ansatz stellt die Kontinuumsmechanik dar. Sie befasst sich mit dem mechanischen Verhalten und der Kinematik von Materialien. Zur Modellierung des makroskopischen Verhaltens, wird für den Körper eine kontinuierliche zeitliche und räumliche Masseverteilung angenommen. Das Kontinuum lässt sich in beliebig kleine Elemente zerlegen, denen man dieselben Eigenschaften zuordnet, wie dem makroskopischen Material. Für den Grenzfall unendlich kleiner Raumelemente lassen sich physikalische Größen wie Impuls und Energie hinreichend genau annähern. Folgende Grundkomponenten sind für die Theorie entscheidend: die Untersuchung der Bewegung und der Verformung (Kinematik), die Untersuchung der Spannungen, die im Kontinuum auftreten durch Einwirkung von Kräften und Momenten und die mathematische Beschreibung der physikalischen Gesetzmäßigkeiten, welche die Bewegung des Kontinuums bestimmen. Die Kinematik selbst ist dabei Folge von oder bewirkt Spannungen.

Die spezielle Theorie der COSSERAT-Stäbe, auf die wir die Simulation der Fasern stützen, ist ein Beispiel für einen kontinuumsmechanischen Ansatz.

3.2 Spezielle Theorie der Cosserat-Stäbe

Im vorliegenden Abschnitt soll die spezielle Theorie der COSSERAT-Stäbe herausgearbeitet werden, wobei wir im Wesentlichen Antman [Antman, 1995] folgen. Diese Theorie kann als „Haupttheorie“ zur Deformationsberechnung an langen dünnen Strukturen angesehen werden. Die klassische KIRCHHOFFSche Stabtheorie ist als spezielle Instanz mit besonderen Materialbedingungen enthalten.

Vier Arten der Deformation können bei einer Faser unterschieden werden; Biegung, Torsion, Dehnung und Scherung. Die spezielle COSSERAT-Theorie der Stäbe ist eine eindimensionale, geometrisch exakte Theorie, die auf der Erhaltung des linearen und des Drehimpulses basiert. Im Gegensatz dazu stehen bspw. die kontinuumsmechanischen Modelle, die den dreidimensionalen Blickpunkt wählen und die Kinematik einzelner Materialpunkte im Kontinuum analysieren.

Wir definieren eine Faser als Gebiet um die Raumkurve $\mathbf{r}(s) : [s_1, s_2] \mapsto \mathbb{R}^3$, wobei s die Bogenlänge sei. Der Bogenlängeparameter s indiziert denjenigen Materialquerschnitt des Körpers, dessen Punkte in der Ebene senkrecht zur Kurve an der Stelle s liegen. Wenn wir annehmen, dass die Kurve glatt ist und durch die Zentren der einzelnen Querschnitte stößt, dann bildet sie die Mittellinie der Faser. Die Dynamik einer Faser oder des speziellen COSSERAT-Stabes kann durch drei vektorwertige Funktionen $\mathbf{r}(s, t) : [s_1, s_2] \times \mathbb{R} \mapsto \mathbb{R}^3$ und $\mathbf{d}_{\{1,2\}}(s, t) : [s_1, s_2] \times \mathbb{R} \mapsto \mathbb{R}^3$ beschrieben werden, wobei die beiden Vektoren $\mathbf{d}_{1,2} \in \mathbb{R}^3$ senkrecht aufeinander stehen ($\mathbf{d}_1^\top \mathbf{d}_2 = 0$) und die Ebene des Materialquerschnitts an der Stelle s aufspannen, vgl. Abb. (3.1). Damit beschreiben sie gleichzeitig die Konfiguration zweier orthogonaler Materialbänder des Stabes zum Zeitpunkt t . Die Konfiguration $\mathcal{C}(s, t)$ der Mittellinie ist zu jedem Zeitpunkt t durch das Tripel $\mathcal{C}(s, t) = \{\mathbf{r}(s, t), \mathbf{d}_1(s, t), \mathbf{d}_2(s, t)\}$ eindeutig festgelegt. Für die nun folgende mechanische Theorie ist es sinnvoll, die Deformation auf eine Referenzkonfiguration der Faser zu beziehen, die durch das Tripel $\mathcal{C}^\circ(s) = \{\mathbf{r}^\circ(s), \mathbf{d}_1^\circ(s), \mathbf{d}_2^\circ(s)\}$ charakterisiert sei. Wir staten die Mittellinie

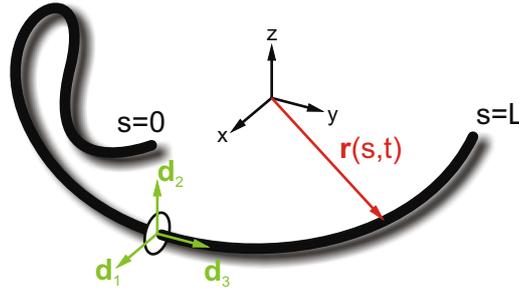


Abbildung 3.1: Die Vektoren $\mathbf{d}_{1,2,3}$ bilden ein rechtshändiges, orthogonales Koordinatensystem an jedem Punkt der Kurve. Dabei spannen $\mathbf{d}_{1,2}$ den lokalen Materialquerschnitt auf, \mathbf{d}_3 bildet die Tangente an die Kurve. Solange wir Scherdeformationen vernachlässigen, steht die Tangente immer senkrecht auf dem Materialquerschnitt.

der Faser mit einer rechtshändigen, orthonormalen Basis $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\} \in \mathbb{R}^3$ aus. Damit gilt für den dritten Vektor $\mathbf{d}_3 := \mathbf{d}_1 \times \mathbf{d}_2$. Aus der Existenz dieser Basis folgt die Existenz entsprechender vektorwertiger Funktionen $\boldsymbol{\omega}(s, t)$ und $\boldsymbol{\kappa}(s, t)$, so dass die Kinematik des Querschnitts durch folgende zwei Gleichungen beschrieben werden kann,

$$\frac{\partial}{\partial s} \mathbf{d}_k = \boldsymbol{\kappa}(s, t) \times \mathbf{d}_k, \quad (3.1)$$

$$\frac{\partial}{\partial t} \mathbf{d}_k = \boldsymbol{\omega}(s, t) \times \mathbf{d}_k. \quad (3.2)$$

Wir indizieren die Ableitung einer Funktion bezüglich des Bogenlängeparameters s bzw. der Zeit t , sofern uns Platzgründe dazu zwingen, mit $(\cdot)' = \partial/\partial s$ und $(\cdot) = \partial/\partial t$, andernfalls schreiben wir die Ableitungen voll aus. Weiterhin bezeichnen wir $\boldsymbol{\kappa}$ als DARBOUX-Vektor und $\boldsymbol{\omega}$ als Twistvektor. Dieser Vektoren selbst werden auf die orthonormale Basis $\{\mathbf{d}_k\}$ bezogen, so dass man deren Komponenten in der folgenden Form schreiben kann,

$$\boldsymbol{\kappa} = \kappa_k \mathbf{d}_k, \quad \boldsymbol{\omega} = \omega_k \mathbf{d}_k. \quad (3.3)$$

Die Kinematik der Mittellinie sei durch $\mathbf{r}' = v_k \mathbf{d}_k$ gegeben. Die Komponenten von $\mathbf{k} = \{\kappa_1, \kappa_2, \kappa_3\}$ und $\mathbf{v} = \{v_1, v_2, v_3\}$ sind die so genannten Verzerrungen. Dabei geben κ_1 und κ_2 die Biegung bzgl. der beiden Hauptachsen \mathbf{d}_1 und \mathbf{d}_2 des Querschnitts an, κ_3 die Torsion, also die Verdrehung um die Achse \mathbf{d}_3 . Die Komponenten von \mathbf{v} hingegen drücken den Grad der Scherung bzw. die Dehnung der Faser aus. Damit ist die Konfiguration $\mathcal{C}(s, t)$ durch die beiden Funktionen \mathbf{k} und \mathbf{v} eindeutig beschrieben.

Sowohl die zeitliche als auch die räumliche Entwicklung der Funktionen $\mathbf{d}_{\{1,2\}}$ müssen auf dieselbe lokale Basis führen. Deshalb fordern wir die Kompatibilität der Basis $\{\mathbf{d}_k\}$ durch

$$\frac{\partial}{\partial t} \left(\frac{\partial}{\partial s} \mathbf{d}_k \right) = \frac{\partial}{\partial s} \left(\frac{\partial}{\partial t} \mathbf{d}_k \right), \quad (3.4)$$

bzw.

$$\frac{\partial}{\partial t} (\boldsymbol{\kappa} \times \mathbf{d}_k) = \frac{\partial}{\partial s} (\boldsymbol{\omega} \times \mathbf{d}_k), \quad (3.5)$$

woraus wir durch einige algebraische Manipulationen die Kompatibilitätsgleichung der Form

$$\frac{\partial \boldsymbol{\omega}}{\partial s} = \frac{\partial \boldsymbol{\kappa}}{\partial t} + \boldsymbol{\kappa} \times \boldsymbol{\omega} \quad (3.6)$$

erhalten.

Damit die Orientierung der Mittellinie erhalten bleibt, fordern wir weiterhin, dass

$$v_3 = \mathbf{d}_3^\top \frac{\partial}{\partial s} \mathbf{r} > 0. \quad (3.7)$$

Damit können dann zwei pathologische Fälle in der Konfiguration der Mittellinie ausgeschlossen werden: a) Ein lokaler Kollaps der Faser, weil das Verhältnis aus der lokalen Länge im deformierten und jener im Referenzzustand nicht null werden kann, $\|\partial_s \mathbf{r}\| > 0$; b) Totale Scherung, bei welcher der Querschnitt tangential zur Mittellinie liegt.

Ausgerüstet mit den Grundlagen für die Kinematik der Mittellinie, können wir uns jetzt den Bewegungsgleichungen durch eine Gleichgewichtsbetrachtung an der Faser widmen.

3.2.1 Gleichgewichtsbedingungen

Seien $s_1 < a < b < s_2$. Wir schneiden die Faser an der Stelle $[a, b]$ auf und ermitteln die Schnittgrößen. Das Teilsegment $[s_1, a]$ übt eine Kontaktkraft $-\mathbf{n}^-(a, t)$ und ein Kontaktmoment $-\mathbf{r}(a, t) \times \mathbf{n}^-(a, t) - \mathbf{m}^-(a, t)$ bzgl. des Ursprungs $\mathbf{0}$ der Faser auf das Teilsegment $[a, b]$ aus. Das Teilsegment $(b, s_2]$ hingegen übt eine Kontaktkraft $\mathbf{n}^+(b, t)$ und ein Kontaktmoment $\mathbf{r}(b, t) \times \mathbf{n}^+(b, t) + \mathbf{m}^+(b, t)$ bzgl. des Ursprungs $\mathbf{0}$ der Faser auf das Teilsegment $[a, b]$ aus. $\mathbf{m}^\pm(s, t)$ ist das resultierende Kontaktmoment. Die Kontaktkraft $\mathbf{n}^\pm(\cdot, t)$ und das Kontaktmoment $\mathbf{m}^\pm(\cdot, t)$ hängen lediglich vom Querschnitt zwischen den beiden Teilsegmenten ab. Die Resultierenden am Querschnitt infolge von Körperkräften $\mathbf{f}(s, t)$ und Momenten $\mathbf{l}(s, t)$, angegeben pro Einheitslänge, werden hier in ihrer integralen Form angegeben,

$$\boldsymbol{\xi}(a, b, t) = \int_a^b \mathbf{f}(s, t) ds, \quad (3.8)$$

$$\boldsymbol{\eta}(a, b, t) = \int_a^b [\mathbf{r}(s, t) \times \mathbf{f}(s, t) + \mathbf{l}(s, t)] ds. \quad (3.9)$$

In Analogie zum linearen Impuls eines Partikels, $\mathbf{p} = m\mathbf{v}$, kann man den Impuls des Teilsegmentes $[a, b]$ ausdrücken durch die Geschwindigkeit der Mittellinie $\partial_t \mathbf{r}$ und die des Querschnitts $\partial_t \mathbf{d}_1 + \partial_t \mathbf{d}_2$,

$$\mathbf{p}(a, b, t) = \int_a^b [\rho(s)A(s)\partial_t \mathbf{r}(s, t) + \rho(s)I_1(s)\partial_t \mathbf{d}_1(s, t) + \rho(s)I_2(s)\partial_t \mathbf{d}_2(s, t)] ds. \quad (3.10)$$

Im selben Sinne lässt sich der Drehimpuls für das Teilsegment $[a, b]$ aus der Position \mathbf{r} und $\mathbf{d}_1 + \mathbf{d}_2$ mit den Geschwindigkeiten $\partial_t \mathbf{r}$ und $\partial_t \mathbf{d}_1 + \partial_t \mathbf{d}_2$ herleiten,

$$\begin{aligned} \mathbf{L}(a, b, t) = \int_a^b \left\{ \rho(s)\mathbf{I}(s, t)\boldsymbol{\omega}(s, t) + \rho(s)A(s)\mathbf{r}(s, t) \times \partial_t \mathbf{r}(s, t) \right. \\ \left. + \mathbf{r}(s, t) \times [\rho(s)I_1(s)\partial_t \mathbf{d}_1(s, t) + \rho(s)I_2(s)\partial_t \mathbf{d}_2(s, t)] \right. \\ \left. - \partial_t \mathbf{r}(s, t) \times [\rho(s)I_1(s)\mathbf{d}_1(s, t) + \rho(s)I_2(s)\mathbf{d}_2(s, t)] \right\} ds, \end{aligned} \quad (3.11)$$

mit dem Trägheitstensor $\mathbf{I} = \text{diag}(I_1, I_3, I_3)$.

Ferner führt die Forderung nach der Gleichheit des Kraftstoßes, der infolge aller Kräfte über das Zeitintervall $[0, \tau]$ auf das Teilstück $[a, s]$, mit $s_1 < a < s < s_2$ einwirkt, und der Änderung des linearen Impulses bzgl. desselben Zeitintervalls, auf folgende Gleichung für die Impulserhaltung am System,

$$\begin{aligned} & \int_0^\tau [\mathbf{n}^+(s, t) - \mathbf{n}^-(a, t)] dt + \int_0^\tau \boldsymbol{\xi}(a, s, t) dt \\ &= \int_a^s \left\{ \rho(\sigma) A(\sigma) [\partial_t \mathbf{r}(\sigma, \tau) - \partial_t \mathbf{r}(\sigma, 0)] + \rho(\sigma) I_1(\sigma) [\partial_t \mathbf{d}_1(\sigma, \tau) - \partial_t \mathbf{d}_1(\sigma, 0)] \right. \\ & \quad \left. + \rho(\sigma) I_2(\sigma) [\partial_t \mathbf{d}_2(\sigma, \tau) - \partial_t \mathbf{d}_2(\sigma, 0)] \right\} d\sigma. \end{aligned} \quad (3.12)$$

Auch hier kann eine entsprechende Beziehung für die Erhaltung des Drehimpulses bzgl. des Ursprungs $\mathbf{0}$ angegeben werden,

$$\begin{aligned} & \int_0^\tau [\mathbf{m}^+(s, t) - \mathbf{m}^-(a, t) + \mathbf{r}(s, t) \times \mathbf{n}^+(s, t) - \mathbf{r}(a, t) \times \mathbf{n}^-(a, t)] dt \\ & + \int_0^\tau \boldsymbol{\eta}(a, s, t) dt \\ &= \int_a^s \left\{ d\rho(\sigma) \mathbf{I}(\sigma, t) \boldsymbol{\omega}(\sigma, t) + \mathbf{r}(\sigma, t) \times \rho(\sigma) A(\sigma) \partial_t \mathbf{r}(\sigma, t) \right. \\ & \quad \left. + \mathbf{r}(\sigma, t) \times [\rho(\sigma) I_1(\sigma) \partial_t \mathbf{d}_1(\sigma, t) + \rho(\sigma) I_2(\sigma) \partial_t \mathbf{d}_2(\sigma, t)] \right. \\ & \quad \left. - \partial_t \mathbf{r}(\sigma, t) \times [\rho(\sigma) I_1(\sigma) \mathbf{d}_1(\sigma, t) + \rho(\sigma) I_2(\sigma) \mathbf{d}_2(\sigma, t)] \right\} d\sigma \Big|_{t=0}^{t=\tau}. \end{aligned} \quad (3.13)$$

Die Grenzen müssen jetzt lediglich auf $a = s_1$ und $s = s_2$ verschoben werden. Wegen der Kontinuität von \mathbf{n}^+ gilt $\mathbf{n}^+(a, t) = \lim_{s \rightarrow a} \mathbf{n}^+(s, t)$ und Analoges für \mathbf{m}^+ , so dass die Bezeichnungen \pm entfallen. Die Differentiation bzgl. s und τ führt uns schließlich auf die finale Form der Bewegungsgleichungen der speziellen Theorie der COSSERAT-Stäbe,

$$\frac{\partial \mathbf{n}}{\partial s} + \mathbf{f} = \rho A \frac{\partial^2 \mathbf{r}}{\partial t^2} + \rho I_1 \frac{\partial^2 \mathbf{d}_1}{\partial t^2} + \rho I_2 \frac{\partial^2 \mathbf{d}_2}{\partial t^2}, \quad (3.14)$$

und

$$\frac{\partial \mathbf{m}}{\partial s} + \frac{\partial \mathbf{r}}{\partial s} \times \mathbf{n} + \mathbf{l} = (\rho I_1 \mathbf{d}_1 + \rho I_2 \mathbf{d}_2) \times \frac{\partial^2 \mathbf{r}}{\partial t^2} + \rho \frac{\partial}{\partial t} (\mathbf{I} \boldsymbol{\omega}). \quad (3.15)$$

Die Biege- bzw. Torsionsmomente $\mathbf{m} := (m_1, m_2, m_3)$, sowie die Scherkräfte bzw. Spannung $\mathbf{n} := (n_1, n_2, n_3)$ werden bezüglich der Basis \mathbf{d}_k ausgedrückt, also $m_k = \mathbf{m} \cdot \mathbf{d}_k$ und $n_k = \mathbf{n} \cdot \mathbf{d}_k$.

Das finale System gekoppelter, partieller Differentialgleichungen (PDE), welches die zeitliche

Entwicklung der Konfiguration einer Faser beschreibt, nimmt damit die folgende Form an,

$$\frac{\partial \mathbf{d}_k}{\partial t} = \boldsymbol{\omega} \times \mathbf{d}_k, \quad (3.16)$$

$$\frac{\partial \boldsymbol{\kappa}}{\partial t} = \frac{\partial \boldsymbol{\omega}}{\partial s} - \boldsymbol{\kappa} \times \boldsymbol{\omega}, \quad (3.17)$$

$$\frac{\partial}{\partial t} \left(\frac{\partial \mathbf{r}}{\partial s} \right) = \frac{\partial}{\partial s} \left(\frac{\partial \mathbf{r}}{\partial t} \right), \quad (3.18)$$

$$\frac{\partial}{\partial t} (\rho \mathbf{I} \boldsymbol{\omega}) = \frac{\partial \mathbf{m}}{\partial s} + \frac{\partial \mathbf{r}}{\partial s} \times \mathbf{n} + \mathbf{l}, \quad (3.19)$$

$$\rho A \frac{\partial^2 \mathbf{r}}{\partial t^2} = \frac{\partial \mathbf{n}}{\partial s} + \mathbf{f}. \quad (3.20)$$

Im Prinzip lässt sich das PDE-System bezüglich der Basis $\{\mathbf{d}_k\}$ zerlegen. Wenn wir berücksichtigen, dass die Differentiation einer Größe \mathbf{q} bzgl. des lokalen Systems folgenden Kreuzproduktterm mitführt,

$$\left(\frac{\partial \mathbf{q}}{\partial t} \right)_{\{\mathbf{e}_i\}} = \left(\frac{\partial \mathbf{q}}{\partial t} \right)_{\{\mathbf{d}_k\}} + \boldsymbol{\omega} \times \mathbf{q}_{\{\mathbf{d}_k\}}, \quad (3.21)$$

$$\left(\frac{\partial \mathbf{q}}{\partial s} \right)_{\{\mathbf{e}_i\}} = \left(\frac{\partial \mathbf{q}}{\partial s} \right)_{\{\mathbf{d}_k\}} + \boldsymbol{\omega} \times \mathbf{q}_{\{\mathbf{d}_k\}}, \quad (3.22)$$

erhalten wir ein äquivalentes Gleichungssystem in lokalen Koordinaten,

$$\frac{\partial \boldsymbol{\kappa}}{\partial t} = \frac{\partial \boldsymbol{\omega}}{\partial s} - \boldsymbol{\omega} \times \boldsymbol{\kappa}, \quad (3.23)$$

$$\frac{\partial}{\partial t} \left(\frac{\partial \mathbf{r}}{\partial s} \right) = \frac{\partial}{\partial s} \left(\frac{\partial \mathbf{r}}{\partial t} \right) + \boldsymbol{\kappa} \times \frac{\partial \mathbf{r}}{\partial t} - \boldsymbol{\omega} \times \frac{\partial \mathbf{r}}{\partial s}, \quad (3.24)$$

$$\frac{\partial}{\partial t} (\rho \mathbf{I} \boldsymbol{\omega}) = \frac{\partial \mathbf{m}}{\partial s} + \boldsymbol{\kappa} \times \mathbf{m} + \frac{\partial \mathbf{r}}{\partial s} \times \mathbf{n} - \boldsymbol{\omega} \times (\rho \mathbf{I} \boldsymbol{\omega}) + \mathbf{l}, \quad (3.25)$$

$$\rho A \frac{\partial^2 \mathbf{r}}{\partial t^2} = \frac{\partial \mathbf{n}}{\partial s} + \boldsymbol{\kappa} \times \mathbf{n} - \boldsymbol{\omega} \times \left(\rho A \frac{\partial \mathbf{r}}{\partial t} \right) + \mathbf{f}. \quad (3.26)$$

$$(3.27)$$

Der Vorteil liegt in der Entkopplung des PDE-Systems von den Kinematikgleichungen Gl. (3.1) und Gl. (3.2). Antman [Antman, 1995] weist in diesem Zusammenhang darauf hin, dass diese Entkopplung zum Verlust von geometrischen Informationen führen kann, die an die Basisvektoren $\{\mathbf{d}_k\}$ gekoppelt sind, und damit u.U. zum Verlust der Gültigkeit der Invarianz der konstituierenden Gleichungen unter Starrkörpertransformation, vgl. Abs. (3.2.2). Der Verlust der Invarianz kann signifikante numerische Probleme nach sich ziehen, vereinfacht jedoch auf der anderen Seite den Lösungsprozess ganz erheblich.

3.2.2 Konstituierende Gleichungen

Um die Theorie abzurunden, benötigen wir Funktionen, die es uns erlauben, eine Beziehung zwischen den Deformationen \mathbf{k} bzw. \mathbf{v} und denen im Material wirkenden Momenten \mathbf{m} bzw. Kräften

\mathbf{n} herzustellen. Durch ein solches *Werkstoffgesetz* kann man dann erreichen, dass sich die Faser wie Gummi oder Metall o.ä. verhält. Die Faser wird als *elastisch* bezeichnet, wenn Funktionen $\hat{\mathbf{m}}$ und $\hat{\mathbf{n}}$ existieren, so dass gilt

$$\mathbf{m}(s, t) = \hat{\mathbf{m}}(\mathbf{k}(s, t), \mathbf{v}(s, t), s), \quad \mathbf{n}(s, t) = \hat{\mathbf{n}}(\mathbf{k}(s, t), \mathbf{v}(s, t), s). \quad (3.28)$$

Der gemeinsame Wertebereich von $\hat{\mathbf{m}}$ und $\hat{\mathbf{n}}$ sind diejenigen \mathbf{k} und \mathbf{v} , die Bedingung Gl. (3.7) erfüllen. Wir bezeichnen die Faser als *hyperelastisch*, wenn eine Funktion $W : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R} \mapsto \mathbb{R}$ existiert, so dass für die Ableitung dieser Funktion nach den Verzerrungen gilt,

$$\hat{\mathbf{m}}(\mathbf{k}, \mathbf{v}, s) = \frac{\partial}{\partial \mathbf{k}} W(\mathbf{k}, \mathbf{v}, s), \quad \hat{\mathbf{n}}(\mathbf{k}, \mathbf{v}, s) = \frac{\partial}{\partial \mathbf{v}} W(\mathbf{k}, \mathbf{v}, s). \quad (3.29)$$

W ist die *Verzerrungsenergiefunktion*. Ferner bezeichnen wir die Faser als *viskoelastisch*, falls Funktionen $\hat{\mathbf{m}}$ und $\hat{\mathbf{n}}$ existieren, die auch von den Zeitableitungen der Verzerrungen \mathbf{k} und \mathbf{v} abhängen,

$$\begin{aligned} \mathbf{m}(s, t) &= \hat{\mathbf{m}}(\mathbf{k}(s, t), \mathbf{v}(s, t), \partial_t \mathbf{k}(s, t), \partial_t \mathbf{v}(s, t), s), \\ \mathbf{n}(s, t) &= \hat{\mathbf{n}}(\mathbf{k}(s, t), \mathbf{v}(s, t), \partial_t \mathbf{u}(s, t), \partial_t \mathbf{k}(s, t), s). \end{aligned} \quad (3.30)$$

Neben den oben genannten Werkstoffgesetzen können die Funktionen auch von der Historie abhängen, also $\mathbf{k}(s, t - \tau)$ für $\tau \geq 0$.

Die Verzerrungen \mathbf{k} und \mathbf{v} bestimmen die Konfiguration der Faser bis auf eine Starrkörpertransformation (Rotation und Translation). Man kann zeigen, dass die Verzerrungen und ihre Zeitableitungen, die Verzerrungsgeschwindigkeiten, invariant bzgl. der Starrkörpertransformation sind und deshalb auch nur Aussagen über die Konfigurationsänderungen einer Faser zulassen, vgl. Antman [Antman, 1995].

Ferner ist das PDE-System Gl. (3.16) bis Gl. (3.20) hyperbolisch, falls die Monotoniebedingung in Form der positiven Definitheit $|\mathbf{A}| > 0$ der Matrix

$$\mathbf{A} := \begin{pmatrix} \partial \hat{\mathbf{m}} / \partial \mathbf{k} & \partial \hat{\mathbf{m}} / \partial \mathbf{v} \\ \partial \hat{\mathbf{n}} / \partial \mathbf{k} & \partial \hat{\mathbf{n}} / \partial \mathbf{v} \end{pmatrix} \quad (3.31)$$

erfüllt ist und $\hat{\mathbf{m}}$ und $\hat{\mathbf{n}}$ differenzierbare Funktionen sind. Falls $\hat{\mathbf{m}}$ und $\hat{\mathbf{n}}$ nicht differenzierbar sind, fordern wir, dass die Werkstoffgesetze auf einer kompakten Untermenge $\mathcal{N}(s) \subset \mathcal{M}(s)$ des gemeinsamen Wertebereichs von $\hat{\mathbf{m}}$ und $\hat{\mathbf{n}}$ einheitlich monoton sind. Mit $\hat{\mathbf{m}}_i = (\mathbf{k}_i, \mathbf{v}_i, s)$ und $\Delta \hat{\mathbf{m}} = \hat{\mathbf{m}}_1 - \hat{\mathbf{m}}_2$, $\Delta \mathbf{k} = \mathbf{k}_1 - \mathbf{k}_2$ bzw. $\Delta \hat{\mathbf{n}} = \hat{\mathbf{n}}_1 - \hat{\mathbf{n}}_2$, $\Delta \mathbf{v} = \mathbf{v}_1 - \mathbf{v}_2$ existiert für alle Paare $(\mathbf{k}_1, \mathbf{v}_1)$ und $(\mathbf{k}_2, \mathbf{v}_2) \in \mathcal{N}(s)$ eine Zahl $c(\mathcal{N}(s)) > 0$, so dass gilt,

$$(\Delta \hat{\mathbf{m}} \Delta \mathbf{k} + \Delta \hat{\mathbf{n}} \Delta \mathbf{v}) > c(\mathcal{N}(s)) (\Delta \mathbf{k}^T \Delta \mathbf{k} + \Delta \mathbf{v}^T \Delta \mathbf{v}). \quad (3.32)$$

Die beiden Bedingungen werden als *starke Elliptizitätsbedingung* bezeichnet, weil sie garantieren, dass der stationäre Teil der Bewegungsgleichungen, der von den Werkstoffgesetzen abhängt, stark elliptisch ist. Stationär bedeutet, dass man die Zeitabhängigkeit der Gleichungen fallen lässt. Falls $\hat{\mathbf{m}}$ und $\hat{\mathbf{n}}$ differenzierbare Funktionen sind, folgt die Äquivalenz beider Bedingungen unmittelbar. Unter der starken Elliptizitätsbedingung lassen sich die konstituierenden Gleichungen eindeutig

nach \mathbf{k} bzw. \mathbf{v} auflösen, wobei die Funktionen $\hat{\mathbf{k}}(\mathbf{m}, \mathbf{n}, s)$ bzw. $\hat{\mathbf{v}}(\mathbf{m}, \mathbf{n}, s)$ stetig differenzierbare Funktionen sind, falls die Matrix

$$\mathbf{B} := \begin{pmatrix} \partial \hat{\mathbf{k}} / \partial \mathbf{m} & \partial \hat{\mathbf{k}} / \partial \mathbf{n} \\ \partial \hat{\mathbf{v}} / \partial \mathbf{m} & \partial \hat{\mathbf{v}} / \partial \mathbf{n} \end{pmatrix} \quad (3.33)$$

positiv-definit ist.

Auf der anderen Seite ist für die theoretische Untersuchung der Frage nach der Existenz und Eindeutigkeit von Lösungen bei Gleichgewichtsproblemen, die typische Instabilitätsmuster wie *Buckling* umfassen, bereits die schwache Elliptizitätsbedingung ausreichend, die lediglich fordert, dass die Werkstoffgesetze auf $\mathcal{M}(s)$ monoton sind, also $(\Delta \hat{\mathbf{m}} \Delta \mathbf{k} + \Delta \hat{\mathbf{n}} \Delta \mathbf{v}) \geq 0$.

Materialien, deren Elastizitätseigenschaften nicht von der Richtung abhängen, bezeichnen wir als *isotrop* (z.B. Metall oder Gummi), jene mit bevorzugter Richtung hingegen als *anisotrop* (z.B. Holz, Muskeln, Komposite). Als Sonderfall der Anisotropie muss die *transversale Isotropie* gelten, weil bei Materialien mit dieser Eigenschaft keine Kopplung zwischen Scherdeformation und Dehnung vorliegt und die Elastizitätseigenschaften invariant bzgl. Rotation um die Vorzugsrichtung sind. *Hemitropie* hingegen bezeichnet die Eigenschaft, dass für jedes Kraftsystem, das am Material angreift, die resultierenden Deformationen unabhängig von der Orientierung sind. Als Beispiel sei ein Seil angeführt, das aus helikal umeinander gewundenen Fasern besteht. Auf Beanspruchung durch ein in axialer Richtung wirkendes Drehmoment reagiert das Seil mit einer Straffung oder einer Lösung der Fasern, je nach der Drehrichtung des Momentes. Das Material ist also transversal hemitrop aber nicht transversal isotrop.

Diese Begriffe lassen sich auch mathematisch formulieren. Sei \mathbf{A} eine geeignete orthogonale Transformation mit $\det(\mathbf{A}) = 1$, welche die Richtung \mathbf{d}_3 unverändert lässt, also

$$\mathbf{A} := \begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.34)$$

Dann hat die Faser die Eigenschaft der *transversalen Hemitropie* genau dann, wenn gilt,

$$\hat{\mathbf{m}}(\mathbf{A} \cdot \mathbf{k}, \mathbf{A} \cdot \mathbf{v}, s) = \mathbf{A} \cdot \hat{\mathbf{m}}(\mathbf{k}, \mathbf{v}, s). \quad (3.35)$$

Die Richtung \mathbf{d}_3 wird nicht beachtet, weil sie keinen Beitrag zum Begriff der Isotropie leistet. Eine hyperelastische Faser mit der Verzerrungsenergiefunktion $W(\mathbf{k}, \mathbf{v}, s)$ ist *transversal isotrop*, falls gilt,

$$W(\mathbf{k}, \mathbf{v}, s) = W(\det(\mathbf{A}) \cdot \mathbf{A} \cdot \mathbf{k}, \mathbf{A} \cdot \mathbf{v}, s), \quad (3.36)$$

wobei $\det(\mathbf{A}) = \pm 1$.

Eine Faser bezeichnen wir als *uniform*, falls a) die Materialeigenschaften und b) die Querschnittsfläche nicht vom Bogenlängeparameter s abhängen. Für die numerischen Lösungsansätze werden wir in der Regel Uniformität annehmen, weil sich die Bewegungsgleichungen vereinfachen. Uniformität fordert zudem, dass die Änderung der Krümmung bei Fasern, die der Monotoniebedingung Gl. (3.31) genügen, in der Referenzkonfiguration verschwindet, $\partial_s \mathbf{k}^\circ(s) = \mathbf{0}$. Die Menge der möglichen Referenzkonfigurationen wird damit auf die initial gerade Faser, den Ring und die Helix mit konstanter Krümmung und Torsion reduziert. Das impliziert aber auf der anderen Seite, dass wir für alle anderen Anfangskonfigurationen Uniformität gar nicht fordern dürfen.

Sei $\mathbf{k}(s, 0)$ eine stetige differenzierbare Funktion, $\mathbf{v}(s, 0) = \text{const}$, und die Referenzkonfiguration *natürlich*, d.h.,

$$\hat{\mathbf{m}}(\mathbf{k}^\circ(s), \mathbf{v}^\circ(s)) = \mathbf{0}, \quad \forall s, \quad (3.37)$$

so dass

$$\frac{\partial \hat{\mathbf{m}}}{\partial \mathbf{k}}(\mathbf{k}^\circ(s), \mathbf{v}^\circ(s)) \cdot \partial_s \mathbf{k}^\circ(s) = \mathbf{0}, \quad \forall s. \quad (3.38)$$

Wegen Gl. (3.31) ist $\partial \hat{\mathbf{m}} / \partial \mathbf{k}$ regulär und damit gilt mit Gl. (3.38) sofort $\partial_s \mathbf{k}^\circ(s) = \mathbf{0}$.

3.2.3 Darstellung der lokalen Basis

Zur Darstellung der lokalen Basis wird in der Regel ein anderer Zugang gewählt, als die direkte Verwendung der drei Basisvektoren $\{\mathbf{d}_i\}$. Der Grund dafür liegt in der Zahl der Zwangsbedingungen, die dem System auferlegt werden müssten, damit die Basis ein Element der $\mathcal{SO}(3)$ ist. Das mündet in der Orthogonalitätsbedingung $\mathbf{d}_1^\top \mathbf{d}_2 = 0$ und der Einheitsbedingung $\|\mathbf{d}_{1,2}\| = 1$. Der dritte Vektor muss nicht explizit angegeben werden, sondern ergibt sich als $\mathbf{d}_3 := \mathbf{d}_1 \times \mathbf{d}_2$.

Wir wollen uns im Folgenden wegen ihrer Einfachheit auf EULER-Winkel konzentrieren. Die lokale Basis kann dann durch drei EULER-Winkel (ϕ, θ, ψ) ausgedrückt werden, d.h., $\mathbf{d}_i := \mathbf{d}_i(\phi, \theta, \psi)$. Sei $\{\mathbf{e}_i\}$ ein ortsfestes, orthonormales Koordinatensystem im \mathbb{R}^3 . Für die lokale Basis schreiben wir dann

$$\{\mathbf{d}_k\} = \{\mathbf{e}_k\} \mathbf{R}, \quad (3.39)$$

mit der Rotationsmatrix $\mathbf{R} := \mathbf{R}(\phi, \theta, \psi)$,

$$\mathbf{R} := \begin{pmatrix} \cos \phi \cos \theta \cos \psi - \sin \phi \sin \psi & -\cos \phi \cos \theta \sin \psi - \sin \phi \cos \psi & \cos \phi \sin \theta \\ \sin \phi \cos \theta \cos \psi + \cos \phi \sin \psi & -\sin \phi \cos \theta \sin \psi + \cos \phi \cos \psi & \sin \phi \sin \theta \\ -\sin \theta \cos \psi & \sin \theta \sin \psi & \cos \theta \end{pmatrix}. \quad (3.40)$$

Die Rotationsmatrix \mathbf{R} kann durch die geometrische Betrachtung dreier nacheinander ausgeführter Rotationen um drei Achsen, jeweils um die drei Winkel (ϕ, θ, ψ) , hergeleitet werden. In Bezug auf EULER-Winkel gibt es mehrere Konventionen: Wir verwenden die „y-Konvention“, indem wir zuerst um \mathbf{e}_3 mit Präzessionswinkel ϕ , dann um die resultierende neue y-Achse \mathbf{e}'_2 mit Nutationswinkel θ und anschließend wieder um die z-Achse \mathbf{e}''_3 mit Spin ψ drehen. Jede der drei Rotationen liefert eine Matrix, deren Konkatenation auf die obige Darstellung von \mathbf{R} führt. Dabei ist zu beachten, dass \mathbf{R} nicht unabhängig von der Reihenfolge der Rotationen ist.

Wenn wir für das ortsfeste System die Einheitsmatrix wählen, ist $\{\mathbf{d}_i\} = \mathbf{R}$. Da die Spalten der Rotationsmatrix \mathbf{R} ein orthogonales System bilden, ist der Zwang der Orthogonalität durch die Verwendung von EULER-Winkeln automatisch erfüllt. EULER-Winkel haben allerdings den Nachteil, dass sie eine polare Singularität in das System einführen, den so genannten GIMBAL-LOCK. Solange jedes lokale System direkt auf das ortsfeste (globale) Koordinatensystem bezogen wird, kann das zu erheblichen Problemen führen. Mit der Wahl einer lokalen Darstellung, d.h., wir drücken die lokalen Systeme jeweils in Bezug auf ihren direkten Vorgänger aus, kann das Problem zwar nicht eliminiert, aber wesentlich entschärft werden, weil die lokalen Deformationen in der Regel klein sind, so dass die kritischen Winkelkonfigurationen, die den GIMBAL-LOCK verursachen, nicht auftreten.

Wir formen die Kinematikgleichung Gl. (3.1) um und erhalten $\boldsymbol{\kappa} = \mathbf{d}_k \times \partial_s \mathbf{d}_k$. Wir drücken jetzt den Krümmungsvektor $\boldsymbol{\kappa}$ durch die EULER-Winkel aus. Dazu differenziert man Gl. (3.39) bzgl. s und erhält

$$\frac{\partial}{\partial s} \{\mathbf{d}_k\} = \{\mathbf{e}_k\} \frac{\partial}{\partial s} \mathbf{R}. \quad (3.41)$$

Aus der Kinematikgleichung Gl. (3.1) folgt unmittelbar

$$\begin{aligned} \frac{\partial}{\partial s} \{\mathbf{d}_k\} &= \tilde{\boldsymbol{\kappa}} \{\mathbf{d}_k\} \\ &= \tilde{\boldsymbol{\kappa}} \{\mathbf{e}_k\} \{\mathbf{d}_k\}, \end{aligned} \quad (3.42)$$

was wir in Gl. (3.39) einsetzen und nach $\tilde{\boldsymbol{\kappa}}$ auflösen,

$$\tilde{\boldsymbol{\kappa}} = \mathbf{R}^\top \frac{\partial}{\partial s} \mathbf{R}. \quad (3.43)$$

$\tilde{\boldsymbol{\kappa}}$ ist die schief-symmetrische Matrixdarstellung des Krümmungsvektors $\boldsymbol{\kappa}$. Die Drehmatrix \mathbf{R} ist bekannt und kann analytisch differenziert werden, so dass wir die Komponenten von $\boldsymbol{\kappa}$ direkt in EULER-Winkeln ausdrücken können,

$$\kappa_1 = \frac{d\theta}{ds} \sin \psi - \frac{d\phi}{ds} \sin \theta \cos \psi, \quad (3.44)$$

$$\kappa_2 = \frac{d\theta}{ds} \cos \psi + \frac{d\phi}{ds} \sin \theta \sin \psi, \quad (3.45)$$

$$\kappa_3 = \frac{d\psi}{ds} + \frac{d\phi}{ds} \cos \theta. \quad (3.46)$$

Im selben Sinne lassen sich auch die Komponenten des Twistvektors $\boldsymbol{\omega}$ darstellen.

Die aus der Verwendung von EULER-Winkeln resultierenden Singularitätsprobleme lassen sich durch alternative Darstellung der $\mathcal{SO}(3)$ umgehen, z.B. EULER-Parameter oder *Quaternione*. Quaternione sind eine Singularitäten-freie, vierdimensionale zwei-zu-eins Parametrisierung der $\mathcal{SO}(3)$. EULER-Parameter oder *Einheitsquaternione* werden durch ein vier-Tupel $\mathbf{q} = (q_1, q_2, q_3, q_4)^\top$ von reellen Zahlen repräsentiert, das der folgenden Bedingung genügt,

$$\mathbf{q}^2 = q_1^2 + q_2^2 + q_3^2 + q_4^2 = 1. \quad (3.47)$$

Quaternione bilden eine 4-dimensionale, assoziative Algebra über den reellen Zahlen. Der Einheitsquaternion ist eine Rotation mit der normierten Drehachse $\boldsymbol{\Omega} = \Omega_j \mathbf{e}_j$ und dem Drehwinkel θ zugeordnet, so dass $q_j = \Omega_j \sin(\theta/2)$ mit $j = 1, 2, 3$ und $q_4 = \cos(\theta/2)$. Damit erfüllt sie Bedingung Gl. (3.47). Die Rotationsmatrix $\mathbf{R}(\mathbf{q})$ stellt sich wie folgt dar [Dichmann & Maddocks, 1996],

$$\mathbf{R} := \begin{pmatrix} q_1^2 - q_2^2 - q_3^2 + q_4^2 & 2(q_1 q_2 - q_3 q_4) & 2(q_1 q_3 + q_2 q_4) \\ 2(q_1 q_2 + q_3 q_4) & -q_1^2 + q_2^2 - q_3^2 + q_4^2 & 2(q_2 q_3 - q_1 q_4) \\ 2(q_1 q_3 - q_2 q_4) & 2(q_1 q_4 + q_2 q_3) & -q_1^2 - q_2^2 + q_3^2 + q_4^2 \end{pmatrix}. \quad (3.48)$$

Die Komponenten des Vektors $\boldsymbol{\kappa}$ erhalten wir nach derselben Vorgehensweise wie im Falle der EULER-Winkel,

$$\kappa_j = \frac{2}{\mathbf{q}^\top \mathbf{q}} \mathbf{B}_j \mathbf{q} \frac{\partial}{\partial s} \mathbf{q}. \quad (3.49)$$

Hierbei sind die \mathbf{B}_j schief-symmetrische (4×4) -Matrizen, die wir wie folgt definieren,

$$\mathbf{B}_1 := \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{B}_2 := \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, \quad \mathbf{B}_3 := \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \quad (3.50)$$

Durch die Antisymmetrie dieser Matrizen und die Bedingung Gl. (3.47) bilden die Vektoren $\{\mathbf{B}_j \mathbf{q}\}$ für $j = 1, 2, 3$ zusammen mit \mathbf{q} eine orthogonale Basis im \mathbb{R}^4 .

3.2.4 Zwangsbedingungen

In der Regel benötigen wir nicht das volle Spektrum an Deformationen, wie Biegung, Torsion, Scherung und Dehnung. Das hat zwei Gründe; a) auf der einen Seite vereinfacht sich die Handhabung der Bewegungsgleichungen, wenn wir bestimmte Deformationen vernachlässigen; b) auf der anderen Seite hat sich gezeigt, dass manche Vereinfachungen für bestimmte Klassen von Deformationen eine ausreichende Näherung darstellen. Bspw. sind Scherung und Dehnung unter Einwirkung einer Terminallast typischerweise klein. Um Scherdeformationen und Dehnung der Faser zu verhindern, muss folgende Bedingung erfüllt sein,

$$\mathbf{v} = (0, 0, 1), \quad (3.51)$$

unabhängig von der Belastung bzw. den Momente \mathbf{m} und Kräften \mathbf{n} , die in der Faser vorherrschen. Das bedeutet, dass der Vektor \mathbf{v} die Einheitstangente an die Kurve $\mathbf{r}(s, t)$ bildet, weil $\partial_s \mathbf{r} = \mathbf{v} = (0, 0, 1)^\top$. Er steht in diesem Fall orthogonal auf dem Materialquerschnitt. Diese Art der Beschränkung nennen wir *konstitutiv*. Es ist dann eine weitere Gleichung der Form

$$\mathbf{m} = \hat{\mathbf{m}}(\mathbf{k}(s, t), s) \quad (3.52)$$

für die Momente notwendig, die in diesem Fall allerdings nicht mehr von den internen Kräften $\mathbf{n}(s, t)$ abhängen. Die Funktion $\mathbf{n}(s, t)$ darf dabei jeden Wert annehmen, der Bedingung Gl. (3.51) genügt. Gl. (3.51) und Gl. (3.52) bilden zusammen ein System von konstituierenden Gleichungen für nicht scher- und dehnbare Fasern. Durch die Einführung einer Abhängigkeit von der Resultante $\mathbf{n}(s, t)$ lässt sich Gl. (3.52) generalisieren. Für elastische Fasern, die keinerlei Einschränkungen hinsichtlich der Deformierbarkeit unterliegen, erhält man das folgende System von Werkstoffgleichungen,

$$\mathbf{v} = \hat{\mathbf{v}}(\mathbf{k}(s, t), \mathbf{n}(s, t), s), \quad (3.53)$$

$$\mathbf{m} = \hat{\mathbf{m}}(\mathbf{k}(s, t), \mathbf{n}(s, t), s). \quad (3.54)$$

Bedingung Gl. (3.51) bildet den Grenzfall $\hat{\mathbf{v}} \rightarrow (0, 0, 1)$ für Materialien mit dieser Art von Werkstoffgesetzen. $\mathbf{n}(s, t)$ unterliegt dabei keinerlei Einschränkungen, ganz wie es das *Prinzip der Virtuellen Arbeiten* fordert. Im Grenzfall $\hat{\mathbf{v}} \rightarrow (0, 0, 1)$ reduziert sich bei hyperelastischen Fasern Gl. (3.54) zu Gl. (3.52), vgl. [Antman, 1995].

In allen Fällen ist \mathbf{n} eine grundlegende Unbekannte. Die Bedingung Gl. (3.51) ist dann notwendig, damit für jede Unbekannte eine korrespondierende Gleichung existiert.

Restriktionen der Art Gl. (3.51) lassen sich also über die Werkstoffgesetze bequem in den vorgestellte Rahmen integrieren.

3.2.5 Theorie von Kirchhoff

Im Prinzip lassen sich auch die klassischen Theorien, wie die von KIRCHHOFF [Kirchhoff, 1859], Timoshenko [Timoshenko, 1921] oder EULER-BERNOULLI [Bernoulli, 1728] aus dem dargestellten Framework ableiten. Die Theorie von KIRCHHOFF [Kirchhoff, 1859] nimmt in dieser Hinsicht eine Sonderstellung ein, da sie auf der Annahme einer nicht scherbaren bzw. nicht dehnbaren elastischen Faser basiert und damit Bedingung Gl. (3.51) erfüllt. Zusätzlich haben wir ein Werkstoffgesetz der Form

$$\mathbf{m} = \hat{\mathbf{m}}(\boldsymbol{\kappa}(s, t), s) = \mathbf{K}(s) \Delta \boldsymbol{\kappa}(s, t) = \mathbf{K}(s) (\boldsymbol{\kappa}(s, t) - \boldsymbol{\kappa}^\circ(s)), \quad (3.55)$$

welches die Faser als elastisch charakterisiert und im Prinzip nichts anderes darstellt, als das klassische HOOKSche Gesetz, mit der Steifigkeitsmatrix $\mathbf{K}(s)$,

$$\mathbf{K}(s) = \begin{pmatrix} EI_1(s) & 0 & 0 \\ 0 & EI_2(s) & 0 \\ 0 & 0 & G\mu(s) \end{pmatrix}. \quad (3.56)$$

Ausgeschrieben erhalten wir für das Moment entlang der Faser

$$\mathbf{m}(s, t) = EI_1(s) (\kappa_1 - \kappa_1^\circ) \mathbf{d}_1 + EI_2(s) (\kappa_2 - \kappa_2^\circ) \mathbf{d}_2 + G\mu(s) (\kappa_3 - \kappa_3^\circ) \mathbf{d}_3. \quad (3.57)$$

3.2.6 Bewegungsgleichungen der Faser

Was sind nun die Konsequenzen der Unscherbarkeit bzw. der Nichtdehnbarkeit in Form von Gl. (3.51)? Die Tangente fällt nun mit der Richtung \mathbf{d}_3 zusammen, die wir als zum Materialquerschnitt orthogonal stehend definiert haben, $\partial_s \mathbf{r}(s, t) = \mathbf{d}_3(s, t)$. In unserem numerischen Ansatz zur Lösung der dynamischen COSSERAT-Gleichungen erzwingen wir diese Bedingung explizit durch Einführung einer weiteren partiellen Differentialgleichung. Wegen der Kontinuitätsbedingung

$$\frac{\partial}{\partial s} \left(\frac{\partial}{\partial t} \mathbf{r}(s, t) \right) = \frac{\partial}{\partial t} \left(\frac{\partial}{\partial s} \mathbf{r}(s, t) \right), \quad (3.58)$$

erhalten wir

$$\begin{aligned} \frac{\partial}{\partial s} \left(\frac{\partial}{\partial t} \mathbf{r}(s, t) \right) &= \frac{\partial}{\partial t} \mathbf{d}_3(s, t), \\ &= \boldsymbol{\omega}(s, t) \times \mathbf{d}_3(s, t). \end{aligned} \quad (3.59)$$

Die zeitliche Änderung der Tangente in Bezug auf die lokale Basis muss verschwinden (vgl. Gl. (3.24)),

$$\frac{\partial}{\partial s} \left(\frac{\partial}{\partial t} \mathbf{r}(s, t) \right) + \boldsymbol{\kappa} \times \frac{\partial}{\partial s} \mathbf{r}(s, t) = \boldsymbol{\omega}(s, t) \times \mathbf{d}_3(s, t). \quad (3.60)$$

Wir führen an dieser Stelle die Mittelliniengeschwindigkeit $\mathbf{v} = \partial_t \mathbf{r}(s, t)$ ein. Unter Einbezug der Materialgesetze können wir das finale PDE-System angeben, welches die Dynamik einer nicht

scherbaren und nicht dehnbaren Faser beschreibt. Dieses System bildet gleichzeitig die Grundlage für die folgende Diskussion numerischer Lösungsstrategien für die stationären sowie instationären COSSERAT-Gleichungen. Wir schreiben das System in Komponentendarstellung,

$$\rho A \partial_t v_1 = \partial_s n_1 + \kappa_2 n_3 - \kappa_3 n_2 - \rho A (\omega_2 v_3 - \omega_3 v_2) + f_1, \quad (3.61a)$$

$$\rho A \partial_t v_2 = \partial_s n_2 + \kappa_3 n_1 - \kappa_1 n_3 - \rho A (\omega_3 v_1 - \omega_1 v_3) + f_2, \quad (3.61b)$$

$$\rho A \partial_t v_3 = \partial_s n_3 + \kappa_1 n_2 - \kappa_2 n_1 - \rho A (\omega_1 v_2 - \omega_2 v_1) + f_3, \quad (3.61c)$$

$$\rho I_1 \partial_t \omega_1 = \partial_s m_1 + \kappa_2 m_3 - \kappa_3 m_2 - n_2 - \rho \omega_2 \omega_3 (I_2 + I_3) + l_1, \quad (3.62a)$$

$$\rho I_2 \partial_t \omega_2 = \partial_s m_2 + \kappa_3 m_1 - \kappa_1 m_3 + n_1 - \rho \omega_1 \omega_3 (I_1 + I_3) + l_2, \quad (3.62b)$$

$$\rho I_3 \partial_t \omega_3 = \partial_s m_3 + \kappa_1 m_2 - \kappa_2 m_1 - \rho \omega_1 \omega_2 (I_1 + I_2) + l_3, \quad (3.62c)$$

$$\partial_t \kappa_1 = \partial_s \omega_1 - \omega_2 \kappa_3 + \omega_3 \kappa_2, \quad (3.63a)$$

$$\partial_t \kappa_2 = \partial_s \omega_2 - \omega_3 \kappa_1 + \omega_1 \kappa_3, \quad (3.63b)$$

$$\partial_t \kappa_3 = \partial_s \omega_3 - \omega_1 \kappa_2 + \omega_2 \kappa_1, \quad (3.63c)$$

$$0 = \partial_s v_1 + \kappa_2 v_3 - \kappa_3 v_2 - \omega_2, \quad (3.64a)$$

$$0 = \partial_s v_2 + \kappa_3 v_1 - \kappa_1 v_3 + \omega_1, \quad (3.64b)$$

$$0 = \partial_s v_3 + \kappa_1 v_2 - \kappa_2 v_1. \quad (3.64c)$$

Dabei gilt nach Gl. (3.55) bei nicht uniformen Fasern mit variablen Materialeigenschaften entlang der Faser,

$$\partial_s \mathbf{m} = \partial_s \mathbf{K}(\boldsymbol{\kappa} - \boldsymbol{\kappa}^\circ) + \mathbf{K}(\partial_s \boldsymbol{\kappa} - \partial_s \boldsymbol{\kappa}^\circ). \quad (3.65)$$

Bei konstanten Materialeigenschaften vereinfacht sich der Ausdruck zu $\mathbf{K}(\partial_s \boldsymbol{\kappa} - \partial_s \boldsymbol{\kappa}^\circ)$. Die obigen Gleichungen schreibt man in der kompakteren Standardnotation für strukturmechanische Probleme als

$$\widehat{\mathbf{M}} \partial_t \mathbf{x}(s, t) + \widehat{\mathbf{K}} \partial_s \mathbf{x}(s, t) + \boldsymbol{\Lambda}(s, t) = \mathbf{0}, \quad (3.66)$$

wobei $\mathbf{x} \in \mathbb{R}^{12}$ der Zustandsvektor der Faser ist, mit $\mathbf{x}(s, t) = \{\mathbf{v}(s, t), \boldsymbol{\omega}(s, t), \boldsymbol{\kappa}(s, t), \mathbf{n}(s, t)\}^\top$, der Massematrix $\widehat{\mathbf{M}} = \text{diag}(\rho A, \rho A, \rho A, \rho I_1, \rho I_2, \rho I_3, 1, 1, 1, 0, 0, 0)$ und der Steifigkeitsmatrix $-\widehat{\mathbf{K}} = \text{adiag}(\mathbf{E}_3, \mathbf{E}_3, \mathbf{K}, \mathbf{E}_3)$, mit \mathbf{K} entsprechend Gl. (3.55). Eine gewisse Vorsicht ist geboten, da die Massematrix keinen vollen Rang hat; die Zeitableitung von \mathbf{n} existiert nicht. Alle verbleibenden

Terme sammeln wir in $\Lambda(s, t)$,

$$\Lambda(s, t) = \begin{pmatrix} \kappa_2 n_3 - \kappa_3 n_2 - \rho A (\omega_2 v_3 - \omega_3 v_2) + f_1 \\ \kappa_3 n_1 - \kappa_1 n_3 - \rho A (\omega_3 v_1 - \omega_1 v_3) + f_2 \\ \kappa_1 n_2 - \kappa_2 n_1 - \rho A (\omega_1 v_2 - \omega_2 v_1) + f_3 \\ \kappa_2 m_3 - \kappa_3 m_2 - n_2 - \rho \omega_2 \omega_3 (I_2 + I_3) + l_1 \\ \kappa_3 m_1 - \kappa_1 m_3 + n_1 - \rho \omega_1 \omega_3 (I_1 + I_3) + l_2 \\ \kappa_1 m_2 - \kappa_2 m_1 - \rho \omega_1 \omega_2 (I_1 + I_2) + l_3 \\ -\omega_2 \kappa_3 + \omega_3 \kappa_2 \\ -\omega_3 \kappa_1 + \omega_1 \kappa_3 \\ -\omega_1 \kappa_2 + \omega_2 \kappa_1 \\ \kappa_2 v_3 - \kappa_3 v_2 - \omega_2 \\ \kappa_3 v_1 - \kappa_1 v_3 + \omega_1 \\ \kappa_1 v_2 - \kappa_2 v_1 \end{pmatrix}. \quad (3.67)$$

An dieser Stelle bleibt anzumerken, dass die Nichtdehnbarkeitsbedingung unter numerischen Gesichtspunkten als kritisch angesehen werden muss, da sie das PDE-System „versteift“. Sie übernimmt die Rolle einer unendlich steifen Feder, so dass für die numerische Integration der Einsatz impliziter Verfahren unumgänglich ist.

3.2.7 Allgemeine Theorien

Die spezielle COSSERAT-Theorie der Stäbe, die wir in den vorangegangenen Abschnitten diskutiert haben, beschreibt die Deformation einer Faser anhand der drei vektorwertigen Funktionen $\mathbf{r}(s, t)$, $\mathbf{d}_1(s, t)$ und $\mathbf{d}_2(s, t)$, wobei die beiden letzten den Querschnitt an der Stelle $\mathbf{r}(s, \cdot)$ aufspannen. Sie muss jedoch wegen der Beschränkung auf nur zwei Basisvektoren, auf welche die physikalischen Qualitäten bezogen werden, und der Orthogonalitätsforderung als eingeschränkte Version der allgemeinen COSSERAT-Theorie gelten. Gibt man letztere Bedingung auf, führt das auf die vollständige 2-Direktor COSSERAT-Theorie, die eine Generalisierung des Spezialfalls darstellt. Scherung und Dehnung lassen sich damit einfacher beschreiben.

Eine allgemeine Theorie muss nicht notwendig auf zwei Basisvektoren beschränkt bleiben. Im Grunde genommen lassen sich für jede zusätzliche dreidimensionale Größe des Querschnitts weitere Vektoren hinzufügen. Das Hauptproblem besteht jetzt allerdings darin, dass durch den Wegfall der Orthogonalität auch kein DARBOUX-Vektor $\boldsymbol{\kappa}$ mehr existiert, dem eine so zentrale Rolle in der speziellen COSSERAT-Theorie zugeordnet war, so dass sich die notwendigen Gleichungen zur Beschreibung des Deformationsvorganges nicht einfach in Analogie zur speziellen Theorie herleiten lassen. Um das vollständige System von Bewegungsgleichungen herzuleiten, ist dann entweder der Apparat der dreidimensionalen Kontinuumsmechanik zu bemühen oder man gewinnt die Gleichungen im Kontext des Prinzips der virtuellen Arbeiten. Letzteres entspricht bei konservativen Systemen der Anwendung des HAMILTON-Prinzips. Der grosse Vorteil der speziellen COSSERAT-Theorie ist in diesem Zusammenhang darin zu sehen, dass die physikalische Bedeutung der Basisgrößen \mathbf{n} und \mathbf{m} der Intuition leicht zugänglich sind, wohingegen die aus der Anwendung physikalischer Prinzipien wie dem der virtuellen Arbeiten resultierenden Gleichungen sich oft der Anschauung entziehen.

Wir denken, dass die eindimensionale COSSERAT-Theorie, die wir in den vorangegangenen Abschnitten vorgestellt haben, ein adäquates Instrumentarium zur Bewältigung des vorliegenden Problems, nämlich der akkuraten physikalisch basierten Simulation von Humanhaar darstellt. Um Vorwände zu entkräften, die KIRCHHOFF-Theorie könne doch im genannten Kontext dasselbe leisten, zumal der hier betrachtete Sonderfall einer nicht scherbaren und nicht dehnbaren Faser der speziellen COSSERAT-Theorie genau equivalent dazu sei, möchten wir anführen, dass wir uns mit der Wahl dieses Rahmens die Möglichkeit der Modellierung von nicht linearen Effekten wie bspw. viskoelastischem Verhalten, sowie Scherdeformationen offenhalten. Diese Phänomene sind durch Wahl geeigneter Werkstoffgesetze leicht einzuarbeiten. Bspw. konnte Schuricht in seinem Beitrag zeigen [Schuricht, 1998], dass die Möglichkeit zur Scherdeformationen im Zusammenhang mit Kontaktphänomenen einen wesentlichen Einfluss auf die Kontaktgeometrie hat. Die Relevanz solcher Effekte auf das Gesamterscheinungsbild einer Ansammlung von Haarfasern kann nicht von vornherein ausgeschlossen werden.

Historisch gehen die ersten Formulierungsversuche solcher physikalischer Theorien zurück auf Jacob BERNOULLI [Bernoulli, 1694]. In seiner *Curvatura Laminae Elasticae* beschreibt er die planare Deformation eines elastischen Stabes durch Betrachtung des Momentengleichgewichtes unter Einbeziehung eines nicht linearen Werkstoffgesetzes. Euler [Euler, 1771a] hat die klassische lineare Beziehung zwischen Biegemoment und Änderung in der Krümmung eingeführt. Dazu nahm er an, dass die Dehnung bei der Biegung linear über dem Querschnitt variiert und HOOKS Gesetz folge. Ein Werkstoffgesetz fand er durch Integration der Momente über den Stabquerschnitt. In seiner endgültigen Form, dem BERNOULLI-EULERSchen Biegegesetz, wurde es dann von Daniel BERNOULLI [Bernoulli, 1728] veröffentlicht. Später führte EULER auch die Idee der Scherdeformationen im Kontext seiner vollständigen Theorie zur Behandlung dynamisch-planarer Probleme ein [Euler, 1771b], [Euler, 1774]. Die erste geometrisch exakte Theorie, welche die räumliche Deformation zu beschreiben in der Lage war, ist jene von KIRCHHOFF [Kirchhoff, 1859], auch bekannt als KIRCHHOFFSche kinetische Analogie, die u.a. durch CLEBSCH [Clebsch, 1862] und LOVE [Love, 1893] Verbesserung erfuhr. Eine Generalisierung zur speziellen COSSERAT-Theorie der Stäbe erfolgte dann 1907 durch die Brüder COSSE-RAT [Cosserat & Cosserat, 1907b] unter Anwendung von Prinzipien der Variationsrechnung. Es folgten eine ganze Reihe von Weiterentwicklungen bzw. Neuansätzen. Eine zusammenfassende Darstellung finden wir schließlich bei Antman [Antman, 1995].

3.2.8 Verwandte Ansätze

Wir gehen an dieser Stelle kurz auf einen der neueren Ansätze zur Simulation von Fasern ein, das so genannte Superhelixmodell [Bertails et al., 2006]². Der Name findet seinen Ursprung in der Tatsache, dass die Mittelliniengeometrie der Faser durch Helices mit stückweise konstanter Krümmung und Torsion approximiert wird. An den Übergangsstellen sind sowohl die Funktion der Mittellinie als auch die Orientierung stetig differenzierbar.

In diesem Modell werden nicht die COSSERAT-Gleichungen direkt integriert, sondern man bedient sich zur Herleitung der Bewegungsgleichungen des LAGRANGE-Formalismus, vgl. Abs. (A.2). Mit den generalisierten Koordinaten $\mathbf{q}(s, t)$ seien der DARBOUX-Vektor $\boldsymbol{\kappa} = \boldsymbol{\kappa}(s, t)$

²Der Begriff der Superhelix stammt ursprünglich aus der Molekularbiologie und bezeichnet dort eine helikale Struktur mit aufmodulierter Helix.

und die Mittellinie $\mathbf{r} = \mathbf{r}(s, t)$ von nun an Funktionen von s und \mathbf{q} , also $\boldsymbol{\kappa} = \boldsymbol{\kappa}(s, \mathbf{q})$ und $\mathbf{r} = \mathbf{r}(s, \mathbf{q})$. Die kinetische bzw. potentielle Energie T und U und das Dissipationspotential Y infolge materialinterner Reibung der Faser sind dann durch folgende Ausdrücke gegeben,

$$T(\mathbf{q}, \partial_t \mathbf{q}, t) = \frac{1}{2} \rho A \int_0^L \partial_t \mathbf{r}^2(s, \mathbf{q}) ds, \quad (3.68)$$

$$U(\mathbf{q}, t) = \frac{1}{2} \int_0^L \Delta \boldsymbol{\kappa}^\top(s, \mathbf{q}) \mathbf{K}(s) \Delta \boldsymbol{\kappa}(s, \mathbf{q}) ds - \frac{1}{2} \rho A \int_0^L \mathbf{g}^\top \mathbf{r}(s, \mathbf{q}) ds, \quad (3.69)$$

$$Y(\mathbf{q}, t) = \frac{1}{2} \int_0^L \partial_t \boldsymbol{\kappa}^\top(s, \mathbf{q}) \mathbf{D}^I(s) \partial_t \boldsymbol{\kappa}(s, \mathbf{q}) ds + \frac{1}{2} \int_0^L \partial_t \mathbf{q}^\top \mathbf{D}^E(s) \partial_t \mathbf{q} ds. \quad (3.70)$$

Sie folgen direkt aus der Anschauung. $\mathbf{K}(s)$ ist die Steifigkeitsmatrix, die hier Diagonalform annimmt; $\mathbf{D}^{(\cdot)}(s)$ sind die Matrizen für interne und externe Dämpfung. $\Delta \boldsymbol{\kappa} = (\boldsymbol{\kappa} - \boldsymbol{\kappa}^\circ)$ ist die Differenz aus aktueller Krümmung und der Anfangskrümmung der Faser. Die lineare Dichte $\rho(s)A(s)$ kann als konstant angenommen werden, so dass auch \mathbf{K} nicht von s abhängt; $\partial_t \boldsymbol{\kappa} = \boldsymbol{\kappa}/\Delta t$ sei die zeitliche Änderung der Krümmung und \mathbf{g} der Gravitationsvektor. Da am System nicht konservative Kräfte angreifen, z.B. $\mathbf{Q} = \mathbf{Q}(s, t)$ infolge Kollision mit anderen Körpern, benötigt man noch die generalisierten Kräfte $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\mathbf{q}, t)$,

$$\boldsymbol{\Lambda}(\mathbf{q}, t) = - \int_0^L \mathbf{J}(s, \mathbf{q}, t) \mathbf{Q}(s, t) ds. \quad (3.71)$$

Hierbei ist $\mathbf{J}(s, \mathbf{q}, t) = \partial \mathbf{r}(s, \mathbf{q})/\partial \mathbf{q}$ die Jacobimatrix, die alle Kräfte auf die Richtungen der generalisierten Koordinaten projiziert (diese werden in der Regel in kartesischen Koordinaten angegeben). Ausgehend von Gl. (A.5) für dissipative Systeme können die oben abgeleiteten Energie- und Kraftterme jetzt eingesetzt und die Bewegungsgleichungen für die Faser hergeleitet werden. Sie folgen damit direkt aus dem Prinzip der kleinsten Wirkung. Die resultierenden Gleichungen sind einer analytischen Lösung nicht mehr zugänglich, so dass auf numerische Methoden zurückgegriffen werden muss. Man überführt Gl. (A.5) durch einige algebraische Manipulationen in die Matrixform für lineare Probleme der Strukturmechanik (vgl. Gl. (3.136)),

$$\mathbf{M} \partial_{tt} \mathbf{q}(s, t) + \mathbf{D} \partial_t \mathbf{q}(s, t) + \mathbf{K} \mathbf{q}(s, t) + \boldsymbol{\Lambda}(\mathbf{q}, t) = \mathbf{0}. \quad (3.72)$$

Numerische Lösungsverfahren fordern eine räumliche und zeitliche Diskretisierung von Gl. (3.72). Zur räumlichen Diskretisierung unterteilt man die Faser in N Segmente s_i mit nicht notwendig gleicher Länge, aber konstanter Krümmung. Eine der wesentlichen Ideen des Superhelixmodells ist die Einbindung einer Auswahlfunktion für die Krümmung, in der Form

$$\boldsymbol{\kappa}(s, \mathbf{q}) = \sum_{i=1}^N \mathbf{q}_i(t) \sigma_i(s), \quad \text{mit } \sigma_i(s) = \begin{cases} 1, & \text{falls } s \in s_i \\ 0 & \text{sonst.} \end{cases} \quad (3.73)$$

Die Funktion $\sigma(s)$ nimmt den Wert eins an, falls der Funktionsparameter s im Intervall $[s_i^l, s_i^u]$ liegt, andernfalls den Wert null. D.h., es wird die konstante Krümmung des i 'ten Segments ausgewählt, solange sich der Parameter s in dem zugehörigen Intervall bewegt. Diese Art der Darstellung mag auf den ersten Blick etwas seltsam anmuten, hat aber den Vorteil, dass dieser Ausdruck

direkt in die oben hergeleiteten Gleichungen eingesetzt und in symbolischer Form weiterverarbeitet werden kann. Mit diesem Ausdruck wird Gl. (3.72) in die semidiskrete Form überführt. Für die zeitliche Diskretisierung kann man den klassischen NEWMARK-Integrator verwenden oder das systematischere α -Verfahren, vgl. Abs. (3.4.1).

Weiterhin gelten $\partial_s \mathbf{r} = \int_0^L \mathbf{d}_3(s, t) ds$ und $\partial_s \mathbf{d}_3 = \boldsymbol{\kappa} \times \mathbf{d}_3$, wobei die Komponenten von $\boldsymbol{\kappa}$ auf die lokale Basis $\{\mathbf{d}_i(s, t)\}$ bezogen werden. Mit Gl. (3.162) und Gl. (3.73) lässt sich daraus ein expliziter Lösungsausdruck für $\mathbf{r}(s, \mathbf{q})$ ableiten, auf dessen Angabe wir an dieser Stelle verzichten.

Die numerische Integration liefert den Vektor der generalisierten Koordinaten $\mathbf{q}(s, t)$, die den Krümmungen entsprechen. Die Kurve zum Zeitpunkt t folgt direkt mit dem Lösungsausdruck für $\mathbf{r}(s, \mathbf{q})$.

Die Nichtdehnbarkeitsbedingung, die zu den bekannten Steifigkeitsproblemen während des numerischen Integrationsprozesses führt, ist durch die reduzierte Koordinatendarstellung implizit erfüllt. Allerdings bleibt das Gleichungssystem steif.

So vorteilhaft die Idee mit der Approximation durch Segmente mit stückweise konstanter Krümmung und anschließender expliziter Integration der kinematischen Gleichungen auch scheinen mag, sie birgt das Problem, dass sie auf Umwegen zu einer dicht besetzten Massematrix führt. Diese ist eine direkte Folge der reduzierten Koordinatendarstellung. Eine Alternative besteht darin, die Nichtdehnbarkeitsbedingung als Zwang mit in das System zu integrieren, und dafür die Systemkoordinaten zu maximieren. Die Rechenaufwand beim Lösen der Bewegungsgleichungen ist quadratisch in der Zahl der Segmente, $\mathcal{O}(N^2)$. Die räumliche Auflösung der Kurve muss infolgedessen immer auf sehr grobe Gitter beschränkt bleiben. Das kann zu Problemen führen, wenn Interaktionskräfte infolge Kollision oder elektrostatischer Natur zwischen den Gitterpunkten aufgebracht werden müssen. Die adaptive Verfeinerung kann dann zu indiskutablen Rechenzeiten führen. Diese Probleme werden mit dem von uns propagierten Lösungsverfahren, das auf dem generalisierten α -Verfahren beruht, eliminiert.

Einen anderen Weg beschreiten die Autoren der COSSERAT-Rod-Elemente Methode, vgl. [Spillmann & Teschner, 2007]. Obwohl der Begriff bereits durch Cao et al. [Cao et al., 2006] vorgelegt ist, wird er hier für ein völlig anderes Verfahren umdefiniert, das ganz analog zum Superhelixmodell auf der Anwendung des LAGRANGE-Formalismus beruht. Drei Hauptunterschiede seien hier aufgezeigt: a) die Autoren verwenden zu Parametrisierung der Rotationen Quaternionen und greifen dabei auf Vorarbeiten von Dichmann [Dichmann, 1994] zurück; b) Dehnungen der Faser sind erlaubt, Scherungen hingegen nicht. Damit entledigt man sich elegant der Steifigkeitsprobleme, die beim Verhindern der longitudinalen Dehnungen entstehen; c) Die Nichtscherbarkeitsbedingung wird hier unter Verwendung von LAGRANGE-Multiplikatoren erzwungen. Die Definitionen der Energierterme (kinetische, potentielle und Dissipation) sind im Wesentlichen mit denen aus dem vorangegangenen Abschnitt identisch, müssen hier also nicht wiederholt werden. Die Zwangskräfte an den Knoten hingegen fordern unsere besondere Aufmerksamkeit und wir wollen deren Herleitung hier kurz angeben. Die Nichtscherbarkeitsbedingung lässt sich durch eine Zwangsbedingung der Form

$$\mathbf{C} = \mathbf{d}_3^T(s) \partial_s \mathbf{r}_{\parallel}(s) - 1 = \mathbf{0} \quad (3.74)$$

ausdrücken, wobei $\partial_s \mathbf{r}_{\parallel}(s)$ die Einheitstangente an der Stelle s und \mathbf{d}_3 der Einheitsvektor orthogonal zum Querschnitt der Faser ist. Das Einbeziehen von Orientierungen in die Energierterme

führt außerdem zu der Forderung, dass die Darstellung Element der $\mathcal{SO}(3)$ ist. Bei der Verwendung von Quaternionen ist dadurch eine zusätzliche Zwangsbedingung der Form $\mathbf{C}_2 = \|\mathbf{p}\|^2 = 1$ notwendig. Interessanterweise konnten Healey und Mehta [Healey & Mehta, 2003] in diesem Zusammenhang zeigen, dass - zumindest für die statischen Gleichungen - durch Anwendung des LJAPUNOVschen Zentraltheorems das System so erweitert werden kann, dass bei Einhaltung der Einheitsbedingung auf dem Rand des Integrationsgebietes, diese automatisch an jeder Stelle dazwischen erfüllt ist.

Wenn man den Vektor der Zwangsbedingungen als

$$\partial\mathbf{C}/\partial\mathbf{q}_i = (\partial\mathbf{C}_1/\partial\mathbf{q}_i, \dots, \partial\mathbf{C}_n/\partial\mathbf{q}_i)^\top$$

definiert, nehmen die LAGRANGESchen Gleichungen die folgende finale Form an,

$$\{L\}_{\partial_i\mathbf{q},\mathbf{q}} + \lambda^\top \frac{\partial\mathbf{C}}{\partial\mathbf{q}} = \mathbf{Q}, \quad (3.75)$$

wobei λ die schon bekannten LAGRANGE-Multiplikatoren in vektorieller Notation darstellen. Die generalisierten Koordinaten sind dann gegeben als $\mathbf{q} = \{r_1, r_2, r_3; p_1, p_2, p_3, p_4\}$. Auf der Basis der Energierterme kann man durch Integration über die Länge eines Elementes diskrete Energien berechnen. Dazu macht man die Approximation $\partial_s \mathbf{r}_i \approx (\mathbf{r}_i - \mathbf{r}_{i-1})/\Delta s_i^0$ und $\partial_s \mathbf{p}_i \approx 2(\mathbf{p}_i - \mathbf{p}_{i-1})/(\Delta s_{i+1}^0 + \Delta s_i^0)$, wobei $\Delta s_i^0 = \|\mathbf{r}_i^0 - \mathbf{r}_{i-1}^0\|$ die Anfangslänge zwischen den Knoten (i) und ($i-1$) darstellt. Da die Verschiebungen nur an den Knoten gegeben sind, wählt man, wie in der Finite-Elemente-Methode üblich, Formfunktionen zur Approximation der Schnittgrößen auf dem Gebiet dazwischen. Diese fallen hier vergleichsweise einfach aus, weil lediglich zwischen zwei Knoten linear zu interpolieren ist. Eine konstante Formfunktion der Art $\mathbf{b}(\sigma) = (\mathbf{a}_i + \mathbf{a}_{i-1})/2$, sowohl für die Position \mathbf{r} als auch für die Orientierung \mathbf{p} , ist nach Meinung der Autoren völlig ausreichend. Durch Einsetzen dieser diskreten Energierterme in Gl. (3.75) erhält man die Bewegungsgleichungen für das COSSERAT-Element. Um die Gesamtdeformation der Faser zu berechnen, sind alle Elementgleichungen simultan zu lösen. Leider weichen die Autoren aus Performancegründen von ihrem eingangs beschriebenen Konzept ab, die Nichtscherbarkeit der Faser durch Zwangsbedingungen in den LAGRANGE-Formalismus zu integrieren und wechseln stattdessen zu einer Energiefunktion der Form

$$\mathcal{E} = \frac{c}{2} \int_0^L [\mathbf{d}_3^\top(s) \partial_s \mathbf{r}_{\|\cdot\|}(s) - 1]^2 ds \quad (3.76)$$

über, die sich in den Term der potentiellen Energie integriert. Dabei ist c eine beliebige Federkonstante. Mit solchen physikalisch nicht zu rechtfertigenden Faktoren degeneriert das Modell wieder zum ad-hoc-Ansatz. Die Autoren rechtfertigen diese Entscheidung mit ihrem Ziel eines lediglich physikalisch plausiblen Verhaltens der Struktur. Und in der Tat, Beispiele der Autoren zeigen, dass das Modell an einer Art „Wackelpudding“-Verhalten leidet.

3.3 Lösung der statischen Cosserat-Gleichungen

Im Folgenden entwickeln wir Lösungsverfahren für die statischen KIRCHHOFF-Gleichungen. Diese besitzen für bestimmte Fälle eine geschlossene Lösung, und zwar in Form JACOBIScher

elliptischer Funktionen. Dazu machen wir von der Analogie zwischen den statischen KIRCHHOFF-Gleichungen und unter bestimmten Bedingungen integrierbaren Starrkörpersystemen Gebrauch. Integrierbar bedeutet in diesem Zusammenhang, dass sich die Lösungen der Bewegungsgleichungen auf bekannte Funktionen oder Integrale zurückführen lassen, bspw. die besagten elliptischen Funktionen.

Wir werden zeigen, wie man diese Lösungen herleitet und für die Modellierung von langen, dünnen Strukturen wie Fasern oder Schläuchen in CAD-Systemen im Bereich der Computergraphik nutzen kann. Insbesondere konzentrieren wir uns auf Randwertprobleme.

3.3.1 Symbolische Lösung

Allgemein gilt, dass die statischen Gleichgewichtsbedingungen einer Faser, welche die Uniformitätseigenschaft besitzt, d.h., bei der sowohl die Anfangskrümmungen $\kappa^\circ = 0$ als auch die Abhängigkeit der Biege- bzw. Torsionssteifigkeit vom Parameter s verschwinden, zu den EULER-POISSONschen-Bewegungsgleichungen eines Starrkörpers äquivalent sind, der sich, an einem Punkt frei drehbar gelagert, im Gravitationsfeld bewegt. Aus diesem Grund wird die statische KIRCHHOFF-Theorie auch als KIRCHHOFFsche kinetische Analogie bezeichnet. Für bestimmte Fälle, z.B., wenn zwei der drei Hauptträgheitsmomente identisch sind, lassen sich die Bewegungsgleichungen in geschlossener Form durch JACOBI-Elliptische Funktionen ausdrücken.

Wir betrachten zunächst einen Starrkörper im Gravitationsfeld, der an einem beliebigen Punkt befestigt ist, so dass die Achse vom Schwerpunkt zum Befestigungspunkt mit einer der drei Hauptträgheitsachsen zusammenfällt. Zusätzlich fordern wir Symmetrie bzgl. dieser Achse. D.h., dass die Trägheitsmomente senkrecht dazu identisch sein müssen. Das einfachste Beispiel ist der symmetrische oder LAGRANGESche Drehkreisel. In diesem Fall sind die EULER-POISSON-Gleichungen, welche die Bewegung dieses Körpers beschreiben, äquivalent zu den statischen KIRCHHOFF-Gleichungen. Wir können dann Korrespondenzen zwischen den Variablen beider Systeme von Bewegungsgleichungen definieren. Bspw. findet der DARBOUX-Vektor κ seine Entsprechung in der Winkelgeschwindigkeit ω des Starrkörpers, der Kurvenparameter s entspricht der Zeit, das Moment \mathbf{m} dem Drehimpuls \mathbf{L} , die Tangente \mathbf{d}_3 dem Einheitsvektor vom Schwerpunkt zum Befestigungspunkt, usw. Eine Zusammenstellung aller Korrespondenzen ist in Tab. (3.1) gegeben. Aus der Symmetrieforderung resultiert die Gleichheit der Biegesteifigkeiten bzgl. der beiden Hauptachsen \mathbf{d}_1 und \mathbf{d}_2 . Das entspricht gerade dem Fall einer Faser mit Kreisquerschnitt, der als LAGRANGE-Fall bekannt ist. Ein weiterer berühmter integrierbarer Fall, der auf die russische Mathematikerin Sofja KOWALEWSKAJA zurückgeht und für dessen Entdeckung sie 1888 den BORDIN-Preis der Pariser Akademie der Wissenschaften erhielt, ist jener, bei dem das Trägheitsmoment entlang einer vom Fixpunkt ausgehenden Achse genau halb so groß ist, wie jene senkrecht dazu, also $2I_3 = I_1 = I_2$, und bei dem der Schwerpunkt genau in der von I_1 und I_2 aufgespannten Ebene liegt. Allerdings nimmt die POISSON-Zahl unter diesen Voraussetzungen negative Werte an. Das entspricht einem Material, das sich bei Kompression senkrecht zur Belastungsachse zusammenzieht, auf Dehnung dagegen mit einer Volumenvergrößerung reagiert. Ein solches Verhalten ist in der Regel bei natürlichen Materialien nicht zu beobachten. Ein drittes integrierbares System bildet der Starrkörper im freien Fall (EULER-Fall). Da in den Korrespondenzen die Resultante \mathbf{n} einer Kraft mit der Richtung der Gravitation aber entgegengesetztem Vorzeichen entspricht, der Körper sich aber andererseits im freien Fall befindet, muss die Resultante

Symbol	Bedeutung Faser	Bedeutung Starrkörper
s	Kurvenparameter	Zeit
$\{\mathbf{d}_i\}$	Koordinatensystem entlang der Faser	körperfestes Koordinatensystem
\mathbf{d}_3	Einheitstangente	Einheitsvektor vom Fixpunkt zum Schwerpunkt
$\boldsymbol{\kappa}$	Darboux-Vektor	Vektor der Winkelgeschwindigkeit
\mathbf{n}	Spannung	Gewichtskraft (mit negativem Vorzeichen)
\mathbf{m}	Drehmoment	Drehimpuls
$EI_{1,2}$	Biegesteifigkeit in Richtung der Querschnittshauptachsen	Hauptträgheitsmomente orthogonal zu \mathbf{d}_3
$G\mu$	Torsionssteifigkeit	Hauptträgheitsmoment entlang \mathbf{d}_3
a	Verhältnis der Biegesteifigkeiten	Verhältnis der Hauptträgheitsmomente orthogonal zu \mathbf{d}_3
b	skalierte Torsionssteifigkeit	skaliertes Hauptträgheitsmoment entlang \mathbf{d}_3

Tabelle 3.1: Korrespondenzen zwischen der Faser der Kirchhoffschen kinetischen Analogie und dem starren Körper.

verschwinden. Die Lösungen solcher Systeme beschreiben Helices [Nizette & Goriely, 1999].

3.3.1.1 Symbolische Integration

Zur Herleitung der Lösung für den LAGRANGE-Fall, $I_1 = I_2$, schreiben wir die statischen KIRCHHOFF-Gleichungen in einer „abgespeckten“ Form, ohne extern wirkende Kräfte und Momente,

$$\frac{d\mathbf{n}}{ds} = \mathbf{0}, \quad (3.77)$$

$$\frac{d\mathbf{m}}{ds} + \mathbf{d}_3 \times \mathbf{n} = \mathbf{0}. \quad (3.78)$$

Das Materialgesetz Gl. (3.57) dividiert man wegen $I_1 = I_2$ durch EI und erhält mit $b = G\mu/EI$ die skalierte Form der Momentengleichung Gl. (3.55),

$$\mathbf{m} = (\kappa_1 - \kappa_1^\circ) \mathbf{d}_1 + (\kappa_2 - \kappa_2^\circ) \mathbf{d}_2 + b (\kappa_3 - \kappa_3^\circ) \mathbf{d}_3. \quad (3.79)$$

Kehrbaum und Maddocks [Kehrbaum & Maddocks, 1997] zufolge muss die Initialkrümmung $\boldsymbol{\kappa}^\circ$ nicht zwingend verschwinden. In der Analogie entspräche das der Winkelgeschwindigkeit einer im Kreisel aufgehängten Schwungmasse relativ zur Kreiselgeschwindigkeit. Aus Gl. (3.77) folgt durch Integration, dass die Kraft konstant sein muss. In der KIRCHHOFF-Analogie entspricht die Kraft entlang der Verbindungsachse zwischen Fixpunkt und Schwerpunkt gerade dem Eigengewicht des Körpers $-mg$ (mit negativem Vorzeichen). Als Krafrichtung wählen wir die globale z -Achse \mathbf{e}_z und erhalten durch Projektion auf \mathbf{e}_z , $\mathbf{n} = n \mathbf{e}_z$, woraus durch Einsetzen in Gl. (3.78)

und erneute Projektion auf \mathbf{e}_z die Beziehung

$$\mathbf{e}_z \frac{d}{ds} \mathbf{m} + \underbrace{n (\mathbf{d}_3 \times \mathbf{e}_z) \mathbf{e}_z}_0 = \frac{d}{ds} m_z = 0. \quad (3.80)$$

folgt. Hierbei nutzt man die Tatsache, dass \mathbf{e}_z nicht vom Parameter s abhängt. $d_s m_z$ entspricht der Vertikalkomponente des Drehimpulses des Drehkreisels und bildet ein erstes Integral. Im selben Sinne liefert eine Projektion von Gl. (3.78) auf die lokale Tangente \mathbf{d}_3 ein weiteres erstes Integral,

$$\mathbf{d}_3 \frac{d}{ds} \mathbf{m} + \underbrace{(\mathbf{d}_3 \times \mathbf{n}) \mathbf{d}_3}_0 = \frac{d}{ds} m_3 = 0, \quad (3.81)$$

das ein konstantes Torsionsmoment beschreibt. Beim Drehkreisel entspricht m_3 gerade dem Anteil des Drehmomentes in Richtung der Drehachse. Ein weiteres erstes Integral stellt die HAMILTON-Funktion \mathcal{H} oder Energiedichte (Biege- und Dehnungsenergie) der Faser bzw. die Gesamtenergie des Kreisels (kinetische und potentielle Energie) dar, die durch

$$\mathcal{H} = \frac{1}{2} \mathbf{m} \boldsymbol{\kappa} + \mathbf{n} \mathbf{d}_3 \quad (3.82)$$

gegeben ist.

In Abschnitt Abs. (3.2.3) haben wir die Darstellung der Basisvektoren $\{\mathbf{d}_i\}$ durch EULER-Winkel diskutiert. Auf der Basis von EULER-Winkeln drücken wir jetzt die globale Z-Achse aus, und zwar in Abhängigkeit von der lokalen Basis $\{\mathbf{d}_i\}$,

$$\mathbf{e}_z(\theta, \psi) = \sin \theta (\sin \psi \mathbf{d}_2 - \cos \psi \mathbf{d}_1) + \cos \theta \mathbf{d}_3. \quad (3.83)$$

Auf den ersten Blick scheint \mathbf{e}_z von s abhängig zu sein. Da \mathbf{e}_z per Definition fix ist, muss der Ausdruck für alle s einen konstanten Wert annehmen, \mathbf{e}_z ist ein erstes Integral.

Unter Verwendung von Gl. (3.44) bis Gl. (3.46), Gl. (3.79) und Gl. (3.83) lassen sich die ersten Integrale in ihrer endgültigen Form angeben,

$$m_z = \frac{d\phi}{ds} [1 + (b-1) \cos^2 \theta] + b \frac{d\psi}{ds} \cos \theta, \quad (3.84)$$

$$m_3 = b \left(\frac{d\phi}{ds} \cos \theta + \frac{d\psi}{ds} \right), \quad (3.85)$$

$$\mathcal{H} = \frac{1}{2} \left[\left(\frac{d\theta}{ds} \right)^2 + \left(\frac{d\phi}{ds} \right)^2 \sin^2 \theta + \frac{m_3^2}{b} \right] + n \cos \theta. \quad (3.86)$$

Diese Gleichungen liefern, nach den unbekanntem EULER-Winkeln aufgelöst, das folgende System von Differentialgleichungen erster Ordnung,

$$\frac{d\phi}{ds} = \frac{m_z - m_3 z}{1 - z^2}, \quad (3.87)$$

$$\frac{d\psi}{ds} = \frac{m_3 - m_z z}{1 - z^2} + \left(\frac{1}{b} - 1 \right) m_3, \quad (3.88)$$

$$\left(\frac{dz}{ds} \right)^2 = 2n(h-z)(1-z^2) - (m_z - m_3 z)^2, \quad (3.89)$$

mit den Hilfsvariablen $z = \cos \theta$ und

$$h = \frac{1}{n} \left(\mathcal{H} - \frac{m_3^2}{2b} \right). \quad (3.90)$$

Es wurde eingangs bereits darauf hingewiesen, dass wir nur die Fälle mit $\mathbf{n} \neq 0$ betrachten, so dass der Ausdruck für h stets definiert ist.

Wir werden jetzt zeigen, wie man auf geschlossene Lösungsformeln für alle drei EULER-Winkel kommt. „Geschlossen“ bedeutet lediglich, dass ein Integral durch ein bekanntes anderes ersetzt werden kann. Elliptische Integrale besitzen generell keine explizite Lösung, sondern müssen numerisch ausgewertet werden. Der Vorteil geschlossener Lösungsausdrücke ist primär darin zu sehen, dass sie es uns erlauben, die Gleichungen in symbolischer Form weiterzuverarbeiten. Das Integral über ein elliptisches Integral kann oftmals durch ein anderes ersetzt werden.

Beim Betrachten der rechten Seite von Gl. (3.89) fällt zunächst auf, dass sie als kubisches Polynom in z mit den Wurzeln $\{z_1, z_2, z_3\}$ geschrieben werden kann,

$$\left(\frac{dz}{ds} \right)^2 = 2n(z - z_1)(z_2 - z)(z_3 - z) \quad (3.91)$$

$$= c_1 z^3 + c_2 z^2 + c_3 z + c_4, \quad (3.92)$$

mit den Koeffizienten

$$\{2n, -2nh - m_3^2, -2n + 2m_z m_3, 2nh - m_z^2\}. \quad (3.93)$$

Wir werden z_1, z_2 und z_3 hier nicht als Funktionen von m_z, m_3 und h ausdrücken, sondern die Wurzeln selbst als erste Integrale auffassen, von denen die verbleibenden Größen abhängen. Für kubische Polynome gibt es geschlossene Lösungsformeln, vgl. Abs. (A.3). Eine numerische Auswertung ist deshalb nicht notwendig. Falls die Konstanten c_1, \dots, c_4 reell sind, sind auch die Wurzeln reell. Der komplexe Fall wird hier nicht betrachtet.

Der Wertebereich von z_1, z_2 und z_3 lässt sich abschätzen. Für $n > 0$ und $z \rightarrow +\infty$ geht die rechte Seite gegen $+\infty$, so dass eine der Wurzeln im Intervall $[1, +\infty[$ liegen muss. Weiterhin fordern wir, dass $(dz/ds)^2 > 0$ für alle z im Intervall $[-1, 1]$ liegt, damit der Winkel θ reelle Werte annimmt. Daraus folgt sofort, dass die beiden anderen Wurzeln zwischen -1 und 1 liegen. Die Wurzeln werden so sortiert, dass ihre Werte die folgende Bedingung erfüllen,

$$-1 \leq z_1 \leq z_2 \leq 1 \leq z_3. \quad (3.94)$$

Wir sind nun in der Lage, eine Lösung für Gl. (3.89) anzugeben. Diese Gleichung hat die allgemeine Form

$$\left(\frac{dy}{dx} \right)^2 = (1 - y^2)(1 - k^2 y^2). \quad (3.95)$$

Es ist bekannt, dass die JACOBIsche elliptische Funktion $\text{sn}(x, k)$ eine Lösung von Gl. (3.95) darstellt. Die Integration von Gl. (3.89) nach dem Parameter z liefert deshalb

$$z = z_1 + (z_2 - z_1) \text{sn}^2 [\lambda \cdot (s + s_0), k], \quad (3.96)$$

mit dem Modul $k \in [0, 1]$ und der Konstanten λ , die wie folgt definiert sind,

$$k = \sqrt{\frac{z_2 - z_1}{z_3 - z_1}}, \quad \lambda = \sqrt{\frac{n(z_3 - z_1)}{2}}. \quad (3.97)$$

Das dieses tatsächlich eine Lösung von Gl. (3.89) ist, lässt sich leicht durch Einsetzen überprüfen, bspw. mittels eines Computeralgebraprogramms wie Maple. Die Funktion $\text{sn}(\cdot, k)$ hat die Eigenschaft, dass sie für $k = 0$ auf die Sinusfunktion, $\text{sn}(\cdot, 0) = \sin(\cdot)$, und für $k = 1$ auf den Tangens Hyperbolicus, $\text{sn}(\cdot, 1) = \tanh(\cdot)$, reduziert wird. Nach der Definition für den Modul k ist sein Wert genau dann null, wenn die erste und zweite Wurzel gleich sind, $z_1 = z_2$. Hingegen nimmt er den Wert 1 an, wenn $z_2 = z_3$ ist. Der einzige gemeinsame Wert, den z_2 und z_3 annehmen können, ist nach Gl. (3.94) 1. Es bleibt zu untersuchen, für welche Konstellationen von $\{n, m_3, m_z\}$ die Fälle $z_1 = z_2$ bzw. $z_2 = z_3 = 1$ eintreten. Das Besondere an diesen Spezialfällen ist, dass das Quadrat der Sinusfunktion bzw. des Tangens Hyperbolicus analytisch integrierbar ist und bei der endgültigen Integration zu einfachen Ausdrücken mit transzendenten Funktionen führt. Der Bereich der Faser muss dann nicht diskretisiert werden, wodurch mit einem nicht unerheblichen Geschwindigkeitszuwachs bei der Bestimmung ihrer Mittelliniengeometrie zu rechnen ist.

Wichtig ist für uns die Integrationskonstante s_0 . Für ihre Bestimmung brauchen wir die Umkehrfunktion des Lösungsausdrucks Gl. (3.96) an der Stelle $s = 0$ und erhalten durch Umformung,

$$\eta = \text{sn}(\lambda s_0, k) = \frac{z_0 - z_1}{z_2 - z_1}, \quad (3.98)$$

wobei $z_0 = z(0) = \cos \theta_0 = \cos \theta(0)$. θ_0 ist eine Konstante, die den Wert des Winkels θ an der Stelle 0, also am Anfang der Faser angibt. Wenn wir diesen Winkel einfach auf null setzen, fallen nach der „y-Konvention“ die beiden verbleibenden Rotationen zusammen, weil die Achse \mathbf{e}_3 mit der Achse \mathbf{e}_3'' zusammenfällt. Der Drehwinkel um \mathbf{e}_3 ist dann $(\phi + \psi)$. Die Integrationskonstante s_0 kann jetzt in der folgenden Form angegeben werden,

$$s_0 = \frac{1}{\lambda} F(\eta, k), \quad (3.99)$$

wobei wir mit der Funktion $F(x, k)$ ein *unvollständiges elliptisches Integral der ersten Art* vorliegen haben, mit der folgenden formalen Definition,

$$F(x, k) = \int_0^x \frac{1}{\sqrt{(1-t^2)(1-k^2t^2)}} dt, \quad t = \sin \Phi. \quad (3.100)$$

Weil die Werte der Funktion $\text{sn}(\cdot)$ im Intervall $[0, 1]$ liegen, folgt daraus mit Gl. (3.89) unmittelbar, dass sich θ periodisch zwischen $\arccos z_1$ und $\arccos z_2$ bewegen muss. Während einer Periode wird die periodische Distanz $2\Delta s$ zurückgelegt, die sich direkt aus dem *vollständigen elliptischen Integral der ersten Art* $K(k)$ ableiten lässt als

$$2\Delta s = \frac{2}{\lambda} K(k). \quad (3.101)$$

In der vollständigen Form dieses Integrals nimmt der erste Parameter den Wert 1 oder $\pi/2$ an, je nach Art der Darstellung, vgl. Abs. (A.5).

Bei der Berechnung des Winkels θ ist eine gewisse Vorsicht geboten: Nur für den Parametersatz $z_1 \neq -1$ und $z_2 \neq 1$ ist $z(s)$ eine bijektive Funktion und wir dürfen $\theta = \arccos z$ annehmen. Für alle anderen Fälle untersucht man das Verhalten der Ableitung $d\theta/dt$ im Bereich der Stelle -1 oder 1 . Für den Sonderfall $z_1 \neq -1$, $z_2 = 1$ und $z_3 = 1$ bleibt $\cos \theta$ unterhalb seiner Extremwerte $\{-1, 1\}$, so dass wir auch hier $\theta = \arccos z$ wählen. In den beiden Extremfällen $z \in \{-1, 1\}$ bleibt der Limes der Ableitung erhalten und signalisiert einen Vorzeichenwechsel, und zwar von $\theta - \pi$ in einen und θ im anderen Fall. Auch wenn die Annahme nicht immer gerechtfertigt erscheint, wollen wir im Folgenden postulieren, dass andere mögliche kritische Parameterkonstellationen nicht auftreten.

Wir können uns jetzt den verbleibenden EULER-Winkeln Gl. (3.87) und Gl. (3.88) zuwenden, da zu ihrer Bestimmung die Funktion $z(s)$ bekannt sein muss. Die Gleichungen lassen sich durch Umformen in die folgenden Ausdrücke überführen,

$$\frac{d\phi}{ds} = \frac{m_z}{2} \left(\frac{1}{1+z} + \frac{1}{1-z} \right) + \frac{m_3}{2} \left(\frac{1}{1+z} - \frac{1}{1-z} \right), \quad (3.102)$$

$$\frac{d\psi}{ds} = \frac{m_3}{2} \left(\frac{1}{1+z} + \frac{1}{1-z} \right) + \frac{m_z}{2} \left(\frac{1}{1+z} - \frac{1}{1-z} \right) + \left(\frac{1}{b} - 1 \right) m_3. \quad (3.103)$$

Wir setzen unsere Lösung für $z(s)$, Gl. (3.96), ein und erhalten nach einigen algebraischen Manipulationen,

$$\begin{aligned} \frac{d}{ds} \phi(s) &= \frac{m_z + m_3}{2(1+z_1)} \left\{ \frac{1}{1 - c_a \operatorname{sn}^2[\lambda \cdot (s + s_0), k]} \right\} + \\ &\quad \frac{m_z - m_3}{2(1-z_1)} \left\{ \frac{1}{1 - c_b \operatorname{sn}^2[\lambda \cdot (s + s_0), k]} \right\}, \end{aligned} \quad (3.104)$$

$$\begin{aligned} \frac{d}{ds} \psi(s) &= \frac{m_3 + m_z}{2(1+z_1)} \left\{ \frac{1}{1 - c_a \operatorname{sn}^2[\lambda \cdot (s + s_0), k]} \right\} + \\ &\quad \frac{m_3 - m_z}{2(1-z_1)} \left\{ \frac{1}{1 - c_b \operatorname{sn}^2[\lambda \cdot (s + s_0), k]} \right\} + \left(\frac{1}{b} - 1 \right) m_3. \end{aligned} \quad (3.105)$$

Die Konstanten $c_{a,b}$ definieren wir als $c_a = (z_1 - z_2)/(1 + z_1)$ und $c_b = (z_2 - z_1)/(1 - z_1)$. Diese Formen sind völlig identisch mit Gl. (3.87) und Gl. (3.88), wie sich leicht durch Nachrechnen überprüfen lässt.

Mit diesen Definitionen erhalten wir schließlich die folgenden Integrale für die beiden EULER-Winkel $\phi(s)$ und $\psi(s)$,

$$\begin{aligned} \phi(s) &= \frac{m_z + m_3}{2(1+z_1)} \int_0^s \frac{1}{1 - c_a \operatorname{sn}^2[\lambda \cdot (\sigma + \sigma_0), k]} d\sigma + \\ &\quad \frac{m_z - m_3}{2(1-z_1)} \int_0^s \frac{1}{1 - c_b \operatorname{sn}^2[\lambda \cdot (\sigma + \sigma_0), k]} d\sigma, \end{aligned} \quad (3.106)$$

$$\begin{aligned} \psi(s) &= \frac{m_3 + m_z}{2(1+z_1)} \int_0^s \frac{1}{1 - c_a \operatorname{sn}^2[\lambda \cdot (\sigma + \sigma_0), k]} d\sigma + \\ &\quad \frac{m_3 - m_z}{2(1-z_1)} \int_0^s \frac{1}{1 - c_b \operatorname{sn}^2[\lambda \cdot (\sigma + \sigma_0), k]} d\sigma + \left(\frac{1}{b} - 1 \right) m_3 s. \end{aligned} \quad (3.107)$$

Hierbei nutzen wir die Tatsache, dass m_3 ein erstes Integral ist und damit nicht vom Parameter s abhängt. Ein Blick in den Abramowitz und Stegun [Abramowitz & Stegun, 1970] offenbart uns sofort, dass sich Integrale des Typs $\int (1 - c \operatorname{sn}^2(s, k))^{-1} ds$ durch ein *unvollständiges elliptisches Integral der dritten Art* ausdrücken lassen, womit wir die endgültigen geschlossenen Lösungsausdrücke für $\phi(s)$ und $\psi(s)$ erhalten,

$$\phi(s) = \frac{m_z + m_3}{2(1 + z_1)} \Pi(s, c_a, k) + \frac{m_z - m_3}{2(1 - z_1)} \Pi(s, c_b, k), \quad (3.108)$$

$$\psi(s) = \frac{m_3 + m_z}{2(1 + z_1)} \Pi(s, c_a, k) + \frac{m_3 - m_z}{2(1 - z_1)} \Pi(s, c_b, k) + \left(\frac{1}{b} - 1\right) m_3 s. \quad (3.109)$$

$\Pi(s, n, k)$ ist ein elliptisches Integral der dritten Art, wobei wir, um den Parameter s verwenden zu können (anstatt des Winkels oder dem Sinus des Winkels, wie sonst üblich), die Parametertransformation $s = F(\Phi, k)$ durchführen. $F(\Phi, k)$ ist das schon bekannte elliptische Integral der ersten Art. Für $\Phi = \pi/2$ erhalten wir $\Delta s = F(\pi/2, k) = K(k)$, die semiperiodische Distanz während eines Umlaufs von θ . Daraus lässt sich die Winkeländerung $\Delta\phi$ bzw. $\Delta\psi$ für die Dauer einer Periode wie folgt ermitteln,

$$\Delta\phi = \phi(\Delta s) = \frac{m_z + m_3}{2(1 + z_1)} \Pi(\Delta s, c_a, k) + \frac{m_z - m_3}{2(1 - z_1)} \Pi(\Delta s, c_b, k) \quad (3.110)$$

$$\Delta\psi = \psi(\Delta s) = \frac{m_3 + m_z}{2(1 + z_1)} \Pi(\Delta s, c_a, k) + \frac{m_3 - m_z}{2(1 - z_1)} \Pi(\Delta s, c_b, k) + \left(\frac{1}{b} - 1\right) m_3 \Delta s. \quad (3.111)$$

Da sich θ , wie oben angedeutet, periodisch zwischen $\arccos z_1$ und $\arccos z_2$ bewegt, ergeben sich folgende Konsequenzen für die Bereichsgrenzen in den Integralen Gl. (3.106) und Gl. (3.107). Man integriert von z_1 bis z_2 bzw. im Intervall $[0, s \bmod N(2\Delta s)]$, wobei $2\Delta s$ genau die periodische Distanz von θ ist, oder direkt als Integral ausgedrückt,

$$\phi(s) = N(2\Delta\phi) + \int_0^{s \bmod N(2\Delta s)} \frac{d}{ds} \phi(\sigma) d\sigma, \quad (3.112)$$

$$\psi(s) = N(2\Delta\psi) + \int_0^{s \bmod N(2\Delta s)} \frac{d}{ds} \psi(\sigma) d\sigma + \left(\frac{1}{b} - 1\right) m_3 s. \quad (3.113)$$

N ist die Anzahl der vollen Perioden. Diese Ausdrücke beschreiben die lokale Basis entlang der Faser durch Gl. (3.40) als Funktion von s .

Sind die Funktionen für die EULER-Winkel erst einmal bekannt, kann daraus die Geometrie der Mittellinie $\mathbf{r}(s)$ rekonstruiert werden. Es gibt dazu verschiedene Möglichkeiten. Am einfachsten ist, wegen $d\mathbf{r}/ds = \mathbf{d}_3$, die direkte Integration der Tangentenfunktion (dritte Spalte von Gl. (3.40)), als $\mathbf{r}(s) = \int_0^s \mathbf{d}_3(\sigma) d\sigma$. Das liefert drei Integrale für die drei Koordinaten,

$$X(s) = X_0 + \int_0^s \cos \phi(\sigma) \sin \theta(\sigma) d\sigma, \quad (3.114)$$

$$Y(s) = Y_0 + \int_0^s \sin \phi(\sigma) \sin \theta(\sigma) d\sigma, \quad (3.115)$$

$$Z(s) = Z_0 + \int_0^s \cos \theta(\sigma) d\sigma, \quad (3.116)$$

mit $\mathbf{r}(s) = \{X(s), Y(s), Z(s)\}$ und $\mathbf{r}(0) = \{X_0, Y_0, Z_0\}$.

Offenbar werden für die Rekonstruktion der Mittellinie nur die Funktionen ϕ und θ benötigt. Zur Berechnung des vollständigen lokalen Koordinatensystems brauchen wir auch den dritten Winkel ψ . Das dritte Integral kann durch Einsetzen von Gl. (3.96) in Gl. (3.116) vereinfacht werden,

$$Z(s) = Z_0 + z_1 s + (z_2 - z_1) \int_0^s \operatorname{sn}^2 [\lambda \cdot (s + s_0), k] ds. \quad (3.117)$$

Auch hier lässt uns der Abramowitz und Stegun [Abramowitz & Stegun, 1970] nicht im Stich. Die Quadrate der JACOBIschen elliptischen Funktionen lassen sich integrieren und man erhält

$$Z(s) = Z_0 + z_1 s + (z_2 - z_1) - E(s) + s. \quad (3.118)$$

3.3.1.2 Randwertproblemlöser

Was haben wir bis jetzt erreicht? Wir sind jetzt in der Lage, den statischen Gleichgewichtszustand bzw. die Gleichgewichtskonfiguration einer Faser mit definierten Materialeigenschaften durch Auswertung elliptischer Funktionen und anschließender partieller numerischer Integration zu bestimmen. Partiiell bedeutet hier, dass sich lediglich die dritte Koordinatenfunktion $Z(s)$ durch einmaliges Auswerten einer elliptischen Funktion an beliebigen Stellen s_i berechnen lässt; für die beiden verbleibenden Koordinaten muss mittels eines numerischen Integrationsverfahrens, wie bspw. der GAUSS-Quadratur, das Integral über der elliptischen Funktion ausgewertet werden.

Das entspricht dem Lösen eines Anfangswertproblems bei bekannten Startwerten und Systemkonstanten n, m_3, m_z, z_0 und θ'_0 . Die einzelnen Größen werden in der folgenden Reihenfolge berechnet:

1. Wahl von n, m_3, m_z, z_0 und θ'_0 , wobei die letzten beiden als Integrationskonstanten zu sehen sind und nicht als erste Integrale. Der Grund für die Wahl von θ'_0 anstelle des ersten Integrals \mathcal{H} liegt darin, dass \mathcal{H} nur bestimmte Werte annehmen darf, wohingegen $\theta'_0 \in \mathbb{R}$ keinerlei Beschränkungen hinsichtlich des zulässigen Wertebereichs unterliegt, was besonders den Einsatz numerischer Optimierungsverfahren vereinfacht.
2. Berechnung der HAMILTON-Funktion $\mathcal{H} = \mathcal{H}(z_0, \theta'_0)$, Gl. (3.82), und daraus der Hilfsvariablen h , Gl. (3.90).
3. Berechnung der Wurzeln $\{z_1, z_2, z_3\}$ mittels \mathcal{H}, n, m_3, m_z durch Auswertung des Polynoms Gl. (3.92) mittels der Formeln im Anhang, Abs. (A.3).
4. Berechnung der Integrationskonstante s_0 mittels Gl. (3.99)
5. Auswertung der elliptischen Funktionen an beliebigen Stellen s_i der Kurve:
 - a) Falls nur die Koordinaten des Endpunktes gesucht werden, berechnen wir $\Delta\phi, \Delta s$ und werten anschließend die Integrale Gl. (3.96) und Gl. (3.112) aus;
 - b) falls neben der Tangente auch eine weitere Koordinatenachse benötigt wird (\mathbf{d}_1 oder \mathbf{d}_2), muss auch Gl. (3.113) ausgewertet werden.
6. Integration der Tangentenfunktion durch Auswertung der Integrale Gl. (3.114), Gl. (3.115), Gl. (3.116) an beliebigen Stellen s_i .

7. Die Achsen \mathbf{d}_1 oder \mathbf{d}_2 erhält man durch Auswertung von Gl. (3.40) für beliebige s_i .

Damit können wir die folgenden Fragestellungen bearbeiten.

Problem 3.3.1 *Unter der Annahme, dass die Position $\mathbf{r}(0)$ und die Orientierung $\{\phi_0, \theta_0, \psi_0\}$ am Anfangspunkt der Faser gegeben sind, suchen wir einen Parametersatz $\mathbf{x}^* = \{n, m_3, m_z, \theta'_0\} \in \mathbb{R}^4$, der diejenige Konfiguration der Faser beschreibt, bei der*

1. *der Endpunkt der Faser mit einem gegebenen Zielpunkt \mathbf{p}_t übereinstimmt;*
2. *die Richtung der Tangente $\mathbf{d}_3(l)$ am Ende der Faser mit einer Zielrichtung \mathbf{T}_t zusammenfällt.*

Solche Probleme entstehen regelmäßig in CAD-Applikationen für die Konstruktion von Industrieanlagen, bei denen Anlagenteile über Schläuche miteinander verbunden werden sollen. Das hier vorgestellte Modell erlaubt es uns dann, nicht nur die Konfiguration vorherzusagen, die ein Schlauch mit gegebener Länge einnehmen wird, sondern dient auch gleichzeitig als Dimensionierungshilfe.

Die oben angeführten Aufgabenstellungen führen uns mathematisch auf ein Zwei-Punkt-Randwertproblem (BVP) mit den Randbedingungen

$$\text{RB}_a = \{\mathbf{r}(0), \mathbf{d}_{\{1,2\}}(0) = f(\phi_0, \theta_0, \psi_0), \mathbf{r}(l) = \mathbf{p}_t\}, \quad (3.119)$$

$$\text{RB}_b = \{\mathbf{r}(0), \mathbf{d}_{\{1,2\}}(0) = f(\phi_0, \theta_0, \psi_0), \mathbf{r}(l) = \mathbf{p}_t, \mathbf{d}_3(l) = \mathbf{T}_t\}. \quad (3.120)$$

Die Frage lautet also, welche Werte für die Systemkonstanten $\{n, m_3, m_z, \theta'_0\}$ gewählt werden müssen, um die Randbedingung am Ende der Faser ($s = l$) zu erfüllen. Zusätzlich formulieren wir die folgende Nebenbedingung für die Systemkonstante n , um den Singulärpunkt zu vermeiden: $n > 0$.

Im ersten Problemfall suchen wir den Lösungsvektor \mathbf{x}^* , der den Abstand zwischen dem Endpunkt $\mathbf{r}(l)$ und dem Zielpunkt \mathbf{p}_t minimiert,

$$\mathbf{x}^* : \|\mathbf{r}(l) - \mathbf{p}_t\| \rightarrow \min.$$

Diesem Sachverhalt werden wir durch die folgende Definition der Zielfunktion $f : \mathbb{R} \rightarrow \mathbb{R}^4$ gerecht,

$$f_1(\mathbf{x}) = \|\mathbf{r}(l) - \mathbf{p}_t\|^2. \quad (3.121)$$

Im zweiten Fall soll zusätzlich die Richtung der Tangente am Ende der Faser mit der Vorgabe \mathbf{T}_t übereinstimmen, oder

$$\mathbf{x}^* : \|\mathbf{d}_3(l) - \mathbf{T}_t\| \rightarrow \min.$$

Wir erweitern deshalb Zielfunktion Gl. (3.121) wie folgt,

$$f_2(\mathbf{x}) = \|\mathbf{r}(l) - \mathbf{p}_t\|^2 + \|\mathbf{d}_3(l) - \mathbf{T}_t\|^2. \quad (3.122)$$

Das führt uns auf das folgende Optimierungsproblem,

$$\mathbf{x}^* = \underset{\mathbf{x} \in \Omega}{\operatorname{argmin}} \|\mathbf{r}(l) - \mathbf{p}_t\|^2, \quad n > 0, \quad (3.123)$$

wobei $\Omega \subseteq \mathbb{R}^4$ ein zulässiger Bereich des \mathbb{R}^4 darstellt. Für die Lösung des Optimierungsproblems bietet sich dann der Einsatz eines globalen Optimierungsverfahrens an.

3.3.1.3 Numerische Beispiele

Die zur Beschreibung der analytischen Lösung notwendigen Ausdrücke leitet man mit einem Computeralgebraprogramm her, bspw. Maple. Solche Programme sind in der Lage, daraus optimierten Quellcode zu generieren, der sich leicht in einen bereits bestehenden Softwarerahmen integrieren lässt. Für die Lösung des „Anfangswertproblems“ sind lediglich die JACOBI-Elliptischen Funktionen, sowie die Funktionen Gl. (3.114), Gl. (3.115) und Gl. (3.116) numerisch zu integrieren.

Abb. (3.2) zeigt einige typische Beispielkonfigurationen, die man als Lösungen des Anfangswertproblems mit dem vorgestellten analytischen Ansatz erhält.

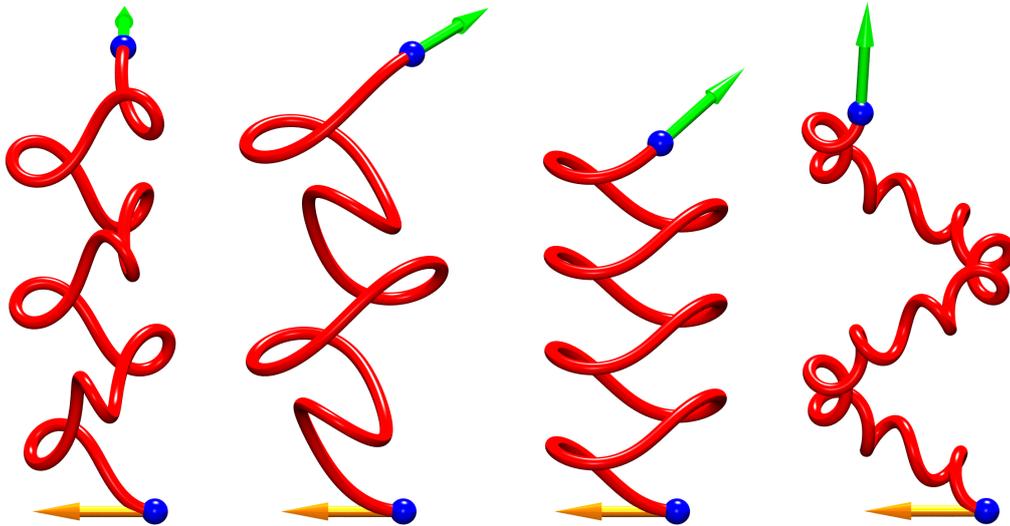


Abbildung 3.2: Reguläre Konfigurationen, die man bei der Vorwärtsrechnung mit dem analytischen Ansatz erhält. Sie entsprechen den Kurven, welche die Extremität des Lagrangeschen Kreisels unter gleichen Bedingungen auf die Oberfläche der Einheitskugel zeichnen würde.

Wenden wir uns nun den Randwertproblemen zu. Als Testfall generieren wir zunächst eine Konfiguration, hier mit den beliebig gewählten Parametern $\{n = 1, 0; m_3 = 2, 0; m_z = 1, 1; \theta'_0 = 1, 0\}$. Die Koordinaten des Endpunktes stellen den zu matchenden Zielpunkt dar. Damit haben wir sicher gestellt, dass für den Zielpunkt ein gültiger Parametersatz existiert. Das Optimierungsverfahren startet man bei einer zufälligen Perturbation dieser Anfangswerte.

Die Bestimmung der gesuchten Parameter auf der Basis des Powellverfahrens oder *konjugierter Richtungen* erweist sich als problematisch, da diese Ansätze regelmäßig in lokalen Minima steckenbleiben. Als in dieser Hinsicht deutlich robuster erweist sich ein „simuliertes Ausglühen“ (engl.: simulated Annealing). Das Verfahren findet für den Testfall bei zufällig gewählten Startparametern und nach hinreichend langer Suchdauer (etwa 1.300 Iterationen) eine Lösung, vgl. Abb. (3.3). Allerdings entspricht das Resultat – ganz den Erwartungen gemäß – den typischen schleifen-ähnlichen Konfigurationen, welche die Extremität des analogen LAGRANGE-Kreisels bei seiner Bewegung auf die Oberfläche der Einheitskugel zeichnen würde, vgl. 3.4. Die Konfiguration hat damit wenig Ähnlichkeit zur Form eines zwischen zwei Punkten eingespannten elastischen Schlauches, die man intuitiv erwarten würde. Hierbei gilt es jedoch zu beachten, dass keine

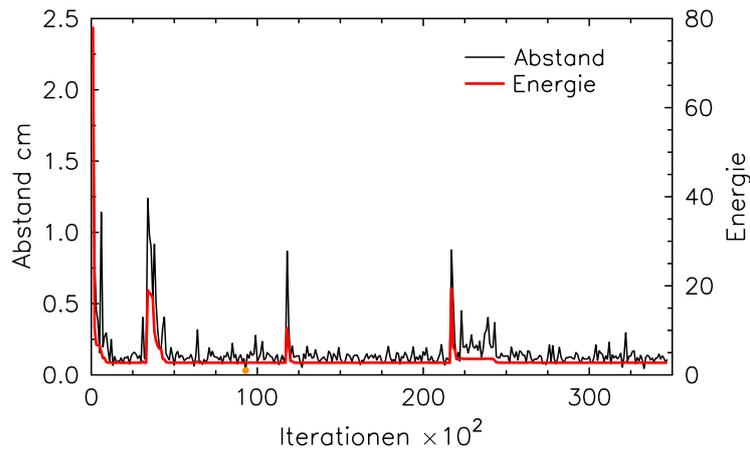


Abbildung 3.3: Lösung des BVPs mittels des Simulated Annealing Ansatzes. Die Abszisse zeigt die Anzahl der Schmelz- und Ausglühzyklen. Jeder Ausglühzyklus hat 100 Iterationen. Die Schrittweite beträgt 0,001. Die Ordinate zeigt den Abstand des Endpunktes vom Zielpunkt.

Gravitation auf das Elastikum einwirkt.

Der Grund für die fehlende Ähnlichkeit liegt auf der Hand: da wir keinerlei Anforderungen an die Funktionsparameter bezüglich des zulässigen Wertebereichs gestellt haben, neigt das Verfahren zur Wahl unangemessen großer Kräfte und Momente, die beim realen Experiment bereits zum Materialversagen führen würden. In der Folge nehmen die Krümmungsfunktion bzw. die Biegeenergie extreme Werte an. Wir können dem entgegenwirken, indem wir die Biegeenergie mit in der Optimierungsprozess einbeziehen. Dazu addiert man einen weiteren Term \mathcal{E}_B in der Energiefunktion, der die Lösung in Richtung möglichst minimal gekrümmter Konfigurationen zwingt. Die Biegeenergie \mathcal{E}_B ist das Integral über eine geeignete Energiefunktion $W(s)$, in der Regel eine quadratische Form, die wir wie folgt definieren,

$$W(\boldsymbol{\kappa}(s)) = \frac{1}{2} \Delta \boldsymbol{\kappa}^T(s) \mathbf{K}_N \Delta \boldsymbol{\kappa}(s) = \frac{1}{2} (\kappa_1^2 + \kappa_2^2 + b \kappa_3^2), \quad (3.124)$$

wobei \mathbf{K}_N die normierte Steifigkeitsmatrix mit den Einträgen $\mathbf{K}_N = \text{diag}(1, 1, b)$ ist. Daraus folgt mit Gl. (3.44) bis Gl. (3.46),

$$W(s) = \frac{1}{2} \left[\left(\frac{d\theta}{ds} \sin \psi - \frac{d\phi}{ds} \sin \theta \cos \psi \right)^2 + \left(\frac{d\theta}{ds} \cos \psi + \frac{d\phi}{ds} \sin \theta \sin \psi \right)^2 + \left(\frac{d\psi}{ds} + \frac{d\phi}{ds} \cos \theta \right)^2 \right], \quad (3.125)$$

und für die Biegeenergie schließlich

$$\mathcal{E}_B(s) = \frac{1}{2} \int_0^s W(\sigma) d\sigma. \quad (3.126)$$

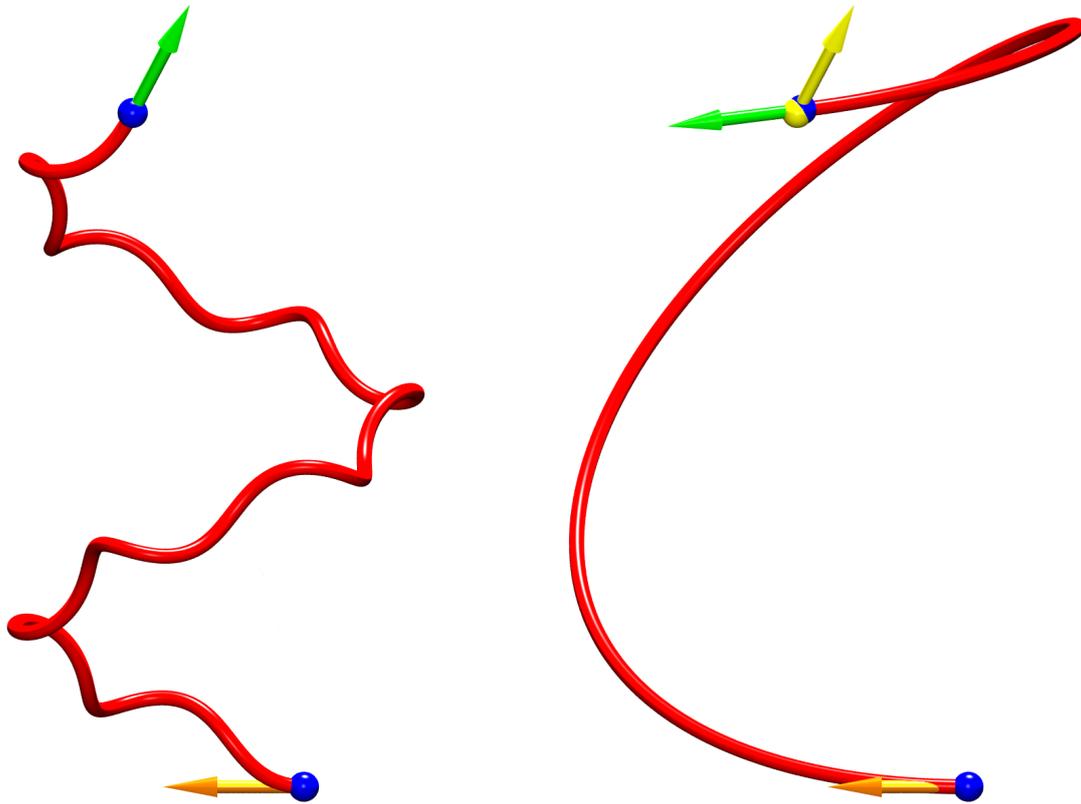


Abbildung 3.4: Stab der Länge 25 cm. Links: Das linke Bild zeigt das Ergebnis des Optimierungsprozesses ohne Einbezug der Biegeenergie. Rechts: Das Resultat des Optimierungsprozesses. Die Krümmung ist minimal und der Zielpunkt wird mit einer Genauigkeit von 0,03 cm getroffen. Die gelbe Kugel bzw. der gelbe Pfeil entsprechen der Zielposition bzw. Zieltangente.

Die Zielfunktion wird entsprechend angepasst zu,

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \Omega} \|\mathbf{r}(l) - \mathbf{p}_t\|^2 + \varepsilon \mathcal{E}_B(s), \quad n > 0. \quad (3.127)$$

Wir geben hier für die Biegeenergie \mathcal{E}_B lediglich die Abhängigkeit vom Parameter s an, wobei klar sein sollte, dass sie indirekt auch eine Funktion von den Systemkonstanten $\{n, m_3, m_z, \theta'_0\}$ ist. ε hingegen ist eine „magische“ Konstante, für die wir in unseren Beispielen $\varepsilon = 1, 8$ angenommen haben. Sie reguliert den Einfluss der Biegeenergie auf das Gesamtergebnis.

Das Resultat des Optimierungsprozesses unter Einbezug der Biegeenergie ist in Abb. (3.4) dargestellt. Die Konfiguration ist offensichtlich glatt, und die Krümmung entlang des Stabes fällt minimal aus. Der Zielpunkt wird mit einer Genauigkeit von 0,03 cm getroffen. Wir können an dieser Stelle allerdings sofort den Nachteil des analytischen Verfahrens sehen: Der Schlauch sollte infolge Eigengewichts an der Ansatzstelle nach unten gebogen werden. Die Berücksichtigung gravitativer Effekte erfordert jedoch asymmetrische Lösungen, die im analytischen Ansatz von vornherein ausgeschlossen sind. Kräfte wirken hier nur an den Endpunkten. Jede Kraft zwischen

den Endpunkten induziert eine Störung im Verlauf der Lösungskurve. Will man solche Effekte modellieren, sind mehrere solcher „KIRCHHOFF-Stab-Elemente“ aneinander zu reihen und entsprechende Kompatibilitätsbedingungen an den Knoten zu formulieren. Ob sich der Aufwand lohnt, ist jedoch fraglich. In den folgenden Abschnitten stellen wir rein numerische Ansätze vor, die auf einer Diskretisierung der COSSERAT-Gleichungen beruhen.

In einem nächsten Schritt wollen wir nun auch die Tangente in den Optimierungsprozess mit einbeziehen. Dazu suchen wir das Minimum der Funktion Gl. (3.122), die ebenfalls um den Biegeenergieterm erweitert wird. Das Resultat nach ca. 1.500 Iterationen simulierten Ausglühens ist in Abb. (3.5) dargestellt. Die Position wird mit einer Genauigkeit von 0,3 cm getroffen, die Tan-

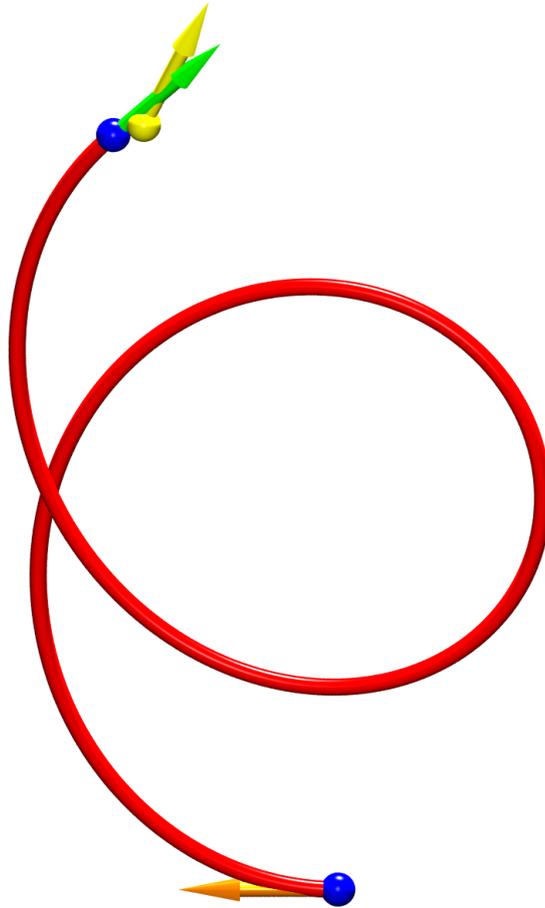


Abbildung 3.5: Erweiterung der Randbedingungen auf Orientierungsvorgaben am Zielpunkt. Solche Randwerte bereiten dem Verfahren offensichtlich mehr Probleme. Der Zielpunkt wird mit einer Genauigkeit von 0,3 cm, die Tangente hingegen nur mit 0,3 rad getroffen.

gente hingegen nur mit 0,3 rad. Die Lösung ist damit unbefriedigend. Zum tieferen Verständnis der Probleme, die der Einbezug der Orientierung in den Optimierungsprozess offenbar aufwirft, kann Abb. (3.6) beitragen. Dargestellt sind die Potentialfunktionen Gl. (3.121) bzw. Gl. (3.122). Auf den beiden Abszissenachsen sind die beiden Momente m_3 und m_z abgetragen, auf der Hochachse

hingegen der Wert der Potentialfunktion ohne bzw. mit Tangente (obere bzw. untere Reihe), für drei feste Werte von n , bei konstantem θ'_0 . Im ersten Fall zeigt sich der Funktionsverlauf unspektakulär, allerdings mit einer deutlichen Störung, die sich mit steigender Kraft n auf der Bodendiagonale durch das Basin bewegt. Die Diagonale indiziert dabei die Stellen, an denen $m_3 = m_z$. Dort verschwindet eines der Integrale in Gl. (3.106) bzw. Gl. (3.107) und die Lösung zeigt einen Sprung, im untersuchten Fall offensichtlich näher zum Zielpunkt hin. Sobald wir die Tangente mit einbeziehen, weist die Potentialfläche neben einer starken Faltung einen Wirbel auf, dessen Zentrum offenbar mit der wandernden „Störungspyramide“ koinzidiert. Diese extremen Wechsel von Minima und Maxima, die stark gekrümmt verlaufende Täler ausformen, führen zu Irritationen bei der Suche des Optimums, besonders mit Verfahren, die sich am lokalen Gefälle der Funktion orientieren.

Abb. (3.7) zeigt dieselbe Funktion, allerdings in einer Flächendarstellung, für drei Wertetripel $\{n, m_3, m_z\}$. Auch hier werden die Manifestation der Störung in Form eines Sprungs bzw. der starke Wechsel von Minima und Maxima deutlich.

Auf der anderen Seite kann das Randwertproblem auch mittels des NEWTON-Verfahrens gelöst werden. Wenn wir die Minimierungsfunktion Gl. (3.121) benutzen, müssen wir nach den Größen n, m_3, m_z, θ'_0 und s ableiten, so dass die JACOBI-Matrix durch $\{\partial f/\partial n, \partial f/\partial m_3, \partial f/\partial m_z, \partial f/\partial \theta'_0, \partial f/\partial s\}$ gegeben ist. Auch zu ihrer Herleitung bietet sich der Einsatz eines Computeralgebraprogramms an. Es hat sich allerdings herausgestellt, dass das NEWTON-Verfahren bei der Suche nach einer Lösung in der Regel keine Konvergenz zeigt (die global konvergente Strategie des „gedämpften“ NEWTON-Verfahrens mit eingeschlossen). Der Grund dafür mag in den eben diskutierten Unstetigkeiten der Funktion liegen.

Wir folgern also, dass der hier vorgestellte semianalytische Randwertproblemlöser sich durchaus für einfache Randbedingungen, wie bspw. die Vorgabe eines Zielpunktes eignet. Die interessanteren Problemklassen jedoch, welche die Orientierung mit einbeziehen, lassen sich momentan nicht mit der erforderlichen Genauigkeit in vertretbarer Zeit lösen. Hier sei auf alternative Ansätze verwiesen, wie bspw. dem von Healey und Mehta [Healey & Mehta, 2003], der sich der Methode der orthogonalen Kollokation bedient.

3.3.2 Numerischer Lösungsansatz

Ein alternativer Ansatz zur Behandlung der statischen COSSERAT-Gleichungen besteht in der Anwendung eines numerischen Integrationsverfahrens als Teil einer Shooting- oder Relaxationsprozedur. Shootingverfahren lösen eine Folge von Anfangswertproblemen durch wiederholtes Anwenden des Integrationsverfahrens auf die Differentialgleichungen. Dabei wird in jeder Iteration versucht, den Abstand zu den gewählten Randbedingungen mittels einer Nullstellensuche, überwiegend auf der Basis des NEWTON-Verfahrens, zu minimieren. Ein solches einfaches Shootingverfahren haben wir schon im vorangegangenen Abschnitt auf der Basis der analytischen Lösung der statischen Form der KIRCHHOFF-Gleichungen entwickelt. Die weiter oben eingeführte Entkopplung der Kinematikgleichungen von den Gleichgewichtsbedingungen hat in Bezug auf die stationären Gleichungen eine weitreichende Konsequenz. Sowohl bei der Kraft-, als auch bei der Momentengleichung sind die Randwerte auf derselben Seite bekannt und auf der anderen Seite unbekannt. Aus dem ursprünglichen Randwertproblem werden nun zwei aufeinander folgend zu lösende Anfangswertprobleme. Wie sieht das bei den dynamischen Gleichungen aus? Nach

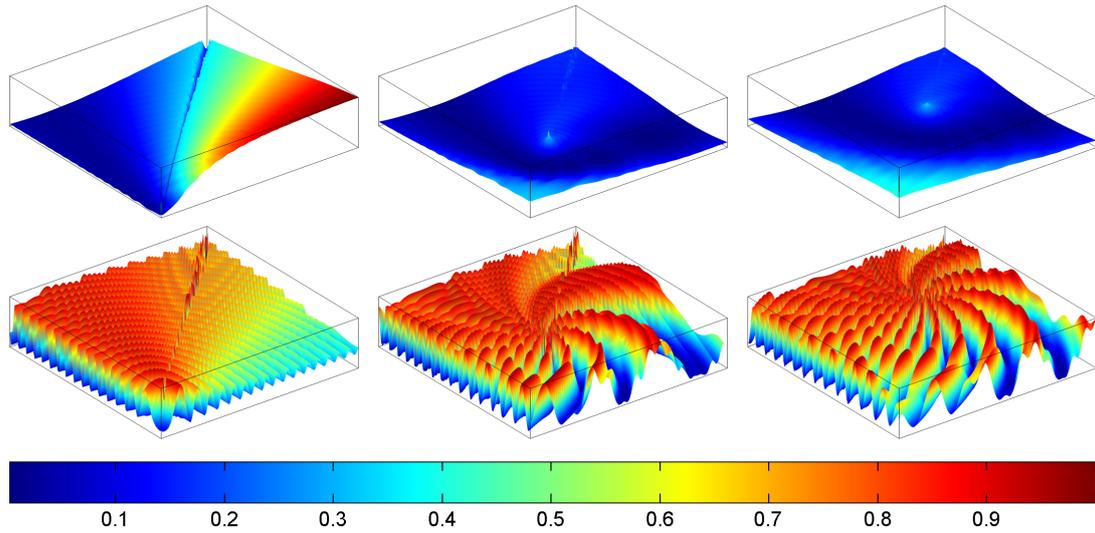


Abbildung 3.6: Die Bilder zeigen eine Höhendarstellung der Potentialfunktion in Abhängigkeit von m_z und m_3 (auf dem Intervall $[0; 5]$) für konstante Werte von n (von links nach rechts: $n = \{0, 001; 2, 5; 5, 0\}$; $\theta'_0 = \text{const.}$). In der oberen Reihe betrachten wir nur das Positionsmatching, in der unteren Reihe dagegen Position und Tangente. Der Einbezug der Orientierung führt offenbar zu einer extremen Faltung der Potentialfläche.

der Entkopplung degeneriert der Behandlungsprozess der Kinematikgleichungen zum Anfangswertproblem. Die verbleibenden vier Gleichungen haben Randwertvorgaben auf beiden Seiten. Die Variablen $(\mathbf{v}, \boldsymbol{\omega}, \boldsymbol{\kappa}, \mathbf{n})$ bleiben voneinander abhängig, so dass sie durch Lösen eines Randwertproblems bestimmt werden müssen, vgl. Abb. (3.8).

Der erste Beitrag im Bereich der Computergraphik diesbezüglich wurde von Pai 2002 vorgestellt, vgl. [Pai, 2002]. Pai ermittelt die Gleichgewichtskonfiguration der Faser durch den logisch aus der Entkopplung folgenden „Zwei-Phasen-Integrationsprozess“. Die Gleichgewichtsbedingungen werden mittels des expliziten EULER-Verfahrens integriert. Das führt auf die folgende diskrete Form der Kraft- bzw. Krümmungsfunktion,

$$\mathbf{n}_{i-1} = \mathbf{n}_i + \Delta s_{i-1}(\mathbf{n}_i \times \boldsymbol{\kappa}_i - \mathbf{f}_i) \quad (3.128)$$

$$\boldsymbol{\kappa}_{i-1} = (\boldsymbol{\kappa}_{i-1}^\circ - \boldsymbol{\kappa}_i^\circ) + \boldsymbol{\kappa}_i + \Delta s_{i-1} \mathbf{K}^{-1} \{ [\mathbf{K}(\boldsymbol{\kappa}_i - \boldsymbol{\kappa}_i^\circ)] \times \boldsymbol{\kappa}_i + \mathbf{n}_i \times \mathbf{d}_3 - \mathbf{l}_i \}, \quad (3.129)$$

wobei man die Kräfte bzw. Krümmungen vom Ende zum Anfang propagiert, wie man leicht an der Indizierung ausmachen kann. Die Kinematikgleichung hingegen löst man anschließend durch Anwendung der RODRIGUES-Formel, welche die Rotationen durch eine Reihenentwicklung linearisiert, vgl. Abs. (3.4.5). Der normierte Krümmungsvektor $\boldsymbol{\kappa}(s)$ wird dabei als Drehachse interpretiert, um welche die Koordinatensysteme entlang der Faser rotieren. Die neuen von den Kräften und Krümmungen abhängigen Orientierungen ergeben sich dann durch Propagation vom Anfang zum Ende der Faser. Die Funktion der Mittellinie $\mathbf{r}(s)$ folgt aus der Integration der Tangentenfunktion.

Das Hauptproblem bei der direkten Verwendung dieses Ansatzes zur Ermittlung der Gleich-

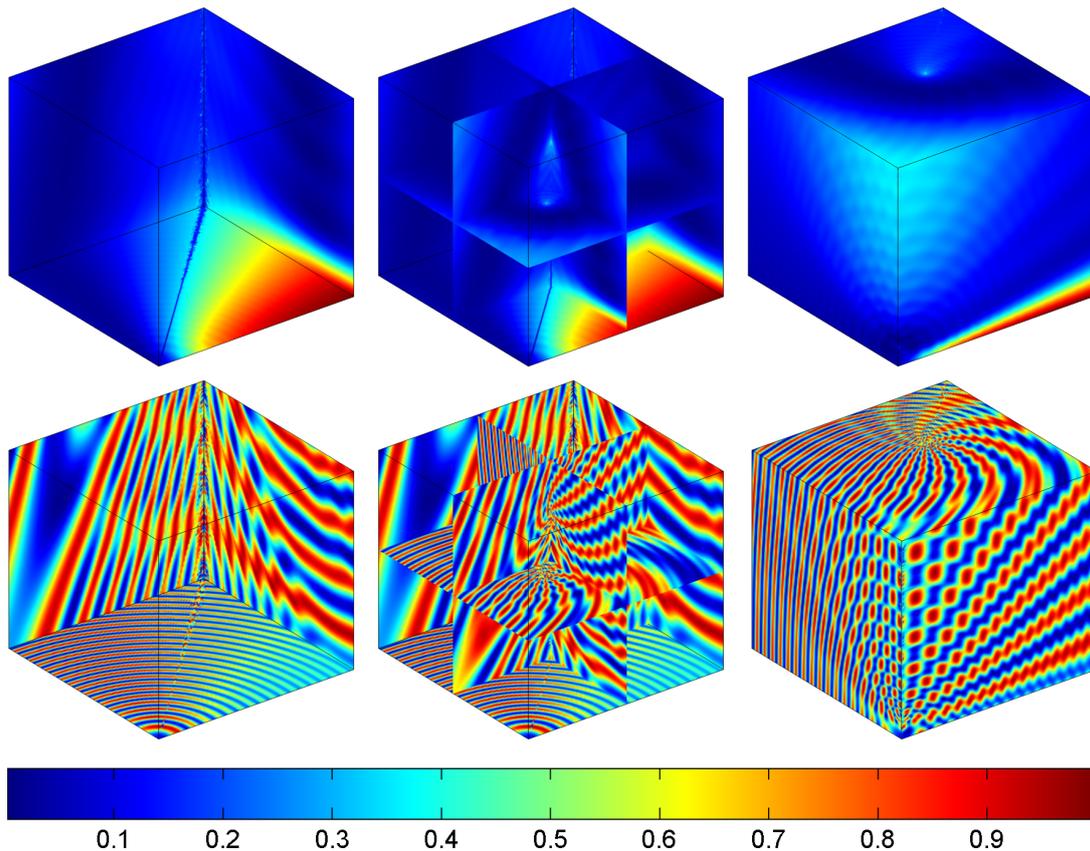


Abbildung 3.7: Die Bilder zeigen eine Flächendarstellung der Potentialfunktion in Abhängigkeit von m_z , m_3 und n (auf dem Intervall $[0;5]$ für m_z und m_3 bzw. $[0;5]$ für n) an den Stellen (von links nach rechts): $\{0, 001; 0; 0\}$, $\{2, 5; 2, 5; 2, 5\}$ und $\{5, 0; 5, 0; 5, 0\}$. In der oberen Reihe betrachten wir nur das Positionsmatching, in der unteren Reihe dagegen Position und Tangente.

gewichtskonfiguration infolge äußerer Lasten besteht in der Tatsache, dass alle Größen auf die lokalen Basen bezogen werden. Infolgedessen werden Kräfte zu Folgelasten, weil sie von der Konfiguration der Faser abhängen. Zur Veranschaulichung der Auswirkungen einer Folgelast auf die Form eines elastischen Stabes stellen wir uns diesen in einer vertikalen Startposition vor. Auf ihn wirke eine Terminallast senkrecht zur Stabachse ein. Intuitiv erwartet man eine Konfiguration, bei der die Mittellinie des Stabes maximal parallel zur Richtung der Kraft verläuft. Wenn die Kraft aber auf das lokale Koordinatensystem des Endpunktes bezogen wird und damit immer senkrecht zur Stabachse bleibt, erfolgt die Deformation über die vertikale Konfiguration hinaus zu einem Halbkreis, vgl. Abb. (3.9). Erhöht man den Betrag der Last weiter, geht der Halbkreis in eine s-Form über, usw. Jeder Umschlag auf die gegenüberliegende Seite erzeugt einen neuen „Buckel“ in der Konfiguration.

Folgelasten haben durchaus Anwendungsgebiete, bspw. bei der Simulation eines Gartenschlauchs, der Wasser ausstößt. Hier bleibt die Reaktionskraft stets tangential zur Stabachse am Endpunkt. Auch bei der Deformation der Wirbelsäule infolge der Lasten, die durch das Kraniaum

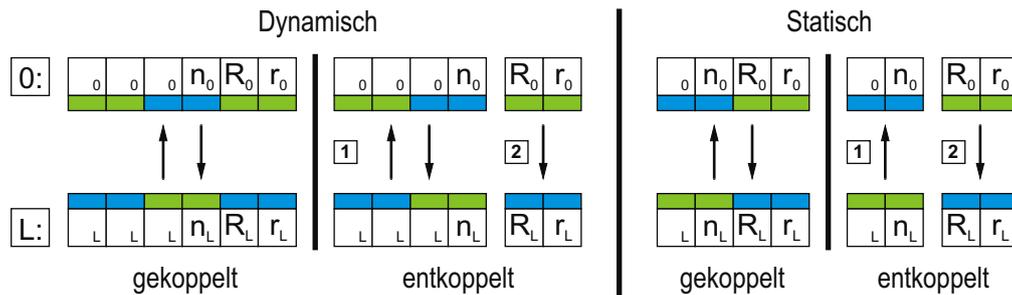


Abbildung 3.8: Die Behandlung der dynamischen Cosserat-Gleichungen führt auf ein Randwertproblem, das durch die Entkopplung von den Kinematikgleichungen vereinfacht werden kann. Der Behandlungsprozess der statischen Gleichungen degeneriert dagegen, bedingt durch die Entkopplung, zu zwei aufeinander folgenden Anfangswertproblemen. Die Randwerte der gekoppelten Kraft-Momenten-Gleichungen sind auf derselben Seite bekannt. Wir gehen hierbei von Standardrandbedingungen aus: Faser an der Stelle $s = 0$ fest eingespannt und Kraft- und Moment- bzw. Krümmungsvorgabe an der Stelle $s = L$. Die grünen Markierungen indizieren bekannte, die blauen unbekannte Größen.

auf sie ausgeübt werden, liegt ein Folgelastproblem vor.

Pai [Pai, 2002] verwendet den Ansatz zur schnellen numerischen Abschätzung der JACOBI-Matrix in Verbindung mit Standardlösern für Randwertprobleme, um die Konfiguration des Stabes bei Vorgabe der Position bzw. Orientierung des Endpunktes zu berechnen. In diesem Fall reichen bereits kleine Lasten aus und der Einfluss der Transformation in die lokalen Systeme bleibt vernachlässigbar.

Wir werden im Folgenden einen Ansatz vorstellen, der die Ermittlung der Faserkonfiguration auf ein Optimierungsproblem zurückführt. Lasten werden dabei stets in Bezug auf das globale System behandelt, so dass die oben diskutierten Probleme nicht auftreten können. Ein solches Modell bietet sich insbesondere für die Behandlung großer Kräfte an, wie sie bspw. bei der Simulation elektrostatischer Effekte zwischen Haarfasern auftreten.

3.3.2.1 Energie der Faser

Unser Lösungsansatz für die statischen KIRCHHOFF-Gleichungen basiert auf der Minimierung des Energiefunktionals der Faser. Das erlaubt die Verwendung von Belastungen, die zwischen den Endpunkten der Faser angreifen. Gesucht wird dabei eine approximative Lösung der Krümmungsfunktion $\kappa^*(s)$, welche die energieminimierende Konfiguration der Faser unter Berücksichtigung aller äußeren Lasten beschreibt. Der geometrische Ansatz von Wakamatsu und Hirai [Wakamatsu & Hirai, 2004] ist unserem sehr ähnlich. Dieser basiert auf der Differentialgeometrie der Kurven, die mit einem rechtshändigen Orthonormalsystem ausgestattet sind. Die Parametrisierung der Rotationen erfolgt auf der Basis globaler EULER-Winkel. Diese sind, wie wir im Zusammenhang mit der analytischen Lösung bereits gesehen haben, Funktionen der Bogenlänge und können durch eine Linearkombination adäquater Basisfunktionen approximiert werden (RITZ-Ansatz). Entsprechende Gewichtungskoeffizienten ergeben sich durch das Lösen eines unbeschränkten Optimierungsproblems, bei dem die interne Energie des Stabes minimiert wird. Die Wahl geeigneter Basisfunktionen ist jedoch kein triviales Problem, da sie die durch die Eu-

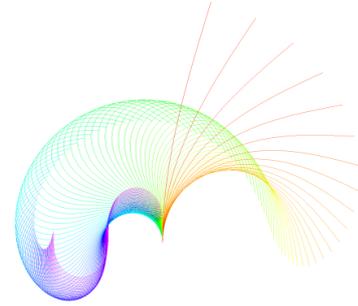


Abbildung 3.9: Sequenz einer Stabdeformation infolge einer terminalen Folgelast mit zunehmender Größe. Die Richtung der Terminalkraft bleibt immer senkrecht zur Mittellinie des Stabes.

LER-Winkel induzierten Singularitäten handhaben müssen. Aus dem oben Gesagten wissen wir jedoch schon, dass damit ein geeigneter Approximant für die JACOBIschen elliptischen Funktionen gesucht wird. Die mit dem Ansatz von Wakamatsu und Hirai [Wakamatsu & Hirai, 2004] erzielten Berechnungszeiten sind selbst für einfachste Probleme, wie dem geraden Stab im Gravitationsfeld, vergleichsweise lang. Unser Ansatz operiert dagegen direkt auf den diskreten Daten. Das zu minimierende Energiefunktional der Faser ist durch die Summe der gespeicherten inneren Energie \mathcal{E}_E , der potentiellen Energie \mathcal{E}_B infolge von Körperkräften und der Arbeit aller externen Belastungen, die am Stab angreifen, in der Form

$$\{\mathcal{E}_T = \mathcal{E}_E - \mathcal{E}_B - \mathcal{W}\} \xrightarrow{!} \min \quad (3.130)$$

gegeben. Die negativen Vorzeichen in Gl. (3.130) sind durch die Tatsache begründet, dass äußere Kräfte dem System Energie zuführen.

Wir postulieren für die hyperelastische Faser die Existenz einer Verzerrungsenergiefunktion $W(\mathbf{v}, \boldsymbol{\kappa}, s)$, mit $\mathbf{n} = \partial W / \partial \mathbf{v}$ und $\mathbf{m} = \partial W / \partial \boldsymbol{\kappa}$. Die elastische Energie ergibt sich dann als Integral über das gesamte Gebiet Ω ,

$$\mathcal{E}_E = \int_{\Omega} W(\boldsymbol{\kappa}) \, ds. \quad (3.131)$$

Die gespeicherte Energie ist als Differenz aus der aktuellen Konfiguration und der Referenzkonfiguration gegeben, $W(\mathcal{C}^C(\boldsymbol{\kappa}) - \mathcal{C}^o(\boldsymbol{\kappa}))$. Wenn wir für die Verzerrungsenergiefunktion den bekannten Ansatz

$$W(\boldsymbol{\kappa}(s)) = \frac{1}{2} \Delta \boldsymbol{\kappa}^T(s) \mathbf{K}_N \Delta \boldsymbol{\kappa}(s) = \frac{1}{2} (\kappa_1^2 + \kappa_2^2 + b\kappa_3^2), \quad (3.132)$$

wählen, ergeben sich die konstituierenden Gleichungen als Ableitungen dieser Funktion. Die Verzerrungsenergiefunktion nimmt ihr globales Minimum notwendig in der Referenzkonfiguration ein. Folglich ist diese ungespannt. \mathbf{v} liefert keinen Beitrag zur Gesamtenergie, weil wir Unscherbarkeit und Nichtdehnbarkeit der Faser annehmen. Die K_{ij} 's sind Materialkonstanten, wobei $K_{11} = EI_1$, $K_{22} = EI_2$, $K_{33} = G\mu$, und $K_{ij} = 0$ für $i \neq j$. E und G sind Materialparameter, auch bekannt als YOUNG'sche Zahl bzw. Torsions- oder Schubmodul. Die Größen $I_{1,2}$ bzw. μ stellen die entsprechenden Flächenträgheitsmomente des Querschnitts dar. Die Achsen \mathbf{d}_1 bzw. \mathbf{d}_2 werden so gewählt, dass sie mit den Hauptträgheitsachsen I_1 bzw. I_2 des Querschnitts zusammenfallen. Die Geometrie des Faserquerschnitts kann beliebige Formen annehmen. Wir müssen uns hier also nicht auf den Kreisquerschnitt beschränken, wie etwa bei der analytischen Lösung des LAGRANGESchen Falls.

Auf die Abhängigkeit der Materialeigenschaften des Keratins von der Luftfeuchte bzw. der Temperatur wurde bereits ausführlich eingegangen, vgl. Abs. (2.4.4) und Abs. (2.4.6). Die Befunde bieten eine einfache Möglichkeit zur Integration von Umweltbedingungen in das mechanische Modell. Die Werte der Materialparameter können dann bei gegebener Luftfeuchte bzw. Temperatur den entsprechenden Diagrammen durch Interpolation entnommen werden.

Die potentielle Energie infolge von Körperkräften wie Gravitation, ist durch die Beziehung

$$\mathcal{E}_B = \|\mathbf{g}\| \int_{\Omega} \rho(s) A(s) \mathbf{r}^T(s) \mathbf{g}_{\|\cdot\|} ds, \quad (3.133)$$

gegeben, wobei $\rho(s)A(s)$ das Gewicht der Faser pro Einheitslänge und $\mathbf{g}_{\|\cdot\|}$ den Kraftfeldeinheitsvektor darstellt.

Bei großen Kräften sind diese in Bezug auf das globale System zu behandeln, um das oben diskutierte Problem der Folgelasten zu umgehen. Große Kräfte treten bspw. bei der Interaktion elektrisch geladener Fasern auf.

Die durch eine externe Kraft \mathbf{f} verrichtete Arbeit ist durch $\mathcal{W} = \mathbf{f} \cdot \delta \mathbf{r}$ gegeben. Hierbei tragen nur Verschiebungen in Richtung der Kraft zur Gesamtenergie bei. Der Ansatzpunkt der Kraft bzw. seine Verschiebung werden durch die Integration der Kinematikgleichung erhalten. Prinzipiell kann \mathbf{f} jede Art von Kraft repräsentieren, COULOMB-, Kontakt- oder Adäsionskräfte, etc. Die durch die Summe der äußeren Kräfte geleistete Gesamtarbeit ist durch

$$\mathcal{W} = \sum_i \mathbf{f}_i \cdot \delta \mathbf{r}_i \quad (3.134)$$

gegeben.

3.3.2.2 Diskretisierung

Die zu minimierende Energiefunktion ist durch Gl. (3.130) gegeben. Gesucht ist also ein Minimierer $\boldsymbol{\kappa}^*$ dieses Variationsproblems, welches denjenigen Zustand der Faser beschreibt, bei dem die externen Kräfte infolge Gravitation oder COULOMB-, Kontakt- oder Adhäsionskräften mit den materialinternen Spannungen im Gleichgewicht stehen,

$$\boldsymbol{\kappa}^* = \operatorname{argmin}_{\boldsymbol{\kappa} \in \mathbb{R}^{3N}} \left\{ \mathcal{E}_T(\boldsymbol{\kappa}) \right\}^2. \quad (3.135)$$

Da jede Kraft eine lokale Störung induziert, diskretisieren wir das gesamte Gebiet.

Zur Auffindung des Minimierers benutzen wir einen Löser, der auf der Methode der konjugierten Richtungen basiert. Die Dimension des zu lösenden Optimierungsproblems hängt von der Diskretisierung ab, weil die lokalen Verdrehungen an jedem Knoten durch den Vektor $\boldsymbol{\kappa}_i$ gegeben sind. Die JACOBI-Matrix $\partial \mathcal{E}_T / \partial \boldsymbol{\kappa}$ des Energiefunktional lässt sich effizient mittels finiter Vorwärtsdifferenzen approximieren. Wir beschränken uns in den folgenden Beispielen auf den Einsatz einer äquidistanten Diskretisierung, mit einer festen Auflösung von 3 Knoten/cm. Bei einer Faser der Länge 30 cm hat der gesuchte Lösungsvektor 300 Dimensionen.

In jeder Iteration ist aus dem aktuellen Lösungsvektor $\boldsymbol{\kappa}^k$ die aktuelle Konfiguration $\mathcal{C}(\boldsymbol{\kappa}^k) = \{\mathbf{r}(\boldsymbol{\kappa}^k), \mathbf{d}_1(\boldsymbol{\kappa}^k), \mathbf{d}_2(\boldsymbol{\kappa}^k)\}$, d.h., die Positionen und Orientierungen entlang der Faser, durch Integration der Kinematikgleichung zu ermitteln. Dazu kann man prinzipiell auf das Verfahren aus Abs. (3.4.5) zurückgreifen.

3.3.2.3 Numerische Beispiele

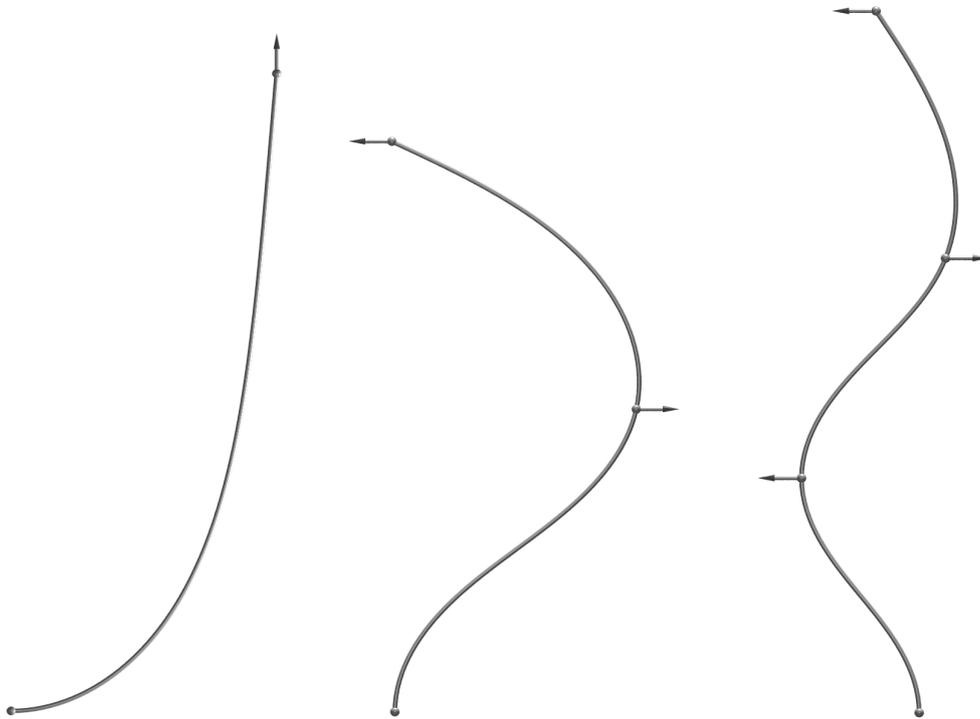


Abbildung 3.10: Fasern unter Einwirkung externer Lasten. Das linke Bild zeigt die Deformation eines horizontalen Stabes bei Einwirkung einer vertikalen Terminallast, die bzgl. des globalen Koordinatensystems wirkt. Das Problem der Folgelasten ist offenbar behoben.

Abb. (3.10) zeigt die auf der Basis des vorgestellten Verfahrens ermittelten Konfigurationen von Fasern ohne Anfangskrümmung. Im linken Bild ist die Deformation eines horizontalen Stabes unter Einwirkung einer vertikalen Terminallast dargestellt, die bzgl. des globalen Koordinatensystems wirkt. Das Problem der Folgelasten ist dadurch behoben. Einige Beispiele von Fasern mit unterschiedlichen Anfangskrümmungen unter Einfluss der Gravitation illustriert Abb. (3.11). Der Ansatz eignet sich auch zum Lösen von Randwertproblemen, bspw. auf der Basis von Positionsvorgaben. Abb. (3.12) zeigt das Ergebnis des Optimierungsprozesses für den Fall einer 50 cm langen gewundenen Faser unter Einfluss der Gravitation. Die Faser soll sich möglichst dicht an die beiden Vorgabepunkte anschmiegen.

Die Berechnung der Lösungen bei 100 Segmenten benötigt auf einem Intel Pentium M mit 1,86 GHz an die 100 ms. Wir ersparen uns an dieser Stelle die Präsentation weiterer Beispiele bzw. eine breitere Diskussion und verweisen auf den Abschnitt über die Elektrostatikberechnung auf Haaren, bei dem dieses Modell zum Einsatz kommt.

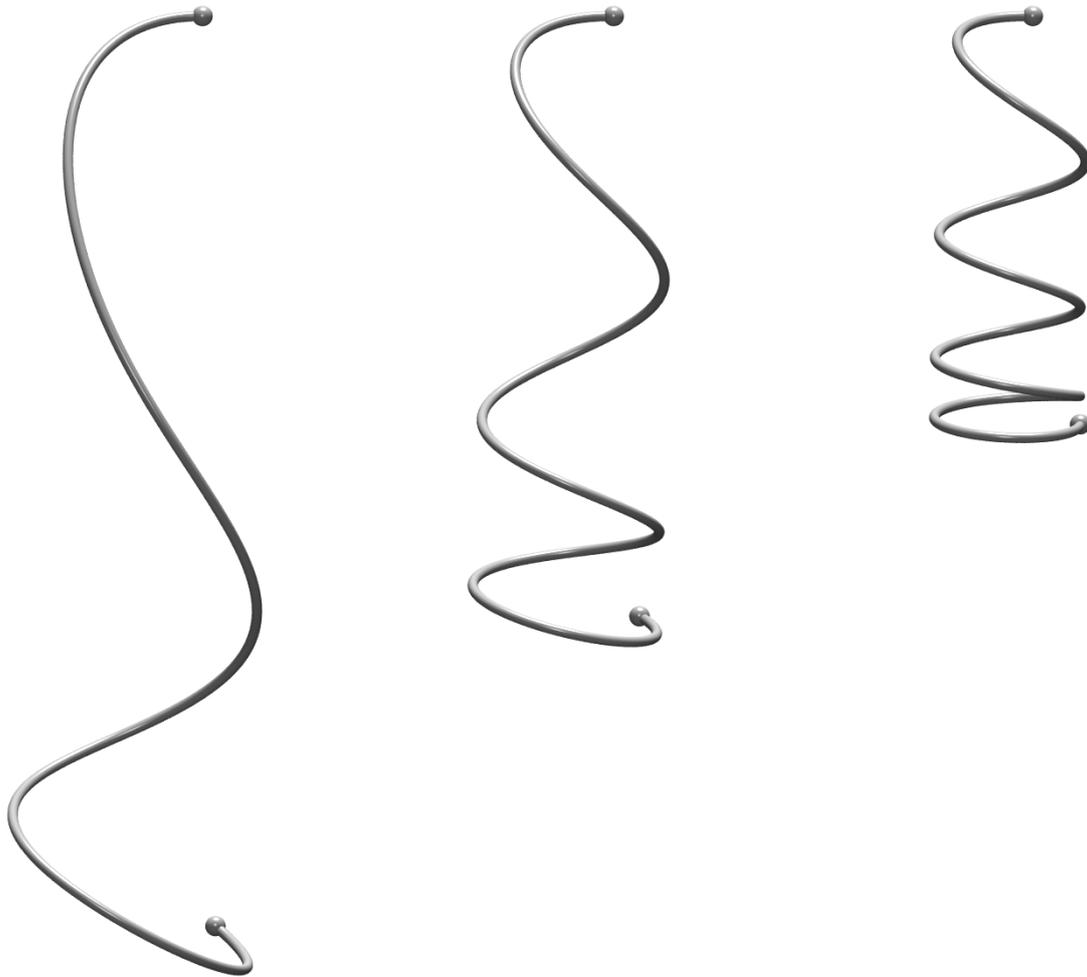


Abbildung 3.11: Fasern mit unterschiedlichen Anfangskrümmungen κ° im Gravitationsfeld. Die Konfigurationen entsprechen den Minimalenergiekurven.

3.4 Lösung der dynamischen Cosserat-Gleichungen

Die dynamischen COSSERAT-Gleichungen stellen recht hohe Anforderungen an potentielle numerische Lösungsverfahren. Das PDE-System, welches das dynamische Verhalten einer nicht dehnbaren Faser beschreibt, ist ein kanonisches Beispiel für ein steifes Differentialgleichungssystem. Die Zwangsbedingung, die einer longitudinalen Deformation entgegenwirkt, erfüllt die Funktion einer unendlich steifen Feder. Ihre Wirkung auf die Lösung kann man sich anhand einer Kugel veranschaulichen, die an einer Feder mit definierter Steifigkeit befestigt ist und unter Einwirkung der Gravitation eine ungedämpfte vertikale Schwingung vollführt. Erhöht man die Steifigkeit der Feder, wird die Auslenkung der Schwingung kleiner, ihre Frequenz dagegen größer. Im Grenzfall einer unendlich großen Federkonstante degeneriert die Feder zum unelastischen Stab, der die Kugel mit der Aufhängung starr verbindet. Die Frequenz ist dann theoretisch un-

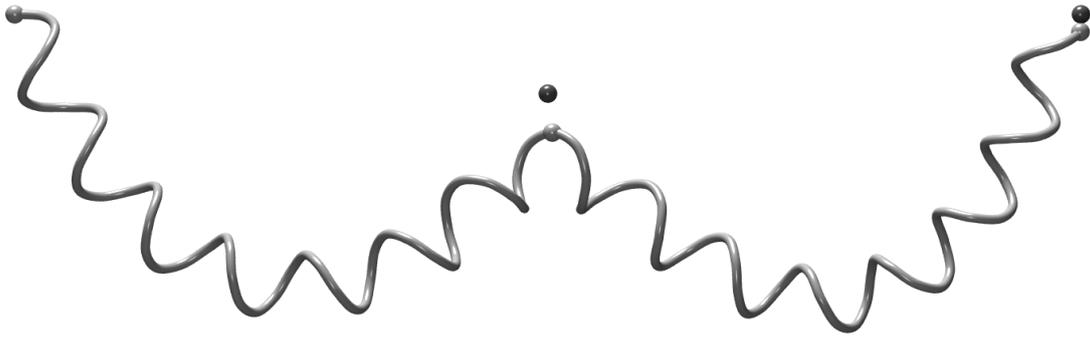


Abbildung 3.12: Randwertproblem: Stab der Länge 50 cm mit Anfangskrümmung unter Einfluss der Gravitation. Der Stab soll sich möglichst dicht an die beiden Vorgabepunkte anschmiegen.

endlich groß. Da der Begriff eines starren Körpers aber lediglich ein Modell ist, findet man auch hier ein Resonanzverhalten mit endlicher Frequenz. Auf die Lösung übertragen bedeutet das, dass ihre niederfrequenten Anteile, welche die globale Deformation der Faser beschreiben, von einem hochfrequenten Anteil überlagert werden. Da explizite Integratoren sich an der kleinsten Schwingung orientieren, würde die Schrittweite beim numerischen Lösungsprozess extrem klein ausfallen. Implizite Verfahren dagegen mitteln zwischen der neuen und der aktuellen Lösung und filtern auf diese Weise hochfrequente Anteile aus der Lösung. Wir greifen deshalb zur räumlichen und zeitlichen Integration der dynamischen COSSERAT-Gleichungen auf ein Verfahren aus dem Bereich der Strukturmechanik zurück, das sich durch besonders interessante Eigenschaften auszeichnet. Das *Generalisierte α -Verfahren* von Chung und Hulbert [Chung & Hulbert, 1993] ist ein implizites finites Differenzenschema mit Koeffizientenmittelung und Genauigkeit zweiter Ordnung. Zwei weitere Eigenschaften zeichnen dieses Verfahren aus: 1.) Es bietet eine steuerbare numerische Dämpfung und 2.) es ist, zumindest unter theoretischen Gesichtspunkten, *unbeschränkt* stabil. Erstere Eigenschaft bedeutet, dass die durch das implizite Schema induzierte numerische Dämpfung des Systems steuerbar ist. Damit kann eine kontrolliert gedämpfte Schwingung des System auf rein numerischem Wege erreicht werden. Das schließt die Integration einer zusätzlichen geschwindigkeitsabhängigen Dämpfungsfunktion $f(\omega, v)$ in die Systemgleichungen nicht aus.

Die zweite Eigenschaft hingegen garantiert eine stabile Lösung, unabhängig von der Wahl der zeitlichen und räumlichen Schrittweiten. An praktischen Beispielen lässt sich zeigen, dass diese Aussagen auch in der Praxis noch Gültigkeit besitzen.

Das generalisierte α -Verfahren wurde bereits erfolgreich zur Simulation von Unterseekabeln [Gobat, 2000, Gobat & Grosenbaugh, 2001, Gobat et al., 2002] und DNA-Molekülen [Goyal et al., 2003, Goyal et al., 2005, Goyal, 2006] im Bereich des Mechanical Engineering eingesetzt. Das Verfahren von Goyal et al. [Goyal et al., 2003, Goyal et al., 2005, Goyal, 2006] basiert ebenfalls auf den dynamischen KIRCHHOFF-Gleichungen. Der numerische Lösungsansatz entspricht allerdings einem Shootingverfahren und unterscheidet sich damit von unserem Ansatz.

3.4.1 Generalisiertes Alpha-Verfahren

Im Jahre 1993 haben Chung und Hulbert [Chung & Hulbert, 1993] das Generalisierte α -Verfahren als neues Integrationsverfahren für Probleme aus dem Bereich der Strukturmechanik vorgestellt. Es zeichnet sich primär durch eine kontrollierbare numerische Dissipation störender hochfrequenter Anteile in der numerischen Lösung strukturmechanischer Probleme aus. Solche treten bspw. bei Einsatz der Finite-Elemente-Methode bei zu grober räumlicher Auflösung der hochfrequenten Zustände auf. Solche Verfahren verbessern i.d.R. auch das Konvergenzverhalten von iterativen Lösungsstrategien für nicht lineare Probleme.

Die Idee der Einführung einer kontrollierbaren numerischen Dämpfung in den Integrationsprozess ist nicht neu und fand u.a. Verwirklichung im α -HHT [Hilber et al., 1977] oder WBZ- α -Verfahren [Wood et al., 1981]. Das Ziel bei der Entwicklung solcher Verfahren besteht darin, die Dämpfung hochfrequenter Anteile zu maximieren und gleichzeitig die wichtigen niederfrequenten Anteile zu erhalten. Beim α -Verfahren ist dieses Verhältnis optimal, d.h., für eine vorgegebene Dämpfung der hochfrequenten Anteile wird die Dämpfung der niederfrequenten Anteile minimal.

Zur Erläuterung des Verfahrens beginnen wir mit der Standardgleichung für lineare Probleme der Strukturmechanik,

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{D}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) + \mathbf{\Lambda}(t) = \mathbf{0}, \quad (3.136)$$

wobei \mathbf{M} , \mathbf{D} und \mathbf{K} die Masse-, Dämpfungs- bzw. Steifigkeitsmatrix darstellen. $\mathbf{x}(t)$ ist der Verschiebungsvektor, dessen erste und zweite Zeitableitung hier als Geschwindigkeit bzw. Beschleunigung auftreten. Der Vektor $\mathbf{\Lambda}(t)$ beschreibt die externen Kräfte, die zum Zeitpunkt t auf das System einwirken. Gesucht werden Funktionen $\mathbf{x}(t)$, $\mathbf{v}(t) = \dot{\mathbf{x}}(t)$ und $\mathbf{a}(t) = \ddot{\mathbf{x}}(t)$, die Gleichung Gl. (3.136) für alle $t \in [0, t_N]$ erfüllen, wobei die Anfangsbedingungen $\mathbf{x}(0) = \mathbf{x}_0$ und $\mathbf{v}(0) = \mathbf{v}_0$ gelten. In seiner Grundform schreiben wir das Integrationsschema wie folgt,

$$\mathbf{M}\mathbf{a}_{1-\alpha_m} + \mathbf{D}\mathbf{v}_{1-\alpha_f} + \mathbf{K}\mathbf{x}_{1-\alpha_f} + \mathbf{\Lambda}(t_{1-\alpha_f}) = \mathbf{0}, \quad (3.137)$$

mit der Ersetzungsregel $(\cdot)_{1-\alpha} = (1-\alpha)(\cdot)_i + \alpha(\cdot)_{i-1}$ und den Approximationen

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \Delta t \mathbf{v}_{i-1} + \Delta t^2 \left[\left(\frac{1}{2} - \beta \right) \mathbf{a}_{i-1} + \beta \mathbf{a}_i \right], \quad (3.138)$$

$$\mathbf{v}_i = \mathbf{v}_{i-1} + \Delta t \left[(1-\gamma) \mathbf{a}_{i-1} + \gamma \mathbf{a}_i \right]. \quad (3.139)$$

α_m , α_f , γ und β sind Integrationskoeffizienten. Wir ersparen es uns an dieser Stelle, die Gleichungen voll zu expandieren.

Die Aktualisierungsschemata der Positionen und Geschwindigkeiten zum Zeitpunkt i entsprechen damit denen des klassischen NEWMARK-Integrators. Das gesamte Integrationsschema resultiert in einem monoton wachsenden maximalem Positions- und Geschwindigkeitsfehler.

3.4.2 Stabilitätsanalyse

Im Folgenden untersuchen wir die Stabilität des α -Verfahrens mittels der VON NEUMANNschen Stabilitätsanalyse und folgen dabei im Wesentlichen Gobat [Gobat, 2000].

Die Stabilität eines numerischen Verfahrens kann für die folgende lineare, homogene Differentialgleichung erster Ordnung,

$$\dot{y} + \omega y = 0, \quad (3.140)$$

auch bekannt als DAHLQUISTsche Testgleichung, mit der Anfangsbedingung $y(0) = y_0$ analysiert werden, indem man das zu untersuchende Verfahren darauf anwendet. Mit dem α -Verfahren und der generalisierten Trapezregel für die Approximation von y^i erhält man das folgende Gleichungssystem,

$$(1 - \alpha)\dot{y}^i + \alpha\dot{y}^{i-1} + \omega[(1 - \beta)y^i + \beta y^{i-1}] = 0, \quad (3.141)$$

$$-y^i + y^{i-1} + \Delta t[\gamma\dot{y}^i + (1 - \gamma)\dot{y}^{i-1}] = 0. \quad (3.142)$$

Das Gleichungssystem lässt sich in Matrixform schreiben. Durch Auflösen von Gl. (3.141) und Gl. (3.142) nach y^i bzw. \dot{y}^i und Gleichsetzen erhält man sofort,

$$\begin{bmatrix} y \\ \dot{y} \end{bmatrix}^i = \mathbf{A} \begin{bmatrix} y \\ \dot{y} \end{bmatrix}^{i-1}, \quad (3.143)$$

wobei \mathbf{A} die so genannte Verstärkungsmatrix darstellt, die hier wie folgt definiert ist,

$$\mathbf{A} = \frac{1}{D} \begin{bmatrix} (1 - \alpha) - \gamma\beta\Omega & (1 - \gamma - \alpha)\Delta t \\ -\omega & -\alpha - (1 - \gamma)(1 - \beta)\Omega \end{bmatrix}, \quad (3.144)$$

mit $D = (1 - \alpha) + \gamma(1 - \beta)\Omega$ und $\Omega = \omega\Delta t$.

Zu einer Aussage über die Stabilität gelangen wir jetzt durch Betrachtung der Eigenwerte λ_i bzw. des Spektralradius $\rho(\mathbf{A}) = \max(|\lambda_i|)$ der Verstärkungsmatrix. Letzterer bestimmt das Anwachsen oder Abfallen der Lösung von einem Zeitschritt zum nächsten. Das Verfahren bleibt stabil, falls der Spektralradius $\rho(\mathbf{A}) \leq 1$ ist. Für $\rho > 1$ wächst die Lösung von einem Zeitschritt zum nächsten und das Verfahren wird instabil. Für das α -Verfahren ergeben sich die folgenden Eigenwerte λ_i ,

$$\lambda_{1,2} = -\frac{1}{2D} \left[2\alpha + (1 - \beta - \gamma + 2\gamma\beta)\Omega - 1 \pm \sqrt{Q} \right], \quad (3.145)$$

wobei $Q = [\gamma(\gamma + 2\beta - 2) + (\beta - 1)^2]\Omega^2 + 2(2\alpha - \beta + \gamma - 2)\Omega + 1$ ist. Daraus leiten wir ab, dass das Verfahren stabil bleibt, falls für beliebige Ω gilt,

$$\alpha \leq \frac{1}{2}, \quad \beta \leq \frac{1}{2}, \quad \gamma \geq \frac{1}{2}. \quad (3.146)$$

Die Anforderung an ein Verfahren zweiter Ordnung folgt direkt aus Gl. (3.141) und Gl. (3.142), wenn diese für zwei aufeinander folgende Zeitschritte (Mehrschrittverfahren) formuliert und anschließend in eine TAYLOR-Reihe entwickelt werden. Es ergibt sich,

$$\alpha - \beta + \gamma = \frac{1}{2}. \quad (3.147)$$

Hier stellt sich die Frage, wie der Zusammenhang zwischen einem gewünschten Dissipationsgrad im hochfrequenten Bereich und den drei Faktoren der α -Methode (α, β, γ) hergestellt werden

kann. Der Spektralradius selbst ist ein Maß für den Grad der numerischen Dissipation. Je kleiner der Spektralradius, desto größer ist die numerische Dämpfung. Damit die niederfrequenten Anteile von der Dämpfung im hochfrequenten Bereich unbeeinflusst bleiben, fordern wir, dass der Spektralradius im niederfrequenten Bereich etwa bei eins liegt und gleichmäßig abfällt, wenn $\Omega = \omega \Delta t$ zunimmt, vgl. Chung und Hulbert [Chung & Hulbert, 1993]. D.h., hochfrequentes Rauschen wird durch einen für $\Omega \rightarrow \infty$ monoton fallenden Spektralradius kontrolliert. Diese Anforderung an die Glattheit beschränkt den zulässigen Bereich der Eigenwerte der Verstärkungsmatrix und reduziert das α -Verfahren auf eine Einparametermethode, die nur von dem Eigenwert λ^∞ abhängt. Dieser Wert bestimmt den Anteil der Dissipation im hochfrequenten Bereich, woraus sich die drei Faktoren des Verfahrens für das Problem erster Ordnung dann wie folgt herleiten lassen. Für $\Omega = 0$ sind die beiden Eigenwerte $\lambda_{1,2}$ der Verstärkungsmatrix reell, werden dann für $\Omega_c < \Omega$ oberhalb eines Grenzwertes Ω_c komplex Konjugierte, um anschließend für $\Omega \rightarrow \infty$ wieder reelle Werte anzunehmen. Im Grenzfall $\Omega \rightarrow \infty$ gilt $\lambda_1^\infty = \beta/(\beta - 1)$ und $\lambda_2^\infty = (\gamma - 1)/\gamma$. Die monoton fallende Eigenschaft des Spektralradius wird erreicht, indem die kritische Frequenz der zweiten Verzweigung, also vom Komplexen zurück zum Reellen ins Unendliche verschoben wird, d.h., $\lambda_{1,2}^\infty = \lambda^\infty$. Damit wird die zweite Verzweigung verhindert. Wegen $\lambda_{1,2}^\infty = \lambda^\infty$ und mit der Bedingung für die Genauigkeit zweiter Ordnung erhalten wir dann schließlich,

$$\beta = \frac{\lambda^\infty}{\lambda^\infty - 1}, \quad \gamma = \frac{1}{1 - \lambda^\infty}, \quad \alpha = \frac{3\lambda^\infty + 1}{2\lambda^\infty - 2}. \quad (3.148)$$

Daraus resultiert also ein Verfahren von der Genauigkeit zweiter Ordnung, das nur noch von dem Eigenwert λ^∞ abhängt. Der Parameter $\lambda^\infty \in [-1, 0]$ wird negativ gewählt, so dass $\alpha, \beta, \gamma \in [1, 0]$

Durch eine geeignete Wahl der Parameter α, β und γ können eine ganze Reihe bekannter Integratoren aus dem α -Verfahren abgeleitet werden, bspw. die Box-Methode ($\alpha = 1/2, \beta = 1/2, \gamma = 1/2$), das Rückwärts-Differenzen-Verfahren ($\alpha = 0, \beta = 0, \gamma = 1$) oder das WBZ- α -Verfahren ($\alpha, \beta = 0, \gamma = 1/2 - \alpha$).

Das PDE-System, das wir hier betrachten, unterscheidet sich grundlegend von der einfachen DAHLQUIST-Testgleichung und von daher lassen sich diese Auswahlregeln nicht einfach auf den COSSERAT-Fall übertragen. Momentan ist keine Strategie bekannt, die eine systematische Wahl der Koeffizienten zulässt. Sie sind problemabhängig und damit an die jeweiligen Testfälle anzupassen. Allerdings zeigen unsere Erfahrungen, dass einerseits die sinnvollen Wertebereiche von Problemfall zu Problemfall nur geringen Schwankungen unterworfen sind und dass sich andererseits über die auf der Basis der Stabilitätsanalyse hergeleiteten Auswahlregeln oftmals akzeptable Einstiegswerte finden lassen.

3.4.3 Diskretisierung der Bewegungsgleichungen

Wir wenden nun das α -Verfahren auf die COSSERAT-Gleichungen an, sowohl zur zeitlichen als auch zur räumlichen Diskretisierung. Die Ersetzungsregel ist also zweimal in Folge anzuwenden, einmal für den Zeitparameter t und einmal für den räumlichen Parameter s .

Die Koeffizienten der zeitlichen Diskretisierung werden mit dem Index $(\)_t$, die der räumlichen mit $(\)_s$ gekennzeichnet. Wir diskretisieren Gl. (3.66) zunächst wie folgt in der Zeit,

$$\mathbf{M}^{1-\alpha_t} \partial_t \mathbf{x}^{1-\alpha_t} + \mathbf{K}^{1-\beta_t} \partial_s \mathbf{x}^{1-\beta_t} + \mathbf{\Lambda}^{1-\beta_t} = \mathbf{0}, \quad (3.149)$$

mit $(\cdot)^{1-\epsilon} = (1-\epsilon)(\cdot)^j + \epsilon(\cdot)^{j-1}$. Hierbei sind α_t und β_t Parameter zur Mittlung der Masse bzw. der Steifigkeit über zwei aufeinander folgende Zeitschritte. Ein expliziter Dämpfungsterm, wie in Gl. (3.136), wird nicht mitgeführt.

Wir erhalten unter Anwendung der Ersetzungsregel die folgende semidiskrete Gleichung,

$$\begin{aligned} & [(1-\alpha_t)\mathbf{M}^i + \alpha_t\mathbf{M}^{i-1}][(1-\alpha_t)\partial_t\mathbf{x}^i + \alpha_t\partial_t\mathbf{x}^{i-1}] \\ & + [(1-\beta_t)\mathbf{K}^i + \beta_t\mathbf{K}^{i-1}][(1-\beta_t)\partial_s\mathbf{x}^i + \beta_t\partial_s\mathbf{x}^{i-1}] \\ & + [(1-\beta_t)\mathbf{\Lambda}^i + \beta_t\mathbf{\Lambda}^{i-1}] = \mathbf{0}. \end{aligned} \quad (3.150)$$

Es ist sinnvoll, die Masse- und Steifigkeitsmatrix als zeitlich konstant anzunehmen, so dass die Mittlung entfällt. Das scheint gerechtfertigt, weil die Faser bei der Deformation weder ihre Masse noch ihren Querschnitt ändert. Die Biege- und Torsionssteifigkeit hingegen sind in der Regel von der Historie abhängig. Solche nicht linearen Effekte sollen hier aber vernachlässigt werden, weil eine der Grundannahmen ja die eines hyperelastischen Stabes war.

Unter diesen Voraussetzungen vereinfachen sich die Gleichungen dann entsprechend zu,

$$\begin{aligned} & \mathbf{M}[(1-\alpha_t)\partial_t\mathbf{x}^i + \alpha_t\partial_t\mathbf{x}^{i-1}] + \mathbf{K}[(1-\beta_t)\partial_s\mathbf{x}^i + \beta_t\partial_s\mathbf{x}^{i-1}] \\ & + [(1-\beta_t)\mathbf{\Lambda}^i + \beta_t\mathbf{\Lambda}^{i-1}] = \mathbf{0}. \end{aligned} \quad (3.151)$$

Diese Gleichung wird dann durch eine anschließende räumliche Diskretisierung in die vollständig diskrete Form der dreidimensionalen COSSERAT-Gleichungen überführt,

$$\begin{aligned} & \mathbf{M}[(1-\alpha_t)\partial_t\mathbf{x}_{1-\alpha_s}^i + \alpha_t\partial_t\mathbf{x}_{1-\alpha_s}^{i-1}] + \mathbf{K}[(1-\beta_t)\partial_s\mathbf{x}_{1-\beta_s}^i + \beta_t\partial_s\mathbf{x}_{1-\beta_s}^{i-1}] \\ & + [(1-\beta_t)\mathbf{\Lambda}_{1-\beta_s}^i + \beta_t\mathbf{\Lambda}_{1-\beta_s}^{i-1}] = \mathbf{0}. \end{aligned} \quad (3.152)$$

Dazu wendet man dieselbe Termersetzungsregel wie oben an, nur mit der Bogenlänge s als Diskretisierungsparameter, also $(\cdot)_{1-\epsilon} = (1-\epsilon)(\cdot)_j + \epsilon(\cdot)_{j-1}$ und erhält die finale Form des diskreten Gleichungssystems,

$$\begin{aligned} & \mathbf{M}\{(1-\alpha_t)[(1-\alpha_s)\partial_t\mathbf{x}_j^i + \alpha_s\partial_t\mathbf{x}_{j-1}^i] + \alpha_t[(1-\alpha_s)\partial_t\mathbf{x}_j^{i-1} + \alpha_s\partial_t\mathbf{x}_{j-1}^{i-1}]\} \\ & + \mathbf{K}\{(1-\beta_t)[(1-\beta_s)\partial_s\mathbf{x}_j^i + \beta_s\partial_s\mathbf{x}_{j-1}^i] + \beta_t[(1-\beta_s)\partial_s\mathbf{x}_j^{i-1} + \beta_s\partial_s\mathbf{x}_{j-1}^{i-1}]\} \\ & + \{(1-\beta_t)[(1-\beta_s)\mathbf{\Lambda}_j^i + \beta_s\mathbf{\Lambda}_{j-1}^i] + \beta_t[(1-\beta_s)\mathbf{\Lambda}_j^{i-1} + \beta_s\mathbf{\Lambda}_{j-1}^{i-1}]\} = \mathbf{0}. \end{aligned} \quad (3.153)$$

Auch hier nehmen wir vereinfachend an, dass die Masse und die Steifigkeit vom Diskretisierungsparameter s unabhängig sind. Eine entsprechende Mittlung über aufeinander folgende Zeitschritte entfällt also.

Die Approximation der zeitlichen und räumlichen Ableitungen von \mathbf{x} , $\partial_t\mathbf{x}$ und $\partial_s\mathbf{x}$ erfolgt auf der Basis der Trapezregel. Man erhält dann die folgenden Gleichungen, die dem klassischen NEWMARK-Schema ähneln,

$$\partial_t\mathbf{x}^i = \frac{1}{\gamma_t\Delta t} [(\mathbf{x}^i - \mathbf{x}^{i-1}) - \Delta t(1-\gamma_t)\partial_t\mathbf{x}^{i-1}], \quad (3.154)$$

$$\partial_s\mathbf{x}_j = \frac{1}{\gamma_s\Delta s} [(\mathbf{x}_j - \mathbf{x}_{j-1}) - \Delta s(1-\gamma_s)\partial_t\mathbf{x}_{j-1}]. \quad (3.155)$$

Die Wahl der impliziten Trapezregel ist optimal, weil es nach der zweiten DAHLQUIST-Testbarriere kein A-stabiles Verfahren höherer Ordnung als zwei und mit einer besseren Fehlerkonstante als 1/12 geben kann.

Die Trapezregel setzt man in Gl. (3.153) ein und erhält die folgende, etwas unhandliche Gleichung,

$$\begin{aligned}
 \mathbf{M} \left\{ (1 - \alpha_t)(1 - \alpha_s) \left[\frac{1}{\gamma_t \Delta t} (\mathbf{x}_j^i - \mathbf{x}_j^{i-1}) - \frac{(1 - \gamma_t)}{\gamma_t} \partial_t \mathbf{x}_j^{i-1} \right] \right. \\
 + \alpha_s(1 - \alpha_t) \left[\frac{1}{\gamma_t \Delta t} (\mathbf{x}_{j-1}^i - \mathbf{x}_{j-1}^{i-1}) - \frac{(1 - \gamma_t)}{\gamma_t} \partial_t \mathbf{x}_{j-1}^{i-1} \right] \\
 \left. + \alpha_t \left[(1 - \alpha_s) \partial_t \mathbf{x}_j^{i-1} + \alpha_s \partial_t \mathbf{x}_{j-1}^{i-1} \right] \right\} \\
 + \mathbf{K} \left\{ (1 - \beta_t)(1 - \beta_s) \left[\frac{1}{\gamma_s \Delta s} (\mathbf{x}_j^i - \mathbf{x}_{j-1}^i) - \frac{(1 - \gamma_s)}{\gamma_s} \partial_s \mathbf{x}_{j-1}^i \right] \right. \\
 + \beta_t(1 - \beta_s) \left[\frac{1}{\gamma_s \Delta s} (\mathbf{x}_j^{i-1} - \mathbf{x}_{j-1}^{i-1}) - \frac{(1 - \gamma_s)}{\gamma_s} \partial_s \mathbf{x}_{j-1}^{i-1} \right] \\
 \left. + \beta_s \left[(1 - \beta_t) \partial_s \mathbf{x}_{j-1}^i + \beta_t \partial_s \mathbf{x}_{j-1}^{i-1} \right] \right\} \\
 + (1 - \beta_t) \left[(1 - \beta_s) \Lambda_j^i + \beta_s \Lambda_{j-1}^i \right] + \beta_t \left[(1 - \beta_s) \Lambda_j^{i-1} + \beta_s \Lambda_{j-1}^{i-1} \right] = \mathbf{0}. \quad (3.156)
 \end{aligned}$$

Unbekannt sind dabei die Größen \mathbf{x}^i und Λ^i , wobei $\Lambda^i = f(\mathbf{x}^i)$ eine Funktion des Zustandsvektors \mathbf{x}^i ist. Alle anderen Größen sind aus den vorangegangenen Zeitschritten, sowie den Anfangs- und Randbedingungen bekannt.

Wir lösen nach \mathbf{x}^i und Λ^i auf, wobei die Gesamtheit aller Terme auf der rechten Seite mit $\mathbf{H}_k := \mathbf{H}_k(\mathbf{x}_j^i, \mathbf{x}_{j-1}^i)$ bezeichnet wird,

$$\begin{aligned}
 \mathbf{M} \left\{ (1 - \alpha_t)(1 - \alpha_s) \left[\frac{1}{\gamma_t \Delta t} \mathbf{x}_j^{i-1} + \frac{(1 - \gamma_t)}{\gamma_t} \partial_t \mathbf{x}_j^{i-1} \right] \right. \\
 + \alpha_s(1 - \alpha_t) \left[\frac{1}{\gamma_t \Delta t} \mathbf{x}_{j-1}^{i-1} + \frac{(1 - \gamma_t)}{\gamma_t} \partial_t \mathbf{x}_{j-1}^{i-1} \right] - \alpha_t \left[(1 - \alpha_s) \partial_t \mathbf{x}_j^{i-1} + \alpha_s \partial_t \mathbf{x}_{j-1}^{i-1} \right] \left. \right\} \\
 + \mathbf{K} \left\{ \beta_t(1 - \beta_s) \left[\frac{1}{\gamma_s \Delta s} (\mathbf{x}_{j-1}^{i-1} - \mathbf{x}_j^{i-1}) + \frac{(1 - \gamma_s)}{\gamma_s} \partial_s \mathbf{x}_{j-1}^{i-1} \right] - \beta_s \beta_t \partial_s \mathbf{x}_{j-1}^{i-1} \right\} \\
 - \beta_t \left[(1 - \beta_s) \Lambda_j^{i-1} + \beta_s \Lambda_{j-1}^{i-1} \right] := \mathbf{H}_k, \quad (3.157)
 \end{aligned}$$

mit der unbekanntem linken Seite

$$\begin{aligned} & \left[\frac{(1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \mathbf{M} + \frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \mathbf{K} \right] \mathbf{x}_j^i + (1-\beta_t)(1-\beta_s) \Lambda_j^i \\ & + \left[\frac{\alpha_s(1-\alpha_t)}{\gamma_t \Delta t} \mathbf{M} - \frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \mathbf{K} \right] \mathbf{x}_{j-1}^i + \beta_s(1-\beta_t) \Lambda_{j-1}^i \\ & + \left[\beta_s(1-\beta_t) - \frac{(1-\beta_t)(1-\beta_s)(1-\gamma_s)}{\gamma_s} \right] \mathbf{K} \partial_s \mathbf{x}_{j-1}^i = \mathbf{H}_k. \end{aligned} \quad (3.158)$$

Die Terme lassen sich entsprechend ihrer Abhängigkeiten von \mathbf{x}_j^i und \mathbf{x}_{j-1}^i zusammenfassen zu,

$$\begin{aligned} & \underbrace{\frac{(1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \mathbf{M} \mathbf{x}_j^i + (1-\beta_t)(1-\beta_s) \left(\frac{1}{\gamma_s \Delta s} \mathbf{K} \mathbf{x}_j^i + \Lambda_j^i \right)}_{\mathbf{A}(\mathbf{x}_j^i)} \\ & + \underbrace{\frac{\alpha_s(1-\alpha_t)}{\gamma_t \Delta t} \mathbf{M} \mathbf{x}_{j-1}^i - (1-\beta_t) \left[\frac{(1-\beta_s)}{\gamma_s \Delta s} \mathbf{K} \mathbf{x}_{j-1}^i - \beta_s \Lambda_{j-1}^i \right]}_{\mathbf{B}(\mathbf{x}_{j-1}^i)} \\ & + \underbrace{\left[\beta_s(1-\beta_t) - \frac{(1-\beta_t)(1-\beta_s)(1-\gamma_s)}{\gamma_s} \right]}_{\zeta} \mathbf{K} \partial_s \mathbf{x}_{j-1}^i = \mathbf{H}_k. \end{aligned} \quad (3.159)$$

Wir erhalten eine nicht lineare inhomogene Differentialgleichung erster Ordnung der Form,

$$\mathbf{F}_k(\mathbf{x}_j^i, \mathbf{x}_{j-1}^i) := \mathbf{A}(\mathbf{x}_j^i) + \mathbf{B}(\mathbf{x}_{j-1}^i) + \zeta \mathbf{K} \partial_s \mathbf{x}_{j-1}^i - \mathbf{H}_k = \mathbf{0}. \quad (3.160)$$

Die Funktionen $\mathbf{A}(\mathbf{x}_j^i)$ und $\mathbf{B}(\mathbf{x}_{j-1}^i)$ sind nicht linear in \mathbf{x}_j^i und \mathbf{x}_{j-1}^i . Die Gesamtheit aller Gleichungen resultiert in einem nicht linearen Gleichungssystem, $\mathcal{F} : \mathbb{R}^{N \times 12} \mapsto \mathbb{R}^{(N-1) \times 12}$ der Form,

$$\mathcal{F}(\mathbf{x}) = \begin{Bmatrix} \mathbf{F}_1 \\ \mathbf{F}_2 \\ \vdots \\ \mathbf{F}_{N-1} \end{Bmatrix} = \begin{Bmatrix} \mathbf{A}(\mathbf{x}_2) + \mathbf{B}(\mathbf{x}_1) + \zeta \mathbf{K} \partial_s \mathbf{x}_1 - \mathbf{H}_1 \\ \mathbf{A}(\mathbf{x}_3) + \mathbf{B}(\mathbf{x}_2) + \zeta \mathbf{K} \partial_s \mathbf{x}_2 - \mathbf{H}_2 \\ \vdots \\ \mathbf{A}(\mathbf{x}_{N-1}) + \mathbf{B}(\mathbf{x}_{N-1}) + \zeta \mathbf{K} \partial_s \mathbf{x}_{N-1} - \mathbf{H}_{N-1} \end{Bmatrix} = \mathbf{0}. \quad (3.161)$$

Die Funktionen \mathbf{F}_k bilden vom \mathbb{R}^{24} in den \mathbb{R}^{12} ab. Bei N Knoten entspricht das $12(N-1)$ Gleichungen für $12N$ Unbekannte.

Die k -te Gleichung ist durch den Vektor \mathbf{x}_k mit der Vorgängergleichung $k-1$ gekoppelt. Diese Gleichungen müssen simultan gelöst werden, damit sich die Einflüsse an den hinteren Knoten auf die Knoten vorne auswirken können. Das ist insbesondere bei Einsatz eines *Shooting*-Verfahrens zu berücksichtigen.

Abb. (3.13) verdeutlicht, welche Knoteninformationen für die Ermittlung der Größen am Knoten (i, j) bekannt sein müssen. Neben den zeitlich-räumlichen Vorgängern gehen auch die Werte (i, j) mit in die Berechnung ein, da das Verfahren impliziter Natur ist.

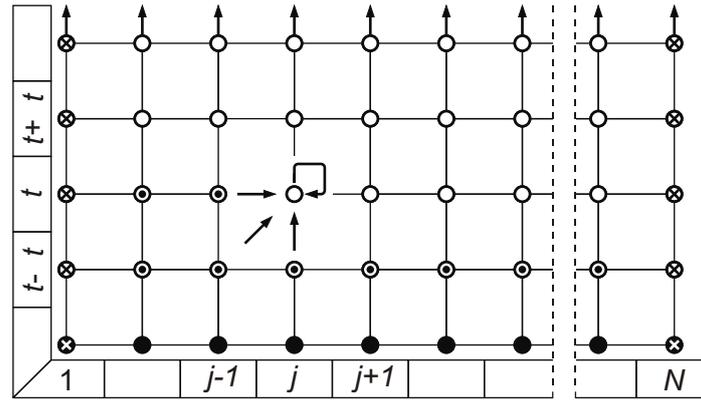


Abbildung 3.13: Gitter mit Anfangs- (schwarze Punkte) und Randbedingungen (gekreuzte Punkte). Die Pfeile, die zum Knoten (t, j) verweisen, verdeutlichen, welche Knoten für die Berechnung von \mathbf{x}_j^t benötigt werden. Man beachte, dass die Werte am Knoten \mathbf{x}_j^t selbst auch in die Berechnung eingehen, da die Formulierung impliziter Natur ist.

3.4.4 Randbedingungen

Wie oben bereits angedeutet, stehen bei N Knoten $12(N - 1)$ Gleichungen für die $12N$ unbekannt Knotengrößen zur Verfügung. Um das Gleichungssystem zu vervollständigen, bedarf es weiterer 12 Gleichungen. Diese sind gegeben durch die Randbedingungen an den Stellen $(s = 0)$ und $(s = L)$.

An der Stelle $(s = 0)$ geben wir im Standardfall die Position, sowie die Orientierung vor, $BC_0 = \{\mathbf{r}(0, \cdot), \mathbf{d}_1(0, \cdot), \mathbf{d}_2(0, \cdot)\}$. Wegen der koordinatenfreien Formulierung drücken wir diese Forderung durch die lineare bzw. Winkelgeschwindigkeit an der Stelle $(s = 0)$ aus, $BC_0 = \{\mathbf{v}(0, \cdot), \boldsymbol{\omega}(0, \cdot)\}$. Diese nehmen einen bestimmten Wert $\mathbf{v}(0, \cdot) = F^v(0, \cdot)$ bzw. $\boldsymbol{\omega}(0, \cdot) = F^\omega(0, \cdot)$ an, wobei die vektorwertigen Funktionen $F^v(0, \cdot) = \mathbf{0}$ und $F^\omega(0, \cdot) = \mathbf{0}$ für alle t den Wert null annehmen, falls die Faser fest eingespannt ist. Bei beweglichen Lagern dagegen liefern sie einen von null verschiedenen Wert. Soll eine Randbedingung in Form einer Beschleunigung vorgegeben werden, müssen wir zusätzlich die Ableitungen der Funktionen $\partial_t F = \mathbf{a}$ betrachten.

An der Stelle $(s = L)$ sind die Kraft und das Drehmoment zu spezifizieren, $BC_L = \{\mathbf{m}(L, \cdot), \mathbf{n}(L, \cdot)\}$. Das drückt man durch die beiden Variablen \mathbf{n} und $\boldsymbol{\kappa}$ als $BC_L = \{\boldsymbol{\kappa}(L, \cdot), \mathbf{n}(L, \cdot)\}$ aus. Mit dem Materialgesetz Gl. (3.55) folgt sofort $\boldsymbol{\kappa}(L, \cdot) = \mathbf{K}^{-1}\mathbf{m}(L, \cdot) + \boldsymbol{\kappa}^o(L)$, mit $\mathbf{m}(L, \cdot) = \mathbf{l}(L, \cdot)$. Da für \mathbf{n} kein äquivalentes Materialgesetz existiert und wir infolgedessen jeden Wert annehmen dürfen, der Bedingung Gl. (3.51) genügt, setzen wir $\mathbf{n}(L, \cdot) = \mathbf{f}(L, \cdot)$.

Die Integration alternativer Randbedingungen bereitet keinerlei Probleme. Soll bspw. die Faser mit einer konstanten Rate λ entlang der globalen Z-Achse gedehnt werden, wählt man die Randbedingung an der Stelle $(s = L)$ zu $BC_L = \{\lambda \mathbf{d}_{13}, \lambda \mathbf{d}_{23}, \lambda \mathbf{d}_{33}, 0, 0, 0\}$. Änderungen der Orientierung sind damit ausgeschlossen. Wir werden diesen Fall bei der Vorstellung der numerischen Beispiele betrachten.

Mit der Definition von sechs Randbedingungen auf jeder Seite der Faser stehen jetzt $12N$ Gleichungen für $12N$ unbekannte Knotengrößen zur Verfügung. Es gilt zu beachten, dass die Knoten-

größen bezüglich des lokalen Koordinatensystems angegeben werden und die Randbedingungen infolgedessen bei der Verwendung iterativer Lösungsstrategien in jedem Iterationsschritt angepasst, d.h., in das lokale System des Endpunktes ($s = L$) transformiert werden müssen.

3.4.5 Integration der Kinematikgleichungen

Bei der Herleitung des PDE-Systems haben wir die Kinematikgleichungen aus Gründen der Komplexitätsreduktion vom PDE-System entkoppelt. Die separate Integration von Gl. (3.2) und Gl. (3.1) liefert ein Matrixexponential $\mathbf{d}_k^i = \exp(\theta \tilde{\xi}) \mathbf{d}_k^{i-1}$, mit $(\theta, \tilde{\xi}) = (\|\boldsymbol{\omega}\| \Delta t, \tilde{\boldsymbol{\omega}} / \|\boldsymbol{\omega}\|)$ bzw. $(\|\boldsymbol{\kappa}\| \Delta s, \tilde{\boldsymbol{\kappa}} / \|\boldsymbol{\kappa}\|)$, das durch die RODRIGUES-Formel,

$$\exp(\theta \tilde{\xi}) = \mathbf{E}_3 + \sin \theta \tilde{\xi} + (1 - \cos \theta)^2 \tilde{\xi}, \quad (3.162)$$

approximiert werden kann. Das Matrixexponential ist orthogonal und erfüllt $\det(\exp(\theta \tilde{\xi})) = 1$, wahrt also die Eigenschaften der $\mathcal{SO}(3)$. Es wird im Sinne einer inkrementellen Änderung der Orientierungen interpretiert, welche an die Rotationsmatrizen anmultipliziert, die neuen Orientierungen an jedem Knoten liefert, vgl. [Simo & Vu-Quoc, 1988]. Das Matrixexponential stellt genau genommen nur dann eine Lösung der Kinematikgleichungen dar, wenn die Spinoren $\tilde{\boldsymbol{\omega}}$ und $\tilde{\boldsymbol{\kappa}}$ nicht explizit von der Zeit abhängen. Weil sich aber sowohl $\boldsymbol{\omega}$, als auch $\boldsymbol{\kappa}$ innerhalb eines Zeitschrittes ändern können, müssen beide Größen über das Zeitintervall Δt bzw. über die Segmentlänge Δs gemittelt werden. Das führt auf $\boldsymbol{\omega} = (\boldsymbol{\omega}^i + \boldsymbol{\omega}^{i-1})/2$ (für $\boldsymbol{\kappa}$ gilt Entsprechendes). Die Drehachse $\theta \tilde{\xi}$ interpretieren wir damit implizit als Integral über den DARBOUX-Vektor $\boldsymbol{\omega}$ bzw. Twistvektor $\boldsymbol{\kappa}$,

$$\theta_{\text{axial}}(\tilde{\xi}) = \int_t^{t+\Delta t} \boldsymbol{\omega}(t) dt. \quad (3.163)$$

$\boldsymbol{\omega}$ und $\boldsymbol{\kappa}$ stehen über die Kinematikgleichungen mit der Änderung der Orientierung in Beziehung und müssen als Konsequenz aus der Kompatibilitätsforderung im Rahmen einer gewissen numerischen Toleranz zum selben Ergebnis führen. Damit folgt für die Aktualisierung der Orientierungen

$$\mathbf{R}_j^i = \exp(\theta \tilde{\xi}) \mathbf{R}_j^{i-1}. \quad (3.164)$$

3.4.6 Rekonstruktion der Mittellinie

Die Mittellinie $\mathbf{r}(s, t)$ folgt in unserem Fall aus der Zwangsbedingung Gl. (3.51), $\partial_s \mathbf{r}(s, t) = \mathbf{d}_3(s, t)$. Da die Tangente mit der Normalen des Querschnitts zusammenfällt, reicht eine Integration der Tangentenfunktion $\mathbf{d}_3(s, t)$, wie bei der analytischen Lösung, aus. Auf der anderen Seite kann die Geschwindigkeitsfunktion auch direkt integriert werden, weil die Nichtdehnbarkeitsbedingung Gl. (3.51) garantiert, dass die neuen Geschwindigkeiten die Zwangsbedingungen erfüllen. Das steht im Gegensatz zu anderen Verfahren, bspw. dem Super-Helix-Modell, bei dem diese Bedingung implizit durch Interpretation des Paares $(\partial_t \mathbf{r}(s, t), \boldsymbol{\kappa})^T$ als Twist oder Schraubenbewegung erfüllt ist. Damit folgt für die Aktualisierung der Positionen

$$\mathbf{r}_j^i = \mathbf{r}_j^{i-1} + \Delta t (\mathbf{R}_j^i \mathbf{v}_j^i + \mathbf{R}_j^{i-1} \mathbf{v}_j^{i-1})/2. \quad (3.165)$$

3.4.7 Relaxationsverfahren

Als numerisches Lösungsverfahren für Systeme nicht linearer Gleichungen wie das aus Gl. (3.161), bietet sich das NEWTON-Verfahren wegen seiner Einfachheit an. Daraus können wir ein einfaches Relaxationsverfahren herleiten. Dazu linearisiert man die Funktionen \mathbf{A} und \mathbf{B} aus Gl. (3.160) im Bereich der aktuellen Lösung \mathbf{x}^* durch Entwicklung in eine TAYLOR-Reihe, die nach dem zweiten Glied abbricht und erhält dann Ausdrücke für $\mathbf{A}^*(\mathbf{x}_j^*)$ und $\mathbf{B}^*(\mathbf{x}_{j-1}^*)$,

$$\mathbf{A}^*(\mathbf{x}_j^*) = \mathbf{A}(\mathbf{x}_j^*) + \frac{\partial}{\partial \mathbf{x}_j^i} \mathbf{A}(\mathbf{x}_j^*) (\mathbf{x}_j^i - \mathbf{x}_j^*), \quad (3.166)$$

$$\mathbf{B}^*(\mathbf{x}_{j-1}^*) = \mathbf{B}(\mathbf{x}_{j-1}^*) + \frac{\partial}{\partial \mathbf{x}_{j-1}^i} \mathbf{B}(\mathbf{x}_{j-1}^*) (\mathbf{x}_{j-1}^i - \mathbf{x}_{j-1}^*). \quad (3.167)$$

Wir können deshalb alle Terme, die von \mathbf{x}^* abhängen, separieren. Man erhält ein lineares Gleichungssystem der Form

$$\mathbf{F}_k^*(\mathbf{x}_j^*, \mathbf{x}_{j-1}^*) = \mathbf{A}^*(\mathbf{x}_j^*) + \mathbf{B}^*(\mathbf{x}_{j-1}^*) + \zeta \mathbf{K} \partial_s \mathbf{x}_{j-1}^* - \mathbf{H}_k = \mathbf{0}, \quad (3.168)$$

das sich mit Standardverfahren lösen lässt. Das entspricht genau der Strategie des NEWTON-Verfahrens. Bei der Iteration wird eine Folge von Approximationen \mathbf{x}^k generiert, bis das Residuum \mathcal{R} der Funktion einen bestimmten Toleranzwert unterschreitet,

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \mathbf{J}^{-1}(\mathbf{x}^k) \mathcal{F}(\mathbf{x}^k). \quad (3.169)$$

Die Iteration startet mit einer geschätzten Lösung $(\mathbf{x}^*)^0$. Hierbei hat es sich als sinnvoll erwiesen, den Zustandsvektor vom letzten Zeitschritt zu wählen ($(\mathbf{x}^*)^0 = \mathbf{x}^{i-1}$), da wir berechtigterweise annehmen können, dass die gesuchte Lösung - zumindest im instationären Fall - dem aktuellen Zustand sehr ähnlich ist (zeitliche Kohärenz der Lösung). Das gilt natürlich nicht, wenn lediglich der statische Gleichgewichtszustand berechnet werden soll, der sich stark vom Ausgangszustand unterscheiden kann.

Das Residuum der Funktion Gl. (3.160) ist

$$\mathcal{R}(\mathbf{x}_j^*, \mathbf{x}_{j-1}^*) = \mathbf{A}(\mathbf{x}_j^*) + \mathbf{B}(\mathbf{x}_{j-1}^*) + \zeta \mathbf{K} \partial_s \mathbf{x}_{j-1}^* - \mathbf{H}_k. \quad (3.170)$$

Als Abbruchkriterium wählt man $\|\mathcal{R}\| \leq \epsilon$. In jeder Iteration berechnen wir $\partial_s \mathbf{x}^*$ und $\partial_t \mathbf{x}^*$ durch Auswertung von Gl. (3.154) und Gl. (3.155). Hierbei ist zu beachten, dass im ersten Schritt, also zum Zeitpunkt $t = 0$, der Vektor \mathbf{x}^{i-1} nicht bekannt ist. In diesem Fall verwenden wir die original Gleichungen und lösen diese nach $\partial_t \mathbf{x}$ auf,

$$\partial_t \mathbf{x} = -\mathbf{M}^{-1} (\mathbf{K} \partial_s \mathbf{x} + \mathbf{\Lambda}). \quad (3.171)$$

Da für $\partial_t \mathbf{n}$ keine Gleichung existiert, ist die Massematrix singulär und das Bilden der Inverse nicht möglich. Die Gleichungen für $(\partial_t \mathbf{v}, \partial_t \boldsymbol{\omega}, \partial_t \boldsymbol{\kappa})$ können einzeln extrahiert werden, für $\partial_t \mathbf{n}$ wählen den Startwert $\partial_t \mathbf{n} = \mathbf{0}$.

Unter der Annahme, dass die erforderlichen Gleichungen richtig eingegeben wurden, läuft der eigentliche Relaxationsalgorithmus wie folgt ab, vgl. Alg. (1). Gegeben seien ρ^∞ (oder α, β, γ),

$\partial_s \kappa_0^\circ$ und $\partial_s \mathbf{x}_0^0$. Daraus berechnet man die Integrationskoeffizienten, sowie die Ableitungen $\partial_s \kappa^\circ$, $\partial_s \mathbf{x}_0$ und $\partial_t \mathbf{x}^0$. Diese Berechnungen sind nur im ersten Zeitschritt durchzuführen. $\partial_t \mathbf{x}^0$ erhält man unter Verwendung von Gl. (3.171), weil $\partial_t \mathbf{x}^{i-1}$ im ersten Zeitschritt nicht bekannt ist. $\partial_t \mathbf{n}_0^0$ wird zu null gewählt, da keine entsprechende Gleichung für die Zeitableitung von $\partial_t \mathbf{n}_0^0$ existiert.

In der anschließenden NEWTON-Iteration erfolgt die Bestimmung eines Lösungsvektors \mathbf{x}^i für das Funktionensystem Gl. (3.161). In jeder Iteration ist dazu die JACOBI-Matrix aufzustellen und die inkrementelle Änderung $\delta \mathbf{x}$ des Vektors durch Lösen eines dünn besetzten Gleichungssystems zu bestimmen. Um nicht über den Lösungspunkt hinauszuschießen, addieren wir nicht das ganze Inkrement, sondern multiplizieren mit einem Relaxationsparameter $\mu \in [0, 1]$. Dieser ist so zu wählen, dass die Funktion $\mathcal{F}(\mathbf{x}^k + \mu \delta \mathbf{x})$ einen minimalen Wert annimmt. Wir werten an dieser Stelle die AMJO-Bedingung

$$\mathcal{F}(\mathbf{x}^k + \mu \delta \mathbf{x}) \leq \mathcal{F}(\mathbf{x}^k) + \iota \mu \delta \mathbf{x}^\top \nabla \mathcal{F}(\mathbf{x}^k), \quad (3.172)$$

mit $\iota = 10^{-4}$ aus. Solange diese Bedingung nicht erfüllt ist, setzt man $\mu := \mu/2$, wobei bei 1 gestartet wird. Diese Backtracking-Technik führt zu einem deutlich verbesserten Konvergenzverhalten. Sie globalisiert das NEWTON-Verfahren, das dadurch auch für Startwerte, die zu weit von der Lösung entfernt sind (was bei zu großen Kräften, die am System wirken, immer der Fall ist), zumindest auf ein lokales Minimum hin konvergiert.

Mit dem aktuellen Lösungsvektor \mathbf{x}^k bzw. den daraus extrahierten lokalen Winkelgeschwindigkeiten $\boldsymbol{\omega}^k$ und Krümmungen $\boldsymbol{\kappa}^k$ integrieren wir zunächst die kinematischen Gleichungen wie oben beschrieben und erhalten die aktuellen Koordinatensysteme an den einzelnen Knoten. Die Kräfte und Momente, die am System wirken, werden dann in die lokalen Knotenkoordinatensysteme transformiert und die ebenfalls von der Winkelgeschwindigkeit abhängigen Randbedingungen aktualisiert. Sie sind dagegen unabhängig von den Koordinaten der Knoten, so dass eine Aktualisierung derselben erst nach Erreichen der Konvergenz notwendig ist. Bei Erweiterung auf geometrische Randbedingungen, bspw. bei Verwendung von Punkt-zu-Punkt Bedingungen, erfolgt die Neuberechnung innerhalb der Iterationsschleife. Kräfte und Momente werden über einen Zeitschritt als konstant angenommen, um ein Springen der Lösung zu verhindern.

Der Pseudocode für den Relaxationsalgorithmus ist in Alg. (1) angegeben. Die hierin verwendeten Bezeichnung $\tilde{\mathbf{x}}$ stellt den reduzierten, d.h., von den Randbedingungen bereinigten Zustandsvektor dar, $\tilde{\mathbf{x}} = \{\mathbf{x}_1 \setminus \text{BC}_0, \mathbf{x}_2, \dots, \mathbf{x}_{N-1}, \mathbf{x}_N \setminus \text{BC}_L\}$. Das Ein- bzw. Ausblenden der Randbedingungen leistet die Funktion $\Xi : \mathbb{R}^{N \times 12} \mapsto \mathbb{R}^{(N-1) \times 12}$ bzw. ihre Umkehrfunktion $\mathbf{x} = \Xi^{-1}(\tilde{\mathbf{x}})$.

Da eine direkte Implementierung der angegebenen Gleichungen in einer Programmiersprache wie C++ ein aufwändiges und fehleranfälliges Unterfangen darstellt, empfiehlt sich der Einsatz eines Computeralgebraprogramms, bspw. Maple. Der Quellcode für die Auswertung der linken Seite von Gl. (3.161) bzw. die JACOBI-Matrix lässt sich damit problemlos und fehlerfrei innerhalb kurzer Zeit generieren. Dazu ist lediglich Gl. (3.160) einzugeben, unter Verwendung der Approximationen Gl. (3.154) bzw. Gl. (3.155), und bzgl. der Variablen \mathbf{x}_j und \mathbf{x}_{j-1} zu differenzieren. Der resultierende symbolische Ausdruck kann in optimierten Quellcode übersetzt und in Form einer Unterfunktion in einen bestehenden Softwarerahmen integriert werden.

Zur Berechnung des Lösungsvektors mittels des NEWTON-Verfahrens brauchen wir die JACOBI-Matrix $\mathbf{J} = \partial \mathcal{F} / \partial \mathbf{x}$ bzw. deren Inverse. Da jedes \mathbf{F}_k in Gl. (3.161) von \mathbf{x}_j^i und \mathbf{x}_{j-1}^i abhängt, müssen lediglich die komponentenweisen Ableitungen der Funktionen bzgl. dieser beiden Größen gebildet werden. Daraus erhält man die einzelnen Untermatrizen der globalen Matrix \mathbf{J} zu (hier nur

Algorithmus 1 : Relaxationsverfahren zur Lösung der diskreten dynamischen COSSERAT-Gleichungen. $\mathcal{F}(\mathbf{x})$ ist die linke Seite des diskretisierten Systems Gl. (3.161).

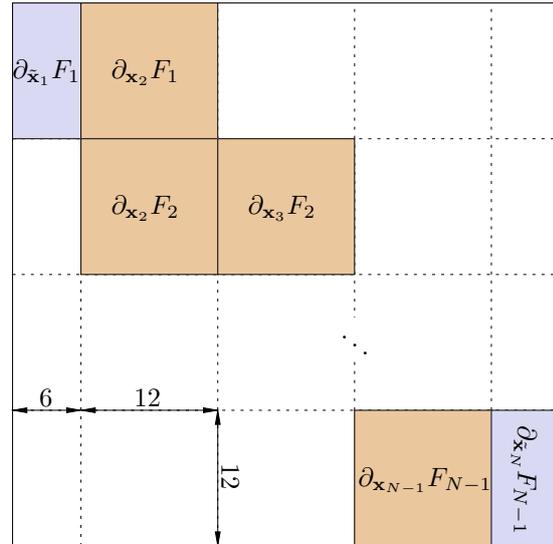
Data : $i, \Delta t, \mathbf{x}^{i-1}, \mathbf{r}^{i-1}, \mathbf{R}^{i-1}, \mathbf{f}^i, \mathbf{I}^i, \varepsilon$

Result : $\mathbf{x}^i, \mathbf{r}^i, \mathbf{R}^i$

```

1: if  $i = 0$  then
2:    $\{\alpha, \beta, \gamma\} \leftarrow \text{compCoefficients}(\rho^\infty)$ 
3:    $\mathbf{x}^i \leftarrow \text{updateBC}(\mathbf{x}^i)$ 
4:    $\partial_s \boldsymbol{\kappa}^0 \leftarrow \text{compDxDs}(\boldsymbol{\kappa}^0, \partial_s \boldsymbol{\kappa}_0^0, \Delta s, \gamma)$ 
5:    $\partial_s \mathbf{x}^i \leftarrow \text{compDxDs}(\mathbf{x}^i, \partial_s \mathbf{x}_0^i, \Delta s, \gamma)$ 
6:    $\partial_t \mathbf{x}^i \leftarrow -\widehat{\mathbf{M}}^{-1}(\widehat{\mathbf{K}} \partial_s \mathbf{x}^i + \boldsymbol{\Lambda}^i)$ 
7: end
8:  $\mathbf{x}^{i-1} \leftarrow \mathbf{x}^i; \partial_s \mathbf{x}^{i-1} \leftarrow \partial_s \mathbf{x}^i; \partial_t \mathbf{x}^{i-1} \leftarrow \partial_t \mathbf{x}^i$ 
9:  $k \leftarrow 0$ 
10: while  $\|\mathcal{F}(\mathbf{x}^i)\| \geq \varepsilon$  do
11:    $\tilde{\mathbf{x}}^k \leftarrow \Xi(\mathbf{x}^i)$ 
12:    $\mathbf{J} \leftarrow \partial \mathbf{F}_m / \partial \tilde{\mathbf{x}}_n^k$ 
13:    $\delta \tilde{\mathbf{x}} \leftarrow \text{sparseSolve}\{\mathbf{J} \delta \tilde{\mathbf{x}} = -\mathcal{F}(\mathbf{x}^i)\}$ 
14:    $\mu \leftarrow \text{compOptimalMu}()$ 
15:    $\tilde{\mathbf{x}}^{k+1} \leftarrow \tilde{\mathbf{x}}^k + \mu \delta \tilde{\mathbf{x}}$ 
16:    $\mathbf{x}_M^i \leftarrow \Xi^{-1}(\tilde{\mathbf{x}}^{k+1})$ 
17:    $\mathbf{R}^{k+1} \leftarrow \text{updateOrientations}(\mathbf{x}_M^i, \Delta s)$ 
18:    $\{\mathbf{f}, \mathbf{I}\}^{k+1} \leftarrow \Delta \mathbf{R}^{k+1} \{\mathbf{f}, \mathbf{I}\}^k$ 
19:    $\mathbf{x}^i \leftarrow \text{updateBC}(\mathbf{x}_M^i)$ 
20:    $\partial_t \mathbf{x}^i \leftarrow \text{compDxDt}(\mathbf{x}^i, \mathbf{x}^{i-1}, \Delta t, \gamma)$ 
21:    $\partial_s \mathbf{x}_0^i \leftarrow -\widehat{\mathbf{K}}^{-1}(\widehat{\mathbf{M}} \partial_t \mathbf{x}^i + \boldsymbol{\Lambda}^i)$ 
22:    $\partial_s \mathbf{x}^i \leftarrow \text{compDxDs}(\mathbf{x}^i, \partial_s \mathbf{x}_0^i, \Delta s, \gamma)$ 
23:    $k \leftarrow k + 1$ 
24: end
25:  $\mathbf{r}^i \leftarrow \text{updatePositions}(\mathbf{x}^i, \Delta s)$ 
26:  $\{\mathbf{f}, \mathbf{I}\}^{i-1} \leftarrow \{\mathbf{f}, \mathbf{I}\}^i$ 

```


Abbildung 3.14: Aufbau der Jacobi-Matrix.

für \mathbf{x}_j^i dargestellt),

$$\frac{\partial \mathbf{F}_k}{\partial \mathbf{x}_j} = \begin{pmatrix} \frac{\partial F_{1k}}{\partial x_{1j}} & \cdots & \frac{\partial F_{1k}}{\partial x_{12j}} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_{12k}}{\partial x_{1j}} & \cdots & \frac{\partial F_{12k}}{\partial x_{12j}} \end{pmatrix} = \begin{pmatrix} \frac{\partial F_{1k}}{\partial v_j} & \frac{\partial F_{1k}}{\partial \omega_j} & \frac{\partial F_{1k}}{\partial \kappa_j} & \frac{\partial F_{1k}}{\partial \mathbf{n}_j} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{12k}}{\partial v_j} & \frac{\partial F_{12k}}{\partial \omega_j} & \frac{\partial F_{12k}}{\partial \kappa_j} & \frac{\partial F_{12k}}{\partial \mathbf{n}_j} \end{pmatrix} \quad (3.173)$$

Bezüglich der ersten und zweiten Ableitung $\partial_{\mathbf{x}_1} \mathbf{F}_1$ und $\partial_{\mathbf{x}_{N-1}} \mathbf{F}_{N-1}$ ist einige Vorsicht geboten, da wir auf jedem der beiden Zustandsvektoren sechs Randbedingungen vorgegeben haben. Uns interessieren hier nur die Ableitungen bzgl. der freien Variablen, nicht hingegen jene bzgl. der Randbedingungen. Unter Verwendung von Standardbedingungen ergibt sich für $\partial_{\mathbf{x}_1} \mathbf{F}_1 = (\partial \mathbf{F}_1 / \partial \mathbf{v}_1, \partial \mathbf{F}_1 / \partial \omega_1)$ bzw. für $\partial_{\mathbf{x}_{N-1}} \mathbf{F}_{N-1} = (\partial \mathbf{F}_{N-1} / \partial \kappa_{N-1}, \partial \mathbf{F}_{N-1} / \partial \mathbf{n}_{N-1})$. Die erste und die letzte Untermatrix von \mathbf{J} haben folglich die Dimension 12×6 , alle anderen Untermatrizen 12×12 . Die globale JACOBI-Matrix hat damit tridiagonale Blockstruktur, ist weder symmetrisch, noch diagonaldominant und von der Dimension $[(N-1) \times 12]^2$, vgl. Abb. (3.14). Die gesuchte Inverse \mathbf{J}^{-1} berechnet man durch Lösen eines iterativen Gleichungssystems der Art $\mathbf{J} \delta \mathbf{x} = \mathcal{F}(\mathbf{x})$. Aufgrund der Struktur von \mathbf{J} bietet sich eine GAUSS-Elimination an, vgl. [Press et al., 2002]. Wegen der fehlenden Diagonaldominanz sind iterative Ansätze wie das GAUSS-SEIDEL-Verfahren nicht geeignet.

3.4.8 Alternative Verfahren

Vom NEWTON-Verfahren existiert eine Reihe von Derivaten, die teilweise ohne lokale Ableitungen oder die Berechnung der inversen JACOBI-Matrix auskommen, kubisches Konvergenzverhalten aufweisen, oder Optimierungsstrategien mit der Iteration kombinieren, vgl. z.B. *hybrd* in MINPACK, [Moré et al., 1980]. Ein alternativer Ansatz zum NEWTON-Verfahren sind die tensorbasierten Verfahren, vgl. [Bouaricha & Schnabel, 1994]. Die Iteration basiert hier auf einem quadratischen Modell. Das erweist sich besonders dann als vorteilhaft, wenn die JACOBI-Matrix im

Bereich der Lösung singulär oder schlecht konditioniert ist. Solche Modelle haben einen größeren Informationsbedarf als die klassische Linearisierung, bei einem lediglich moderat gesteigertem Berechnungsaufwand von $\leq \mathcal{O}(n^{5/2})$.

Im Prinzip ist es auch möglich, die Gl. (3.161) als multiobjektives Optimierungsproblem aufzufassen, das bspw. mittels evolutionärer Algorithmen behandelt werden kann, vgl. [Grosan et al., 2006]. Es hat sich aber gezeigt, dass die meisten der hier angeführten Verfahren nur für vergleichsweise niedrig dimensionale Gleichungssysteme ausgelegt sind, wobei der Berechnungsaufwand recht hoch ausfällt. Diskretisierungen mit 100 oder mehr Knoten und in der Folge Gleichungssysteme mit mehr als 1.200 Gleichungen sind in unserem Fall keine Ausnahme, so dass die Wahl des global konvergenten NEWTON-Verfahrens als optimal zu bewerten ist.

Der Vollständigkeit halber soll hier auch ein alternatives Integrationsverfahren nicht unerwähnt bleiben, das bereits erfolgreich zur Simulation von Unterseekabeln eingesetzt worden ist, wenn auch nicht auf der Basis der COSSERAT-Gleichungen, vgl. [Gobat, 2000]. Die so genannte *Box-Methode* ist ein Finite-Differenzen-Schema mit Genauigkeit zweiter Ordnung (im Gegensatz zur Rückwärts-Differenzen Methode, die sehr ähnlich aussieht), da eine zeitliche und räumliche Mittelung der Werte an benachbarten Punkten erfolgt, die dann an den Halbmitterpunkten angewendet wird. D.h., an jedem Halbmitterpunkt erfolgt die Mittelung der Werte über alle vier Nachbarn. Unter der Annahme einer zeitlich konstanten Massematrix sowie konstanter Steifigkeit entlang s erhalten wir dann

$$\begin{aligned} \frac{2}{\Delta t} \mathbf{M} [\mathbf{x}_j^i - \mathbf{x}_j^{i-1} + \mathbf{x}_{j-1}^i + \mathbf{x}_{j-1}^{i-1}] \\ + \frac{2}{\Delta s} \mathbf{K} [\mathbf{x}_j^i - \mathbf{x}_{j-1}^i + \mathbf{x}_j^{i-1} + \mathbf{x}_{j-1}^{i-1}] + [\mathbf{\Lambda}_j^i + \mathbf{\Lambda}_{j-1}^i + \mathbf{\Lambda}_j^{i-1} + \mathbf{\Lambda}_{j-1}^{i-1}] = \mathbf{0}. \end{aligned} \quad (3.174)$$

Numerische Dämpfung ist mit diesem Verfahren nicht möglich. Zur Stabilitätsanalyse zieht man das VON NEUMANNsche-Verfahren heran. Für die Eigenwerte ergeben sich $\lambda_1 = (2 - \Omega)/(2 + \Omega)$ und $\lambda_2 = -1$. Das bedeutet, dass der Spektralradius einen konstanten Wert von eins annimmt und das Verfahren unter theoretischen Gesichtspunkten für alle Ω stabil bleibt, vgl. [Gobat, 2000]. In der Praxis zeigt sich jedoch ein anderes Bild. Die hochfrequenten Anteile der Lösung oszillieren mit jedem Zeitschritt, ein Phänomen, das als CRANK-NICHOLSON-Rauschen bekannt ist. Außerdem bereitet die räumliche Diskretisierung in der Regel Probleme, weil die hochfrequenten Anteile zu schlecht aufgelöst sind und damit zu falschen Ergebnissen führen.

3.4.9 Numerische Beispiele

Der mit Maple generierte Quellcode wurde in unsere Simulationsplattform Rapunzel integriert und verschiedene numerische Tests gefahren. In den folgenden Testfällen bestehen die Fasern aus jeweils 100 Segmenten.

Unser Relaxationsalgorithmus konvergiert durchschnittlich innerhalb von 14 Iterationen. Die zum Erreichen der Konvergenz benötigte Zeit beträgt durchschnittlich etwa 135 ms, gemessen auf einem Laptop mit Intel Pentium M (Centrino), 1,86 GHz und 1 GB RAM unter Verwendung von Standardrandbedingungen, wie oben beschrieben. Die Schrittweite beträgt $\Delta t = 1/30$ Sek. in allen Beispielen, wobei $\alpha = 0, 3$, $\beta = 0$ und $\gamma = 0, 7$.

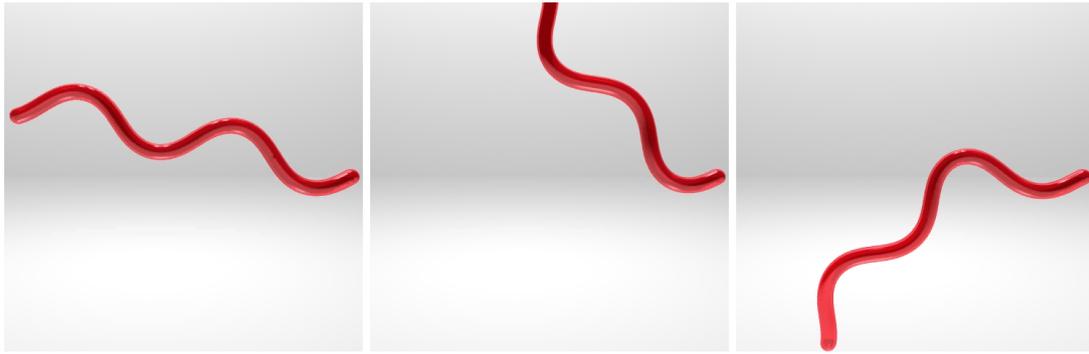


Abbildung 3.15: Sinusförmiger, vorgespannter Stab, der unter Einwirkung der Gravitation aus seiner horizontalen Anfangslage entlassen wird. Die Form bleibt unter Deformation erhalten.

3.4.9.1 Stab unter Gravitation

Das erste Beispiel zeigt einen einseitig eingespannten, sinusförmigen, vorgespannten Stab, der aus seiner horizontalen Anfangslage unter Einwirkung der Gravitation entlassen wird, vgl. Abb. (3.15). Die Bewegung ist ungedämpft. Dieses Experiment zeigt klar, dass eine etwaige Vorkrümmung des Stabes unter Deformation erhalten bleibt. In diesem Fall treten lediglich Biegedeformationen auf, so dass $\kappa_1, \kappa_2 \neq 0$ und $\kappa_3 = 0$.

3.4.9.2 Helix mit Terminallast

Im zweiten Beispiel betrachten wir eine helikale Faser der Länge 30 cm unter Einwirkung einer zeitlich veränderlichen Belastung am Endpunkt, vgl. Abb. (3.16). Die Faser hat eine hohe materialinterne Dämpfung. Das wird durch Setzen der Integrationskoeffizienten auf $\alpha = 0,4$, $\beta = 0$ und $\gamma = 1,0$ erreicht. Das Drehmoment führt je nach Drehrichtung entweder zu einer Verengung oder Aufweitung, die Kraft hingegen zu einer Kompression oder Streckung der Spirale. In beiden Fällen tritt eine Torsion des Querschnitts auf, d.h., $\kappa_3 \neq 0$. Genau diese Verdrehung um den Faserquerschnitt ist einer direkten Modellierung auf der Basis von Masse-Feder-Systemen nicht zugänglich.

3.4.9.3 Gerader Stab mit Drehmoment

Das Beispiel zeigt das Deformationsverhalten eines initial geraden, horizontalen Stabes, der Länge 45 cm, unter Einwirkung eines zeitlich veränderlichen Drehmomentes, das am Ende desselben wirkt, vgl. Abb. (3.17). Nur das Drehmoment kann in dem initial geraden Stab Schlaufen erzeugen.

3.4.9.4 Vorgespannte Spiralfeder

In diesem Beispiel demonstrieren wir das Oszillationsverhalten einer helikalen Keratinfaser, die durch eine am Endpunkt angreifende, parallel zur Spiralachse wirkende Kraft, die nach 0,1 Sek. aussetzt, angeregt wird. Abb. (3.18) zeigt die Entwicklung der Z-Koordinate der Endpunktgeschwindigkeit über die Zeit. Nach Anregung findet der Endpunkt innerhalb von 0,05 Sek. seine

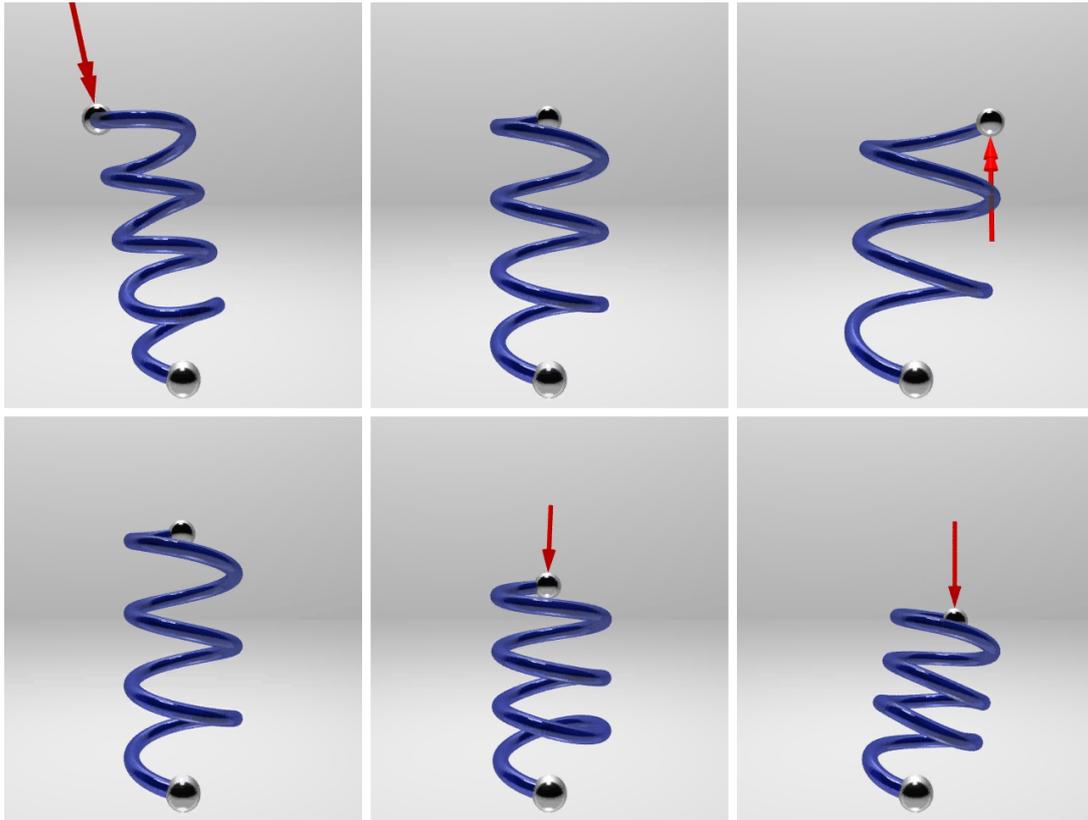


Abbildung 3.16: Helikaler Stab unter Einwirkung eines am Endpunkt angreifenden Drehmomentes mit Wirkungsrichtung parallel zur Stabachse (obere Reihe). Helikaler Stab unter Einwirkung einer Endpunktlast in Richtung der Stabachse (untere Reihe).

Gleichgewichtslage. Die Oszillationsfrequenz liegt bei etwa 160 Hz. Die von der Theorie vorhergesagte unbedingte numerische Stabilität bei frei wählbarer Schrittweite kann hier experimentell überprüft werden. Dazu wurden in einem Experiment die räumlichen Schrittweiten von 100 auf 6 Segmente verringert und in einem zweiten Experiment die zeitlichen von $\Delta t = 10^{-5}$ bis auf $\Delta t = 10^{-2}$ Sek. erhöht, vgl. Abb. (3.19).

Interessanterweise bleibt der Algorithmus stabil, unabhängig, welche zeitlich-räumlichen Schrittweiten wir dabei wählen. Diese Resultate belegen eindrucksvoll, dass die Eigenschaft der unbeschränkten Stabilität, die das Generalisierte α -Verfahren für sich in Anspruch nimmt, nicht nur in der Theorie, sondern auch in der Praxis gültig ist, die richtige Wahl der Integrationskoeffizienten vorausgesetzt.

3.4.9.5 Buckling

Das Buckling-Verhalten einer schlanken Struktur ist eines der kanonischen Testfälle in der Stabmechanik. Deshalb zeigen wir in einem weiteren Experiment, dass unser Modell ebenfalls in der Lage ist, diese Bifurkation im Verlauf der Lösung des zugrunde liegenden PDE-Systems zu be-

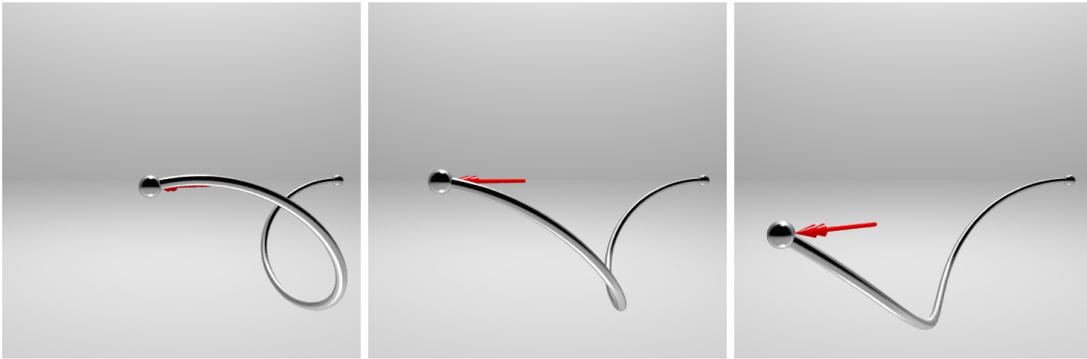


Abbildung 3.17: Gerader Stab mit zeitlich veränderlichem Drehmoment am Endpunkt.

wältigen. Zur Induktion des „Überschlags“ werden die Enden des Stabes entlang der Stabachse aufeinander zu bewegt, unter gleichzeitiger Applikation eines konstanten axialen Drehmomentes. Damit der Stab aus der ursprünglichen Lage ausschlagen kann, muss zu Beginn des Simulationsprozesses der Knickfall durch Aufbringen einer mittig und orthogonal zu Stabachse wirkenden Kraft induziert werden. Die Bewegung des Stabes beschleunigt, je näher die Lösung an den Bifurkationspunkt kommt, vgl. Abb. (3.20). Das erste Bild der unteren Reihe zeigt den Kollaps der Faser. An diesem Punkt schlägt die Konfiguration in die Schleifenform um.

3.4.9.6 Faseranordnung unter Zugbelastung

In diesem Beispiel demonstrieren wir die einfache Integrierbarkeit alternativer Randbedingungen am Beispiel einer Haarsträhne der Länge 25 cm, die aus einer deformierbaren Leitfaser und 180 interpolierten Fasern besteht. Die Leitfaser wird mit einer konstanten Geschwindigkeit am Endpunkt um 30 % der ursprünglichen Länge in Z-Richtung auseinander gezogen, vgl. Abb. (3.21). Dieser Effekt ist dem der kraftgesteuerten Simulation, bspw. bei einer Haarsträhne unter Einwirkung der Gravitation, sehr ähnlich. In diesem Fall wird der Endpunkt allerdings auf eine Linie parallel zur Richtung der Gravitation gezwungen. Die Randbedingungen an der Stelle $s = L$ sind dafür wie folgt anzupassen: $BC_L = \{\mathbf{v}, \boldsymbol{\omega}\} = \{\lambda d_{13}, \lambda d_{23}, \lambda d_{33}, 0, 0, 0\}$, wobei die Elongationsge-

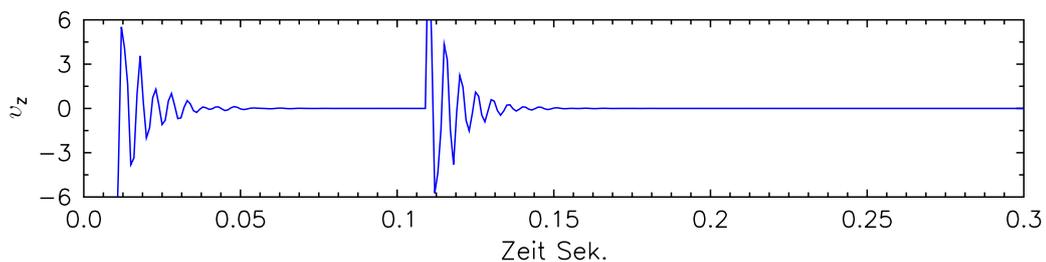


Abbildung 3.18: Entwicklung der Z-Koordinate der Endpunktgeschwindigkeit der Spiralfeder in den unterschiedlichen Phasen des Belastungsspiels: Anregung durch konstante Kraft in Z-Richtung und Entlassen aus dem Zustand nach 0,1 Sek. Dauer.

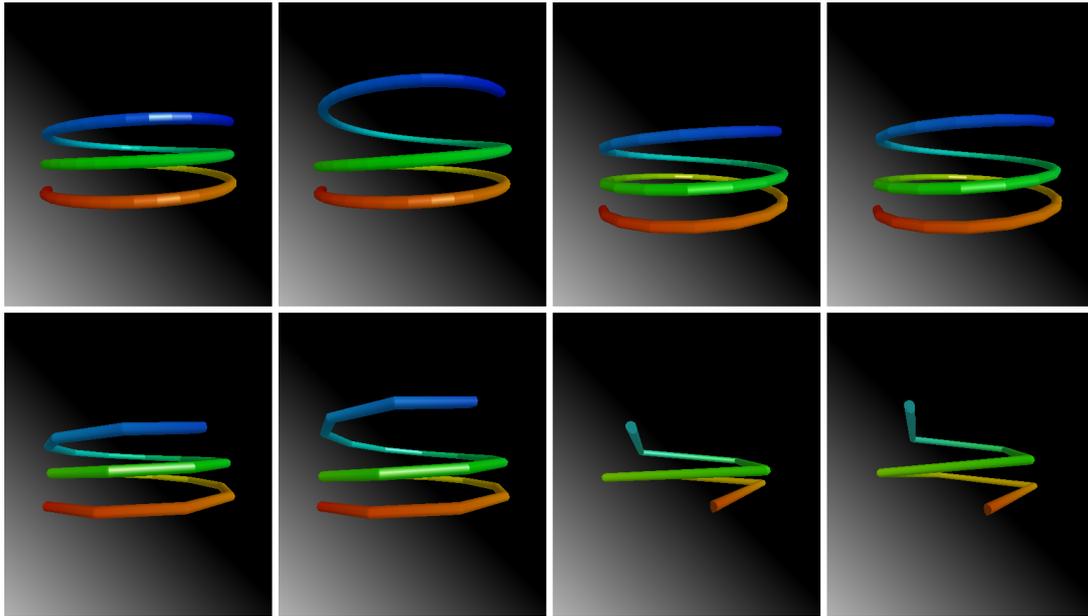


Abbildung 3.19: Demonstration der Stabilitätseigenschaften des generalisierten α -Verfahrens anhand unterschiedlicher räumlicher Schrittweiten (100, 50, 25 und 12 Segmente).

schwindigkeit zu $\lambda = -2, 0$ gewählt wird. Die Faser entlässt man anschließend aus dieser gespannten Lage. Das erreichen wir, indem die Randbedingungen am Ende der Faser auf die Standardbedingungen zurückgesetzt werden. Die Faser schnell mit hoher Geschwindigkeit zurück, über die Anfangslage hinaus, um sich anschließend in einer endlichen Anzahl von Schwingungen wieder auf die Ausgangskonfiguration einzupendeln. Das Beispiel demonstriert eindrucksvoll, dass unser Lösungsansatz auch für große Beschleunigungen, die während des Zurückschnappens auftreten, stabil bleibt.

3.4.9.7 Faseranordnung unter Gravitation

In diesem Beispiel betrachten wir dieselbe Haarsträhne, wie im vorhergehenden Fall, allerdings unter Einwirkung der Gravitation. Wie wirkt sich nun die Wahl der Integrationsparameter auf das Dämpfungsverhalten aus? Abb. (3.22) zeigt die Entwicklung der kinetischen Energie und damit das Dämpfungsverhalten in Abhängigkeit von der Wahl der Koeffizienten α und γ , wobei wir β in allen Fällen zu null wählen. Die Bewegung der Faseransammlung ist in Abb. (3.23) an mehreren charakteristischen Positionen dargestellt. Die Ansammlung streckt sich unter Einwirkung der Gravitation, um dann mehrere gedämpfte Schwingungen zur vollführen. Wie schnell sie dabei zur Ruhe kommt, hängt von der numerischen Dämpfung ab.

3.4.9.8 Faseranordnung unter Bewegung des Lagers

Auch die Integration von geschwindigkeitsabhängigen Randbedingungen auf der Seite der Aufhängung einer Anordnung ist problemlos möglich. Abb. (3.24) illustriert dies. Der Aufhängungs-

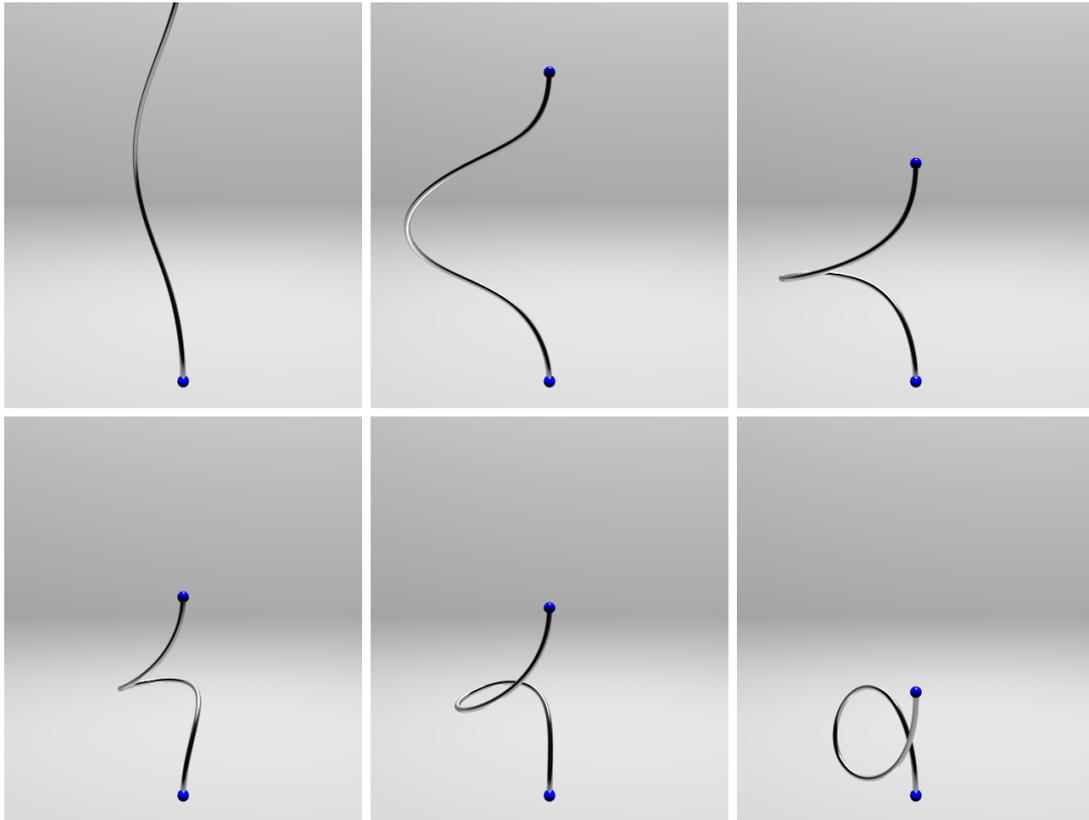


Abbildung 3.20: *Kanonisches Beispiel in der Stabmechanik: Buckling.*

punkt bewegt sich mit konstanter Geschwindigkeit auf einem Bogensegment mit dem Winkel 45° und einem Radius von 30 cm auf und ab. Das Modell bleibt auch hier stabil.

3.4.10 Diskussion

Im vorliegenden Kapitel haben wir unterschiedliche numerische Lösungsansätze für die statischen und dynamischen COSSERAT-Gleichungen vorgestellt. Insbesondere im Hinblick auf die dynamischen COSSERAT-Gleichungen konnten wir, basierend auf dem generalisierten α -Verfahren, einen Relaxationsmechanismus entwickeln, der sich durch eine beeindruckende numerische Robustheit auszeichnet. Unser Ansatz unterscheidet sich damit von allen bisher im Bereich der Computergraphik vorgestellten Verfahren zur Simulation von langen, dünnen Strukturen. Bei der Wahl geeigneter Integrationskoeffizienten sind zeitlich-räumliche Schrittweiten nahezu beliebiger Größe möglich. Es sollte allerdings nicht vergessen werden, dass zu große Schrittweiten zu einer Tiefpassfilterung der Lösung führen, bei der wichtige höherfrequente Anteile verloren gehen. Darüber hinaus sind wir in der Lage, die numerische Dämpfung, die sich bei Anwendung impliziter Integrationsverfahren nicht vermeiden lässt, durch die Wahl der Parameter gezielt zu steuern. Die Theorie garantiert uns ein optimales Verhältnis zwischen den konservierten niederfrequenten und den dissipierten hochfrequenten Anteilen der Lösung.

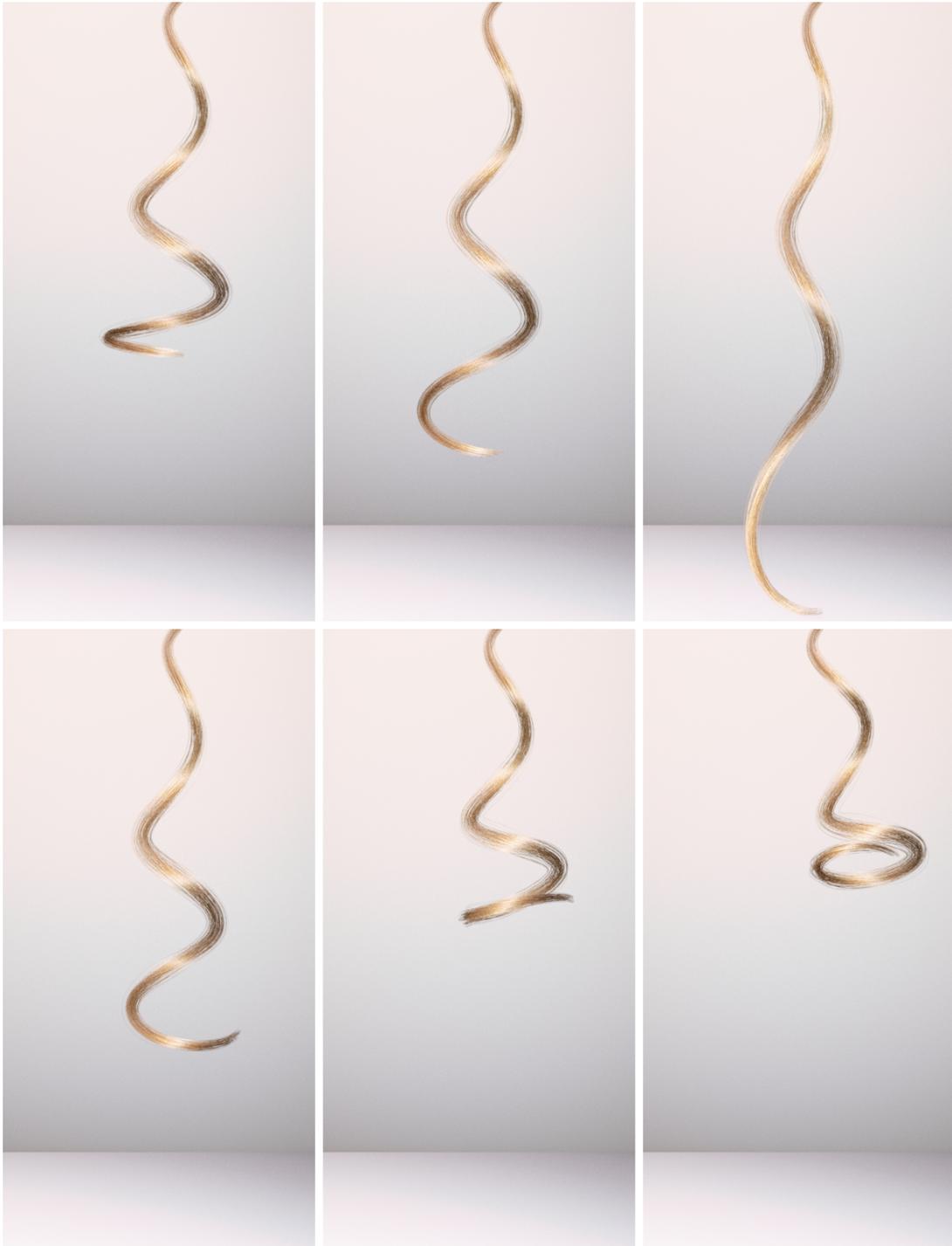
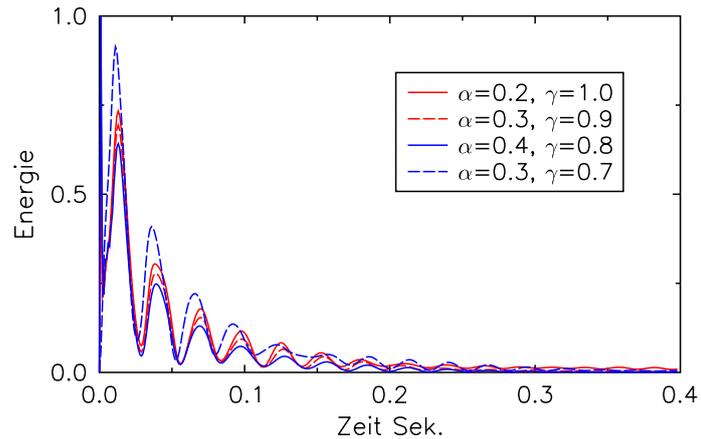


Abbildung 3.21: Eine Faseransammlung bestehend aus 180 einzelnen Fasern wird mit konstanter Rate auseinander gezogen und anschließend aus der gespannten Lage entlassen. Die untere Reihe zeigt das Zurückschnappen der Ansammlung. In dieser Phase treten große Beschleunigungen auf.

Abbildung 3.22: Entwicklung der kinetischen Energie bzw. Dämpfungsverhalten einer Keratinfaser in Abhängigkeit von der Wahl der Integrationskoeffizienten α und γ , wobei $\beta = 0$.



Auch in dieser Hinsicht ist der Ansatz einzigartig. Es gibt zudem keinen Grund, der gegen eine Regelung des Systemverhaltens über die numerische Dämpfung spricht, zumal man bei alternativer Mitführung eines expliziten Dämpfungsterms mit empirisch nur schwer bestimmbar GröÙen hantieren muss. In beiden Fällen ist ein gewisses Fingerspitzengefühl im Hinblick auf die Eichung des Dämpfungsverhaltens erforderlich.

Darüber hinaus konnten wir an einem der kanonischen Beispiele für steife PDE-Systeme schlechthin, eindrucksvoll zeigen, wie derartige Probleme durch Anwendung entsprechender, teilweise im Bereich der Computergraphik unbekannter Verfahren, entschärft werden können, ohne dabei etwa die ursächlichen Faktoren aufzuweichen oder ganz zu eliminieren. Fasern mit hoher Zugfestigkeit, wie sie mit Haarfäsern gegeben sind, stellen damit kein Problem mehr dar. Das steht im Gegensatz zu „Konkurrenzverfahren“, die ad-hoc-Federn durch die Hintertür wieder einführen und dem Modell den Stempel eines „überelastischen“ Masse-Feder-Systems aufdrücken.

An dieser Stelle sollen die wesentlichen Unterschiede im Vergleich zu zwei aktuellen „Konkurrenzmodellen“, dem Superhelixmodell [Bertails et al., 2006] und den COSSERAT-Rod-Elementen [Spillmann & Teschner, 2007], aufgezeigt werden.

Unser Ansatz beruht auf der speziellen Theorie der COSSERAT-Stäbe. Auch wenn andere Verfahren, wie das Superhelixmodell [Bertails et al., 2006] oder die COSSERAT-Rod-Elemente [Spillmann & Teschner, 2007] dies für sich in Anspruch nehmen, sollte beachtet werden, dass beide Verfahren nichts weiter als eine phänomenologische Neuformulierung des „Faserproblems“ auf der Basis der LAGRANGESchen Mechanik darstellen. Sie sind als solche keineswegs weniger für die anstehenden Aufgaben geeignet, stellen damit allerdings höchstens eine Approximation der dynamischen COSSERAT-Gleichungen dar.

Ein weiterer wichtiger Aspekt betrifft die hohe Zugfestigkeit von Haarfäsern. Der hier vorgestellte Ansatz erreicht die Dehnungsresistenz durch eine entsprechende Zwangsbedingung in Form einer zusätzlichen Differentialgleichung. Es wurde weiter oben darauf hingewiesen, dass eine hohe Resistenz gegen Deformationen jeglicher Art, sei es Biegung, Torsion, Dehnung, etc., in der Modellvorstellung durch die steife Feder symbolisiert wird. Im Hinblick auf den Lösungsprozess des zugrunde liegenden PDE-Systems bedeutet das eine Induktion hochfrequenter Schwingungen, die dann in den wichtigen niederfrequenten Anteil der Lösung einkoppeln und diese

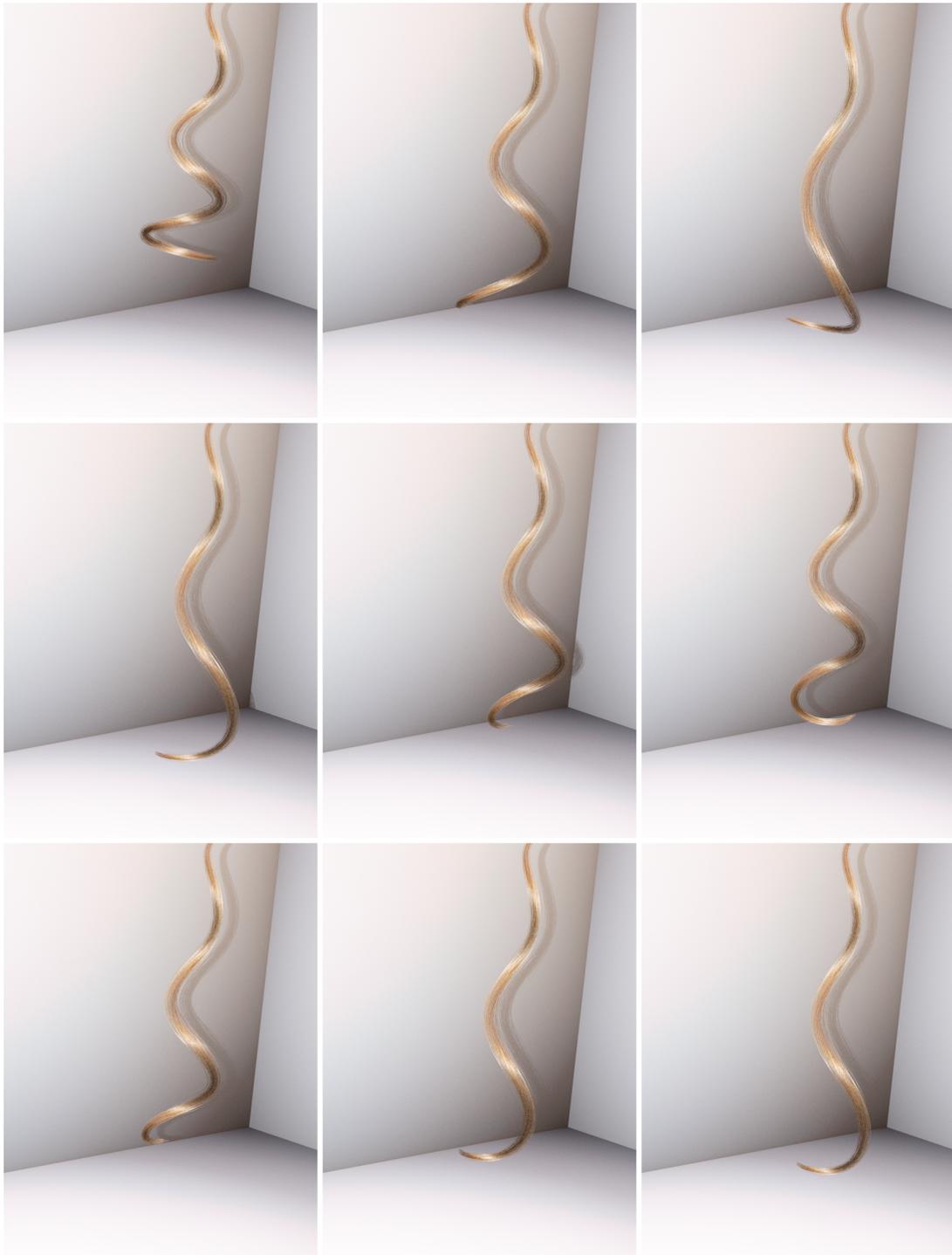


Abbildung 3.23: Faseransammlung bestehend aus 180 einzelnen Fasern unter Einwirkung der Gravitation. Die Dynamik wurde auf der Basis unseres Ansatzes berechnet.

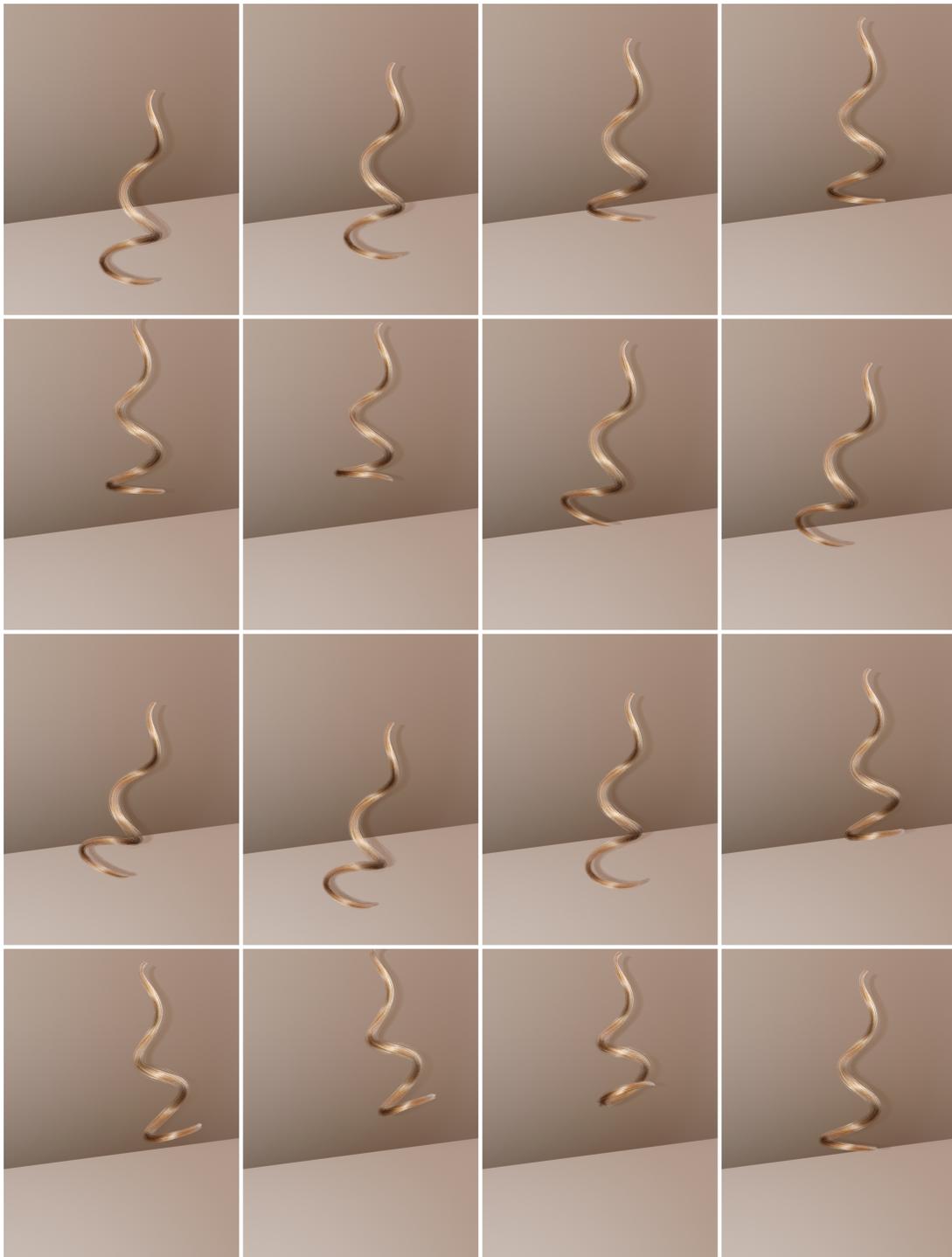


Abbildung 3.24: Faseranordnung unter Gravitation. Der Aufhängungspunkt bewegt sich mit konstanter Geschwindigkeit auf einem Bogensegment mit dem Winkel 45° und einem Radius von 30 cm auf und ab.

teilweise unbrauchbar machen. Die hochfrequenten Anteile können wir durch das eingesetzte α -Verfahren kontrolliert herausfiltern, so dass das Hauptproblem steifer Differentialgleichungssysteme, die unbrauchbar kleinen Schrittweiten oder Instabilität der Lösung, nicht mehr auftreten. Dass das nicht nur Theorie ist, haben wir oben in praktischen Beispielen gezeigt, in denen die temporalen und räumlichen Schrittweiten unrealistisch groß gewählt wurden, ohne dass Einbußen hinsichtlich der Stabilität zu verzeichnen wären.

Im Superhelixmodell [Bertails et al., 2006] ist die Nichtdehnbarkeitsforderung implizit durch die Wahl einer reduzierten Koordinatendarstellung erfüllt. Allerdings führen auch diese Formulierungen bekanntermaßen auf steife Differentialgleichungssysteme. Steifigkeitsprobleme hinsichtlich der räumlichen Auflösung werden hier offenbar durch die geschickte Wahl einer Diskretisierung auf der Basis von helikalen Segmenten mit stückweise konstanter Krümmung und Torsion innerhalb gewisser Grenzen abgefangen.

Bei den COSSERAT-Rod-Elementen [Spillmann & Teschner, 2007] existiert keine solche Nichtdehnbarkeitsbedingung. Dehnung und Stauchung sind also zulässig, Scherung hingegen nicht, so dass sich dieser Ansatz nicht für unsere Zwecke eignet. Der Nachteil der COSSERAT-Rod-Elemente besteht allerdings in der Wahl der Repräsentation der Orientierungen durch Quaternionen. Die Forderung nach der Einheitslänge muss dem System als zusätzlicher Zwang auferlegt werden. Die hier verwendeten Exponential-Maps haben den Vorteil, dass sie selbst Elemente der $SO(3)$ darstellen.

Im Gegensatz zu existierenden Ansätzen, die auf der Anwendung einfacher LAGRANGE-Mechanik beruhen, lösen wir die bestimmenden partiellen Differentialgleichungen unter Anwendung des generalisierten α -Verfahrens direkt. In diesem Sinne kann unser Ansatz als erster Versuch verstanden werden, das Problem in seiner ureigenen Form zu lösen. In Kombination mit geeigneten Kontaktmodellen, die sich elegant in den Rahmen der partiellen Differentialgleichungen integrieren lässt, bspw. durch eine Reformulierung im Sinne von Maßdifferentialinklusionen, scheint eine exakte Modellierung all jener Effekte möglich, die bei der Faser-Faser-Interaktion eine Rolle spielen. Darüber hinaus ist eine Übertragung des Verfahrens auf andere Anwendungsgebiete steifer Differentialgleichungen, wie bspw. das der Kleidermodellierung, möglich und sinnvoll.

4 Elektrostatik auf Haaren

4.1 Einleitung

Im Folgenden stellen wir einen Ansatz zur Untersuchung elektrostatischer Phänomene auf menschlichen Haaren oder in Faseransammlungen vor, der auf der Berechnung paarweiser COULOMB-Fernfeldinteraktionen zwischen punktförmigen über die Fasern verteilten Ladungsträgern basiert. Das Modell integriert sich ideal in das Gesamtkonzept der speziellen Theorie der COSSERAT-Stäbe, kann aber mit beliebigen mechanischen Stabmodellen kombiniert werden. Ferner untersuchen wir den Einfluss von Umweltbedingungen wie Temperatur und Luftfeuchte auf die Ladungsauswirkungen in Haaransammlungen.

Zur Untersuchung elektrostatischer Effekte sind Strähnen-basierte Modelle ungeeignet. Die physikalischen Prozesse spielen sich hier auf der Ebene einzelner Fasern ab, so dass der Einsatz eines faserbasierten Modelles sinnvoll erscheint.

Bis jetzt wurde diese Schicht eher vernachlässigt. Einige wenige Beiträge, die sich dieser Problematik eher beiläufig annehmen, bspw. im Hinblick auf den Einfluss von Wasser und Haarpflegeprodukten auf Haaransammlungen, wurden vorgestellt, vgl. bspw. [Ward et al., 2004]. Die Verfahren haben aber einen stark approximativen Charakter. Andere prominente Effekte, die sich bspw. durch auf die Haaroberfläche aufgebrachte Ladung ergeben, waren bisher noch nicht Gegenstand von Untersuchungen.

In ihrer extremen Form sind die Effekte des Ladungsauftrags auf Haare aus dem Physikunterricht bekannt: Ein Proband legt beide Hände auf die metallische Hülle eines VAN-DE-GRAAFF-Generators. Ein Plastik- oder Gummiband wird über einen Motor oder manuell angetriebene Walzen an einer Art Drahtbürste vorbeigeführt, die an der Metallhülle befestigt ist, und induziert dadurch eine Ladung auf deren Oberfläche. Dieser Effekt ist auch als Triboelektrizität bekannt. Die Ladung geht durch Berühren der Metalloberfläche auf die Oberfläche des Probanden über. Bei trockenen, sauberen und nicht zu stark aneinander haftenden Haarfasern nehmen diese den größtmöglichen Abstand zueinander ein. Die Fasern folgen den Flußlinien des elektrischen Feldes. Im Extremfall kommt es dann zum bekannten „Struwelpetereffekt“, vgl. Abb. (4.1).

Auch in weniger extremen Fällen können die Auswirkungen statischer Ladung sichtbaren Einfluss auf das Gesamterscheinungsbild haben, insbesondere das Volumen des Haarkörpers. Auch Kämmen und Bürsten verursacht statische Elektrizität auf der Haaroberfläche und trägt zu einer Änderung des Gesamthaarvolumens in sehr subtiler, aber doch sichtbarer Weise bei. Da die Anteile an der Gesamtladung, die durch mechanische Belastungen wie Kämmen und Bürsten induziert werden, nicht empirisch quantifiziert werden können, sind Modelle erforderlich, die den gesteigerten Genauigkeitsansprüchen genügen.

In diesem Zusammenhang scheint der faserbasierte COSSERAT-Ansatz geeignet, da er sich ideal mit der Technik der Multipolentwicklung von über die Haaroberfläche verteilten Punktladungen kombinieren lässt.



Abbildung 4.1: Berühmter Versuch: Die Testperson lädt sich an einem Van-de-Graaff-Generator elektrisch auf. Die Haare stellen sich in der Folge entlang der Feldlinien des sich ausbildenden elektrischen Feldes auf (Quelle: American Museum of Science & Energy).

4.2 Berechnung elektrostatischer Einflüsse

Die Kraft, die zwischen zwei Punktladungen an den Positionen \mathbf{x}_1 und \mathbf{x}_2 wirkt, wird durch das klassische COULOMBSche Gesetz beschrieben¹,

$$\mathbf{f}^C = c_0 \cdot \left(q_1 \cdot q_2 \cdot \frac{\mathbf{x}_2 - \mathbf{x}_1}{\|\mathbf{x}_2 - \mathbf{x}_1\|^3} \right). \quad (4.1)$$

Bei mehreren Partikeln ergeben sich die COULOMB-Interaktionen aus dem Superpositionsprinzip, das auch bei elektrostatischen Feldern seine Gültigkeit behält,

$$\mathbf{f}^C = c_0 \cdot \sum_{j=1, i \neq j}^n \left[q_i \cdot q_j \cdot \frac{\mathbf{x}_i - \mathbf{x}_j}{\|\mathbf{x}_i - \mathbf{x}_j\|^3} \right]. \quad (4.2)$$

Das Hauptproblem bei der Berechnung des elektrostatischen Feldes besteht offensichtlich darin, dass alle wechselseitigen Interaktionen einer Menge von Punktladungen in Betracht gezogen werden müssen, selbst dann, wenn einzelne Quellen weit voneinander entfernt sind. Die Komplexität ist quadratisch in der Anzahl der Partikel. Dieser Sachverhalt ist als *N-Körperproblem* bekannt. Modelle, die Einflüsse weit entfernter Ursachen berücksichtigen, werden i.d.R. als *Fernfeldmodelle* bezeichnet.

Probleme, welche die Interaktion mehrerer Körper zum Gegenstand haben, sind in der Computergraphik wohlbekannt, insbesondere im Zusammenhang mit der Kollisionserkennung. Allerdings sollte dies nicht mit dem klassischen N-Körperproblem verwechselt werden, weil sich der Wirkungsbereich der Körper nicht über größere Distanzen erstreckt, sondern lokal beschränkt

¹Die Konstante c_0 hängt von der Wahl der Einheiten ab und ist gleich 1 im cgs-System.

ist. Kollisionen und die aus ihnen folgenden Antworten sind lokale Phänomene. Wir dürfen deshalb nicht erwarten, dass die Lösung zur Beschleunigung der Berechnung paarweiser COULOMB-Interaktionen allein in der Anwendung wohlbekannter Beschleunigungsstrukturen aus der Kollisionserkennung, wie Raum- oder Körpersubdivision, besteht. Solche Techniken erlauben uns, den Wirkungsradius einzuschränken und uns auf lokale Interaktionen zu konzentrieren. Das Einfache Abschneiden von Fernfeldinteraktionen oberhalb einer vorgegebenen Distanz kann zu einem unphysikalischen Verhalten führen, wie durch Darden et al. [Darden et al., 1993] eindrucksvoll in Kristallwachstumsexperimenten gezeigt werden konnte.

Was wir benötigen, ist eine Technik zur Approximation der Fernfeldinteraktionen, denn intuitiv sollte der Einfluss zweier benachbarter Körper auf einen weit entfernten dritten sehr ähnlich sein. Das Missing Link, mit dem wir genau das erreichen, ist in der Physik als *Multipolentwicklung* bekannt.

4.3 Fernfeldapproximation durch Multipolentwicklung

Zum besseren Verständnis des Konzepts der Multipolentwicklung des elektrostatischen Potentials nehmen wir eine Ladung q an der Position $\mathbf{x}_0 \in \mathbb{R}^3$. Das Potential in der Umgebung des Partikels ist dann durch $\Phi(\mathbf{x}) = 1/\tilde{r}$ gegeben, wobei $\tilde{r} = \|\mathbf{x} - \mathbf{x}_0\|$ der Abstand zum Partikel ist. Die Multipolentwicklung ist eine Reihenentwicklung, welche das Potential $\Phi(\mathbf{x})$ in Abhängigkeit vom Abstand r zum Ursprung ausdrückt. Die Einführung von Kugelkoordinaten $\mathbf{x} = (r, \theta, \varphi)$ und $\mathbf{x}_0 = (r_0, \theta_0, \varphi_0)$, führt direkt auf einen Ausdruck mit LEGENDRE-Polynomen,

$$\Phi = 1/\tilde{r} = \sum_{n=0}^{\infty} r_0^n / r^{n+1} P_n(\cos \gamma), \quad (4.3)$$

wobei γ der Winkel zwischen \mathbf{x}_0 und \mathbf{x} ist. Dieser Sachverhalt kann wiederum durch eine bestimmte Klasse von Funktionen ausgedrückt werden, die als *Kugelflächenfunktionen* (engl.: *spherical harmonics*) bekannt sind. Diese bilden eine Menge von orthogonalen Lösungen der zugrunde liegenden LAPLACE-Gleichung $\nabla^2 \Phi = 0$, wenn das Problem durch Kugelkoordinaten parametrisiert wird.

Im Folgenden erweitern wir die Untersuchung auf k Punktladungen der Stärke q_i , die innerhalb einer Kugel mit dem Radius R an den Positionen $\mathbf{x}_i = (r_i, \theta_i, \varphi_i)$ mit $i = 1..k$ lokalisiert sind, vgl. Abb. (4.2). Das Potential an jedem Punkt $\mathbf{x} = (r, \theta, \varphi)$ außerhalb der Kugel ist gegeben durch

$$\Phi(\mathbf{x}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{M_n^m}{r^{n+1}} \cdot Y_n^m(\theta, \varphi), \quad (4.4)$$

wobei gilt,

$$M_n^m = \sum_{i=1}^k q_i \cdot r_i^n \cdot Y_n^{-m}(\theta_i, \varphi_i), \quad (4.5)$$

das Moment der Entwicklung darstellt, durch das die Verteilung der Ladungen innerhalb der Kugel berücksichtigt wird. $Y_n^m(\theta, \varphi)$ sind die Kugelflächenpolynome vom Grade n und sind wie folgt definiert,

$$Y_n^m(\theta, \varphi) = \sqrt{\frac{(n - |m|)!}{(n + |m|)!}} P_n^{|m|}(\cos \theta) e^{im\varphi}, \quad (4.6)$$

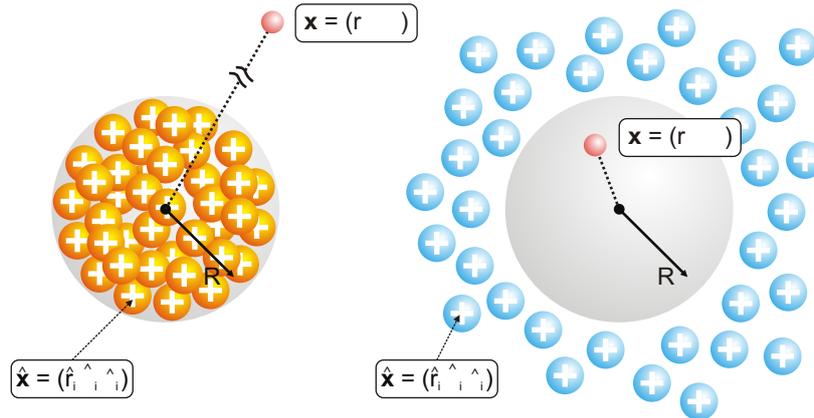


Abbildung 4.2: Links: Die äußere Multipolentwicklung drückt das Potential $\Phi(\mathbf{x})$ außerhalb einer Kugel mit Radius R infolge aller Partikel innerhalb der Kugel aus. Rechts: Die innere oder lokale Multipolentwicklung hingegen, drückt das Potential $\Phi(\mathbf{x})$ an einem Punkt \mathbf{x} innerhalb der Kugel infolge aller Ladungen außerhalb derselben aus.

wobei $P_n^m(x)$ die assoziierten LEGENDRE-Polynome sind.

Diese Gleichungen sind nur gültig, wenn eine ausreichende Separierung der Partikel gegeben ist. Darüber hinaus ist es von entscheidender Wichtigkeit, dass man alle Nahfeldinteraktionen weiterhin explizit berechnet, weil die gegebenen Approximationen bei kleinen Abständen nicht mehr konvergieren. Nur wenn die Distanz r eines Partikels groß ist im Vergleich zum Radius der umgebenden Kugel, $r \gg R$, für die eine Multipolentwicklung durchgeführt werden soll, approximieren wir die COULOMB-Interaktionen mit allen Partikeln innerhalb der Kugel durch Auswertung einer endlichen Anzahl von p Termen der Multipolentwicklung. Die Gesamtkraft, die auf einen Partikel einwirkt, setzt sich aus dem Anteil für die Nahfeld- (nf) und dem für die Fernfeldinteraktionen (ff) zusammen, $\mathbf{f}_\Sigma = \mathbf{f}_{nf} + \mathbf{f}_{ff} + \mathbf{f}_{ext}$, wobei ext externe Interaktionskräfte infolge Gravitation, Kollision, etc. indiziert.

In Analogie zur äußeren Multipolentwicklung lässt sich auch eine innere oder lokale Entwicklung angeben. Nehmen wir dazu an, dass die k Ladungen der Stärke q_i außerhalb der Kugel mit dem Radius R an den Stellen $\mathbf{x}_i = (r_i, \theta_i, \varphi_i)$ lokalisiert sind. Das Potential an jedem Punkt $\mathbf{x} = (r, \theta, \varphi)$ innerhalb der Kugel ist jetzt gegeben durch

$$\Phi(\mathbf{x}) = \sum_{n=0}^{\infty} \sum_{m=-n}^n L_n^m \cdot Y_n^m(\theta, \varphi) \cdot r^n, \quad (4.7)$$

wobei

$$L_n^m = \sum_{i=1}^k q_i \cdot \frac{Y_n^{-m}(\theta_i, \varphi_i)}{r_i^{n+1}}. \quad (4.8)$$

Bei der numerischen Auswertung muss die unendliche Summe in der Potentialfunktion durch eine endliche mit p Termen ersetzt werden. Die Fehlerabschätzung für die äußere Entwicklung

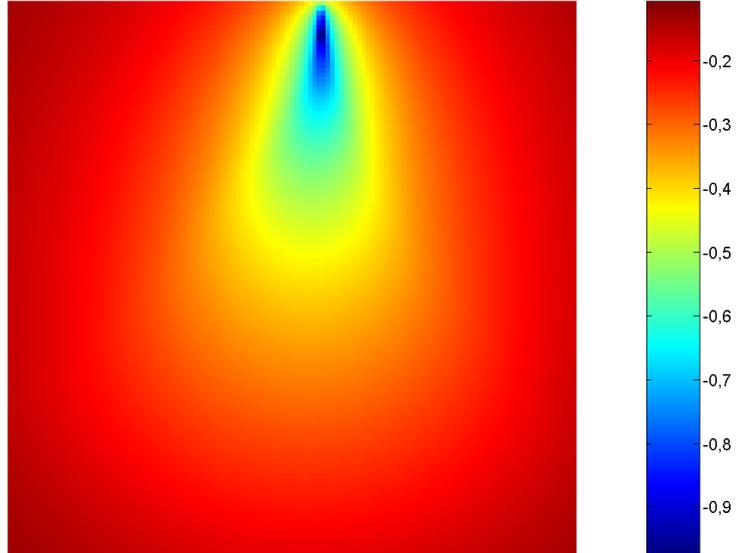


Abbildung 4.3: Potentialfeld einer Faseranordnung, berechnet auf einem 128^3 -Gitter.

ist dann bei $p \geq 1$ durch die Beziehung

$$|\Phi(\mathbf{x}) - \tilde{\Phi}(\mathbf{x})| = \quad (4.9)$$

$$\left| \sum_{n=0}^{\infty} \sum_{m=-n}^n \frac{M_n^m}{r^{n+1}} \cdot Y_n^m(\theta, \varphi) - \sum_{n=0}^p \sum_{m=-n}^n \frac{M_n^m}{r^{n+1}} \cdot Y_n^m(\theta, \varphi) \right| \leq \left(\frac{\sum_{i=1}^k |q_i|}{r - R} \right) \left(\frac{R}{r} \right)^{p+1} \quad (4.10)$$

gegeben. Das zugehörige elektrostatische Feld wird durch den Gradienten des Potentials,

$$\mathbf{E} = -\nabla\Phi(\mathbf{x}) = -\partial\Phi/\partial x_i, \quad (4.11)$$

eindeutig beschrieben. Aus dem elektrostatischen Feld erhalten wir die auf jeden Partikel einwirkenden Ladungskräfte in der Form $\mathbf{F} = q \cdot \mathbf{E}$. Das Beispiel eines Potentialfeldes, das auf der Basis unseres Ansatzes für eine computergenerierte Haarsträhne berechnet wurde, ist in Abb. (4.3) dargestellt.

4.3.1 Abtastung des elektrostatischen Feldes

Die Multipolentwicklung wurde für Punktladungen hergeleitet und prinzipiell kann man das auch für andere Primitive wie Zylinder oder Tori wiederholen. Das ist i.d.R. bei anderen Geometrien recht aufwendig und normalerweise auch nicht kompatibel mit Standard-Multipolbibliotheken. Zur Approximation der Oberflächenladungen eines dünnen generalisierten Zylinders gegebener Länge wird jede Faser mit (negativen) Punktladungen ausgestattet. Zur Approximation der Fernfeldinteraktionen aller Partikel berechnet man die Multipolentwicklung des Potentials. Daraus ergeben sich die COULOMB-Kräfte, die auf die einzelnen Partikeln wirken, als Gradient der Potentialfunktion.

Zunächst ist unklar, welcher Wert für die Partikeldichte ρ_p zu wählen ist. Die COULOMB-Kräfte an jedem Partikel setzen sich aus einem tangentialen und einem Anteil normal zur Faser zusammen. Der tangentielle Anteil bewirkt eine Verschiebung der Partikel entlang der Faser, bis

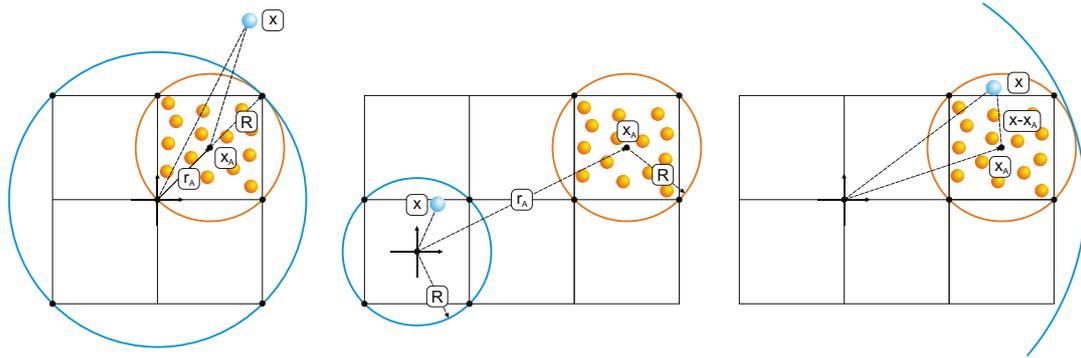


Abbildung 4.4: Links: Translation der Multipolentwicklung. Mitte: Konversion der Multipol- in eine lokale Entwicklung. Rechts: Translation einer lokalen Entwicklung.

ein Gleichgewichtszustand erreicht ist. In der Realität würde sich dieser Einfluss mit Lichtgeschwindigkeit ausbreiten. Die Umverteilung der Ladungen macht die Neuberechnung der COULOMB-Interaktionen notwendig, solange, bis die Partikelverteilung gegen ein stabiles Gleichgewicht konvergiert. Diese Vorgehensweise scheint bei steigenden Partikelzahlen nicht praktikabel. Für unsere Zwecke wählen wir deshalb eine konstante Verteilungsdichte. Dieser Wert fällt nicht notwendigerweise mit der Knotendichte des diskreten COSSERAT-Stabes zusammen und ist damit unabhängig von der zugrunde liegenden Diskretisierung der Fasern. Damit die Positionen der Partikel auch noch nach der Deformation der Fasern gültig sind, werden sie anhand des Bogenlängeparameters s angegeben. Bei Positionierung eines Partikels zwischen zwei Knoten wird linear zwischen den Koordinaten interpoliert.

4.4 Schnelles Multipolverfahren (FMM)

Zur Lösung des N -Körperproblems wurden im Bereich der Physik eine Reihe von Ansätzen vorgestellt. Diese Verfahren kombinieren Standardraum- oder Objektunterteilungsverfahren wie Okt-Bäume oder Hüllkörperhierarchien für die Partitionierung des Berechnungsraumes mit der Multipolentwicklung. Der baumbasierte Algorithmus von Barnes und Hut [Barnes & Hut, 1986] findet seinen Ursprung in der Astrophysik und berechnet die paarweisen Gravitationswechselwirkungen zwischen den Sternen einer Galaxie in $\mathcal{O}(n \log n)$ Zeit. Das schnelle Multipolverfahren (engl.: *Fast Multipole Method (FMM)*) von Greengard and Rokhlin [Greengard & Rokhlin, 1987] ist eine der erfolgreichsten Ansätze, da sie auf der Analyse der paarweisen Interaktionen von Punktladungen basiert und a priori Fehlerschranken zulässt. Der Berechnungsaufwand wird hier durch die Einführung lokaler Entwicklungen auf $\mathcal{O}(n)$ gedrückt. Der Hauptunterschied zwischen der FMM und dem Ansatz von Barnes und Hut [Barnes & Hut, 1986] besteht darin, dass ersterer mit Potentialen operiert, wohingegen der von Barnes und Hut Kräfte berechnet. Unsere Elektrostatikberechnungen basieren auf der FMM. Dazu erstellen wir zunächst einen Oktalbaum über der Menge der Ladungspartikel, die mit einer vorgegebenen Dichte ρ_P über die Fasern verteilt seien. An den Blättern des Baumes werden die Multipolentwicklungen direkt aus allen darin enthaltenen Partikeln bzw. deren Positionen und

Stärken ermittelt. Die Berechnung des Potentials erfolgt dann in zwei Schritten: dem Aufwärtspass, bei dem die Multipolentwicklungen aller Zellen akkumuliert werden, und dem Abwärtspass, bei dem die Potentiale für jeden Partikel innerhalb einer Zelle infolge aller Partikel außerhalb derselben berechnet werden. Die Berechnung der Nahfeldeinflüsse erfolgt nebenläufig, weil sie lediglich von den unmittelbar benachbarten Partikeln in benachbarten Zellen des Oktalbaumes abhängt. Zur Propagation der Multipolentwicklungen sind spezielle analytische Mechanismen notwendig, vgl. Abb. (4.4).

1. *Translation der Multipolentwicklung.* Durch Einführung eines linearen Operators gelingt die Verschiebung einer um den Punkt \mathbf{x}_A gebildeten Multipolentwicklung auf einen beliebigen Punkt \mathbf{x}_B . Sei $\Phi_A(\mathbf{x} - \mathbf{x}_A)$ die Entwicklung, die den Fernfeldeinfluss aller Partikel innerhalb einer Kugel mit Radius R und Zentrum $\mathbf{x}_A = (r_A, \theta_A, \varphi_A)$ an jedem Punkt $\mathbf{x} = (r, \theta, \varphi)$ außerhalb der Kugel ausdrückt, wie durch Gl. (4.4) gegeben. Dann gilt für jeden Punkt \mathbf{x} außerhalb einer Kugel mit Radius $R + r_A$ und Zentrum im Ursprung das Potential $\Phi_B(\mathbf{x})$ mit den Koeffizienten $(M_n^m)_B$. Die Koeffizienten $(M_n^m)_A$ und $(M_n^m)_B$ der beiden Funktion $\Phi_A(\mathbf{x} - \mathbf{x}_A)$ und $\Phi_B(\mathbf{x})$ hängen voneinander ab und definieren den linearen Operator \mathcal{L}_{TM} , so dass $(M_n^m)_B = \mathcal{L}_{TM}(M_n^m)_A$. Dieser Operator konvertiert also die Koeffizienten $(M_n^m)_A$ der einen Multipolentwicklung in die der anderen, $(M_n^m)_B$. Wir verzichten an dieser Stelle bewusst auf die Angabe des linearen Operators.
2. *Translation der lokalen Entwicklung.* Sei $\Phi_A(\mathbf{x} - \mathbf{x}_A)$ eine endliche lokale Entwicklung um den Punkt $\mathbf{x}_A = (r_A, \theta_A, \varphi_A)$, wie durch Gl. (4.7) gegeben. Die Funktion kann durch Wahl geeigneter Koeffizienten $(M_n^m)_B$ auf den Ursprung verschoben werden, $\Phi_B(\mathbf{x})$. Dabei wird der Zusammenhang zwischen den Koeffizienten durch den linearen Operator \mathcal{L}_{TL} ausgedrückt, mit $(M_n^m)_B = \mathcal{L}_{TL}(M_n^m)_A$.
3. *Konversion der Multipolentwicklung in eine lokale Entwicklung.* Sei $\Phi_A(\mathbf{x} - \mathbf{x}_A)$ die Multipolentwicklung um den Punkt $\mathbf{x}_A = (r_A, \theta_A, \varphi_A)$ infolge aller Partikel innerhalb einer Kugel K_{X_A} mit Radius R und Zentrum \mathbf{x}_A , wobei $r_A > 2R$. Dann konvergiert die Multipolentwicklung $\Phi_A(\mathbf{x} - \mathbf{x}_A)$ (vgl. Gl. (4.4)) innerhalb einer Kugel K_0 mit dem Radius R und Zentrum im Ursprung. Weiterhin gilt für jeden Punkt $\mathbf{x} = (r, \theta, \varphi)$ mit $r < R$ die innere Entwicklung $\Phi_B(\mathbf{x})$ mit den Koeffizienten $(M_n^m)_B$ infolge aller Partikel in K_{X_A} , vgl. Gl. (4.7). Auch hier existiert ein linearer Operator \mathcal{L}_{KM} , so dass $(M_n^m)_B = \mathcal{L}_{KM}(M_n^m)_A$. D.h., durch Anwendung dieses Operators lassen sich die Koeffizienten der lokalen Entwicklung aus denen der Multipolentwicklung ermitteln. Die Multipolentwicklung um \mathbf{x}_A wird also in eine lokale Entwicklung um den Ursprung überführt.

Die analytische Mechanismen lassen sich durch fundamentale Theoreme der Theorie der Kugelflächenfunktionen rechtfertigen. Der interessierte Leser sei auf die Originalarbeit von Greengard und Rokhlin [Greengard & Rokhlin, 1987] verwiesen.

4.4.1 Interaktionslisten

Zum besseren Verständnis der einzelnen Berechnungsschritte einer $\mathcal{O}(n \log n)$ -Prozedur führen wir zunächst die Begriffe des nächsten Nachbarn einer Zelle, den der Wohlsepariertheit von Zellen und die Interaktionsliste ein. Zwei Zellen des Oktalbaumes sind *nächste Nachbarn*, wenn sie

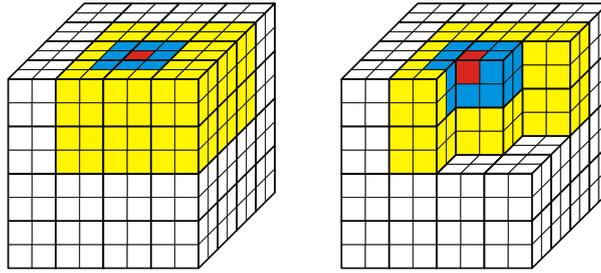


Abbildung 4.5: Zum Problem der Interaktionslisten: Jede Zelle i hält eine Interaktionsliste vor, die genau diejenigen von i wohlseparierten Zellen enthält, die Kinder der nächsten Nachbarn des Vaters von i sind. Die aktuelle Zelle ist rot markiert, ihre nächsten Nachbarn blau und die wohlseparierten Zellen der Interaktionsliste gelb.

dieselbe Rekursionstiefe haben und sich mindestens einen Eckpunkt teilen. Zwei Zellen sind hingegen *wohlsepariert*, wenn sie nicht nächste Nachbarn sind. Jede Zelle i verfügt über eine *Interaktionsliste*, welche genau diejenigen von i wohlseparierten Zellen enthält, die Kinder der nächsten Nachbarn des Vaters von i sind, vgl. Abb. (4.5). Die Interaktionsliste enthält maximal 189 Zellen, wie man leicht nachrechnen kann. Die Auswertung einer Multipolentwicklung kann nur für wohlseparierte Zellen erfolgen, wegen fehlender Konvergenz hingegen nicht zwischen nächsten Nachbarn. Alle Zellen außerhalb der nächsten Nachbarn des Vaters werden auf der jeweils übergeordneten Ebene behandelt. Die Multipolentwicklung erfolgt für alle Zellen einer Ebene über alle darin enthaltenen Partikel, ihre Auswertung hingegen für alle Zellen in der entsprechenden Interaktionsliste.

Offenbar muss beim Ansatz mit Interaktionslisten auf jeder Ebene auf alle Partikel genau einmal zugegriffen werden, was bei einer gleichförmigen Verteilung und $\log N$ Leveln auf $\mathcal{O}(n \log n)$ Aufwand führt. Es scheint gerechtfertigt, diesen Aufwand auch für praktische Verteilungen anzunehmen.

Zur Reduktion der Laufzeitkomplexität auf $\mathcal{O}(n)$ bedient sich die FMM der oben bereits eingeführten Translation bzw. Konversion der Multipolentwicklung. Die Translation erlaubt uns eine direkte Berechnung der äußeren Entwicklung einer Zelle durch den Zusammenschluss (*merge*) der Beiträge ihrer Kinder. Umgekehrt kann durch die Translation aus der inneren Entwicklung einer Zelle die ihrer Kinder ermittelt werden. Ein direkter Zugriff auf alle Partikel ist nicht mehr notwendig.

Damit lassen sich die einzelnen Schritte wie folgt zusammenfassen. Im Aufwärtspass erfolgt die Berechnung der Entwicklungen aller Zellen durch Zusammenschluss der Kindentwicklungen (Translation der Multipolentwicklung). Lediglich an den Blättern ist eine direkte Berechnung aus den Partikeln notwendig. Im Abwärtspass werden die Multipolentwicklungen der aktuellen Zelle bzgl. der Zentren aller Zellen in der Interaktionsliste in eine lokale Entwicklung konvertiert (Konversion der Multipolentwicklung in eine lokale Entwicklung). Diese drückt das Potential aller Partikel der aktuellen Zelle infolge aller Zellen außerhalb ihrer nächsten Nachbarn aus. In einem zweiten Abwärtspass werden diese lokalen Expansionen auf die Level der Kindzellen verschoben und zu deren lokalen Entwicklungen addiert (Translation der lokalen Entwicklung). Durch diesen Mechanismus propagiert man die lokalen Expansionen bis auf die Blattebenen des Baumes.

Nur diese Entwicklungen, also jene auf dem untersten Level, müssen ausgewertet werden, um das Potential der einzelnen Partikel zu erhalten.

Die Gesamtzahl der benötigten Operationen beläuft sich auf $189(n/s)p^4 + 2np^2 + 27ns$, wobei s die durchschnittliche Anzahl an Partikeln pro Blattknoten und p die Anzahl auszuwertender Terme in der Multipolentwicklung darstellt. Zur Erreichung einer gewünschten Genauigkeit ϵ sind $p = \log_{\sqrt{3}} 1/\epsilon$ Terme auszuwerten. Die Berechnungskosten für die lokale Entwicklung pro Zelle belaufen sich auf $189(n/s)p^4$ Operationen, lassen sich aber auf $189(n/s)3p^3$ reduzieren, wenn das lokale Koordinatensystem so gedreht wird, dass die z-Achse mit dem Verbindungsvektor zwischen Quell- und Zielbox zusammenfällt. Die Entwicklung wird dann entlang dieser Achse verschoben und anschließend wieder auf das Ursprungssystem zurückgedreht.

4.4.2 Interaktion zwischen Partikeln und Oberfläche

Die Kopfoberfläche, oder die Objektoberfläche im Allgemeinen, übt einen Einfluss auf die Fasern aus, da sie selbst eine Ladung trägt. Die abstoßenden Kräfte zwischen der Kopfoberfläche und den Fasern, die sich infolge der gleichnamigen Ladungen einstellen, fungieren als eine Art Kollisionsantwortkraft, da sie das Eindringen der Fasern in das Objekt verhindern.

Zur schnellen Berechnung der Interaktionen ermitteln wir für die Objektoberfläche \mathcal{S} eine implizite, lineare Approximation oder Distanzfeld $\Phi_{\mathcal{S}} : \mathbb{R}^3 \rightarrow \mathbb{R}$, so dass

$$\Phi_{\mathcal{S}}(\mathbf{x}) := \min_{\mathbf{q} \in \mathcal{S}} \{ \|\mathbf{x} - \mathbf{q}\| \}, \quad \forall \mathbf{x} \in \mathbb{R}^3. \quad (4.12)$$

Genau genommen gilt das COULOMBSche Gesetz nur für punktförmige Ladungsträger ohne räumliche Ausdehnung, aber wir wenden es in diesem Zusammenhang auch auf die Berechnung der Kräfte zwischen Partikeln und der Objektoberfläche an. Die Partikel in der Umgebung der Oberfläche erfahren eine abstoßende Kraft, die mit $1/r^2$ abnimmt. Die Kraft ist dann gegeben durch

$$\mathbf{f}^{\mathcal{S}}(\mathbf{x}) = \frac{|q_i| \cdot |Q_{\mathcal{S}}|}{r_x^2} \cdot \frac{\nabla \Phi_{\mathcal{S}}(\mathbf{x})}{\|\nabla \Phi_{\mathcal{S}}(\mathbf{x})\|}, \quad (4.13)$$

wobei $\nabla \Phi_{\mathcal{S}}(\mathbf{x})$ der Gradient der impliziten Funktion an der Stelle \mathbf{x} und r_x der kleinste Abstand von \mathbf{x} zu \mathcal{S} ist. $Q_{\mathcal{S}}$ gibt die Oberflächenladung an, mit $q_i \cdot Q_{\mathcal{S}} > 0$. Diese Bedingung besagt, dass die beiden Ladung gleiche Vorzeichen haben und sich infolgedessen abstoßen. Die implizite Approximation wird mit der Charakteristikenmethode berechnet, vgl. Abs. (7.2). Die einzelnen Distanzen werden an der Knoten eines kartesischen Gitters gespeichert. Der Abstand eines Punktes an der Position \mathbf{x} lässt sich durch trilineare Interpolation der Werte an den Eckpunkten der Zellen berechnen, in welcher der Punkt liegt. Der Gradient hingegen ergibt sich durch Approximation mittels finiter Differenzen. Der Gesamtaufwand für die Berechnung der Approximation aller COULOMB-Interaktionen ist demnach linear in der Anzahl der Segmente.

Das Auftreten singulärer Kräfte kann nur verhindert werden, wenn alle Ladungsträger wohlsepariert sind, d.h., keine Ladungsträger aufeinander liegen.

4.4.3 Randbedingungen

Randbedingungen infolge in das elektrostatische Feld eingeführter Objekte bedürfen einer speziellen Behandlung, weil die Wirkungslinie zweier Partikel sich mit einer solchen Objektberandung

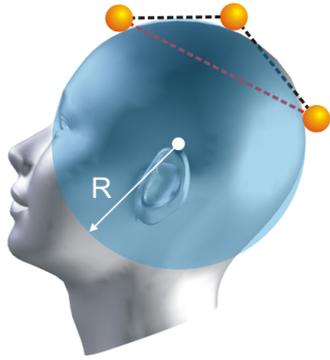


Abbildung 4.6: Zum Problem der Randbedingungen: Interaktion der Wirkungslinie zweier Partikel auf gegenüberliegenden Seiten des Kopfes mit der Kopfoberfläche.

schneiden kann. In unserem Fall wird der Rand durch die Oberfläche eines Kopfmodells gebildet. Es ist i.d.R. nicht möglich, solche Randbedingungen mit Standardmultipolbibliotheken unter Aufrechterhaltung des $\mathcal{O}(n)$ Schemas zu handhaben, da die Multipolentwicklungen auf der Basis einzelner Zellen berechnet werden. Prinzipiell können solche Interaktionen direkt berechnet und anschließend von der Gesamtladung der betroffenen Partikel subtrahiert werden. Vereinfachend wollen wir jedoch annehmen, dass die induzierten Berechnungsfehler klein sind, da lediglich Partikel direkt an der Kopfoberfläche betroffen sind, vgl. Abb. (4.6). Das dadurch im Bereich der Haarwurzel induzierte Biegemoment ist vernachlässigbar. Wegen der quadratisch mit dem Abstand abnehmenden Kräfte sind auch die Auswirkungen zwischen Partikeln mit großem Abstand zueinander vernachlässigbar.

4.5 Numerische Beispiele

Zur Überprüfung unseres Ansatzes wurden unterschiedliche Experimente durchgeführt. Zur Ladungsberechnung kam die frei erhältliche C-Bibliothek DPMTA [Rankin, 2002] zum Einsatz. Hierbei wurde ausschließlich vom Einzelprozessormodus Gebrauch gemacht, obwohl die DPMTA-Bibliothek ihre Stärken besonders beim Einsatz in parallelen Umgebungen ausspielt. Zur Senkung des Speicherbedarfs wurde der Bibliotheksquellcode voll in unsere Simulationsumgebung Rapunzel integriert, so dass sich alle Berechnung direkt auf den internen Datenstrukturen von Rapunzel durchführen lassen. Als Hardware diente ein Pentium IV, 2.4 GHz mit 2 GB RAM. Im Einzelnen betrachten wir die folgenden beiden Testfälle:

1. Eine vorgekrümmte Faseranordnung bestehend aus 300 Einzelfasern der Länge 30 cm, zu je 100 Segmenten. Die Faserenden sind gleichmäßig über einen elliptischen Querschnitt mit einer Dichte von 200 Fasern/cm² verteilt.
2. Als komplexeres Beispiel dient eine computergenerierte Frisur aus über 65 K einzelnen Fasern, die eine Hälfte eines menschlichen Kopfmodells bedeckt.

Die Anfangskonfigurationen wurden auf der Basis unseres COSSERAT-Ansatzes generiert. Die geometrischen Eigenschaften entsprechen den realen, bei Humanhaaren auftretenden Werten.

Der Algorithmus verteilt zunächst Ladungspartikel mit einer gewünschten Ladungsdichte über die Fasern und berechnet anschließend die COULOMB-Interaktionen mittels der FMM. In einem

zweiten Schritt folgen alle Interaktionen zwischen den Partikeln und der Kopfoberfläche durch Auswertung des Distanzfeldes. Auf der Basis dieser Kräfte kann dann eine Neuberechnung der Faserkonfigurationen durch Lösen eines unbeschränkten Optimierungsproblems erfolgen, wie in Abs. (3.3.2.1) erläutert. Der Pseudocode ist in Alg. (2) dargestellt. Die Ladungsträgerdichte be-

Algorithmus 2 : Berechnung der Konfiguration \mathcal{C}^+ aus einer gegebenen Anordnung \mathcal{C}^- infolge COULOMB-Interaktion zwischen den Fasern nach Aufbringen einer Ladung Q . \mathcal{S} ist die Kopfoberfläche in Form eines Dreiecksnetzes, n die Anzahl der Simulationsschritte, T die Temperatur, RH die Luftfeuchte und ρ_P die Ladungspartikeldichte. \mathcal{M} ist eine Menge von Materialparametern.

```

Data :  $\mathcal{C}^-$ ,  $\mathcal{S}$ ,  $Q$ ,  $n$ ,  $\rho_P$ ,  $T$ ,  $RH$ 
Result : Neue Konfiguration  $\mathcal{C}^+$  der Anordnung
1:  $q_{ij} \leftarrow 0$ 
2:  $\Phi_S \leftarrow \text{COMPOTENTIALFIELD}(\mathcal{S})$ 
3:  $\mathcal{M} := (E, G, \rho, \Delta d) \leftarrow \text{LOOKUPMECHPROPS}(T, RH)$ 
4: for  $h : 1 \rightarrow n$  do
5:    $\mathbf{x}_{ij} \leftarrow \text{COMPARTICLEDISTRIBUTION}(\mathcal{C}^-, \rho_P)$ 
6:    $dq \leftarrow \text{COMPSTEP}(Q, h)$ 
7:    $q_{ij} \leftarrow q_{ij} + dq$ 
8:    $\mathbf{f}_{ij}^C \leftarrow \text{COMPFMM}(\mathbf{x}_{ij}, q_{ij})$ 
9:    $d_{ij} \leftarrow \Phi_S(\mathbf{x}_{ij})$ 
10:   $\mathbf{f}_{ij}^S \leftarrow |q_{ij}| \cdot |Q_S| / d_{ij}^2 \cdot \nabla \Phi_S(\mathbf{x}_{ij}) / \|\nabla \Phi_S(\mathbf{x}_{ij})\|$ 
11:   $\mathbf{F}_{ij} = \mathbf{f}_{ij}^C + \mathbf{f}_{ij}^S + \mathbf{f}_{ij}^{ext}$ 
12:   $\mathcal{C}^+ \leftarrow \text{COMPEQUILIBRIUMCONFIGURATION}(\mathcal{C}^-, \mathbf{F}_{ij}, \mathbf{g}, \mathcal{M})$ 
13: end

```

trägt konstant $\rho_P = 330$ Partikel/cm. Die Simulationstemperatur betrug bei allen Tests 25 °C, bei veränderlicher Luftfeuchte (65 %-100 %). Die Linienladungsdichte ist $-7,7 \cdot 10^{-4}$. Die Simulationsergebnisse sind in Abb. (4.7), Abb. (4.8) und Abb. (4.9) dargestellt.

Abb. (4.7) und Abb. (4.8) zeigen die finale Konfiguration nach Aufbringen der Ladung auf die vorgekrümmte Faseranordnung bei einer Luftfeuchte von 65 % bzw. 100 %. Infolge des erhöhten Eigengewichtes bzw. der verringerten Steifigkeit der Faser bei 100 % Luftfeuchte fällt der Effekt des Ladungsauftrags deutlich geringer aus. Darüber hinaus führt das erhöhte Eigengewicht zu einer Verringerung der Krümmung, die Faser verliert infolgedessen ihre Form. Dieser Effekt ist auch bei realen Haaren zu beobachten.

Die Ergebnisse für das Halbkopfmodell nach Ladungsauftrag sind in Abb. (4.9) dargestellt. Die Endkonfiguration entspricht den Erwartungen. Für den Ladungsauftrag gibt es prinzipiell zwei Möglichkeiten:

1. Der volle Ladungsbetrag wird in einem einzigen Schritt aufgebracht, oder
2. der Ladungsbetrag wird über n Schritte verteilt, um den Aufbau des elektrostatischen Feldes zu simulieren. Dafür wird ein logarithmisch abnehmender Ladungsbetrag angenommen.

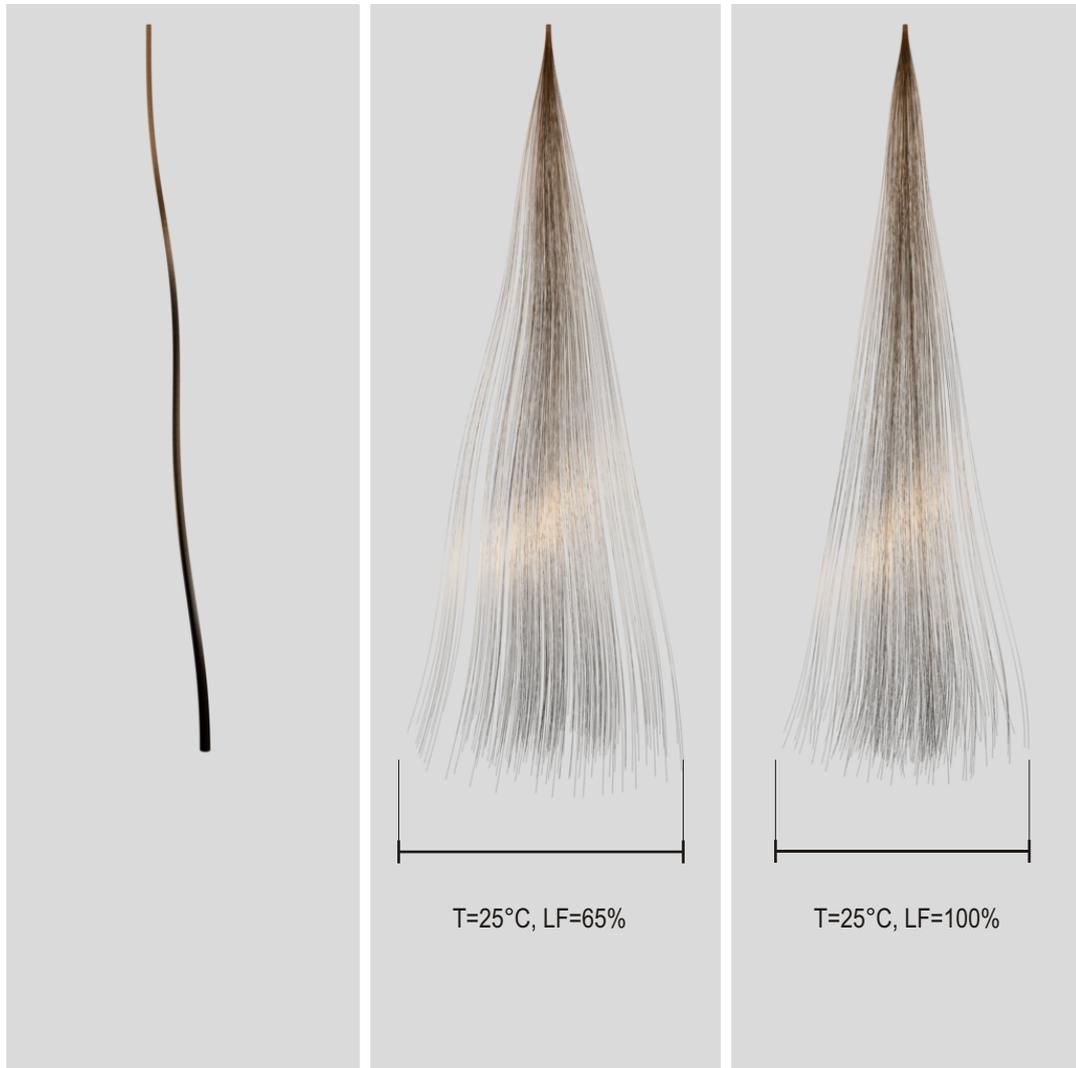


Abbildung 4.7: Die Bilder zeigen die Ergebnisse unseres Ansatzes für die Elektrostatikberechnung auf Humanhaaren. Links: Vorgekrümmte Faseranordnung bestehend aus 300 Einzelfasern der Länge 30 cm zu je 100 Segmenten. Mitte: Nach der Applikation einer Ladung ($T=25\text{ }^{\circ}\text{C}$, $RH=65\%$). Rechts: Nach Anheben der Luftfeuchte auf 100 %. Infolge der Gewichtszunahme bzw. der Abnahme der Steifigkeitsparameter neigen die Fasern zum Verlust ihrer Form, ein Effekt, der auch bei realen Haaren zu beobachten ist.



Abbildung 4.8: Das Bild zeigt die vorgekrümmte Faseranordnung bei $T=25\text{ °C}$ und $RH=65\%$ (bei selber Ladung), jedoch ohne den Effekt der Gravitation.

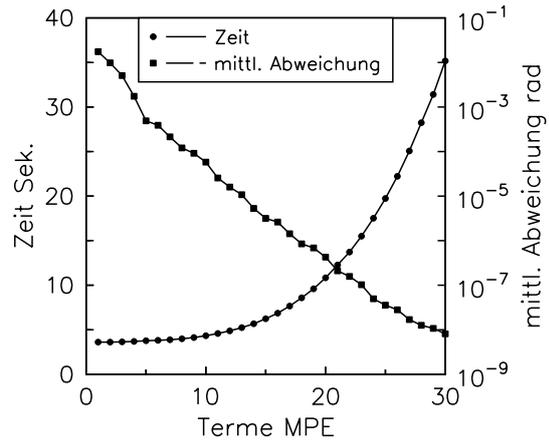


Abbildung 4.9: Die Bilder zeigen die Ergebnisse unseres Ansatzes für die Elektrostatikberechnung auf Humanhaaren beim Halbkopfmodell bestehend aus 65 K Fasern gleicher Länge (100 Segmente pro Faser), vor und nach Applikation einer Ladung auf der Kopfoberfläche und den Haarfasern.

Beide Strategien führen zu völlig unterschiedlichen Ergebnissen. Die Applikation des vollen Ladungsbetrages in einem einzigen Schritt bewirkt nicht dieselbe Endkonfiguration, wie das Aufbringen über mehrere aufeinander folgende Schritte. Im ersten Fall ergibt sich eine asymmetrische Kraftverteilung, mit größeren Kraftbeiträgen auf den äußeren Fasern der Anordnung. Im zweiten Fall konvergiert das System jedoch gegen eine symmetrische Konfiguration. Der Grund dafür liegt darin, dass Asymmetrien in der Kraftverteilung in den folgenden Integrationschritten der COSSERAT-Gleichungen numerisch geglättet werden. Der Integrationsprozess fungiert als Filter. Zudem gilt es zu beachten, dass jeder Integrationsschritt einen weiteren numerischen Fehler induziert. Darüber hinaus ist, wie in Abs. (3.3.2.1) dargelegt, zur Bestimmung der Gleichgewichtskonfiguration das Lösen eines unbeschränkten Optimierungsproblems notwendig. Dieser Optimierungsprozess lenkt die Systemenergie in Richtung eines Minimums und solche Minima entsprechen bekanntermaßen symmetrischen Systemzuständen.

Auf der Basis dieser Berechnungen lassen sich die elektrischen Feldstärken der Anordnung auswerten, die durch die Beziehung $E = F/q$ gegeben sind. Zum Vergleich mit real auftretenden Feldstärken bedienen wir uns einer der wenigen Publikationen, die sich mit solchen Messungen an menschlichen Haaren befassen. Dabei handelt es sich um ein US Patent zur Reduktion statischer Elektrizität mittels spezieller Haarkonditionierer [Patel, 1988]. Die Effektivität solcher Konditionierer quantifizieren die Autoren an einer Gruppe von Haarsträhnen, die durch 50-maliges Kämmen mit einem Gummikamm elektrisch geladen wurden. Der Sensor eines Feldmeters ermittelt die Stärken in einem Abstand von 12 cm von der Oberfläche der Strähne. Die Werte rangieren zwischen 20 kV/m und 120 kV/m und korrespondieren mit denen aus unseren Berechnungen, die im Bereich von etwa 13 kV/m liegen. Darüber hinaus wurde die Anzahl der Terme in der Multipolentwicklung systematisch variiert und die Auswirkung auf die Rechengenauigkeit gemessen. Wie aus Abb. (4.10) leicht zu ersehen ist, steigt die Berechnungszeit in etwa quadratisch

Abbildung 4.10: Berechnungszeiten (in Sek.) der Fast-Multipole-Method bei veränderlicher Anzahl an Termen in der Multipolentwicklung (MPE). Die zweite Kurve entspricht der mittleren Richtungsabweichung der Kraft von der exakten Lösung, gemittelt über alle Partikel.



mit der Anzahl der ausgewerteten Terme. Die Kurve zeigt die mittlere Richtungsabweichung von der exakten Lösung basierend auf dem naiven Ansatz einer direkten Interaktionsauswertung. Die Winkel sind über alle Partikel der Anordnung gemittelt. Die Abweichung liegt bei 10^{-2} rad bei zwei Termen und fällt auf 10^{-9} rad bei 30 Termen ab. Die größte beobachtete Abweichung von der exakten Kraft beträgt 0,3 %. Das stimmt mit dem für zwei Terme vorhergesagten Fehler von $3^{-p/2}$ überein. Schlussendlich belegen unsere Untersuchungen, dass die Benutzung von mehr als zwei Termen für die Multipolentwicklung keine sichtbaren Auswirkungen auf die Konfigurationen der hier getesteten Beispiele hat. Der Grund dafür mag darin liegen, dass die geringfügigen Schwankungen in den Kraftamplituden bzw. Richtungen bei Auswertung einer größeren Anzahl an Termen durch den Integrations- bzw. Optimierungsprozess geglättet werden.

Mit dem vorgestellten Modell konnten die Möglichkeiten der faserbasierten Haarsimulation am Beispiel der Elektrostatikberechnung an Humanhaaren aufgezeigt werden. Die COULOMB-Interaktionen lassen sich durch Kopplung einer Multipolentwicklung über auf den Fasern verteilten Punktladungen und einem statischen COSSERAT-Modell effizient berechnen. Auf der Basis dieses Modells scheint eine Vorhersage der durch Kämmen und Bürsten in Faseranordnungen induzierten Ladungen möglich. Eine numerische Untersuchung der Frage, wie groß der Anteil einer so induzierten Ladung bzw. der durch die mechanischen Belastungen verursachten plastischen Deformation an der Änderung des Volumenkörpers ist, scheint ebenfalls in greifbare Nähe gerückt.

5 Rekonstruktion von Frisuren aus groben Oberflächeninformationen

Im folgenden Kapitel stellen wir einen neuen Ansatz für die Rekonstruktion von expliziten Frisurmodellen aus groben Oberflächeninformationen vor, wie sie bspw. nach der Laserabtastung von Skulpturen vorliegen. Unser Ansatz basiert auf einem Shapematchingalgorithmus, der einzelne Leitfasern unter Beachtung bestimmter Randbedingungen und zusätzlicher Haardichteinformationen in ein vorgegebenes Volumen einpasst. Die Basis bildet ein mechanisches Stabmodell, das durch die speziellen Theorie der COSSERAT-Stäbe beschrieben wird. Solche Modelle können zur effizienten Simulation der Deformation einzelner Fasern in aufeinander folgenden Bearbeitungsschritten herangezogen werden.

5.1 Einleitung

Das Erzeugen realistisch aussehender, expliziter Frisuren stellt ein weitgehend ungelöstes Problem im Bereich der Computergraphik dar. Standardansätze erfordern immer noch einen nicht unerheblichen manuellen Einsatz, da mindestens die Hauptflussrichtungen von Frisuren anhand von Leitkurven vorgegeben werden müssen. Während das Problem der Rekonstruktion lokaler Haargeometrien aus Sequenzen von Photographien jüngst angegangen worden ist, fand die Generierung von Frisuren aus groben Oberflächeninformationen in bereits existierenden Arbeiten noch keine Beachtung. Solche Oberflächendaten können bspw. dem Laserabtastungsvorgang von Skulpturen entstammen (vgl. hierzu [Levoy, 1999]). Die Ränder der zu erzeugenden Frisuren werden zum einen durch die Kopfoberfläche bzw. die Region des Haaransatzes und zum anderen durch die Grenzen des Haarvolumenkörpers gebildet. Möglichkeiten zur Generierung plausibler Frisuren unter Beachtung eben solcher Randbedingungen sind von großem Interesse, bspw. im Kontext virtueller Altertumsprojekte. Die Ergebnisse eines Frisurenrekonstruktionsversuchs auf der Basis der Oberflächendaten einer Beethovenbüste aus dem Bonner Beethovenhaus unter Anwendung des hier vorgestellten Algorithmus, ist in Abb. (5.1) dargestellt.

Der Ansatz ist voll automatisiert. Vom Benutzer ist lediglich eine handvoll Parameter zu spezifizieren. Interaktionen manueller Natur sind hingegen nicht erforderlich. Die notwendige Deformation einzelner Fasern im Rahmen des Einpassungsvorganges in das Zielvolumen soll auch hier durch ein statisches COSSERAT-Modell beschrieben werden.

5.2 Verwandte Ansätze

In den letzten Jahren wurden einigen Ansätze für das Problem der Frisurengenerierung im Allgemeinen vorgestellt. Lee et al. [Lee et al., 2002] haben in diesem Rahmen ein System zur in-

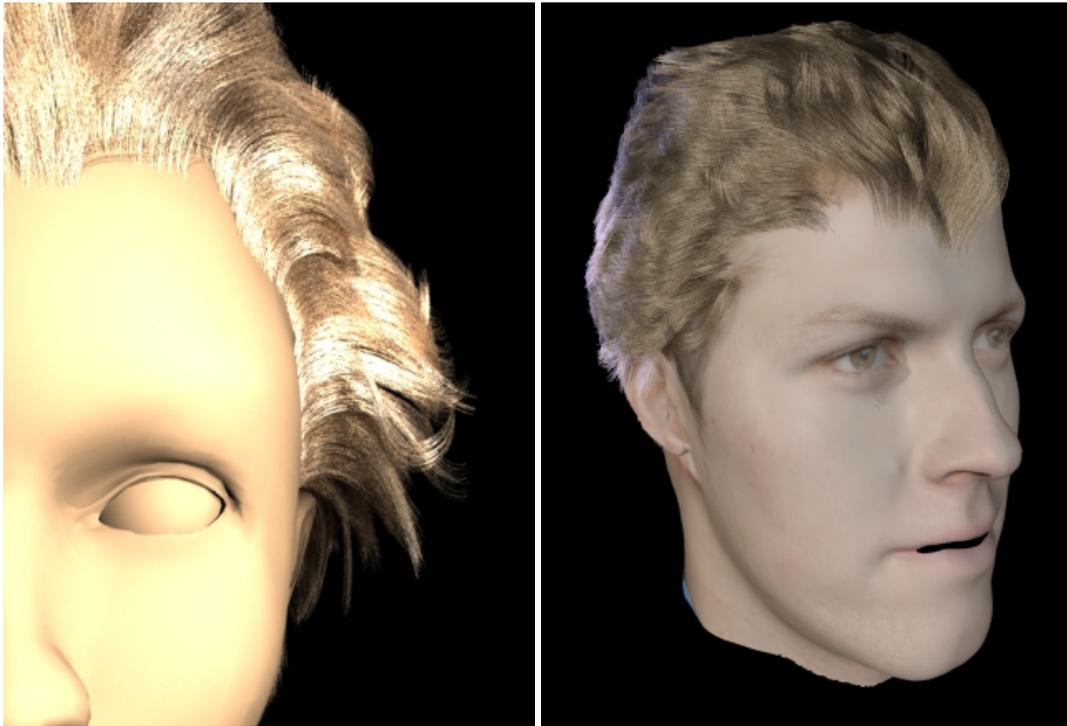


Abbildung 5.1: Ergebnisse eines Frisurenrekonstruktionsversuchs auf der Basis der Oberflächendaten einer Beethovenbüste aus dem Bonner Beethovenhaus unter Anwendung des hier vorgestellten Algorithmus. Das Gesamtergebnis nach Aufsetzen der rekonstruierten Beethovenfrisur auf ein Kopfmodell ohne und mit Texturierung.

teraktiven Manipulation computergenerierter Frisuren mittels eines Motion-Tracking-Gerätes entwickelt. Der Ansatz von Hadap und Thalmann [Hadap & Magnenat-Thalmann, 2000] fasst Haargeometrien als idealisierte Strömung viskoser Fluide auf. Quellen werden entlang der Kopfoberfläche positioniert und die Stromlinien so am Eintreten in die Peripherie gehindert. Allerdings können selbst hohe Quelldichten keine Interaktionsfreiheit der generierten Frisuren garantieren. Dieser Ansatz ist mit unserem insofern verwandt, als wir lineare Approximationen der Objektberandungen auf kartesischen Gittern verwenden, um sowohl das Problem des Eindringens von Fasern in die Peripherie als auch deren Durchdringung untereinander zu vermeiden. Darüber hinaus bleibt der Ansatz von Hadap und Thalmann, bedingt durch die Strömungsanalogie, auf Kurvengeometrien mit weichen Verläufen und kleinen Krümmungen beschränkt. Yu [Yu, 2001] verfolgt mit seiner Arbeit einen ähnlichen Ansatz unter Ausnutzung der Analogie von „natürlichen“ Vektorfeldern und dem zumindest bei oberflächlicher Betrachtung glatt verlaufenden Fluss natürlicher Haare. Prozedural generierte Vektorfeldprimitive können dann zu komplexen Frisuren überlagert werden. Die Methode von Xu und Yang [Xu & Yang, 2001] basiert hingegen auf der manuellen Anordnung von Haarclustern, die durch das wiederholte Anwenden von Elementaroperationen bis zur vollständigen Frisur multipliziert werden. Im gleichen Lichte erscheint der Beitrag von Kim und Neumann [Kim & Neumann, 2002]. Deren Multiresolution-

Haarmodellierungssystem liegt erneut das Clusterparadigma zugrunde. Eine ganze Frisur stellt sich dann als Ergebnis einer Serie von Kopier-, Einfüge-, Verdrehungs- und Verschiebeoperationen dar. Darüber hinaus lassen sich einzelne Cluster bis zum Erreichen der gewünschten Detailgenauigkeit individuell nachbearbeiten. Besondere Würdigung verdient die Arbeit von Choe und Ko [Choe & Ko, 2005] im Hinblick auf die beeindruckenden Ergebnisse, die sich damit erzielen lassen. Ihr Haarmodellierungswerkzeug basiert, ähnlich dem unsrigen, auf einem pseudophysikalischen Modell, in dem Kraft- und Dichtefelder die Basis für die Annäherung physikalischer Einflüsse wie Gravitation und Haar-Haar-Interaktion bilden. Die Faseranordnung wird als Kontinuum aufgefasst, dessen lokale Dichtefunktion ein Maß für die regionale Kollisionswahrscheinlichkeit darstellt. Orte hoher Dichte signalisieren dem Anpassungsalgorithmus bereits belegte Bereiche, so dass folgende Fasern durch Aufbringen adäquater Kräfte aus dem verbotenen Bereich gedrückt werden können. Berechnet man solche Dichtefelder auf kartesischen Gittern, lassen sich auch die Gradienten der Dichtefunktion bestimmen. Dieser gibt Aufschluss darüber, in welche Richtung eine einzupassende Leitfaser verschoben werden muss, um in eine Region geringerer Dichte zu gelangen.

In Choe und Kos Ansatz definiert ein dreidimensionales Kraftfeld $\Phi(\mathbf{x})$ mit $\mathbf{x} \in \mathcal{A}$ und $\mathcal{A} := \{\mathbf{x} \in \mathbb{R}^3 | x_l < x_1 < x_u, y_l < x_2 < y_u, z_l < x_3 < z_u\}$ die einzuhaltenden Flussrichtungen infolge Stylingoperationen oder Gravitation an jedem Punkt des Raumes. Die Biegewinkel zwischen den Segmenten einer weniggliedrigen kinematischen Kette werden dann in einem Optimierungsprozess so gewählt, dass sie sich, entsprechend der Steifigkeit ihrer Gelenke, möglichst entlang des Kraftfeldes ausrichten. Beschrieben wird das durch die Funktion

$$\mathcal{U}(\mathbf{d}) = \left[\left(\Phi(\mathbf{x}) + c \cdot \frac{\mathbf{d}_0}{\|\mathbf{d}_0\|} \right) \cdot \frac{\mathbf{d}}{\|\mathbf{d}\|} \right] \rightarrow \max, \quad (5.1)$$

in der \mathbf{d} und \mathbf{d}_0 die Richtungsvektoren des Segmentes sind, und zwar in der aktuellen Lage bzw. in Anfangslage, also wenn der Biegewinkel Null ist. c ist eine beliebige Steifigkeitskonstante. Diese Funktion nimmt genau dann ihr Maximum ein, wenn das Segment sich parallel zum lokalen Fluss stellt und die Segmentrichtung der Ruhelage entspricht. Es wird die Segmentrichtung gesucht, welche die Funktion $\mathcal{U}(\mathbf{d})$ maximiert,

$$\mathbf{d}^* = \operatorname{argmax}_{\mathbf{d} \in \mathbb{R}^3} \mathcal{U}(\mathbf{d}). \quad (5.2)$$

Räumliche Zwänge infolge von Stylingoperationen lassen sich auf lokale Vektorfeldprimitive $\Phi_C(\mathbf{x})$ mit $\mathbf{x} \in \mathcal{A}' \subseteq \mathcal{A}$ reduzieren, die in Form einer konvexen Kombination $\alpha\Phi(\mathbf{x}) + (1 - \alpha)\Phi_C(\mathbf{x})$ mit dem Kraftfeld zu superponieren sind. Ein Beispiel für ein solches Vektorfeldprimitiv ist bspw. die lokale Vorgabe der Fließrichtung durch eine Trajektorie und einen Einflussradius, innerhalb dessen die Richtungen tangential zur Trajektorie verlaufen. Durch Einführung eines Dichtefeldes kann vor Einbau von Segmenten mit Wirkungsradius geprüft werden, ob die lokale Dichte zuzüglich der durch das neue Segment verursachten Änderung einen Vorgabewert überschreitet. In solchen Fällen ist der Biegewinkel in Richtung des lokalen Dichtefeldgradienten zu modifizieren.

Alle genannten Ansätze haben gemeinsam, dass Frisuren manuell unter zur Hilfenahme von interaktiven Werkzeugen erstellt werden müssen. Auf der anderen Seite steht der Computer Vision Ansatz, dessen primäres Ziel, die vollautomatische Rekonstruktion von Frisuren aus Photos, in den letzten Jahren in den Mittelpunkt des Interesses gerückt ist. Haider und Kaneko

[Haider & Kaneko, 2000] haben bspw. Algorithmen für die Rekonstruktion des Haarvolumens aus einer Sequenz von Videobildern von Patienten entwickelt. Ein haarloses, dreidimensionales Kopfmodell aus dem CT-Scanner dient zur Registrierung der Kameraparameter. Nach anschließender Segmentierung der Haarregion in den Bildern und Registrierung der Kamerapositionen kann ein dreidimensionales Haarvolumen errechnet werden. Die extrahierten Haartexturen aus dem Segmentierungsprozess lassen sich dann auf die Oberfläche des Volumens aufbringen. Dieser Ansatz ist rein volumetrisch, es werden keine Haarfasern generiert. Die Methode von Paris et al. [Paris et al., 2004] ist hingegen in der Lage, aus einer Serie von unter kontrollierten Beleuchtungsbedingungen aufgenommenen Bildern von Frisuren durch die systematische Anwendung einer Kaskade von Filteroperationen dreidimensionale, explizite Haarmodelle zu erzeugen. Die Filteroperationen liefern Orientierungsinformationen für die einzelnen Pixel eines Bildes. Diese Richtungsinformationen, die visuelle Hülle des Haarvolumens und zusätzliche Informationen wie Kamerapositionen bilden die Datenbasis für den Rekonstruktionsalgorithmus. Die beeindruckendsten Resultate konnten bis jetzt von Wei et al. [Wei et al., 2005] erzielt werden. Nur eine Folge von Bildern, die im Gegensatz zum Ansatz von Paris et al. [Paris et al., 2004] unter unkontrollierten Beleuchtungsbedingungen aufgenommen werden können, ist notwendig. Die zur expliziten Generierung von Fasern notwendigen Orientierungen ergeben sich durch Triangulation von Orientierungsinformationen aus unterschiedlichen Perspektiven. Die Kopfoberfläche wird durch eine Offsetfläche aus der visuellen Hülle generiert. Diese Ansätze erfordern zur Rekonstruktion lokale Orientierungsinformationen, die bei Oberflächenabtastungen i.d.R. nicht verfügbar sind. Mit der Oberfläche liegt lediglich die visuelle Hülle vor.

5.3 Der Algorithmus

Die Eingabespezifikation unseres Algorithmus erfordert die Bereitstellung der Kopfoberfläche \mathcal{S}_S in Form eines Dreiecksnetzes und zusätzlich die Berandung oder visuelle Hülle \mathcal{S}_H des zugeordneten Haarvolumens, vgl. Abb. (5.2). Ausgehend von einem Oberflächenscan einer Skulptur besteht also die grundlegende Aufgabe zunächst in der Zerlegung des Skulpturvolumens in die notwendigen Eingabedaten. Nach Extraktion der Haaroberfläche aus dem Skulpturvolumen bleibt ein zu füllendes Loch. Über die Kopfoberfläche ist i.d.R. nichts bekannt und entsprechend schwierig gestaltet sich diese Aufgabe. Eine Möglichkeit besteht darin, die Kopfoberfläche durch eine Offsetfläche aus der Berandung zu approximieren und anschließend zu glätten. Dieses Vorgehen liefert bestenfalls eine sehr grobe Annäherung. Die Unsicherheit bezüglich der realen Verhältnisse unter dem Haarvolumen bleibt aber weiter bestehen. Ein vielversprechender Ansatz ist die Suche nach ähnlichen haarlosen Kopfmodellen in einem großen Datenbestand und anschließender Subtraktion vom Skulpturvolumen zwecks Gewinnung der Haaroberfläche bzw. des Hinterkopfes. Dieser Ansatz weist starke Ähnlichkeiten zu dem *Morphable Head Models* von Blanz and Vetter [Blanz & Vetter, 1999] auf.

Momentan haben wir keine voll automatisierte Lösung für dieses Teilproblem im globalen Haarrekonstruktionsprozess anzubieten. Für unsere Beispiele wurden nicht vorhandene Daten vielmehr semimanuell erzeugt. Der gesamte Preprocessingschritt ist deshalb im Folgenden als Blackbox aufzufassen.

Auf der Basis der Berandung, die durch die Kopfoberfläche und die Oberfläche des Haarvo-

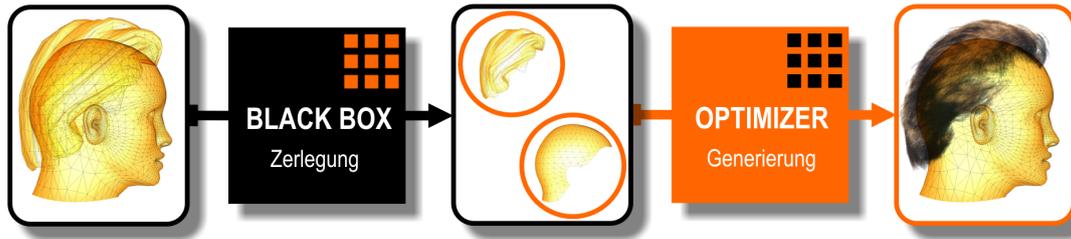


Abbildung 5.2: Die Kopfoberfläche und die Oberfläche des Haarvolumens bilden die Eingabedaten für den Rekonstruktionsalgorithmus. Die Extraktion dieser Informationen aus der Oberfläche eingescannter Skulpturen ist nicht Gegenstand des Algorithmus und wird hier als Blackbox betrachtet.

lumen gebildet wird, lösen wir ein Optimierungsproblem, um optimale Kraft-Moment-Paare zu finden, die Fasern in eine optimale Konfiguration deformieren, entsprechend ihrer Abstände von den Berandungen und unter Einbeziehung lokaler Dichteinformationen. Auf der technischen Seite erfolgt die Lösung eines uneingeschränkten Optimierungsproblems, indem alle notwendigen Bedingungen in Form schwacher Nebenbedingungen formuliert und in eine eindimensionale Zielfunktion kollabiert werden. Da wir von einem leeren Volumen ausgehen, werden zunächst so genannte Vorableithaare erzeugt, welche die Hauptflussrichtungen grob für eine anschließende Generierung der Hauptleithaare abbilden. Nach diesem Optimierungsprozess werden auf der Basis dieser Hauptleithaare mittels eines geeigneten Algorithmus Strähnen generiert. Wir bezeichnen das von der Kopfoberfläche und dem Haarvolumen eingeschlossene Volumen als *zulässige Region*. Der Übergang zwischen zulässiger und unzulässiger oder verbotener Region ist glatt.

In den folgenden Abschnitten erfolgt eine detaillierte Besprechung der einzelnen Komponenten unseres Systems.

5.3.1 Repräsentation von Randbedingungen

Ein zentrales Problem für den Algorithmus ist die effiziente Darstellung der Gebietsberandung, die aus der Oberfläche des Kopfes bzw. der des Haarvolumens gebildet wird, so dass Haare möglichst einfach am Austreten aus der zulässigen Region gehindert werden können. Hier, wie bei ähnlich gelagerten Problemen in dieser Arbeit auch, bedienen wir uns einer impliziten, linearen Approximation oder eines Distanzfeldes Φ_S der Flächen \mathcal{S}_S und \mathcal{S}_H . Diese ist eine reellwertige Funktion $\Phi_S : \mathbb{R}^3 \rightarrow \mathbb{R}$, so dass

$$\Phi_S(\mathbf{x}) := \min_{\mathbf{q} \in \mathcal{S}} \{\|\mathbf{x} - \mathbf{q}\|\}, \quad \forall \mathbf{x} \in \mathbb{R}^3. \quad (5.3)$$

Diese Approximation wird auf einem kartesischen Gitter berechnet. Da die Dreiecksnetzdarstellungen der beiden Oberflächen nicht notwendig geschlossen oder zusammenhängend sein muss, wird lediglich eine Distanzhülle mit vordefinierter Dicke berechnet.

5.3.2 Parametrisierung der Wachstumsregion

Vor der Generierung von Frisuren sind zunächst die Wachstumsregion der Haare bzw. deren Dichten, die Auswuchsrichtungen und die Haarlängen in Abhängigkeit von der Position auf der

Kopfoberfläche zu spezifizieren. Die Angabe dieser Informationen soll möglichst unabhängig von der Diskretisierung der Kopfoberfläche erfolgen. Dazu führt man zweckmäßigerweise eine Parametrisierung des in Frage kommenden Bereiches durch. Die Parametrisierung der diskretisierten Kopfoberfläche \mathcal{S}_3 erfolgt durch eine konformale Abbildung $C : \mathbb{R}^3 \mapsto \mathbb{R}^2$, vgl. Floater [Floater, 1997], so dass alle Dreiecke möglichst verzerrungsfrei bzw. winkeltreu auf die Ebene projiziert werden. Die Parametrisierung garantiert uns eine eins-zu-eins Abbildung. Die Abbildung des Parameterbereiches auf dem Einheitskreis scheint uns besonders sinnvoll. Sind die Dreiecke erst einmal auf die Ebene übertragen, lassen sich alle notwendigen Informationen in Form einer Textur, hier ein einfaches Grauwertbild, welches die Dichte an jedem Pixel durch einen Wert zwischen 0 und 255 darstellt, auf die zweidimensionale Darstellung der Kopfoberfläche übertragen. Die Parametrisierung liefert die Transformationen für die Abbildung der Dreiecke vom dreidimensionalen Bereich in den zweidimensionalen Bereich. Durch Invertierung lassen sich folglich die Koordinaten der Texturpixel in die dreidimensionale Darstellung zurückrechnen. Diese Koordinaten bilden die Poren der Kopfhaut. Grauwertbilder lassen sich mittels einfacher Bildbearbeitungsprogramme erstellen und manipulieren.

Sei $\mathcal{M} \in \mathbb{R}^3$ ein Dreiecksnetz und $\mathcal{M}' \in \mathbb{R}^2$ das parametrisierte Gegenstück. Dann ist jedem Dreieck $\tau = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\} \in \mathcal{M}$ im dreidimensionalen Raum genau ein korrespondierendes $\tau' = \{\mathbf{p}'_1, \mathbf{p}'_2, \mathbf{p}'_3\} \in \mathcal{M}'$ in der Parameterebene zugeordnet. Die Menge der Dreiecke $\{\tau'\}$ definiert also ein ebenes Dreiecksnetz $\mathcal{M}' \in \mathbb{R}^2$. Die Abbildung zwischen beiden Darstellungen ist bijektiv. Die Korrespondenz der Punkte von τ und τ' definiert eine affine Abbildung $P : \tau \rightarrow \tau'$, deren Dehnung durch die Eigenwerte des Metrikensors $\delta(\tau') = \sqrt{(\lambda_{max}^2 + \lambda_{min}^2)}/2$ der konformalen Abbildung $\mathcal{M} \rightarrow \mathcal{M}'$ beschrieben wird.

Bei der Parametrisierung, wie bspw. bei Floater [Floater, 1997] beschrieben, werden die Randknoten des Netzes \mathcal{M} auf denen von \mathcal{M}' abgebildet, die ein Polygon in der Parameterebene aufspannen, eine Nichtgeschlossenheit des Netzes ist also Voraussetzung. Alle inneren Knoten nehmen Positionen innerhalb des Polygons ein, so dass der folgende Energieausdruck minimal wird,

$$\mathcal{U}(\mathcal{M}') := \left[\sum_i \sum_j w_{jk} (\mathbf{p}'_k - \mathbf{p}'_j)^2 \right] \stackrel{!}{\rightarrow} \min. \quad (5.4)$$

Der Index k bezeichnet alle direkten Nachbarn des Knotens j auf \mathcal{M} . Das Netz wird also als System von durch Federn elastisch verbundenen Massepunkten aufgefasst, deren elastische Gesamtenergie in der „optimalen“ Position minimal ist. Die Position eines jeden Knoten ist die konvexe Summe $\mathbf{p}_i = \sum_j w_{ij} \mathbf{p}_j$ aller seiner Nachbarn mit $\sum_j w_{ij} = 1$. So gesehen, garantiert dieser Ansatz für alle vom Knoten j ausgehenden Dreiecke Überschneidungsfreiheit. Die Positionen aller Knoten bestimmt man durch Lösen des folgenden linearen Gleichungssystems,

$$\mathbf{W}\mathbf{p} = \mathbf{u}. \quad (5.5)$$

$\mathbf{W} \in \mathbb{R}^{n \times n}$ ist die Gewichtungsmatrix, die alle Kantengewichte $w_{ij} > 0 \Rightarrow i \neq j \wedge \mathbf{p}_i \in \tau_j$ festlegt. Falls die Berandung von \mathcal{M}' ein konvexes Polygon bildet, ist die Lösung von Gl. (5.5) eine gültige Parametrisierung von \mathcal{M} . Allerdings beinhaltet das nicht auch notwendigerweise eine lokal optimale Lösung im Sinne möglichst kleiner lokaler Verzerrungen. Nach [Yoshizawa et al., 2004] kann durch Einführung eines geeigneten Dehnungsmaßes $\delta(\mathbf{p}'_i)$ die lokale Dehnung gemessen und anschließend durch Änderung der Gewichte in der Form $w'_{ij} = w_{ij}/\delta_j$ neu verteilt werden.

Ein geeignetes Dehnungsmaß erhält man bspw. unter Einbeziehung des Flächeninhaltes aller zu einem Punkt benachbarter Dreiecke in der Form

$$\delta(\mathbf{p}'_i) = \sqrt{\frac{\sum_{j=\{j|\mathbf{p}_i \in \tau_j\}} A(\tau_i) \cdot \delta(\tau'_i)^2}{\sum_{j=\{j|\mathbf{p}_i \in \tau_j\}} A(\tau_i)}}, \quad (5.6)$$

wobei $A(\tau_i)$ der Flächeninhalt des Dreiecks ist.

Zunächst führt man also die formerhaltende Parametrisierung durch, um anschließend die lokalen Verzerrungen schrittweise durch Lösen von Gl. (5.5) mit neu berechneten Gewichten zu entspannen. Man bricht ab, wenn die Änderung der Gesamtverzerrung $\sum \delta(\mathbf{p}'_i)$ von einem Schritt des Relaxationsprozesses zum nächsten unter einen gewünschten Wert fällt.

Für unsere Zwecke greifen wir auf die Implementierung von [Yoshizawa et al., 2004] zurück, die uns hinsichtlich ihrer Integrierbarkeit in unseren Simulationscode angemessen erscheint.

Die Haardichte-, Wachstumsrichtungs- und Haarlängenverteilungen werden durch ein Grauwertbild kodiert, das dann auf die zweidimensionale Fläche abgebildet wird. Leider liegen in der Literatur nur wenige empirische Befunde über echte Verteilungen der Wachstumsrichtungen auf dem Kopf vor. Aus unseren eigenen Beobachtungen wissen wir jedoch, dass die Haare der Temporal- bzw. Occipitalregion des Kopfes eher in tangentialer Richtung zur Oberfläche als normal dazu auswachsen. Dieser Befund ist nicht unwesentlich, da seine Berücksichtigung die Erstellung realistischerer Frisuren erlaubt, wie unsere Ergebnisse klar zeigen.

5.3.3 Leithaargenerierung

Auf der Basis des weiter oben vorgestellten COSSERAT-Modells und mittels der linearen Approximation der Grenzflächen kann nun ein Algorithmus entwickelt werden, der die Leithaare innerhalb der zulässigen Region hält und dabei dem durch die Vorableithaare festgelegten Fluss folgt. Wir definieren dazu die terminale Endlast $\eta_L = \{\mathbf{f}_L, \mathbf{l}_L\}$, die an einer Faser an der Stelle $s = L$ angreift. Gesucht ist diejenige terminale Endlast η_L^* , welche eine später noch zu bestimmende Zielfunktion minimiert. Folglich ist unser Optimierungsproblem sechsdimensional.

An dieser Stelle möchte man sich Fragen, aus welchem Grunde ein terminales Kraft-Moment-Paar für die Lösung des Einpassungsproblems von Fasern in die vorgegebenen Begrenzungen ausreichend ist. Der Grund ist offensichtlich: die Verwendung eines Kraft-Moment-Paares erlaubt uns die Modellierung beliebiger Faserkonfigurationen, Helices für Locken eingeschlossen. Eine terminale Kraft hingegen vermag die Faser lediglich in die Form eines Bogens zu überführen. Die zusätzliche Berücksichtigung eines terminalen Drehmomentes erweitert das Spektrum möglicher Konfigurationen. Prinzipiell ist die Einbeziehung intermediärer Lasten, die also zwischen den Endpunkten wirken, ebenso möglich. Die Dimension des Optimierungsproblems wird dadurch allerdings vergrößert. Eine direkte Kontrolle der Biegewinkel, wie in [Choe & Ko, 2005] vorgeschlagen wurde, kann im Gegensatz zum Einsatz eines Stabmodells keinen kontinuierlichen Verlauf der Fasergeometrie garantieren.

Generierung der Vorableithaare: Die Generierung von Vorableithaaren erfolgt nach dem folgenden Schema: Zunächst werden die Ansatzpunkte mit einer geringen Dichte von einem Punkt pro cm oder weniger über die Wachstumsregion verteilt, vgl. Abs. (5.3.2). Anschließend werden initial gerade Leithaare mit der für den jeweiligen Ansatzpunkt vorgegebenen Länge in Richtung

der zugehörigen Wachstumsrichtungsvektoren (an diesem Punkt der Wachstumsregion) gedreht. Ausgehend von dieser Anfangsposition wird eine terminale Endlast iterativ erhöht, bis die gesamte Faser bis zu einer vorgegebenen Tiefe unter die Begrenzung des Haarvolumens taucht. Die Kraft wirkt dabei normal zur Stabachse und liegt in der Ebene, die durch die Normale des Dreiecks, auf welcher der Ansatzpunkt liegt, und die dort vorgegebene Wachstumsrichtung aufgespannt wird. Die Wahl der Eintauchtiefe erfolgt so, dass die frontalen Haare näher an der Oberfläche des Haarvolumens liegen, und dann mit abnehmender geodätischer Höhe der Ansatzpunkte zunimmt. Untere Haarschichten werden also in Richtung Kopfoberfläche dirigiert, wohingegen die oberen näher zur Haaroberfläche tendieren. Die maximal mögliche Eintauchtiefe erhält man durch Superposition der beiden Distanzfelder, dem der Kopfoberfläche und dem der Haarvolumenoberfläche.

Generierung der Hauptleithaare: Für die Erzeugung der Hauptleithaare lösen wir ein sechsdimensionales Optimierungsproblem. Die zu minimierende Zielfunktion ist eine gewichtete Summe von n Termen, von denen jeder einen bestimmten Aspekt des zugrunde liegenden Problems steuert. Die Zielfunktion ist gegeben durch

$$f(\eta) = \sum_{i=1}^n w_i \cdot \mathcal{F}_i \rightarrow \min, \quad (5.7)$$

wobei $\mathcal{F}_i : \mathbb{R}^3 \mapsto \mathbb{R}$ skalarwertige Funktionen und $w_i \in \mathbb{R}$ entsprechende Gewichte mit $\sum w_i = 1$ sind.

Der Erste Term \mathcal{F}_1 steht für die Richtungsentsprechung der zu generierenden Leitfaser und ihren direkt benachbarten Vorableitfasern. Sie wird wie folgt ausgedrückt,

$$\mathcal{F}_1 = \int_0^1 \left(1 - \left\langle \frac{\mathbf{r}'_G}{\|\mathbf{r}'_G\|}, \frac{\mathbf{r}'_C}{\|\mathbf{r}'_C\|} \right\rangle \right) ds. \quad (5.8)$$

Die Richtungskorrespondenz an der Stelle s wird durch den Cosinus des Winkels $\varphi(s)$ zwischen den lokalen Tangenten (erste Ableitung der Kurvenfunktion $\mathbf{r}'(s)$) von Leit- und Vorableitfaser, $\mathbf{r}'_G(s)$ bzw. $\mathbf{r}'_C(s)$, gemessen. Die Funktion folgt also $(1 - \cos \varphi(s))$. D.h., dass bei Parallelität der lokalen Tangenten die Funktion den Wert 0, bei Orthogonalität 1 und bei antiparallelem Verlauf den Wert -1 annimmt. Die Funktion wird für eine veränderliche Anzahl von Vorableitfasern in der Nachbarschaft der aktuellen Faser ausgewertet. Es hat sich gezeigt, dass drei Vorableitfasern ausreichen. Der Beitrag jeder Vorableitfaser wird mit dem Abstand ihres Basispunktes zum Basispunkt der Leitfaser auf der Kopfoberfläche gewichtet. Das führt zu glatteren Übergängen. An dieser Stelle sei darauf hingewiesen, dass hier $\mathbf{r}'(s) = \mathbf{d}_3(s)$.

Der zweite Term \mathcal{F}_2 der Zielfunktion bewertet die Ähnlichkeit der aktuellen Faser mit den umgebenden Vorableitfasern. Sie ist definiert als

$$\mathcal{F}_2 = \left[\int_0^1 (\Delta \boldsymbol{\kappa}^2) ds \right]^{\frac{1}{2}}. \quad (5.9)$$

Die lokale Biegung und Torsion drücken dabei die Ähnlichkeit zweier Fasern aus. Die Größen sind durch den DARBOUX-Vektor $\boldsymbol{\kappa}$ gegeben. Der Vektor $\Delta \boldsymbol{\kappa} = \boldsymbol{\kappa}_G - \boldsymbol{\kappa}_C$ beschreibt die Verdrehungsdifferenz von Vorableit- und aktueller Faser. Da zwei Fasern unterschiedliche Längen haben können, vergleichen wir die normalisierten Kurven, so dass $s \in [0, 1]$.

Der dritte Term \mathcal{F}_3 der Zielfunktion behandelt das Problem der Durchdringung der Grenzflächen durch die Fasern und ist gegeben durch

$$\mathcal{F}_3 = \begin{cases} |d|, & \text{falls } d < 0, \\ c_d + |d|, & \text{falls } d > 0. \end{cases} \quad (5.10)$$

Wie bereits erwähnt, wird die Berandung durch die Kopf- und die Haaroberfläche gebildet. Wir berechnen den Abstand jedes Punktes zu der Berandung mittels ihrer Distanzfelder. Wie aus der Definition von \mathcal{F}_3 ersichtlich ist, wird das Austreten aus der Berandung durch Addition eines Strafzuschlags c_d härter bestraft, als das Annähern der Fasern an die Berandung von innerhalb des Volumens. Die Funktion wird für beide Flächen ausgewertet, die Kopf- und die Haaroberfläche.

Die Beiträge der einzelnen Funktionen lässt sich bequem durch die Gewichtungparameter w_i steuern. Je größer bspw. der Einfluss des Ähnlichkeitsterms ist, desto besser fällt die Angleichung der Hauptleitfasern an die Form der entsprechenden Leitfasern aus.

Bei jeder Hauptleitfaser starten wir den Optimierungsprozess von einer geraden Konfiguration ($\boldsymbol{\kappa}^\circ = \mathbf{0}$) in Richtung der entsprechenden Auswachsrichtung.

5.3.4 Einsatz lokaler Dichteinformationen

Zur Behandlung von Haar-Haar-Interaktionen und zur Ausrichtung des Gesamtprozesses hin auf eine homogenere Verteilung der Fasern innerhalb der zulässigen Region, beziehen wir auch lokale Dichteinformationen mit ein. Für diese Zwecke wird iterativ ein Dichtefeld aufgebaut, das die lokale Dichte an jedem Punkt der zulässigen Region approximiert. Die Erzeugung erfolgt auf einem kartesischen Gitter. Falls die lokale Dichte innerhalb der Region, die eine Leitfaser und ihre Strähne, die in einem abschließenden Interpolationsprozess über jeder Hauptleitfaser erzeugt wird, voraussichtlich besetzen wird, einen bestimmten Toleranzwert überschreitet, beaufschlagen wir den Energiewert mit einem hohen Strafzuschlag. Wenn eine Hauptleitfaser ihre finale Position einnimmt, werden alle Dichten innerhalb einer zylindrischen Region um die Faser angepasst. Zwei unterschiedliche Dichtemodelle sind möglich:

- **Binäres Belegungsmodell (BBM):** Dieses Modell weist allen Voxeln, die eine Strähne belegt, den Wert 1 zu. Eine hohe Auflösung des Dichtefeldes ist Voraussetzung dafür, dass diejenigen Regionen, die nur von einem Segment ausgefüllt werden, möglichst klein bleiben. Wir werden diesem Modell durch die Addition eines weiteren Terms zur Zielfunktion gerecht,

$$\mathcal{F}_4^{BBM} = c_{BOM} \cdot \sum \delta(\mathbf{s}_i). \quad (5.11)$$

Die Funktion δ evaluiert zu 1, falls das Segment s_i einen belegten Voxel schneidet, sonst zu 0. D.h., dass für jedes Segment der Faser, die einen belegten Voxel schneidet, der Wert c_{BBM} addiert wird. Zur Bestimmung der Belegung benutzen wir einen einfachen Voxelisierungsalgorithmus, der die Überschneidung mit Gitterzellen mittels des Separating-Axis-Tests (SAT, vgl. [Gottschalk et al., 1996]) detektiert. Dafür wird jedes Segment des generalisierten Zylinders (Kegelstumpf) der Strähne in einen orientierten Quader eingebettet. Der SAT kann dann für alle Voxel innerhalb des achsenparallelen Quaders des Kegelstumpfs und dem orientierten Quader ausgeführt werden.

- **Kontinuierliches Dichtemodell (KDM):** Für das zweite Modell berechnen wir ein skalares Dichtefeld $\Phi_\rho : \mathbb{R}^3 \mapsto \mathbb{R}$ auf einem rechtwinkligen Gitter. Für jede Faser ermitteln wir dazu alle Faser-Voxel-Überschneidungen mittels eines Voxel-Traversieralgorithmus. Die lokale Dichte eines Voxel wird genau um das Verhältnis der im Voxel liegenden Segmentlänge und der Zelldiagonale erhöht. Die Idee besteht darin, die Faser in Richtung gleicher Dichte zu zwingen. Dafür muss der Dichtegradient $\nabla\Phi_\rho$ entlang der Leitkurve berechnet werden. Der entsprechende zur Zielfunktion zu addierende Term ist

$$\mathcal{F}_4^{KDM} = c_\rho \cdot \int_0^L \left\langle \nabla\Phi_\rho, \frac{\mathbf{r}'}{\|\mathbf{r}'\|} \right\rangle^2 ds. \quad (5.12)$$

Falls die Richtung der lokalen Tangente parallel zum Gradienten verläuft, wird der Energiewert mit $c_\rho \cdot \|\nabla\Phi_\rho\|^2$ beaufschlagt. Falls die Tangente orthogonal zum Dichtegradienten verläuft, ist der Zuschlag 0, weil die lokale Richtung der Faser in optimaler Ausrichtung, also tangential zu einer Dichteäquipotentialfläche verläuft. Der Gradient des Dichtefeldes wird hier durch zentrale finite Differenzen approximiert. Das Vorgehen ist dasselbe wie bei der Gewinnung der Informationen aus Distanzfeldern.

Zum Vergleich des Homogenitätsgrades vor und nach der Einbeziehung der Dichtefunktion haben wir für beide Fälle ein Dichtehistogramm berechnet, das im Abschnitt über die Ergebnisse zu finden ist.

5.3.5 Finaler Schnitt und Strähneninterpolation

Zur Erzeugung der endgültigen Frisur aus der Menge der Hauptleithaare setzen wir einen einfachen Strähnengenerator ein. Die Gruppierung der Haare lässt sich dadurch rechtfertigen, dass reale Haare unter natürlichen Bedingungen zur Ausformung von Strähnen tendieren. Alle Leitfasern, die nach Abschluss des Optimierungsprozesses die Berandung penetrieren, werden am ersten Austrittspunkt aus der Hülle abgeschnitten. Die Penetration tritt vorzugsweise in der Occipitalregion auf, weil hier der Abstand von den Ansätzen auf der Kopfoberfläche zur Berandung



Abbildung 5.3: Die beiden Referenzhaaroberflächen und die von unserem Ansatz daraus generierten Frisuren. Links: Kugelförmige Oberfläche vom Scan einer Beethovenbüste und die daraus rekonstruierte Frisur. Rechts: Eine mit der Software PoserTM exportierte, gerade Oberfläche einer Langhaarfrisur und das daraus rekonstruierte Modell.



Abbildung 5.4: Das Bild zeigt die Überlagerung der visuellen Hülle der Haaroberfläche und der rekonstruierten Frisur aus unterschiedlichen Perspektiven (Obere Reihe: Beethovenoberfläche; Untere Reihe: Poseroberfläche). Die Bilder verdeutlichen eindrucksvoll die strukturerhaltenden Eigenschaften unseres Verfahrens. Leere Bereiche innerhalb der visuellen Hülle haben ihre Ursache in einer zu geringen Haarlänge.

sehr klein ausfällt. Zur Auffindung des ersten Segmentes, das die Berandung durchdringt, benutzen wir unseren Distanzfeldansatz. Falls die Länge einer Faser außerhalb des Volumens einen bestimmten Toleranzwert überschreitet, bevor sie erneut in das Volumen eintritt, erfolgt der Schnitt am ersten Segment, das die Berandung durchdringt. Ausgehend von der Form des Hauptleithaars generiert ein Strähneninterpolationsalgorithmus eine Anzahl zusätzlicher Fasern innerhalb einer zylindrischen Region um die Leitfaser. Diese Region wird durch einen generalisierten Zylinder mit veränderlichem Radius entlang der Faser geformt. Unser Ansatz diesbezüglich ist sehr ähnlich zu dem aus [Choe & Ko, 2005], aber erlaubt den Interpolanten zusätzlich ein Winden um die Leitfaser.

5.4 Numerische Beispiele

Die Effizienz unseres Ansatzes zeigen wir anhand zweier Frisuren, die mit unterschiedlichen Haaroberflächen auf der Basis veränderlicher Parametersätze erhalten wurden. Dabei betrachten wir die folgenden beiden Testfälle:

1. Ein kugelförmiges Volumen aus dem Oberflächenscan einer Beethoven-Skulptur und

2. ein gerades Volumen, das mit der Software PoserTM exportiert wurde,

vgl. Abb. (5.3).

Beide Haarstile bedecken eine Hälfte eines menschlichen Kopfmodells. Die Resultate sind aus unterschiedlichen Perspektiven zusammen mit den visuellen Hüllen in Abb. (5.4) dargestellt.

Die Faseranzahl beträgt 90 K beim ersten und 60 K beim zweiten Modell. Die Anzahl der Hauptleitfasern zum Abtasten der Hauptflussrichtungen betrug 3 K. Die zur Generierung der Gesamtfrisur benötigte Zeit betrug auf einem Pentium IV mit 2.2 GHz etwa 2,5 Min. beim ersten und 1,5 Min. beim zweiten Modell. Abb. (5.4) zeigt deutlich, dass unser Algorithmus in der Lage ist, alle strukturellen Eigenheiten, die durch das Volumen diktiert werden, akkurat abzubilden. Leere Regionen innerhalb des Haarvolumens haben ihre Ursache in unzureichender Haarlänge, ein Problem, das im Rahmen dieser Arbeit nicht weiter vertieft werden soll. Abb. (5.5) zeigt die beiden Referenzfrisuren aus unterschiedlichen Perspektiven und verdeutlicht zudem den Einfluss des Ähnlichkeitsterms sowie den Effekt einer größeren Haarlänge auf das Resultat. Das Haare nicht unwesentlich zum Gesamterscheinungsbild digitaler Avatare beitragen, verdeutlicht hingegen Abb. (5.1).

Der Einfluss der beiden Dichtemodelle KDM und BBM auf die Endresultate, bei veränderlichen Werten der Gewichtungparameter, ist mit Abb. (5.5) (untere Reihe) klar herausgearbeitet. Diese Beispiele verdeutlichen, dass ein zunehmender Anteil der Frisur ein strähnigeres Aussehen verleiht. Darüber hinaus ist die Verteilung belegter Voxel homogener mit dem KDM oder BBM Modell, als ohne Berücksichtigung von Dichteinformationen, vgl. Abb. (5.6).

Die Effizienz unseres Ansatzes hinsichtlich der Verhinderung des Eindringens von Fasern in den Kopfraum ist mit Abb. (5.7) nachgewiesen. Diese zeigt einen Schnitt durch das Kopfmodell mit Frisur entlang der mittleren Sagittalebene. Im Inneren des Kopfmodells sind keine Fasern zu beobachten.

5.5 Diskussion

Im vorangegangenen Kapitel haben wir ein neues, vollautomatisches Verfahren zur Erzeugung realistischer, expliziter Frisuren aus groben Oberflächendaten, wie sie bspw. nach der Laserabtastung von Skulpturen vorliegen, vorgestellt. Dieses Modell kann zur Generierung von Anfangswerten für eine sich anschließende physikalisch basierte Simulation der Haardynamik herangezogen werden. Einige Fragen bleiben allerdings weiterhin unbeantwortet, z.B. jene nach einer sinnvollen Abschätzung der Haarlängen, sowie der Hauptflussrichtungen direkt aus der Oberflächenstruktur. Eine Lösung dieser Probleme würde die Zahl der durch den Benutzer zu handhabenden Parameter weiter reduzieren.

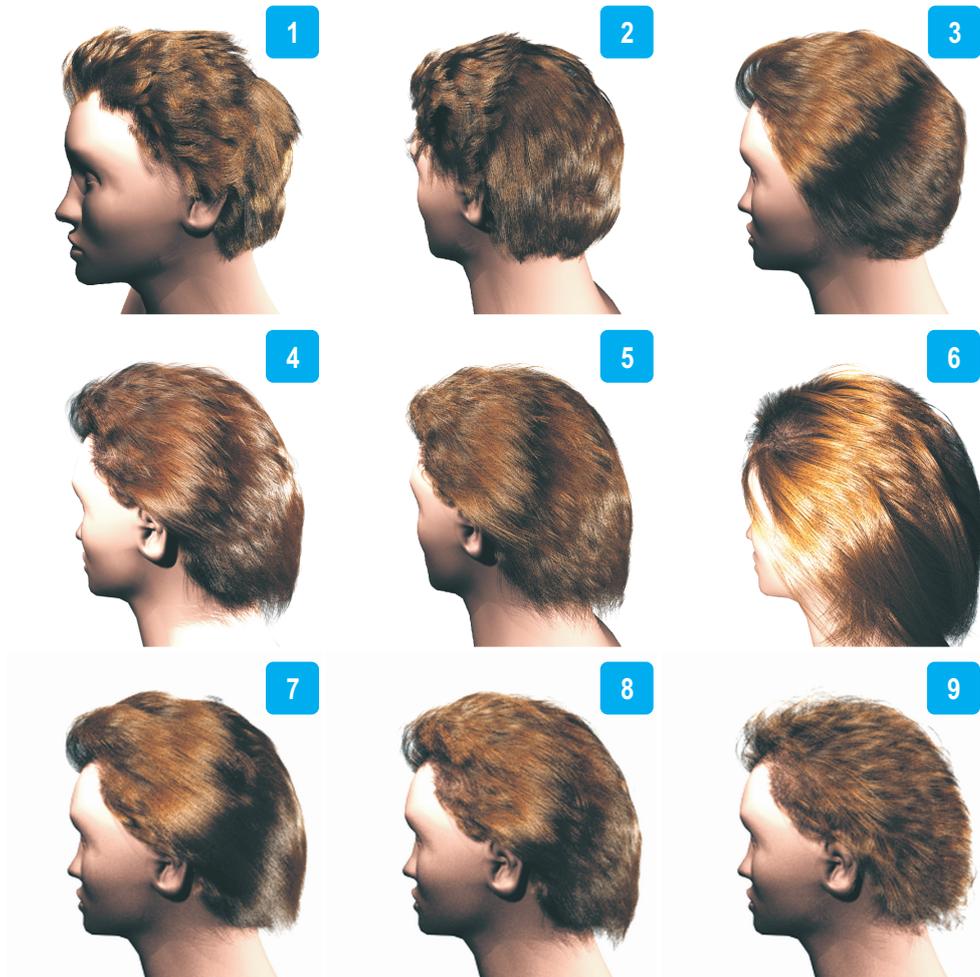


Abbildung 5.5: Die Referenzbeispiele aus unterschiedlichen Perspektiven. 1) und 2): Die aus dem Beethovenvolumen hervorgegangene Frisur aus unterschiedlichen Ansichten. 3): Dieses Modell wurde mit einem größeren Gewichtungparameter für den Ähnlichkeitsterm und größeren Haarlängen erhalten. Die Frisur wirkt insgesamt geglättet. 4) und 5): Die aus dem Poservolumen hervorgegangene Frisur aus unterschiedlichen Ansichten. 6) Dasselbe Modell mit vergrößerten Haarlängen. 7), 8) und 9): Die mit dem Poservolumen erzeugte Frisur: Die Beiträge des KDM wurden hier verändert. Eine Vergrößerung des Beitrages des Dichtekontrollterms resultiert in einem strähnigeren Gesamteindruck.

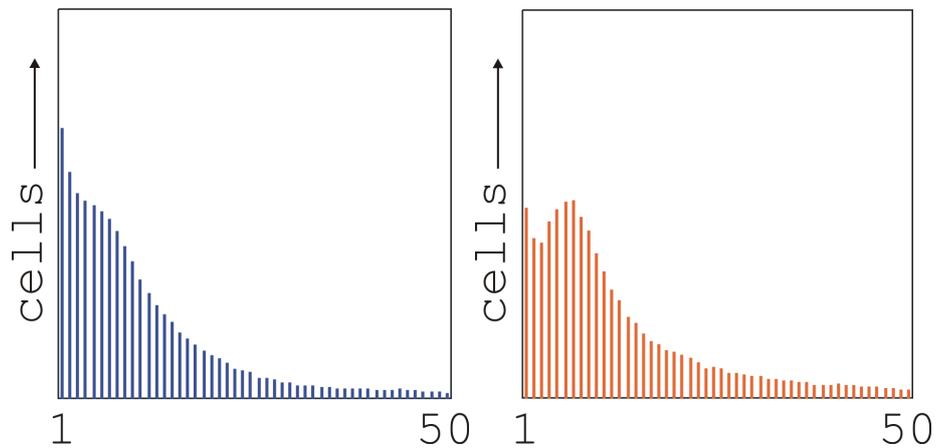


Abbildung 5.6: Histogramm der Dichteverteilung, mit dem KDM (links) und ohne Einbeziehung eines Dichtemodells (rechts). Abszisse: Anzahl der Segmente (1-50) pro Zelle. Ordinate: Anzahl der Zellen.

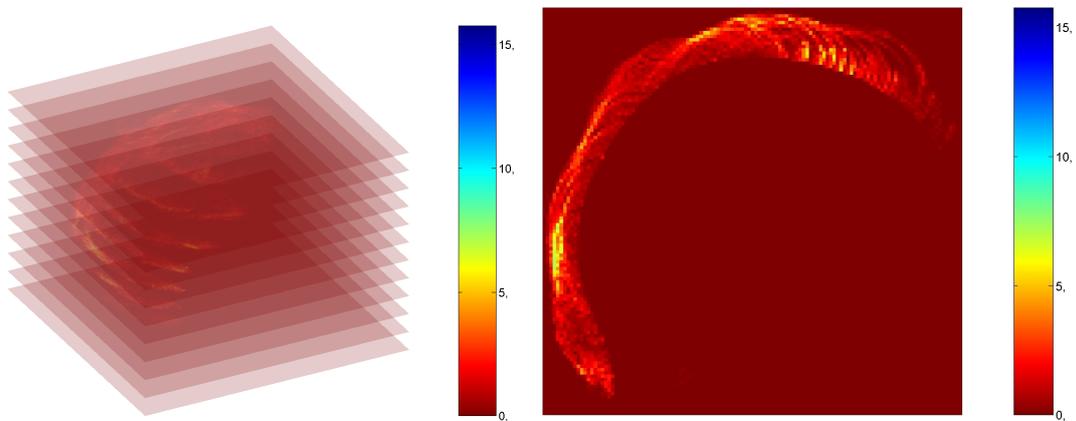


Abbildung 5.7: Links: Mehrere Schnitte durch das Modell. Rechts: Das Bild zeigt die Haardichteverteilung innerhalb des Haarvolumens auf der Basis des KDM. Der Schnitt erfolgte entlang der mittleren Sagittalebene. Man beachte, dass die Haardichte homogen im Bereich des Haarvolumens und null innerhalb des Kopfes ist.

6 Photorealistische Darstellung von Faseransammlungen

Alle in den vorangegangenen Kapiteln vorgestellten Algorithmen operieren auf abstrakten Kurvenrepräsentationen. Die Kluft zwischen abstrakter Repräsentation und photorealistische Darstellung wird überbrückt durch die physikalisch motivierte Simulation des Lichttransportes an der Faser unter wohl definierten Beleuchtungsbedingungen.

6.1 Einleitung

Die Zahl der Beiträge auf dem Gebiet der photorealistischen Darstellung von Faseransammlungen ist überschaubar. Neben Kajiyas und Kays [Kajiya & Kay, 1989] phänomenologischem Ansatz - ein lineares Highlight senkrecht zum Verlauf der Fasern - und den Deep Shadowmaps von Lokovic und Veach [Lokovic & Veach, 2000] hat sich besonders das Fernfeldmodell von Marschner et al. [Marschner et al., 2003] hervorgetan, das man zurecht als Meilenstein auf diesem Gebiet bezeichnen kann. Das Konzept beruht hier, unter Annahme hinreichend weit entfernter Lichtquellen bzw. Betrachter, auf der konsequenten Analyse des Strahlengangs an der Haarfaser. Die Theorie der Lichtstreuung an dielektrischen Zylindern mit ideal glatter Oberfläche rechtfertigt eine getrennte Betrachtung der Streuteile normal zur Faser und senkrecht dazu. Normalstreuung, auch azimuthale Streuung, lässt sich dann auf Basis der Theorie dielektrischer Kreise abhandeln. Die longitudinale Streuung hingegen beschränkt sich, zumindest der Theorie nach, auf den spekularen Kegel. Allerdings erfährt dieses Prinzip bei Haarfasern eine Aufweichung,



Abbildung 6.1:
Nahaufnahme blonder Haare. Bei den hellen Bereichen handelt es sich um Kaustiken (©Universität Bonn).

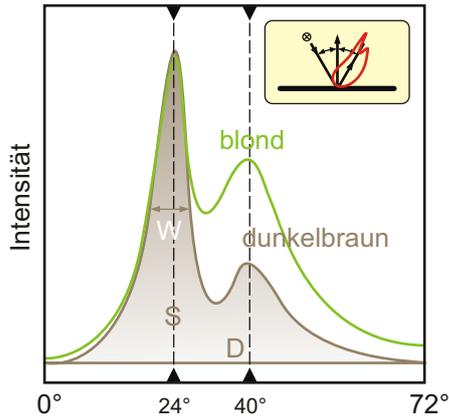


Abbildung 6.2: Goniophotometrische Untersuchungen [Stamm et al., 1977a, Stamm et al., 1977b, Guiolet et al., 1987, Bustard & Smith, 1990, Reich & Robbins, 1993] einzelner blonder und dunkelbrauner Haarfasern. Die Faser wird in axialer Richtung (Ansatz-Spitze) unter einem Winkel von 30° gemessen gegen die Lotrechte angestrahlt. Das Maximum ist um etwa 6° zum Ansatz hin (24°) verschoben. Ein weiteres lokales Maximum, das von der Haarfarbe abhängt, folgt etwa 16° rechts davon, bei 40° .

die ihre Ursache in den morphologischen Eigenheiten der Haarkeratinstruktur findet, zum einen in der Cuticula bzw. der typisch geschuppten Oberfläche und zum anderen in der Inhomogenität der Haarmatrix selbst. Diese bewirkt eine Verschiebung oder Verzerrung des spekularen Streukegels. Marschner et al. [Marschner et al., 2003] werden diesem Umstand durch Substitution des δ -Funktionenmodells mit skalierbaren, energierhaltenden Einheitsintegralen gerecht. Neben dem wird ein praktisches Modell zur Handhabung von Kaustiken (vgl. Abb. (6.1)) und elliptischer Querschnitte entwickelt.

Nicht weniger beachtenswert sind die Beiträge von Zinke [Zinke & Weber, 2007], [Zinke, 2008] auf diesem Gebiet, der die Theorie der Lichtstreuung an Haaren auf ein solides mathematisches Fundament stellt.

Das im Folgenden vorgestellte Nahfeldmodell ist im Wesentlichen eine Generalisierung des Ansatzes von Marschner et al. [Marschner et al., 2003]. Deshalb rekapitulieren wir in den nachfolgenden Abschnitten in stark geraffter Form alle wesentlichen Merkmale dieses Ansatzes, verwenden dabei aber eine an unser Problem angepasste Notation und stellen einige Fehler richtig. Zudem zeigen wir die Bedeutung der indirekten Beleuchtung für die photorealistische Darstellung heller Haartypen auf.

6.2 Streuphänomene bei Haarfasern

Die systematische Untersuchung der Lichtstreuung an natürlichen Haarfasern hat gezeigt, dass sich die morphologischen Eigenheiten des Haarkeratins offenbar auch im Streuverhalten manifestieren.

Die Bestimmung der Streueigenschaften von Haarfasern erfolgt unter Anwendung goniophotometrischer Techniken [Stamm et al., 1977a]. Hierbei wird die Faser in einer planaren Anordnung unter einem bestimmten Winkel beleuchtet und die Intensität des reflektierten Lichtes als Funktion des Winkels gemessen. Das Maximum dieser Funktion ist der spekulare Anteil der Reflexion. Ein gewisser Teil des Lichtes, der diffuse Anteil, wird dagegen in unterschiedliche Richtungen gestreut. Ursachen für die Streuung sind die Rauigkeit der Haaroberfläche sowie Inhomogenitäten der Haarmatrix.

Goniophotometrische Untersuchungen [Stamm et al., 1977a, Stamm et al., 1977b,

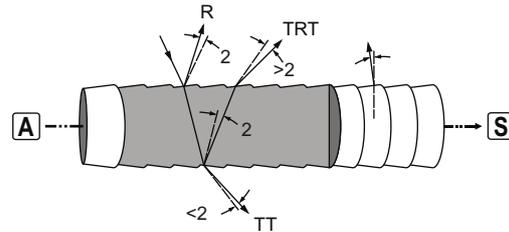


Abbildung 6.3: Schematisierter Längsschnitt durch eine Haarfaser mit Strahlengängen (A: Ansatz, S: Spitze).

Guiolet et al., 1987, Bustard & Smith, 1990, Reich & Robbins, 1993] an einzelnen Haarfasern haben gezeigt, dass das Maximum der spekularen Reflexion bei natürlichen Haare um etwa 6° gegenüber dem von Kunstfasern verschoben ist, wenn solche Fasern in axialer Richtung (Ansatz-Spitze) in einen bestimmten Einfallswinkel (hier 30° gemessen gegen die Lotrechte) angestrahlt werden, vgl. Abb. (6.2). Das ist zunächst erstaunlich, weil die Theorie für Einfalls- und Ausfallswinkel gleiche Werte vorhersagt. Der Grund für die Verschiebung liegt in den morphologischen Gegebenheiten des Haarkeratins. Die geschuppte Struktur der Cuticula induziert eine mehr oder weniger regelmäßige Oberflächenrauigkeit. Die Schuppen schließen einen Winkel von etwa 3° mit der Haarachse ein. Dieser Winkel variiert entlang der Haarachse, mit leicht abnehmender Tendenz zur Spitze hin [Guiolet et al., 1987].

Interessanterweise lässt sich ein weiteres lokales Intensitätsmaximum, dessen Wert von der Haarfarbe abhängt, etwa 16° rechts des globalen Maximums lokalisieren [Stamm et al., 1977b]. Bei braunen Haaren fällt dieses Maximum, bedingt durch die größere Absorption, kleiner aus, als bei blonden Haaren. Bei schwarzen Haaren existiert in der Regel kein zweites Maximum. Beim Beleuchten der Haarfaser in entgegengesetzter Richtung (Spitze-Ansatz) ist das Maximum der Streufunktion erwartungsgemäß um 6° (hier 36°) in die entgegengesetzte Richtung verschoben. Der Kurvenverlauf ist quasi gespiegelt, die Intensitäten weichen leicht voneinander ab [Stamm et al., 1977b].

Die Farbe des spekularen Anteils ist weiß, die des diffusen Maximums entspricht der Haarfarbe [Stamm et al., 1977a]. Die spekulare Reflexion ist ein Oberflächenphänomen, während die diffusen Anteile der Reflexion von der inneren Struktur bzw. von der Haarfarbe abhängen. Werden Haare in einer Schicht exakt parallel angeordnet, wird der diffuse Anteil der Reflexion für diese Anordnung minimal. In diesem Zusammenhang taucht in der kosmetischen Literatur immer wieder der Begriff des Haarglanzes (engl.: *luster*) auf. Glänzendes Haar ist aus Sicht des Konsumenten eine der wünschenswertesten kosmetischen Eigenschaften. Um so verständlicher sind die Bemühungen der kosmetischen Industrie, die physikalischen Ursachen des Phänomens zu verstehen und zu quantifizieren. Allgemein gilt, dass der Glanz mit zunehmendem spiegelnden Anteil größer und mit zunehmendem diffusen Anteil kleiner wird. Robbins und Reich [Reich & Robbins, 1993] schlagen folgende Beziehung zur Ermittlung des Haarglanzes vor: $L = S/DW(1/2)$. D ist der Wert der diffusen Reflexion, der durch Abtrennen der Basisreflexion aus der Gesamtkurve erhalten wird¹. S ist das Integral der spiegelnden Reflexion ohne den diffusen Anteil, $W(1/2)$ ist die Breite des Maximumpeaks bei halber Intensität. Diese Gleichung gilt nur, solange ein diffuser Anteil vorhanden ist. Das ist bei Haaren strukturell bedingt immer der Fall.

¹Verbindung der Streulichtintensitäten bei 0° und 75°

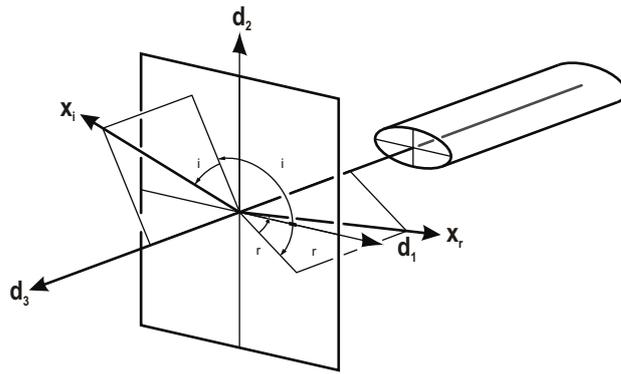


Abbildung 6.4: Zur Einführung des Inklinationswinkels θ bzw. des Azimutalwinkels φ .

Die Messungen von [Stamm et al., 1977a] beschränken sich allerdings auf koplanare Fälle, in denen Lichtquelle, Detektor und Haarfaser in einer Ebene liegen. Zudem bleibt der Winkel der Lichtquelle fix. Marschner et al. [Marschner et al., 2003] haben diese Messungen auf das volle Spektrum möglicher Raumrichtungen erweitert. Neben der Tatsache, dass die Messungen von [Stamm et al., 1977a] bestätigt werden, findet man auch eine starke VorwärtsstreuKomponente, die rückwärtig angeleuchtete Haare mit hellen Pigmenten sehr hell erscheinen lässt. Darüber hinaus lässt sich feststellen, dass das von der Haarfarbe abhängige Intensitätsmaximum zwei helle Glanzpunkte einschließt, deren Örtlichkeiten vom Einfallswinkel abhängen. Je geringer der mit der Faserachse eingeschlossene Winkel ist, desto näher liegen diese Punkte an der Einfallsebene.

6.3 Fernfeldmodell für Haare

Der Verlauf des Lichtstrahls kann nach Stamm et al. [Stamm et al., 1977a] durch ein stark vereinfachtes Modell beschrieben werden, vgl. Abb. (6.3). Bei jedem Auftreffen auf die Grenzschicht wird ein Teil des einfallenden Lichtes reflektiert und ein anderer passiert die Grenzschicht (Transmission) unter Änderung seiner Richtung. Drei charakteristische StreuKomponenten lassen sich anhand des vom Lichtstrahl zurückgelegten Weges unterscheiden: 1) R: Einfache Reflexion an der Oberfläche; 2) TT: Eintritt in das Medium (1. Transmission), Austritt aus dem Medium (2. Transmission); 3) TRT: Eintritt in das Medium (1. Transmission), interne Reflexion, Austritt aus dem Medium (2. Transmission). Der Anteil des an der Grenzschicht reflektierten bzw. transmittierten Lichtes kann über die FRESNELSchen Formeln abgeschätzt werden. Die Richtungsänderung des Strahls beim Passieren der Grenzschicht zwischen zwei Medien unterschiedlicher Brechzahl folgt aus dem bekannten Gesetz von SNELLIUS zu $n_1/n_2 = \sin \gamma / \sin \alpha$ und beträgt für Luft $n_1 \approx 1.0$ und für Keratin $n_2 \approx 1,55$. Die Winkel α bzw. γ messen den Winkel des einfallenden bzw. gebrochenen Strahls gegen das Lot.

6.3.1 Die Streufunktion

Sinn der Streufunktion ist die Vorhersagbarkeit der in eine bestimmte Raumrichtung von der Faser abgestrahlten und damit mess- oder wahrnehmbaren Lichtintensität unter bekannten Beleuchtungsbedingungen. Zur Herleitung der Streufunktion ist die Wahl eines festen, hier ortho-normalen, Bezugssystems ($\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$) auf der Faser sinnvoll, vgl. Abb. (6.4). Dabei spannen \mathbf{d}_1

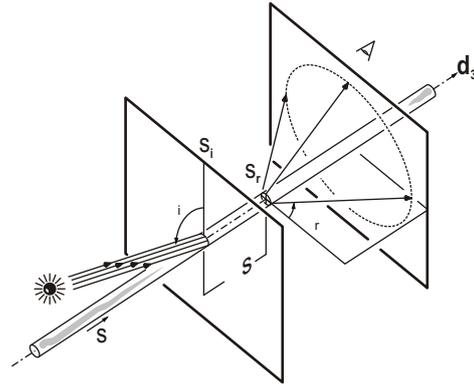


Abbildung 6.5: Bündel paralleler Strahlen werden am perfekten dielektrischen Zylinder in Form eines Kegels gestreut.

und \mathbf{d}_2 die Normalebene auf und \mathbf{d}_3 entspreche der Tangente. Die Richtung des einfallenden Strahls sei, unter Verwendung von Kugelkoordinaten, $\mathbf{x}_i = (\theta_i, \varphi_i)$ und die des austretenden Strahls $\mathbf{x}_r = (\theta_r, \varphi_r)$. θ misst die Inklination in Bezug auf die Normalebene (0 liege in der Normalebene, $\pi/2$ entspreche der Tangente) und φ ist der Azimuthalwinkel, der um die Haarachse dreht (0 entspreche \mathbf{d}_1 , $\pi/2$ entspreche \mathbf{d}_2). Neben dem werden folgende Winkeldefinitionen gebraucht: Differenzwinkel $\theta_d = (\theta_r - \theta_i)/2$, relativer Azimuthalwinkel $\Delta\varphi = \varphi_r - \varphi_i$, Halbwinkel $\theta_h = (\theta_i + \theta_r)/2$ und $\varphi_h = (\varphi_i + \varphi_r)/2$.

Auf der Basis dieser Definitionen führt man in Analogie zur BRDF für Flächen eine bidirektionale Streufunktion

$$S(\mathbf{x}_i, \mathbf{x}_r) = \frac{dL_r(\mathbf{x}_r)}{dE_i(\mathbf{x}_i)} \quad (6.1)$$

für Fasern ein. $dE_i(\mathbf{x}_i)$ ist die differentielle Kurvenbestrahlungsstärke (Leistung pro dl) und $dL_r(\mathbf{x}_r)$ die differentielle Kurvenintensität (Energie pro dl , also ein infinitesimaler Energiefluss), die von der Faser gestreut wird. Differentiell bedeutet, dass sich die Größen auf ein infinitesimal dünnes Faserscheibchen der Dicke dl beziehen. Die Bestrahlungsstärke ist proportional zur Leuchtdichte, $dE_i(\mathbf{x}_i) = 2RL_i(\mathbf{x}_i) \cos \theta_i d\mathbf{x}_i$. Aufintegrieren über alle einfallenden Richtungen der Kugel liefert die gesuchte Kurvenintensität an der Stelle (θ_r, φ_r) ,

$$L_r(\theta_r, \varphi_r) = 2R \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} S(\theta_i, \varphi_i; \theta_r, \varphi_r) L_i(\theta_i, \varphi_i) \cos \theta_i d\theta_i d\varphi_i. \quad (6.2)$$

Gesucht ist die vierdimensionale Streufunktion $S(\theta_i, \varphi_i; \theta_r, \varphi_r)$. Beachtet man die symmetrischen Verhältnisse am Zylinder, lässt sich S als Produkt lediglich zweier zweidimensionaler Funktionen ausdrücken: Eine von φ abhängige Normalkomponente N und einen von θ abhängigen Longitudinalanteil M . Diese Faktorisierung lässt sich aus der Theorie der Streuung an perfekt glatten Zylindern rechtfertigen. Folgende zwei Beobachtungen sind dabei von elementarer Bedeutung. 1) Ein unter dem Winkel $90 - \theta_i$ zur Zylinderachse eintretender Strahl verlässt diesen unter demselben Winkel, unabhängig von der Anzahl an Brechungen und Reflexionen, die er im Innern des Zylinders erfährt. Deshalb beschreiben die Streurichtungen eines Bündels paralleler Strahlen mit Einfallsrichtung \mathbf{x}_i die Oberfläche eines Kegels, dessen Zentrum auf der Zylinderachse liegt und der die Richtung $-\mathbf{x}_i$ einschließt, vgl. Abb. (6.5). Das gestreute Licht wird also nur dann wahrgenommen, wenn für den Augstrahl gilt, $\theta_r = -\theta_i$.

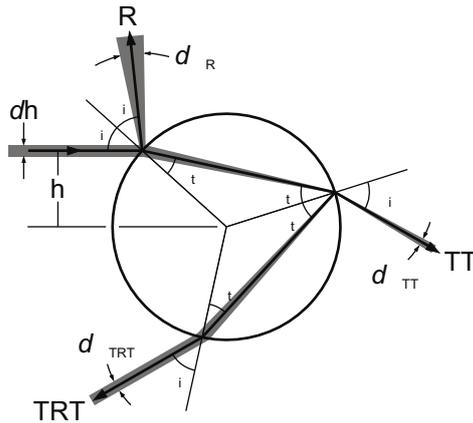


Abbildung 6.6: Komponenten der Streufunktion, R , TT und TRT am dielektrischen Kreis.

2) Die Analyse des von φ_r abhängigen Streuteils vereinfacht sich durch Projektion der einfallenden bzw. ausgehenden Richtungen auf die Normalebene. Nach BRAVAIS behält das Brechungsgesetz von SNELLIUS seine Gültigkeit, wenn man die Brechzahl n des Mediums durch eine entsprechend modifizierte Version $n'(n, \theta) > n$ ersetzt. Durch die azimuthale Unabhängigkeit von n' reicht die Betrachtung des zweidimensionalen Falles völlig zur Beschreibung des dreidimensionalen Problems aus.

Die vorläufige Streufunktion als Produkt des Anteils N normal und des Anteils M longitudinal zur Faser ist deshalb

$$S(\theta_i, \varphi_i; \theta_r, \varphi_r) = M(\theta_i, \theta_r)N(n'(\theta_d); \varphi_i, \varphi_r) / \cos \theta_d^2, \quad (6.3)$$

mit der longitudinalen Streufunktion $M = \delta(\theta_r + \theta_i)$. Der Ausdruck für die Normalkomponente N selbst folgt wegen Beobachtung 2) aus einer weitergehenden Analyse der Lichtstreuung am dielektrischen Kreis, vgl. Abb. (6.4). Man beachte, dass N wegen der alleinigen Abhängigkeit vom relativen Azimuthalwinkel $\Delta\varphi = \varphi_r - \varphi_i$ eindimensional ist. Der Faktor $\cos \theta_d^{-2}$ folgt unmittelbar aus der projektiven Natur der Normalkomponente.

6.3.2 Normalstreuung - Streuung am dielektrischen Kreis

Wir entwickeln im Folgenden den Ausdruck der Normalstreuung am dielektrischen Kreis. Sei $h \in] -1, 1[$ der Abstand eines einfallenden Strahls vom Mittelpunkt des Kreises, vgl. Abb. (6.6). Der Winkel des einfallenden Strahls, gemessen gegen den Radiusvektor am Auftreffpunkt, sei $\sin \gamma_i = h$, der des gebrochenen Strahls $n' \sin \gamma_t = h$. Der Differenzwinkel zwischen ausgehender und einfallender Richtung $\Delta\varphi = \varphi_r - \varphi_i$ folgt daraus,

$$\begin{aligned} \Delta\varphi(p, h) &= 2p\gamma_t - 2\gamma_i + p\pi \\ &= 2p \arcsin\left(\frac{h}{n'}\right) - 2 \arcsin h + p\pi, \end{aligned} \quad (6.4)$$

wie man leicht anhand von Abb. (6.4) nachvollziehen kann; p gibt die Anzahl der inneren Pfadsegmente an, wobei für die einzelnen Komponenten gilt: R : $p = 0$, TT : $p = 1$ und TRT : $p = 2$. Umgekehrt lassen sich die Pfade $\{\rho_r\} = h(p, \Delta\varphi)$, die zur Streuung in einer bestimmten Richtung

$\Delta\varphi^*$ beitragen, durch eine Nullstellensuche in der Funktion $\Delta\varphi(p, h) - \Delta\varphi^* = 0$ bestimmen. Für die R- und TT-Komponente existiert genau eine, für die TRT-Komponente hingegen eine oder drei Nullstellen bzw. interne Pfade. Der Übergang von einer zu drei Nullstellen tritt genau bei $d\Delta\varphi/dh = 0$ bzw. $h^2 = (4 - (n')^2)/3$ auf. Aus der Symmetrie des Ausdrucks folgt die Existenz zweier symmetrischer Extrema.

Wenn man annimmt, dass ein Strahl der Dicke dh nach Reflexion an oder Austritt aus der Faser mit dem Winkel $d\Delta\varphi$ divergiert, vgl. Abb. (6.4), folgt die Intensität des gestreuten Lichtes unter der Bestrahlungsstärke $E(h)$ aus der Funktion $\Delta\varphi(p, h)$ als

$$L(p, h) = \frac{1}{2} A(p, h) T(p, h) \left| \frac{d}{dh} \Delta\varphi(p, h) \right|^{-1} E(h), \quad (6.5)$$

oder für die Streuung am dielektrischen Kreis

$$S(p, h) = \frac{1}{2} A(p, h) T(p, h) D(p, h), \quad (6.6)$$

mit $D(p, h) = |d(\Delta\varphi(p, h))/dh|^{-1}$. An der Stelle $d\Delta\varphi/dh = 0$ nimmt die Intensitätsverteilung unendlich große Werte an. Wegen der Symmetrie der Nullstelle gibt es genau zwei solche Singularitäten, so genannte Kaustiken, welche sich in Form zweier heller Glanzpunkte manifestieren. Mit steigendem Einfallswinkel θ_i rücken diese immer näher an die Normalebene heran, um schließlich bei Erreichen eines Grenzwertes von zwei durch die Brechzahl $n'(n, \theta)$ zu koinzidieren.

Genau genommen erfährt der Lichtstrahl bei der Interaktion mit der Grenzschicht und beim Durchlaufen des Mediums auch eine Intensitätsabschwächung. Unter Beachtung der FRESNEL-Reflektion, die den jeweils an der Grenzfläche reflektierten Anteil des Lichtes angibt, hat man folgende Terme für die Abschwächung anzusetzen²,

$$A(0, h) = F(n, \gamma_i), \quad (6.7)$$

$$A(p, h) = (1 - F(n, \gamma_i)) (1 - F(1/n, \gamma_t)) F(1/n, \gamma_t)^{p-1}. \quad (6.8)$$

Für die TT-Komponente sind zwei Transmissionsterme $(1 - F(n, \gamma))$ bzw. $(1 - F(1/n, \gamma))$ anzusetzen, bei der zweiten Transmission jedoch mit reziproker Brechzahl, weil vom Medium mit höherer spezifischer Dichte in das mit geringerer Dichte gebrochen wird. Der Ausdruck für die TRT-Komponente enthält zusätzlich einen Term $F(1/n, \gamma_t)$, welcher der internen Reflexion an der Rückwand der Faser gerecht wird.

Unter Annahme einer exponentiell mit der im Medium zurückgelegten Wegstrecke abklingenden Abschwächungsfunktion

$$T(p, h) = \exp(-\sigma_a s_M(h))^p = \exp\left(-2\sigma_a R \sqrt{1 - \left(\frac{h}{n'}\right)^2}\right)^p, \quad (6.9)$$

mit der Pfadsegmentlänge $s_M(h) = 2R\sqrt{1 - (h/n')^2}$ ³ und dem Abschwächungskoeffizienten σ_a , bleiben auch die volumetrischen Streueffekte nicht länger unberücksichtigt. Was hier zunächst

²Die Angaben in [Marschner et al., 2003] diesbezüglich sind nicht korrekt.

³Die Angabe in [Marschner et al., 2003] muss korrekt lauten: $\sqrt{2 + 2 \cos(2\gamma_t)}$. Das lässt sich zum o.a. Ausdruck vereinfachen.

nur für den zweidimensionalen Fall gilt, also wenn der Winkel des einfallenden Strahls gegen die Normalebene null ist, darf nach dem Gesetz von BRAVAIS auch auf den dreidimensionalen Fall angewendet werden, wenn Brechzahlen und Abschwächungskoeffizienten eine entsprechende Modifikation erfahren. Dazu substituiert man in Gl. (6.8) alle FRESNEL-Terme $F(n, \cdot)$ und $F(1/n, \cdot)$ durch die generalisierten Formen $F(n', n'', \cdot)$ bzw. $F(1/n', 1/n'', \cdot)$, mit $n'(\theta_i) = \sqrt{n^2 - \sin^2 \theta_i} / \cos \theta_i$ und $n''(\theta_i) = n^2 \cos^2 \theta_i / \sqrt{n^2 - \sin^2 \theta_i}$. Im selben Sinne modifiziert man σ_a in Gl. (6.9) durch Multiplikation mit der Verlängerung der Wegstrecke um $1/\sin \theta_i$ im dreidimensionalen Fall, $\sigma'_a = \sigma_a / \sin \theta_i$.

Die analytischen Ausdrücke für die Normalstreuung nehmen damit die folgende Form an,

$$N(\Delta\varphi) = \sum_p N_p(p, \Delta\varphi), \quad (6.10)$$

$$N_p(p, \Delta\varphi) = \frac{1}{2} \sum_{r=1}^N A(p, \rho_r(p, \Delta\varphi)) T(p, \rho_r(p, \Delta\varphi)) D(p, \rho_r(p, \Delta\varphi)), \quad (6.11)$$

wobei $\rho_r(p, \Delta\varphi^*)$ die r -te Nullstelle der Funktion $\Delta\varphi(p, h)$ an der Stelle $\Delta\varphi^*$ ist und $N = 3$, falls $p = 2$ und $n' < 2$, sonst $N = 1$.

Die Kaustiken der TRT-Komponente erzeugen an der Stelle φ_c der Streufunktion eine Singularität, so dass eine Auswertung der entsprechenden Formel alleine nicht ausreichend ist. Die Kaustiken sind vielmehr zu entfernen und durch eine angemessene Funktion zu ersetzen, so dass die Energie erhalten bleibt.

Die Ausdrücke für die Normalstreuung gelten genau genommen nur für kreisrunde Querschnitte. Allerdings ist der Einfluss selbst geringer Exzentrizität auf die TRT-Komponente nicht vernachlässigbar. Die Herleitung äquivalenter Formeln für elliptische Querschnitte ist i.d.R. nicht möglich. Marschner et al. [Marschner et al., 2003] schlagen deshalb eine entsprechende Anpassung der Brechzahl vor. Wenn φ_h an einer der Hauptachsen der Ellipse ausgerichtet ist, folgt die Änderung der TRT-Komponente in etwa $n_1^* = 2(n-1)a^2 - n + 2$. Für alle Werte dazwischen interpoliert man sinusförmig,

$$n^*(\varphi_h) = \frac{1}{2} ((n_1^* + n_2^*) + \cos(2\varphi_h)(n_1^* - n_2^*)), \quad (6.12)$$

mit $n_2^* = 2(n-1)/a^2 - n + 2$, wobei a die Exzentrizität darstellt. Die modifizierte Brechzahl n' berechnet man dann für die TRT-Komponente aus n^* .

6.3.3 Einfluss der Cuticula - Longitudinale Streuung

Zur Erweiterung des Modells ideal glatter Zylinder bzw. zur Berücksichtigung aller Effekte, die zu einer Abweichung vom idealen Verhalten führen, wie Oberflächenrauheit und Inhomogenität des Mediums, werden longitudinale Streufunktionen hergeleitet. Die mit Gl. (6.3) eingeführte longitudinale Streufunktion M drückt die DIRAC-Funktion $\delta(\theta_i + \theta_r)$ aus, welche die Streuung auf den Modellfall oder spekularen Kegel ($\theta_i = -\theta_r$) beschränkt. Dieses Prinzip wird in der Praxis durch die Inhomogenität des Mediums und volumetrische Streuung aufgeweicht. Für die Funktion M_p setzt man dann vereinfachend eine GAUSS-Normalverteilung $g(\beta, x)$ mit der Standardabweichung β und Mittelpunkt an der Stelle θ_h an.

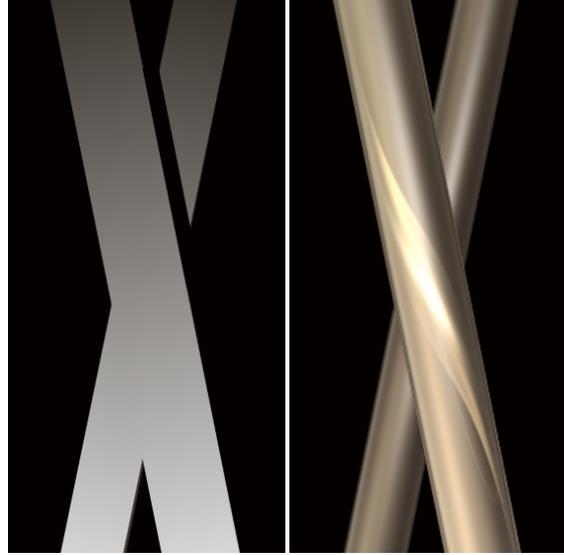


Abbildung 6.7: Das Bild verdeutlicht den Effekt eines Nahfeldmodells bei Nahaufnahmen. Dasselbe Bild wurde dazu einmal mit dem Fernfeldmodell nach Marschner [Marschner et al., 2003] und dem Nahfeldansatz berechnet. Ganz offensichtlich fehlen alle Details in der Fernfelddarstellung, die Farbe bleibt konstant über den Querschnitt.

Zudem bedingt die Cuticula durch ihre Struktur eine Änderung der Interaktionswinkel an der Grenzschicht. Bei der R-Komponente finden wir infolgedessen eine Verschiebung um 2α in Richtung Ansatz, bei der TT-Komponente um maximal 2α und bei der TRT-Komponente um mehr als 2α zur Spitze hin, vgl. Abb. (6.3). Die Abhängigkeit der Verschiebungen von θ_i und $\Delta\varphi$ wird hier vernachlässigt. Den Einfluss der Cuticula approximiert man durch die Verschiebung der Normalverteilungsfunktion um den Winkel α_p , so dass wir folgende Ausdrücke für die longitudinale Streuung erhalten,

$$M_p(\theta_h) = g(\beta_p; \theta_h - \alpha_p). \quad (6.13)$$

Daraus folgt die eigentliche Fernfeldstreuungsfunktion für die Lichtstreuung an Fasern mit nicht glatter Oberfläche, moderat elliptischem Querschnitt und hinreichend weit entferntem Betrachter bzw. hinreichend weit entfernter Lichtquelle zu

$$\begin{aligned} S(\varphi_i, \theta_i, \varphi_r, \theta_r) = & M_R(\theta_h) N_R(n'(n, \theta_d), \Delta\varphi) / \cos^2 \theta_d + \\ & M_{TT}(\theta_h) N_{TT}(n'(n, \theta_d), \Delta\varphi) / \cos^2 \theta_d + \\ & M_{TRT}(\theta_h) N_{TRT}(n'(n^*(\varphi_h), \theta_d), \Delta\varphi) / \cos^2 \theta_d. \end{aligned} \quad (6.14)$$

6.4 Entwurf eines Nahfeldmodells

Der Ansatz von Marschner et al. [Marschner et al., 2003] betrachtet die vom Querschnitt gestreute Intensität als von der lokalen Fasergeometrie unabhängig. Das manifestiert sich einer konstanten Farbe über den Querschnitt, vgl. Abb. (6.7). Für einen hinreichend weit entfernten Betrachter mag die Annahme gerechtfertigt sein. Für unsere Beispiele benötigen wir jedoch oft Nahaufnahmen von Faseransammlungen, um bspw. lokale Kontaktverhältnisse zu verdeutlichen. In solchen Fällen ist die Annahme eines weit entfernten Betrachters nicht mehr zulässig. Deshalb skizzieren wir im Folgenden den ersten Versuch eines adäquaten Nahfeldmodells, das in einem späteren Beitrag von Zinke [Zinke & Weber, 2007] eine mathematische Fundierung erhalten hat.

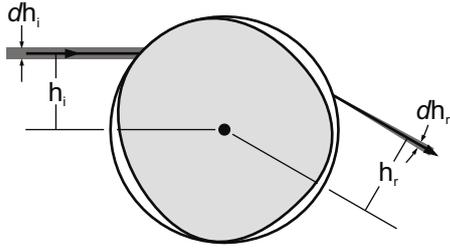


Abbildung 6.8: Schematische Zeichnung zum Nahfeldmodell. Die Parametrisierung erfolgt am Umkreis des lokalen Faserquerschnitts.

Bei der Entwicklung des Nahfeldmodells folgen wir im Wesentlichen dem oben vorgestellten und bereits an unsere Notation angepassten Ansatz von Marschner et al. [Marschner et al., 2003]. D.h., dass sich auch die Streufunktion für das Nahfeldmodell als Superposition der Produkte aus Normal- und Longitudinalanteil der drei Hauptstreuanteile formulieren lässt. Normal- und Longitudinalanteil bedürfen allerdings einer Anpassung an die neue Problemstellung.

Im Nahfeldmodell berücksichtigen wir die lokale Fasergeometrie durch Parametrisierung der Ein- und Austrittsorte des Lichtstrahls am Umkreis. Neben dem Parameter $h_i = h$ führen wir dazu $h_r \in]-1, 1[$ ein, der den Abstand des austretenden Strahls vom Mittelpunkt des Umkreises angibt. Die im Medium entlang der Faserachse zurückgelegte Wegstrecke sei $\Delta s = s_r - s_i$, vgl. Abb. (6.5). Die Streufunktion für das Nahfeldmodell ist demnach siebendimensional,

$$S(\mathbf{x}_i, h_i, \mathbf{x}_r, h_r, \Delta s) = \frac{dL_r(\mathbf{x}_r, h_r, s_r)}{dE_i(\mathbf{x}_i, h_i, s_i)}. \quad (6.15)$$

Nach dem LAMBERTSchen Kosinusetz für diffuse Flächen ist die Bestrahlungsstärke proportional zum Kosinus des Einfallswinkels,

$$dE_i(\mathbf{x}_i, h_i, s_i) = L_i(\mathbf{x}_i, h_i, s_i) \cos(\mu_i(\theta_i, \varphi_i, h_i)) dh_i ds_i. \quad (6.16)$$

Aufintegrieren über alle Dimensionen liefert die gesuchte Leuchtdichte

$$L_r(\theta_r, \varphi_r, h_r, s_r) = 2R \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} \int_{-1}^{+1} \int_{s_i}^{s_r} S(\theta_i, \varphi_i, h_i, \theta_r, \varphi_r, h_r, \Delta s) L_i(\theta_i, \varphi_i, h_i, s_i) \cos(\mu_i(\theta_i, \varphi_i, h_i)) d\theta_i d\varphi_i dh_i ds_i. \quad (6.17)$$

Während die Parametrisierung am Umkreis stattfindet, wird die eingeschlossene lokale Fasergeometrie durch eine spezifische Funktion $\mu_i(\theta_i, \varphi_i, h_i)$ berücksichtigt. Der Einfachheit halber wählen wir hier den Kreisquerschnitt, mit

$$\cos \mu_i = \cos \gamma_i \cos \theta_i = \sqrt{1 - h_i^2} \cos \theta_i. \quad (6.18)$$

Ein Strahl der Dicke dh_i divergiert nach der Interaktion mit der Grenzschicht in ein Bündel von Strahlen der Dicke dh_r , vgl. Abb. (6.8). Wenn man die Approximation $dh_r \approx d\Delta\varphi$ zulässt, erhält man für die Strahldichte

$$D^{\text{NF}}(p, h_i, h_r) \approx \left| \frac{dh_r}{dh_i} \right|^{-1} \approx \left| \frac{d}{dh_i} \Delta\varphi(p, h_i) \right|^{-1} = D(p, h_i). \quad (6.19)$$

Die Streufunktion kann auch beim Nahfeldmodell als Superposition der ersten drei Streukomponenten dargestellt werden. Die Normalanteile setzen sich neben den schon bekannten Termen für die Strahldichte, die FRESNEL-Reflektion und die Abschwächung zusätzlich aus zwei Deltafunktionen der Art $\delta(h_i + h_r)$ und $\delta(\varphi_r - \varphi_i - \Delta\varphi(p, h))$ zusammen. Die Normalkomponente ist also nur dann von Null verschieden, wenn ein- und ausfallender Strahl denselben Abstand zum Mittelpunkt des Umkreises haben und der Differenzwinkel $\Delta\varphi$ gleich $\Delta\varphi(p, h)$ ist,

$$N_p(p, \Delta\varphi, h_i, h_r) = \frac{1}{2} \sum_{r=1}^N A(p, \rho_r(p, \Delta\varphi)) T(p, \rho_r(p, \Delta\varphi)) D(p, \rho_r(p, \Delta\varphi)) \delta(h_i + h_r) \delta(\varphi_r - \varphi_i - \Delta\varphi(p, h_i)). \quad (6.20)$$

Wie bereits oben angedeutet, wird das Prinzip der δ -Funktionen durch Inhomogenitäten des Mediums aufgeweicht, so dass diese durch entsprechende größenhaltende Einheitsintegrale wie bspw. der GAUSS-Funktion ersetzt werden müssen. Die Kaustiken der TRT-Komponente entfernt man mit der von Marschner et al. [Marschner et al., 2003] vorgestellten Methode.

Die longitudinalen Streufunktionen M müssen um einen Term zur Berücksichtigung der Wegstrecke Δs , die der Strahl im Innern des Zylinders zurücklegt, ergänzt werden. Ihr Beitrag ist nur dann von Null verschieden, wenn für den Punkt s_i auf der Zylinderachse gilt,

$$s_i = \begin{cases} s_r & \text{R} \\ s_r + s_M \tan \theta_t \operatorname{sgn} \theta_i & \text{TT} \\ s_r + 2s_M \tan \theta_t \operatorname{sgn} \theta_i & \text{TRT}, \end{cases} \quad (6.21)$$

mit der Absorptionslänge $s_M(h) = 2R\sqrt{1 - (h/n')^2}$. Die vollständigen Ausdrücke für die longitudinalen Streufunktionen sind demnach,

$$M_R(h) = \delta(\theta_i + \theta_r) \delta(\Delta s) / \theta_i^2, \quad (6.22)$$

$$M_{\text{TT}}(h) = \delta(\theta_i + \theta_r) \delta(\Delta s + 2R\sqrt{1 - (h/n')^2} \tan \theta_t \operatorname{sgn} \theta_i) / \theta_i^2, \quad (6.23)$$

$$M_{\text{TRT}}(h) = \delta(\theta_i + \theta_r) \delta(\Delta s + 4R\sqrt{1 - (h/n')^2} \tan \theta_t \operatorname{sgn} \theta_i) / \theta_i^2. \quad (6.24)$$

Auch hier ersetzen wir alle δ -Funktionen durch entsprechende Gaussverteilungen. Den Winkel θ_i in M substituiert man durch θ_d .

6.4.1 Einfluss indirekter Beleuchtung

Zur Darstellung heller Haartypen wie hellem Braun, Rot oder Blond ist die Simulation der direkten Beleuchtung alleine nicht ausreichend. Besonders blondes Haar zeichnet sich dadurch aus, dass das an den Fasern gestreute Licht andere Fasern illuminiert. Dieses Phänomen der Mehrfachstreuung bezeichnet man als indirekte Beleuchtung, weil Fasern indirekt zu aktiven Lichtquellen werden. In diesen Fällen scheint die Annahme einer hinreichend weit entfernten Lichtquelle nicht mehr gerechtfertigt, so dass der Einsatz eines adäquaten Nahfeldmodells bessere Ergebnisse erwarten lässt. Die erforderliche Ordnung der Mehrfachstreuung hängt primär von der Art der Pigmente in den Haaren bzw. von den Absorptionseigenschaften (Haarfarbe) ab. Den Ein-

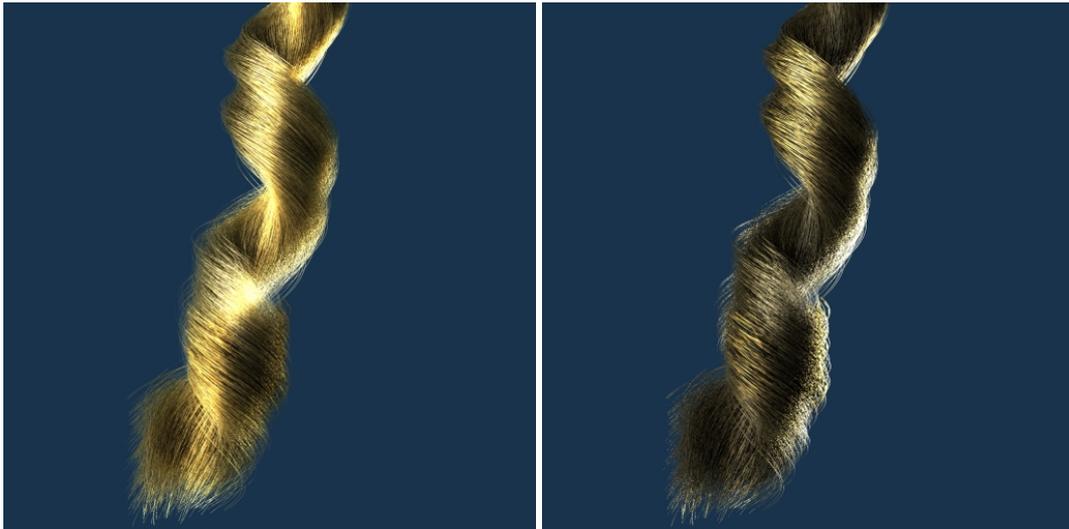


Abbildung 6.9: Links: Die Strähne wurde mit dem Nahfeldmodell unter Beachtung der Haar-Haar-Streuung bis zur Ordnung vier berechnet. Rechts: Dieselbe Strähne, allerdings ohne Haar-Haar-Streuung.

fluss der indirekten Beleuchtung bei blondem Haar wird besonders in Abb. (6.9) deutlich. Das linke Bild wurde mit dem Nahfeldmodell unter Beachtung der Mehrfachstreuung bis zur Ordnung vier, das rechte hingegen ohne Mehrfachstreuung berechnet. Der Leser möge sich sein eigenes Urteil bilden. Wie sich die Haarfarbe mit der Ordnung der Mehrfachstreuung entwickelt, zeigt hingegen Abb. (6.10). Die Streuung erster Ordnung, also das Einbeziehen von einmalig an den Fasern gestreutem Licht in den Illuminationsprozess, bringt den Sprung von nicht identifizierbarer, dunkel wirkender Haarfarbe zum Blond. Die Simulation der Streuung höherer Ordnung zeitigt dann weitere wahrnehmbare, wenn auch deutlich abgeschwächte Aufhellungseffekte und zwar bis zur Ordnung drei oder vier. Abb. (6.11) verdeutlicht den Einfluss der Mehrfachstreuung an einer komplexen Frisur.

6.5 Diskussion

Im vorliegenden Kapitel wurde ein systematischer Streuformalismus für die Lichtstreuung an Humanhaar entwickelt und die Signifikanz der Mehrfachstreuung bei der Darstellung primär heller Haartypen aufgezeigt. Unser Nahfeldansatz stellt damit eine Generalisierung des Fernfeldmodells von Marschner et al. [Marschner et al., 2003] dar. Neben der hochwertigen Detaildarstellung bei Nahaufnahmen erlaubt unser Ansatz die Verwendung nahezu beliebiger Faserquerschnitte.

Eine der Basiseinsichten unserer Untersuchungen ist, dass die Art des erforderlichen Renderingmodells von der darzustellenden Haarfarbe abhängt. Während für schwarzes Haar Kajiyas und Kays [Kajiya & Kay, 1989] phänomenologischer Ansatz befriedigende Ergebnisse liefert, ist für die photorealistische Darstellung brauner Haare bereits eine detaillierte Analyse des Lichttransportes an der Haarfaser unumgänglich, vgl. Marschner et al. [Marschner et al., 2003]. Helle Haartypen, und hier insbesondere helles Blond, sind erst unter Berücksichtigung der Mehrfach-



Abbildung 6.10: Einfluss der Ordnung der Mehrfachstreuung auf die Haarfarbe, Ordnung 0 bis 3.

streuung realisierbar. Dadurch ist jedoch noch nichts über die Umsetzung des letzten Faktors gesagt. Unterschiedliche Ansätze sind denkbar. Während für die Realisierung der Mehrfachstreuung durch Monte-Carlo Path Tracing selbst bei Einsatz von Importance Sampling noch ein recht hoher Preis zu zahlen ist, beschreiten aktuelle Ansätze, welche die Signifikanz der Mehrfachstreuung bei Haaren erkannt haben, neue Wege, z.B. durch Einsatz von volumetrischem Photonemapping [Moon & Marschner, 2006], [Zinke & Weber, 2006] oder Kugelflächenfunktionen [Moon et al., 2008] zur Abschätzung der indirekten Beleuchtung aus einer bestimmten Richtung. Allerdings konnten Zinke et al. [Zinke et al., 2008] zeigen, dass sich der Einfluss der indirekten Beleuchtung sehr effizient statistisch abschätzen lässt. Unter Einsatz moderner Graphikhardware ist die Berechnung der Mehrfachstreuung dann selbst bei komplexen Frisurmodellen mühelos in Echtzeit möglich.



Abbildung 6.11: Darstellung einer komplexen Frisur (Beethoven) mit dem Nahfeldmodell, mit (links) und ohne (rechts) Mehrfachstreuung.

7 Kollisionserkennung für 1D-Strukturen

7.1 Einleitung

Bei der physikalisch basierten Simulation von Objekten sehen wir uns unweigerlich mit einer immanenten Eigenschaft der Natur konfrontiert: Zwei Materiepunkte können nicht zur selben Zeit denselben Ort besetzen. Ist diese Eigenschaft in der Natur quasi von selbst erfüllt, muss sie auf dem Computer mit teilweise nicht unerheblichem Aufwand erzwungen werden. Dieses Gebiet wird in der physikalisch basierten Simulation, einem Teilgebiet der Computergrafik, als Kollisionsbehandlung bezeichnet.

Die Kollisionserkennung ist ein integraler Bestandteil dieses Prozesses bei der physikalisch basierten Simulation. Bei der Simulation komplexer Faseranordnungen mit hoher Packungsdichte tun sich zwei Hauptprobleme auf: die Fasern können miteinander in Wechselwirkung treten oder mit Objekten der Umgebung. Beide Aspekte erfordern unterschiedliche Ansätze zur Erkennung etwaiger Kollisionen. Für die Interaktion der Fasern untereinander setzen wir Hüllkörperhierarchien ein, für die Faser-Umgebungs-Kollisionen hingegen Distanzfelder.

Faseranordnungen mit hoher Packungsdichte liegen mit Haarsträhnen oder Kabelbündeln vor. Eine Faser oder Filament ist eine eindimensionale deformierbare Struktur und kann durch eine verbundene Abfolge von Zylindern mit Polkappen repräsentiert werden, die wir als Segmente bezeichnen. Der Durchmesser ist eine Funktion der Bogenlänge. Ihm kommt bei der Kollisionserkennung eine gewichtige Rolle zu, obwohl er im Vergleich zur Länge der Struktur vernachlässigbar erscheint.

Die Wahrscheinlichkeit für das Auftreten von Kollisionen innerhalb dicht gepackter Faseranordnungen ist sehr hoch (sonst könnten solche Strukturen ihre Volumen nicht aufrechterhalten).

7.2 Detektion von Faser-Umgebungs-Kollisionen

Für die schnelle Kollisionserkennung von Körpern mit einer nicht deformierbaren Umgebung eignen sich Distanzfelder besonders gut. Distanzfelder haben im Bereich der Computergraphik eine weite Verbreitung gefunden, u.a. für die Objektmetamorphose [Breen et al., 2001], zur beschleunigten Darstellung von Volumendaten, zur Bewegungsplanung und zur Abschätzung der Eindringtiefe bei der Kollision von Objekten im Bereich der physikalisch basierten Simulation [Fisher & Lin, 2001].

Zur Berechnung von Distanzfeldern existieren Standardverfahren. Jones und Satherley [Jones & Satherley, 2001] setzen Distanztransformationen ein, um voxelisierte Minimalflächen¹

¹Entspricht der Minimalmenge an voxelisierten Informationen, die zur Rekonstruktion der Isofläche o notwendig sind.

durch das Volumen zu propagieren. Von Sethian [Sethian, 1999] wurde bereits 1999 die Klasse der so genannten Level-Set-Methoden im Bereich der Computergraphik eingeführt, die eine Analogie zwischen dem Propagationsproblem und einer sich ausbreitenden Wellenfront herstellen und den Ausbreitungsvorgang durch ein System partieller Differentialgleichungen beschreiben. Voxalisierte Minimalflächen dienen als Anfangswerte für Level-Set-basierte Fast-Marching-Methoden, welche die Differentialgleichungen auf einem Gitter numerisch lösen.

7.2.1 Distanzfelder

Das Distanzfeld einer Fläche \mathcal{S} ist eine skalare Abbildung

$$\Phi : \mathbb{R}^3 \longrightarrow \mathbb{R},$$

mit der Eigenschaft, dass

$$\Phi(\mathbf{p}) = \min_{\mathbf{q} \in \mathcal{S}} \|\mathbf{p} - \mathbf{q}\|, \quad \forall \mathbf{p} \in \mathbb{R}^3. \quad (7.1)$$

Sie liefert für beliebige Punkte $\mathbf{p} \in \mathbb{R}^3$ im Raum den kleinsten Abstand zur Fläche \mathcal{S} zurück. Der Nachteil einer punktbasierten Definition offenbart sich unmittelbar. Da wir nur den Abstand zwischen Punkten und einer Fläche messen können, müssen nicht punkthafte Objekte, die auf Kollision zu testen sind, punktuell abgetastet werden.

Des Weiteren führt man für geschlossene Flächen \mathcal{S} eine binäre Funktion $\text{sgn} : \mathbb{R}^3 \longrightarrow \{-1, 1\}$ ein, die durch ihr Vorzeichen die Lage des Punktes, innen oder außen, indiziert,

$$\text{sgn}(\mathbf{p}) = \begin{cases} 1, & \text{falls } \mathbf{p} \text{ außerhalb,} \\ -1, & \text{falls } \mathbf{p} \text{ innerhalb.} \end{cases}$$

Zusammen mit Gl. (7.1) erhalten wir ein vorzeichenbehaftetes Distanzfeld. Das Konzept der Vorzeichenfunktion macht nur Sinn, wenn die Fläche geschlossen ist. Wir können diese Definition aufweichen, indem der Begriff des Innen und Außens an den Gradienten $\nabla\Phi$ des Distanzfeldes gekoppelt wird,

$$\text{sgn}^+(\mathbf{p}) = \begin{cases} 1, & \text{falls } (\mathbf{p} - \mathbf{q})^\top \nabla\Phi(\mathbf{q}) \geq 0, \\ -1, & \text{falls } (\mathbf{p} - \mathbf{q})^\top \nabla\Phi(\mathbf{q}) < 0. \end{cases}$$

Dazu müssen wir lediglich fordern, dass der Gradient in der Umgebung von \mathbf{q} existiert. Liegt die Fläche in diskreter Form vor, bspw. in Form eines Dreiecksnetzes, so lassen sich approximative Normalen immer berechnen. Die einzige Forderung ist die nach der Stetigkeit des Vektorfeldes $\nabla\Phi$. Das setzt voraus, dass das Dreiecksnetz orientiert ist.

Der Gradient am Punkte \mathbf{q} wird in der Regel bei der Berechnung einer Kollisionsantwort benötigt, weil er in die Richtung weist, in welche die Auflösung zu erfolgen hat, um ein weiteres Eindringen in das Objekt zu verhindern. Der Gradient folgt bspw. direkt aus der Anwendung finiter Vorwärtsdifferenzen als

$$\nabla\Phi(\mathbf{q}) = \left\{ \frac{\Phi(\mathbf{q} + \Delta\mathbf{x}) - \Phi(\mathbf{q})}{\Delta s}, \frac{\Phi(\mathbf{q} + \Delta\mathbf{y}) - \Phi(\mathbf{q})}{\Delta s}, \frac{\Phi(\mathbf{q} + \Delta\mathbf{z}) - \Phi(\mathbf{q})}{\Delta s} \right\},$$

mit der Schrittweite Δs und $\Delta\mathbf{x} = (\Delta x, 0, 0)^\top$ usw.

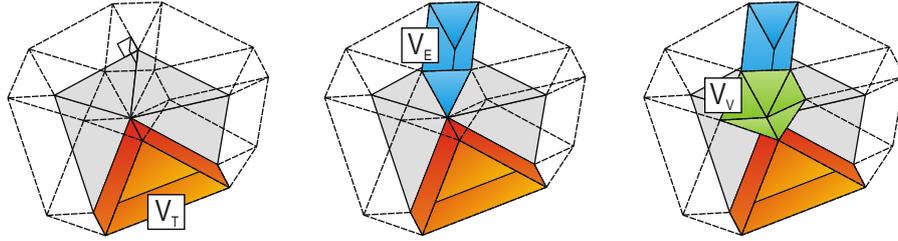


Abbildung 7.1: Voronoi-Regionen der Dreiecke V_T , Kanten V_E und Punkte V_V . Die Voronoi-Regionen werden durch den Schnitt mehrere Halbräume gebildet und sind unbegrenzt. Für die Berechnung des Distanzfeldes beschneidet man sie in einer bestimmten Höhe.

Der Aufwand zur Bestimmung der Distanz zwischen einem Punkt und der Oberfläche hängt offenbar von der Art der Implementierung der Distanzfunktion zusammen. Ein etabliertes Verfahren, von dem auch wir Gebrauch machen werden, ist die Vorabberechnung eines diskreten Distanzfeldes durch Abtastung der Oberflächenabstände von den Punkten eines kartesischen Gitters aus. Für jeden beliebigen Punkt innerhalb dieses Gitters folgt die Distanz zur Oberfläche des Körpers durch trilineare Interpolation der Distanzen an den acht Gitterpunkten der Zelle, in welcher der Punkt lokalisiert ist.

7.2.2 Charakteristikenmethode

Der Vorteil des Verfahrens liegt in der Möglichkeit, den Hauptaufwand auf das Preprocessing des punktuell abgetasteten Distanzfeldes zu verlegen. Die Abtastung erfolgt unter Anwendung der Charakteristikenmethode. Neben der Orientierung des Netzes werden auch Nachbarschaftsinformationen benötigt.

Wenn die Oberfläche in Form eines Dreiecksnetzes vorliegt, werden dazu alle Dreiecke, Kanten und Punkte desselben traversiert und der Abstand jedes dieser Features zu denen am nächsten gelegenen Gitterpunkten berechnet. Zur Reduktion des Aufwandes kann man die Suche der nächsten Punkte auf die VORONOI-Regionen der jeweiligen Features reduzieren. Die VORONOI-Region des Features X_i ist die Menge der Punkte, die näher an X_i liegen, als an allen anderen Features des Dreiecksnetzes. Die Konstruktion der VORONOI-Regionen für die Dreiecke, Kanten und Punkte gestaltet sich denkbar einfach. Für ein Dreieck ist sie als unendlich ausgedehntes Prisma definiert, dessen Querschnitt durch die Dreiecksfläche gebildet wird. Die VORONOI-Region einer Kante hat die Form eines Keils und die des Punktes die Form eines diskreten Kegels, vgl. Abb. (7.1). Die VORONOI-Regionen werden normal zum jeweiligen Feature ab einer bestimmten Höhe $\pm \varepsilon$ beschnitten. ε definiert damit die Dicke des Distanzgürtels. Die Berechnung von Abständen erfolgt nur für diejenigen Gitterpunkte, die innerhalb des Distanzgürtels liegen.

Sei $\mathbf{T}_i = \{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ das aktuelle Dreieck mit der Normalen \mathbf{n}_i und den Kanten $\mathbf{e}_1 = (\mathbf{p}_2 - \mathbf{p}_1)/\|\cdot\|$, $\mathbf{e}_2 = (\mathbf{p}_3 - \mathbf{p}_2)/\|\cdot\|$, $\mathbf{e}_3 = (\mathbf{p}_1 - \mathbf{p}_3)/\|\cdot\|$. Dann kann die VORONOI-Region bzw. das Prisma von \mathbf{T}_i als Menge aller Punkte beschrieben werden, die folgende Bedingungen erfüllen,

$$V_{T_i}^\varepsilon(\mathbf{T}_i, \varepsilon) = \left\{ \mathbf{q} \in \mathbb{R}^3 \mid -\varepsilon < \mathbf{n}^T(\mathbf{q} - \mathbf{p}_1) < \varepsilon \wedge \bigwedge_{\beta} \mathbf{d}_{\beta}^T(\mathbf{q} - \mathbf{p}_{\beta}) > 0 \right\}, \quad (7.2)$$

mit den Seitenflächennormalen $\mathbf{d}_\beta = (\mathbf{n}_i \times \mathbf{e}_\beta) / \|\cdot\|$. Alle Normalen sind hier nach innen gerichtet. Um alle Punkte des Gitters $\mathbf{k}_j \in V_{T_i}^\varepsilon$ zu finden, berechne man die achsenparallele Integerboundingbox von $V_{T_i}^\varepsilon$ auf dem Gitter, traversiere alle Gitterpunkte innerhalb der Box und prüfe per Halbraumtest (implizit mit der Definition nach Gl. (7.2) gegeben) auf Zugehörigkeit zur VORONOI-Region $V_{T_i}^\varepsilon$. Falls der aktuelle Punkt \mathbf{k}_j innerhalb derselben liegt, ist der Abstand zur Dreiecksfläche elementargeometrisch zu bestimmen und mit dem entsprechenden Vorzeichen versehen, am betreffenden Gitterpunkt zu speichern. Das Verfahren entspricht damit einem primitiven Scankonvertierungsalgorithmus. Ein etwas effizienteres Vorgehen bei der Scankonvertierung unter Einbezug von Graphikhardware findet man in [Breen et al., 2001] und [Mauch, 2003].

Mit den VORONOI-Regionen $V_{E_i}^\varepsilon$ und $V_{V_i}^\varepsilon$ der Kanten bzw. Punkte verfährt man genauso. Zur Berechnung der Kantenregionen bedarf es zusätzlich der Normale des benachbarten Dreiecks. Die Region wird nicht einfach gerade abgeschnitten, sondern mit einem Dach versehen, dessen Grat parallel zur Kante $\mathbf{E}_i = \{\mathbf{p}_1, \mathbf{p}_2\}$ mit der Richtung $\mathbf{e}_i = (\mathbf{p}_2 - \mathbf{p}_1) / \|\cdot\|$ und der Höhe $\varepsilon \mathbf{n}_e$ verläuft. $V_{E_i}^\varepsilon$ definieren wir entsprechend als

$$V_{E_i}^\varepsilon(\mathbf{E}_i, \varepsilon) = \left\{ \mathbf{q} \in \mathbb{R}^3 \left| 0 < \mathbf{e}_i^\top (\mathbf{q} - \mathbf{p}_1) < \|\mathbf{p}_2 - \mathbf{p}_1\| \wedge \bigwedge_{\beta=\{1,2\}} \mathbf{d}_\beta^\top (\mathbf{q} - \mathbf{p}_1) > 0 \wedge \bigwedge_{\beta=\{3,4\}} \mathbf{d}_\beta^\top (\mathbf{q} - \mathbf{p}_1 - \varepsilon \mathbf{n}_e) > 0 \right. \right\}, \quad (7.3)$$

mit $\mathbf{d}_1 = (\mathbf{e}_i \times \mathbf{n}_1) / \|\cdot\|$ und $\mathbf{d}_2 = (\mathbf{n}_2 \times \mathbf{e}_i) / \|\cdot\|$, $\mathbf{n}_e = (\mathbf{n}_1 + \mathbf{n}_2) / \|\cdot\|$, $\mathbf{d}_3 = -(\mathbf{n}_1 + \mathbf{n}_e) / \|\cdot\|$ und $\mathbf{d}_4 = -(\mathbf{n}_2 + \mathbf{n}_e) / \|\cdot\|$.

Für die Punktregion sind die Normalen aller inzidenten Dreiecke notwendig. $V_{V_i}^\varepsilon$ definieren wir entsprechend als

$$V_{V_i}^\varepsilon(\mathbf{p}_i, \varepsilon) = \left\{ \mathbf{q} \in \mathbb{R}^3 \left| \bigwedge_{\beta} \mathbf{e}_\beta^\top (\mathbf{q} - \mathbf{p}_i) > 0 \wedge \bigwedge_{\beta} \mathbf{d}_\beta^\top (\mathbf{q} - \mathbf{p}_i - \varepsilon \mathbf{n}_v) > 0 \right. \right\}, \quad (7.4)$$

mit den Kantenrichtungsvektoren $\mathbf{e}_\beta = (\mathbf{T}_\beta \cap \mathbf{T}_{\beta+1}) / \|\cdot\|$, so dass \mathbf{e}_β zum Punkt \mathbf{p}_i zeigt, den Kantennormalen $\mathbf{n}_{e_\beta} = (\mathbf{n}_\beta + \mathbf{n}_{\beta+1}) / \|\cdot\|$, den Deckflächennormalen $\mathbf{d}_\beta = -(\mathbf{n}_{e_\beta} + \mathbf{n}_v) / \|\cdot\|$ und der Vertexnormalen $\mathbf{n}_v = \sum_{\beta} \mathbf{n}_\beta / \|\cdot\|$.

Kanten und Punkt-VORONOI-Regionen existieren nur an „konvexen“ Features. Eine Kante, die außen konvex ist, ist innen konkav und umgekehrt. Bei innenseitig konvexen Kanten, müssen für die Berechnung der Regionen die Normalen der angrenzenden Dreiecke umgedreht werden. Dasselbe gilt für Punktregionen.

Bei benachbarten, nicht koplanaren Dreiecken durchdringen sich die VORONOI-Regionen auf der konkaven Seite notwendigerweise. Einen Sonderfall bilden außerdem Sattelpunkte, also Punkte, an denen mindestens zwei Kanten oberhalb und zwei unterhalb der von der Punktnormalen aufgespannten Ebene liegen. Auch die VORONOI-Regionen der entsprechenden Kanten durchdringen sich hier. Eine Punkt-VORONOI-Region existiert in diesem Fall auf keiner Seite der Ebene, weil jeder Punkt infolge der gegenseitigen Durchdringung von Dreiecks- und Kanten-VORONOI-Regionen immer näher an einem dieser Features liegt, als am gemeinsamen Punkt. Man kann die Durchdringungen bei der Generierung der Regionen mit ins Kalkül ziehen, um

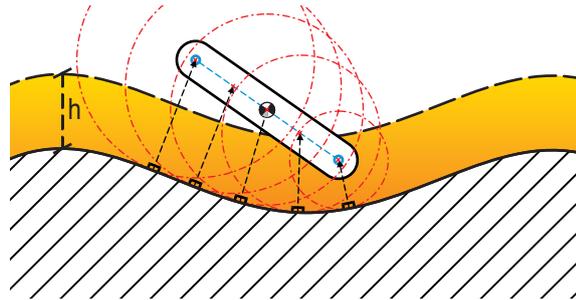


Abbildung 7.2: Die Segmente der Fasern werden bei der Kollisionserkennung gegen die Umgebung punktwise abgetastet.

den Aufwand weiter zu reduzieren. Die Zahl der zu scannenden Punkte verringert sich dadurch. Eine Überschneidung von VORONOI-Regionen stellt auf der anderen Seite kein Problem dar, weil die Punkte innerhalb des Schnittbereiches gegen die erzeugenden Features getestet und dem, mit dem kleinsten Abstand zugeordnet werden.

Mit kartesischen Gittern stößt man bei einer großen Zahl hoch aufgelöster Objekte schnell an die Grenzen des Speicherplatzes. Da wir nur an Abständen innerhalb eines vorgegebenen Distanzgürtels interessiert sind, ist alternativen Methoden wie Oktalbäumen klar der Vorzug zu geben, weil hier je nach Anforderungen lokal adaptiv verfeinert werden kann. Auch sehr fein tesselierte Objekte können sich bei Verwendung von Gittern u.U. als problematisch erweisen, weil zur Abtastung die Auflösung des gesamten Gitters erhöht werden muss.

Es wurde weiter oben schon darauf hingewiesen, dass die Kollisionserkennung von Objekten mittels skalarer Distanzfelder ein punktuell abtasten derselben notwendig macht. Liegt die Geometrie eines Objektes in Form eines Dreiecksnetzes vor, kann man alle Punkte in die Distanzfunktion einsetzen, um das Minimum zu bestimmen. Ein solches vorgehen ist ohne eine vorgeschaltete „broad phase“ zur Eingrenzung potentieller Kandidaten nicht sehr effizient. Im vorliegenden Fall haben wir es mit sehr dünnen Segmenten zu tun, welche die Geometrie der Fasern approximieren. In unserer Implementierung tasten wir die Mittellinie eines jeden Segmentes mit einer vom Benutzer zu spezifizierenden Abtastrate ab und addieren zusätzlich zur Approximation der Segmentdicke den halben Radius dazu, vgl. Abb. (7.2).

7.3 Detektion von Faser-Faser-Kollisionen

Das Ziel dieses Abschnittes ist der systematische Vergleich unterschiedlicher Arten von Hüllkörperhierarchien hinsichtlich ihrer Eignung und Performance bei der Erkennung von Selbstkollisionen im Kontext der faserbasierten Simulation komplexer Anordnungen. Der Begriff der Selbstkollision umfasst sowohl die Kollisionen der Fasern untereinander, als auch die Kollision einer Faser mit sich selbst. Solange man die Glieder einer Kette als unabhängige Entitäten betrachtet (Segmentsuppe), mag hier kein Unterschied bestehen. Sobald wir eine Faser als eindimensionales Kontinuum auffassen, zwischen dessen Punkten gewisse kinematische Gesetzmäßigkeiten herrschen, ergeben sich neue Möglichkeiten für die Erkennung von Selbstkollisionen in solchen Strukturen. Deshalb konzentrieren wir uns neben den konventionellen räumlich angepassten auch auf die topologiebasierten Hüllkörperhierarchien, die unter bestimmten Voraussetzungen eine schnelle Kollisionserkennung und Aktualisierung ermöglichen.

Im Zusammenhang mit den konventionellen Hüllkörperhierarchien stellen wir zudem das

Konzept der parametrischen k -DOPs auf der Basis von GRASSMANN-Räumen vor, der die systematische Generierung von Normalenrichtungen für beliebige k und damit eine empirische Bestimmung optimaler k -Werte für die Kollisionserkennung mittels DOP-Hierarchien erlaubt.

7.3.1 Räumlich und topologisch angepasste Hierarchien

Für die Kollisionserkennung in einer Menge von Objekten unterscheidet man zwei Hauptansätze: Partionierung des Raumes, bspw. durch Gitter, kd-Bäume, BSP-Bäume, Oct-Bäume auf der einen Seite und Hierarchien von Hüllkörpern [Ericson, 2005] auf der anderen. Im Kontext der Haarsimulation finden zur Kollisionserkennung zwischen den einzelnen Fasern oder Clustern expliziter Haarmodelle, sofern man diese nicht völlig ignoriert, bevorzugt Raumpartitionierungsverfahren in Form von Voxelgittern oder Oktalbäumen Verwendung, vgl. Hadap and Thalmann [Hadap & Magnenat-Thalmann, 2001], Anjyo et al. [Anjyo et al., 1992], Plante et al. [Plante et al., 2001, Plante et al., 2002], Chang et al. [Chang et al., 2002], Ward et al. [Ward et al., 2003] sowie Bando et al. [Bando et al., 2003]. Diese zeigen gute Leistung, wobei die Speicheranforderungen allerdings recht hoch ausfallen können. Bei dynamisch verformbaren Objekten werden die teuren Aktualisierungsstrategien der Raumpartitionierung schnell zum Flaschenhals. Hier bieten sich hierarchische Objektrepräsentationen mittels Hüllkörperhierarchien an [Gottschalk et al., 1996, van den Bergen, 1997, Klosowski et al., 1998, Larsson & Akenine-Möller, 2001, Bridson et al., 2002], die traditionell die räumliche Nachbarschaft atomarer Einheiten der Objektoberfläche, z.B. Dreiecke, abbilden. Da diese Nachbarschaft zeitlich veränderlich ist, koppeln neuere Ansätze für die Detektion von Kollisionen in deformierbaren Materialien eine hierarchische Objektrepräsentation an „Invarianten“, wie die Topologie. Die Verwendung von Hüllkörperhierarchien, die über der topologischen Struktur der zugrunde liegenden Geometrie aufgebaut sind, hat einige Vorteile gegenüber traditionellen Ansätzen. Zum einen beinhaltet die topologische Nachbarschaft zweier atomarer Einheiten immer auch deren räumliche Nachbarschaft. Umgekehrt sind zwei räumlich benachbarte atomare Einheiten nicht notwendigerweise auch topologische Nachbarn. Dieser Fall liegt z.B. vor, wenn die beiden Enden einer Faser sich berühren.

Zum anderen ändert sich die räumliche Nachbarschaft der Primitive über die Zeit, wenn das Objekt deformiert wird. Die topologische Struktur hingegen kann sich nicht ändern. Eine Ausnahme bildet Materialfraktur, die wir hier aber nicht betrachten.

Zur Reduktion des Aktualisierungsaufwandes versucht man in der Regel, die Struktur einer räumlich angepassten Hierarchie durch Anpassung der Hüllkörper an die neuen Positionen der atomarer Einheiten über mehrere Zeitschritte aufrechtzuerhalten. Die Rechtfertigung für diese so genannten „träge“ Aktualisierungsstrategien entspringt der Annahme einer zeitlichen Kohärenz aufeinander folgender Zustände. Zurecht weisen Lotan et al. [Lotan et al., 2002] darauf hin, dass bereits kleine lokale Änderungen in eindimensionalen Strukturen große globale Deformationen nach sich ziehen können. Bei der Monte Carlo Simulation von Proteinmolekülen wird eine randomisierte Folge von Gelenkwinkeländerungen erzeugt, die akzeptiert werden, wenn sie das System in einen energetisch günstigeren Zustand überführen. Andernfalls, und dazu zählen auch Selbstkollisionen des Moleküls, wird die Änderung verworfen. Unter diesen Voraussetzungen ist die Annahme zeitlicher Kohärenz nicht mehr aufrechtzuerhalten, so dass „träge“ Aktualisierungsstrategien die Hierarchie in einen ungültigen Zustand überführen können, der eine quadratische

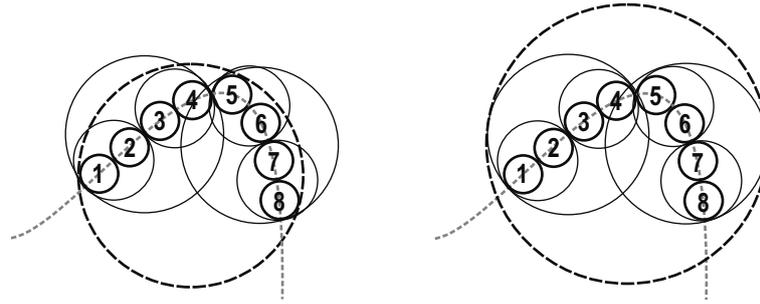


Abbildung 7.3: Topologisch angepasste Hierarchien: direkt \mathcal{H}_{T_D} (rechts) und indirekt einschließende \mathcal{H}_{T_I} (links) Hierarchie einer Faser. Bei einer \mathcal{H}_{T_D} schließen die Hüllkörper die kanonische Teilkette direkt ein.

Anzahl an Kollisionstests nach sich zieht. Zeitliche Kohärenz können wir für unsere Fasern fordern, wenn die bestimmenden Gleichungen dynamisch sind und für eine Folge kleiner zeitlicher Änderungen gelöst werden.

Als Schlüsselbeitrag auf diesem Gebiet, muss die Arbeit von Guibas et al. [Guibas et al., 2002] über die Selbstkollisionserkennung in eindimensionalen, deformierbaren Strukturen mittels einer topologiebasierten Kugelhierarchie gelten. Dabei handelt es sich um einen balancierten Binärbaum, dessen Blätter atomare Teile der Struktur, hier kugelförmige Kettenglieder, umschließen. An jedem inneren Knoten wird die minimale umhüllende Kugel der kanonischen Teilkette vorgehalten, vgl. Abb. (7.3). Im Gegensatz zu den konventionellen räumlich angepassten Hierarchien bilden die eingeschlossenen Primitive ein zusammenhängendes Teilstück der Gesamtstruktur.

Diese Art von Hierarchie, bei der Kugeln an den Knoten die zugehörige Teilgeometrie an den Blättern des Teilbaumes minimal umschließen, wird als *wrapped* oder *direkt einhüllende Hierarchie* (\mathcal{H}_{T_D}) bezeichnet. Auf der anderen Seite steht die *layered* oder *indirekt einhüllende Hierarchie* (\mathcal{H}_{T_I}), bei welcher die Kugeln an jedem Knoten jene der beiden Kindknoten minimal umschließen. Die aus dem unterschiedlichen Anschmiegegrad notwendig resultierenden Größenunterschiede können quantifiziert werden: Der Radius der Kugel am Wurzelknoten der \mathcal{H}_{T_I} beträgt höchstens das $\sqrt{h+1}$ -Fache des Kugelradius an der Wurzel der \mathcal{H}_{T_D} , wobei $h = \lceil \log n \rceil$ die Höhe des balancierten Binärbaumes darstellt, vgl. [Guibas et al., 2002].

Lotan et al. [Lotan et al., 2002] verwenden topologisch angepasste Hierarchien orientierter Quader für die Kollisionserkennung bei der Monte Carlo Simulation langer Proteinmoleküle. Besonders eindrucksvoll sind hier die Kosten der Aktualisierung von $\mathcal{O}(k \log \frac{n}{k})$, wobei k die Zahl der Freiheitsgrade darstellt. Diese werden durch den Einsatz einer separaten Transformationshierarchie erreicht. Insbesondere die Änderungen von nur einem Freiheitsgrad führen auf Aktualisierungskosten von $\mathcal{O}(\log n)$.

In dem Beitrag von Brown et al. [Brown et al., 2003] findet man eine Anpassung topologiebasierter Hierarchien an die Bedürfnisse der Simulation des Knotenbildungsprozesses in Seilen. James und Pai [James & Pai, 2004] haben mit der Einführung des *Bounded Deformation Trees* (*BD-Baum*) die Idee der topologischen Anpassung auf polygonale Modelle übertragen.

Dieses Konzept lässt sich im Prinzip auf andere Hüllkörpergeometrien erweitern.

7.3.2 Subquadratische Kollisionserkennung

Der topologisch basierte Ansatz ist auch unter theoretischen Gesichtspunkten interessant, denn auf der Basis topologischer Hierarchien kann der Beweis erbracht werden, dass der Aufwand für die Kollisionserkennung in kinematischen Ketten, die bestimmten Einschränkungen unterliegen, im \mathbb{R}^3 eine subquadratische obere Schranke von $\mathcal{O}(n^{4/3})$ hat [Guibas et al., 2002, Lotan et al., 2002]. Dabei macht es keinen Unterschied, ob die Hierarchie vom direkt oder indirekt einhüllenden Typ ist [Bereg, 2004]. Ein ähnlicher Beweis wurde von Halperin and Overmars bereits 1994 [Halperin & Overmars, 1994] für deformierbare Kugelketten geführt, bei der die Zahl der Kugeln, die beliebige anderen Kugeln überlappen, durch eine Konstante beschränkt ist. Sie konnten zeigen, dass in diesem Fall die Detektion der Selbstkollisionen in $\mathcal{O}(n)$ Zeit möglich ist. Als Beispiel für Strukturen mit einer solchen inhärenten Eigenschaft wären hier Molekülmodelle (Kalottenmodelle) zu nennen.

In topologisch angepassten OBB-Hierarchien beträgt die maximale Anzahl möglicher überlappender Boxen $\Theta(n^{4/3})$, [Lotan et al., 2002], so dass die Kollisionserkennung hier im schlechtesten Fall $\Theta(n^{4/3})$ Zeit benötigt.

Alle Beweise machen die folgenden Annahmen: Eine Kette bestehe aus einer Folge $\mathcal{B} = \{B_1, B_2, \dots, B_n\}$ von n (geschlossenen) Kugeln im \mathbb{R}^d . Nur benachbarte Kugeln überschneiden sich und keine Kugel ist vollständig in einer anderen enthalten. Daraus lassen sich die folgenden zwei Bedingungen ableiten.

1. **Uniformität:** Es existiert eine Konstante $\rho \geq 1$, so dass das Verhältnis der Radien zweier beliebiger Kugeln im Intervall $[1/\rho, \rho]$ liegt.
2. **Konnektivität:** Zwei beliebige aufeinander folgende Kugeln haben einen Punkt gemeinsam.

7.3.3 Hierarchieaufbau

In unseren Tests betrachten wir die folgenden drei Klassen von Hierarchien:

1. Die kanonische Form der Hüllkörperhierarchie, die wir im Folgenden als *räumlich angepasste Hierarchie* (\mathcal{H}_R) bezeichnen.
2. Die topologisch basierte Hierarchie vom „wrapped“ Typ [Guibas et al., 2002], im Folgenden *direkt einhüllende Hierarchie* (\mathcal{H}_{TD}).
3. Die topologisch basierte Hierarchie vom „layered“ Typ [Guibas et al., 2002, Lotan et al., 2002], im Folgenden *indirekt einhüllende Hierarchie* (\mathcal{H}_{TI}).

All drei Klassen können mit unterschiedlichen Hüllkörpern instanziiert werden. Insbesondere konzentrieren wir uns auf AABBs, OBBs, k -DOPs und Kugeln. Die Hierarchien vom direkten Typ unterliegen einigen Einschränkungen, die später diskutiert werden. Daneben existieren andere Hüllkörpertypen, wie bspw. Quantized Orientation Slabs (QuOSPO's) [Taosong, 1999], Shells [Krishnan et al., 1998b, Krishnan et al., 1998a], Pie Slices (Tetraeder

[Barequet et al., 1996], Zylinder [Bereg, 2004], Zonotopes (Minkowski Summe von Liniensegmenten) [Guibas et al., 2003], etc., die unsere Aufmerksamkeit verdienen würden, die aber eher von untergeordneter Bedeutung sind.

Räumlich angepasste Hierarchie: Die \mathcal{H}_R wird für jeden Hüllkörpertyp in einem Top-down-Prozess konstruiert. Dazu nimmt man eine räumliche Subdivision der Segmentmenge in kleiner werdende Teilmengen vor, bis diese nur noch ein einzelnes Segment enthalten. Das räumlich zu unterteilende Gebiet bildet das minimal einschließende Volumen (MEV) des jeweiligen Hüllkörpertyps über der betrachteten Teilmenge der Segmente. Dieses wird am zugehörigen Knoten der Hierarchie gespeichert. Das Hauptproblem dabei ist die Wahl einer geeigneten Teilungsebene bzw. eines Teilungspunktes, welche zu einem möglichst balancierten Binärbaum führen. Als Heuristik wählen wir den Mittelpunkt der längsten Achse eines Hüllkörpers bzw. die Ebene senkrecht dazu. Der Aufbau des Binärbaums kann dann im ungünstigsten Fall $\mathcal{O}(n^2)$ Zeit beanspruchen, z.B. dann, wenn die Segmente schlecht verteilt sind (bspw. aufeinander liegen), wohingegen bei einer homogenen Verteilung, die wir annehmen, mit $\mathcal{O}(n \log n)$ Aufwand zu rechnen ist. Falls diese Medianteilungsstrategie entlang der längsten Achse versagt, wählen wir nacheinander die zweit bzw. dritt längste Achse. Bei bestimmten Segmentkonfigurationen mag die Medianteilungsstrategie versagen, bspw. wenn die Segmente parallel und ihre Mittelpunkte auf einer Geraden orthogonal zu den Segmentachsen liegen. In diesem Fall ergibt sich der Teilungspunkt durch Projektion des Gesamtschwerpunktes der aktuellen Teilmenge auf die erst (zweit, dritt) längste Achse. Diese Strategie kann nur dann versagen, falls die Schwerpunkte zweier Segmente aufeinander fallen.

Topologisch angepasste Hierarchie: Die topologisch angepassten Hierarchien, \mathcal{H}_{T_D} und \mathcal{H}_{T_I} , erfordern zwei unterschiedliche Strategien zum Aufbau eines balancierten Binärbaums. Ganz in Analogie zur \mathcal{H}_R können wir bei der \mathcal{H}_{T_D} verfahren. Die Konstruktion des Baumes erfolgt in einem Top-down-Prozess, bei dem die Faser rekursiv in zwei gleich große Teile separiert wird. Für jede dieser Teilketten berechnet man das MEV bzw. die aufspannende Basis, vgl. hierzu Abs. (7.3.4). Die Basis besteht aus 1 bis m MEVs an den Blättern des Binärbaums, wobei m je nach Hüllkörpertyp variiert. Bei Kugeln als MEV haben wir 1-4 Basiskugeln, welche das MEV an jedem Knoten aufspannen. Der tatsächliche Aufwand für die Konstruktion der Hierarchie hängt vom Umstand ab, ob ein effizienter Algorithmus zur Berechnung des MEV über einer Menge von Hüllkörpern desselben Typs existiert. Für Kugeln ist dieses Problem nichttrivial. Wir werden auf diese Zusammenhänge bei der Diskussion der einzelnen Hüllkörpertypen näher eingehen.

Die \mathcal{H}_{T_I} hingegen wird in einem Bottom-up-Prozess konstruiert. Dabei berechnen wir an den Blättern des Binärbaums die MEVs der einzelnen Segmente der Faser und arbeiten uns durch paarweise Vereinigung der MEVs an den Kindknoten bis zur Wurzel hoch. Da ein anders gearteter Bottom-up-Konstruktionsprozess hier nicht vorstellbar ist, prägen Lotan et al. [Lotan et al., 2002] für die Ausrichtung der Hüllkörper an den atomaren Einheiten der Kette die Bezeichnung „chain-aligned“, also entlang der Kette ausgerichtet.

Die Anzahl Blattknoten entspricht der Anzahl der Segmente in der Faser. Der Baum ist mindestens eins-balanciert und falls die Gesamtzahl an Segmenten eine Potenz von Zwei darstellt, perfekt balanciert.

Eine Pseudotopologie: Während die \mathcal{H}_R direkt über einer komplexen Faseranordnung aufgebaut werden kann, weil nur die räumliche Nähe ihrer atomaren Einheiten – hier gebildet durch die Segmente der Fasern – Berücksichtigung findet, sind wir bei der Konstruktion der \mathcal{H}_{T_D} bzw.

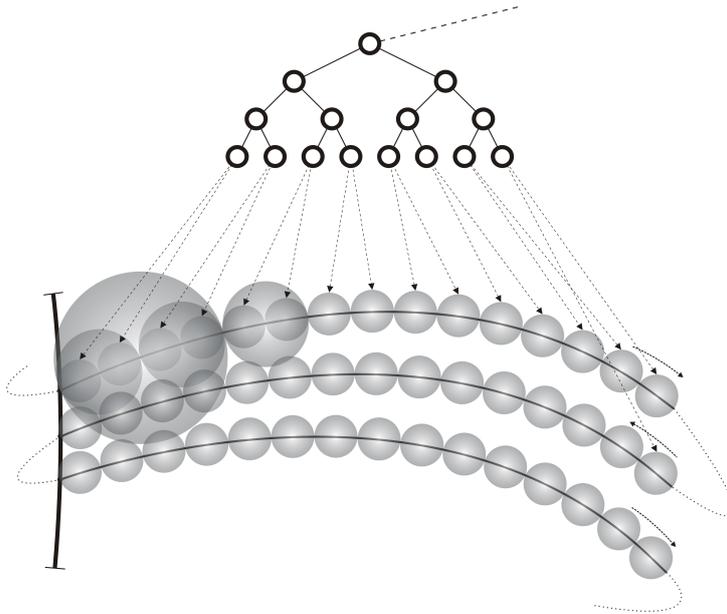


Abbildung 7.4: Zur Konstruktion einer \mathcal{H}_{T_D} bzw. \mathcal{H}_{T_I} über der Fasernanordnung werden alle Fasern durch einen „virtuellen Faden“ zur Gesamtfaser verbunden.

\mathcal{H}_{T_I} an die Topologie der Fasern gebunden. Auf der Basis der $\mathcal{H}_{T_D}/\mathcal{H}_{T_I}$ zweier Fasern lassen sich die Selbstkollisionen innerhalb der Fasern bzw. die Kollisionen zwischen ihnen detektieren. Weil in diesem Sinne keine globale „Gesamtopologie“ über die gesamte Anordnung definiert ist – die Fasern können sich unabhängig von einander bewegen – muss für jede Faser eine eigene Hierarchie erzeugt und diese dann gegen alle anderen auf Kollision getestet werden. Die eindimensionale Struktur der Fasern schließt die Anwendung von Standardverfahren für die so genannte „Broad Phase“ zur Reduktion des damit verbundenen quadratischen Aufwandes von vornherein aus.

Wir nehmen uns hier die Freiheit, eine „Pseudotopologie“ oder eine „Gesamtfaser“ einzuführen und definieren diese als Folge von Fasern, die an ihren Endpunkten paarweise verbunden sind und zwar in der Form, dass das letzte Segment der Faser i mit dem letzten der Faser $i + 1$ und das erste der Faser $i + 1$ mit dem ersten der Faser $i + 2$ usw. verbunden ist. Die Gesamtfaser ergibt sich quasi durch „Einfädeln“ eines virtuellen Fadens durch alle Fasern. Wir bauen dann den Hierarchiebaum \mathcal{H}_B über dieser Folge auf, vgl. Abb. (7.4).

Dieses Vorgehen scheint beliebig. Wenn man aber bedenkt, dass (1) benachbarte Fasern ähnliche geometrische Eigenschaften wie Länge etc. haben, (2) die Fasern in einer Strähne in der Regel durch Kräfte adhäsiver Natur aneinander haften und bei Deformation nicht beliebig weit auseinander scheren und (3) „träge“ Aktualisierungstechniken wie das Refitting einer \mathcal{H}_R ebenfalls zu beliebigen falschen Zuständen führen kann, scheint diese Vorgehensweise für einen Vergleichstest akzeptabel.

In diesem Sinne erfüllt die Gesamtfaser die Uniformitätsbedingung. Der Konnektivitätsanforderung genügt sie hingegen nur indirekt.

7.3.4 Kombinatorische Beschreibung von Hüllkörpern

Während die MEVs einer \mathcal{H}_R immer durch ihre Koordinaten dargestellt werden, bietet sich bei der $\mathcal{H}_{T_D}/\mathcal{H}_{T_I}$ die Möglichkeit, die kinematischen Beziehungen zwischen den atomaren Einheiten zu nutzen. Weil die MEVs immer zusammenhängende Teilstücke einer Faser einschließen, gibt es eine minimale Menge atomarer Einheiten oder Basis-MEVs auf dem kanonischen Teilstück, welche die Ausdehnung des MEVs festlegen. An den Knoten der Hierarchie hält man dann eine entsprechende Liste der Basis-MEVs an den Blättern der Hierarchie vor, welche das Volumen aufspannen. Der Vorteil liegt auf der Hand. Unter Deformation bleibt die Basis stabil oder es tritt ein Wechsel derselben ein, bei dem atomare Einheiten der kanonischen Teilfaser aus dem MEV austreten bzw. andere dafür eintreten. Die aktuelle Basis am Knoten u zum Zeitpunkt t bezeichnen wir mit \mathbb{B}_u^t . Die maximale Zahl der Basis-MEVs hängt dabei von der Art des verwendeten Hüllkörpertyps ab.

7.3.5 Hierarchiebasierte Kollisionserkennung

Gegeben seien die Hierarchieebäume \mathcal{H}_{B_1} und \mathcal{H}_{B_2} . Zum Auffinden aller Kollisionen wird eine simultane Top-down-Traversierung oder Tiefensuche durchgeführt. Das bedeutet, dass die Hüllkörper, beginnend mit den Wurzelknoten der beiden Bäume, paarweise auf Überlappung untersucht werden. Falls eine solche vorliegt, hält man einen der beiden Knoten fest und testet gegen die Hüllkörper an den Kindknoten des anderen. Diese Suche setzt man solange fort, bis ein Blattknoten erreicht ist, wobei in diesem Fall direkt auf Schnitt mit der atomaren Einheit zu testen ist. Kann zwischen zwei Hüllkörpern keine Überlappung festgestellt werden, ist eine Kollision mit atomaren Einheiten an den Blattknoten dieses Teilbaums ausgeschlossen und man fährt mit einem anderen Knoten fort, vgl. Alg. (3).

Die Entscheidungsregel, welche den bei Schnitt zweier Hüllkörper weiter zu traversierenden Teilbaum bestimmt, wird gemeinhin als Split- oder Balancierungsheuristik bezeichnet. Üblicherweise fällt die Wahl auf den Knoten, an dem der Hüllkörper mit dem größeren Volumen hängt. Bei der $\mathcal{H}_{T_D}/\mathcal{H}_{T_I}$ ist hier eine Umdeutung in dem Sinne erforderlich, dass die Wahl auf den Knoten mit den meisten Basis-MEVs fällt. Im speziellen Fall der Kugel- \mathcal{H}_{T_D} garantiert diese Heuristik eine subquadratische Laufzeit [Guibas et al., 2002].

Bei der Selbstkollisionsauffindung in einer einzelnen Faser geht man in analoger Weise vor. Da hier nur ein Hierarchiebaum vorliegt, werden der jeweils linke und rechte Teilbaum an der Wurzel des Baumes \mathcal{H}_B rekursiv auf Überlappungen untersucht.

Beim Kollisionserkennungsprozess werden typischerweise zwei Phasen unterschieden: Die *Broad Phase* und die *Narrow Phase*. In der ersten Phase sollen weit voneinander entfernte Objekte aus dem Erkennungsprozess ausgeschlossen werden. Das wird durch so genannte *Sweep and Prune* Strategien erreicht [Cohen et al., 1995, Coming & Staadt, 2006]. Es ist davon auszugehen, dass durch Umhüllen einer einzelnen Faser einer dicht gepackten Anordnung, auch eine große Zahl weiterer Fasern zumindest teilweise im selben Hüllkörper liegt. Der Standardansatz für Sweep and Prune, wie er bspw. in *I-Collide* [Cohen et al., 1995] verfolgt wird, scheidet deshalb aus. Es bleibt nurmehr, die Gesamtzahl der Segmente als mehr oder weniger in Nachbarschaft stehend zu betrachten und deshalb auf eine gesonderte Broad-Phase zu verzichten.

Algorithmus 3 : Erkennung von Selbstkollisionen in der Faseranordnung auf der Basis des Hierarchiebaums \mathcal{H}_B . s_u und s_v stehen für die Segmente s an den Blattknoten u und v . Die Funktion $\text{adj}(s)$ ermittelt die unmittelbaren Nachbarsegmente von s auf der Faser von s . $l_{u,v}$ bzw. $r_{u,v}$ sind die linken bzw. rechten Kinder der inneren Knoten u bzw. v .

Data : Hierarchiebaum \mathcal{H}_B
 Result : Kollisionsliste C

```

1:  $Q \leftarrow \{(\rho, \rho)\}$ 
2: while  $Q \neq \emptyset$  do
3:    $(u, v) \leftarrow \text{POPFRONT}(Q)$ 
4:   if  $\mathcal{C}_u \cap \mathcal{C}_v \neq \emptyset$  then
5:     if  $\text{leaf}(u) \wedge \text{leaf}(v)$  then
6:       if  $(s_u \notin \{s_v, \text{adj}(s_v)\}) \wedge (s_u \cap s_v \neq \emptyset)$  then
7:          $\text{PUSHBACK}(C, (s_u, s_v))$ 
8:       end
9:     else if  $u = \text{SPLIT}(u, v)$  then
10:       $\text{PUSHBACK}(Q, (l_u, v))$ 
11:       $\text{PUSHBACK}(Q, (r_u, v))$ 
12:    else
13:       $\text{PUSHBACK}(Q, (u, l_v))$ 
14:       $\text{PUSHBACK}(Q, (u, r_v))$ 
15:    end
16:  end
17: end

```

7.3.6 Aufrechterhaltung einer Hierarchie

Vor jeder Kollisionsabfrage muss die Gültigkeit der Hierarchie sichergestellt werden. Die zwei Hierarchietypen $\mathcal{H}_{T_D}/\mathcal{H}_{T_I}$ erfordern unterschiedliche Strategien zur Aufrechterhaltung ihrer Gültigkeit. Bei der \mathcal{H}_{T_I} folgen wir der einfachen Bottom-up-Strategie, die weiter oben beschrieben wurde und traversieren in Nebenreihenfolge (engl.: post-order) unter Aktualisierung der Hüllkörper. An den Blättern werden die Hüllkörper an die Positionen der Segmente angepasst. Alle anderen Hüllkörper folgen dann durch eine Vereinigungsoperation aus den beiden Hüllkörpern an den Kindknoten. Bei der \mathcal{H}_{T_D} sind die Dinge etwas komplizierter. Vor jeder

Algorithmus 4: Top-down-Verifikationskaskade ($\text{CASCADEVERIFICATION}(u)$).

Data : Gültiger Knoten u , dessen Nachkommen bereits verifiziert wurden.
 Result : Aktualisierte Basis \mathbb{B}_u^{t+1}

```

1:  $\mathcal{C}_u^*(t) \leftarrow \text{MEV}(\mathbb{B}_u^t)$ 
2:  $Q \leftarrow \{l_u, r_u\}$ 
3: while  $Q \neq \emptyset$  do
4:    $v \leftarrow \text{POP}(Q)$ 
5:   if  $\mathcal{C}_v(t) \not\subseteq \mathcal{C}_u^*(t)$  then
6:     if leaf( $v$ ) then
7:        $\mathbb{B}_u^{t+1} \leftarrow \text{UPDATE}(u, \mathbb{B}_u^t)$ 
8:     else
9:       PUSH( $Q, l_v$ )
10:      PUSH( $Q, r_v$ )
11:    end
12:  end
13: end
```

Kollisionsabfrage wird die Korrektheit der Hierarchie bzw. die kombinatorischen Beschreibungen ihrer Hüllkörper durch eine *Verifikationskaskade* (engl.: cascade verification) validiert, vgl. [Guibas et al., 2002]. Dazu traversiert man den Baum in Bottom-up-Richtung und prüft an jedem Knoten u , ob die Basis \mathbb{B}_u^t vom letzten Zeitschritt noch aktuell ist. Das ist genau dann der Fall, wenn die Hüllkörper an den Blättern des aktuellen Teilbaums im Hüllkörper \mathcal{C}_u^* des aktuellen Knotens u enthalten sind, vgl. Alg. (4). Das lässt sich direkt oder indirekt testen, indem man den aktuellen Teilbaum in Top-down-Richtung traversiert und prüft, ob der Käfig \mathcal{C}_v am aktuellen Knoten v eine Teilmenge von \mathcal{C}_u^* ist. Falls nicht, setzt man die Prüfung mit den Kindknoten fort. Handelt es sich um einen Blattknoten, bedeutet das, dass $\mathbb{B}_u^{t+1} \neq \mathbb{B}_u^t$ und eine Aktualisierung der kombinatorischen Beschreibung erforderlich ist.

Die ursprünglich für die Kugel- \mathcal{H}_{T_D} entwickelte Verifikationskaskade, ist direkt auf andere Hüllkörperhierarchien übertragbar. Voraussetzung ist dabei lediglich eine kombinatorische Beschreibung der Hüllkörper. Das wirft sofort die Frage nach den Vorteilen einer solchen Aktualisierungsstrategie auf. Die Verifikationskaskade benötigt im schlechtesten Fall $\Theta(n \log n)$ Zeit, wohingegen die einfache Bottom-up-Aktualisierung immer linear ist, bei kleiner Konstante. Zudem würde der Einsatz dieses Verfahrens die Speicherung der kombinatorischen Beschreibung

der Käfige an den inneren Knoten notwendig machen.

Bei der \mathcal{H}_R hingegen ist eine vollständige Aktualisierung recht teuer, weil sie eine rekursive Neupartitionierung der Segmente notwendig macht. Eine als „träge“ Aktualisierungstechnik bekannte Refittingstrategie verfährt analog zur Bottom-up-Strategie bei der \mathcal{H}_{T_1} . Hier gilt es zu beachten, dass ein bloßes Refitting aus o.a. Gründen zu völlig korrupten Hierarchiezuständen führen kann. Eine fällige Neupartitionierung lässt sich nicht dauerhaft umgehen. Die vereinheit-

Algorithmus 5 : Aktualisierung, Traversierung in Post-Order

```

Data : Hierarchiebaum  $\mathcal{H}_B^t$ 
Result : Aktualisierter Hierarchiebaum  $\mathcal{H}_B^{t+1}$ 
1:  $Q \leftarrow \{(\rho, 1)\}$ 
2: while  $Q \neq \emptyset$  do
3:    $\{u, i\} \leftarrow \text{POP}(Q)$ 
4:   if  $(i = 3)$  then
5:     if leaf( $u$ ) then
6:       | UPDATE( $\mathcal{C}_u(s_u)$ )
7:     else if  $\mathcal{H}_B \in \{\mathcal{H}_{T_1}, \mathcal{H}_R\}$  then
8:       |  $\mathcal{C}_u \leftarrow \text{MERGE}(\mathcal{C}_{l_u}, \mathcal{C}_{r_u})$ 
9:     else if  $\mathcal{H}_B = \mathcal{H}_{T_D}$  then
10:      | CASCADEVERIFICATION( $u$ )
11:    end
12:  else
13:    | PUSH( $Q, (u, i + 1)$ )
14:  end
15:  if  $(i = 1) \wedge (!\text{leaf}(u))$  then
16:    | PUSH( $Q, (r_u, i)$ )
17:    | PUSH( $Q, (l_u, i)$ )
18:  end
19: end

```

lichte Aktualisierung ist in Alg. (5) dargestellt.

7.3.7 Hüllkörper

Ein wichtiges Designkriterium für Hierarchien zur Kollisionserkennung ist die Wahl geeigneter Hüllkörper für eine optimale Faserabdeckung. Wir besprechen hier kurz die vier wichtigsten Typen im Hinblick auf ihre Eignung für die \mathcal{H}_R , \mathcal{H}_{T_D} bzw. \mathcal{H}_{T_1} , vgl. Abb. (7.5). Zudem zeigen wir eine Methode zur Parametrisierung von k -DOPs auf.

7.3.7.1 Achsenparallele Quader (AABBs)

Achsenparallele Quader (engl.: *axis aligned bounding boxes (AABBs)*), finden wegen ihrer Einfachheit weiten Einsatz, werden aber i.d.R. durch die effizienteren orientierten Quader (engl.:

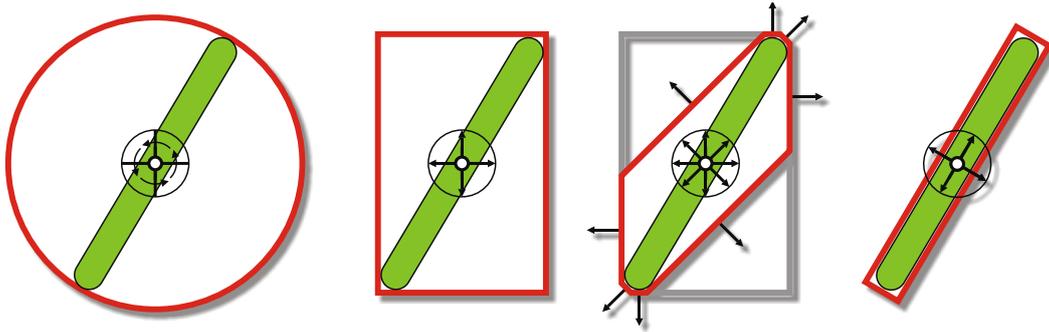


Abbildung 7.5: Die unterschiedlichen Hüllkörper, die wir in Kombination mit den drei Hierarchietypen \mathcal{H}_R , \mathcal{H}_{T_D} bzw. \mathcal{H}_{T_I} verwenden: Kugeln, achsenparallele Quader, k -DOPs und orientierte Quader.

oriented bounding boxes (OBBs)), abgewertet. Die Berechnung der AABB eines einzelnen Segmentes gestaltet sich einfach: man nimmt die beiden Endpunkte, sortiert deren Koordinaten nach der Größe, so dass die erste die linke untere Ecke bzw. die zweite die rechte obere Ecke der Box im kartesischen Koordinatensystem repräsentiert und vergrößert die Box in alle sechs Koordinatenrichtungen um den Radius. Der letzte Schritt rechtfertigt sich aus der Tatsache, dass wir den Zylindern zur Vereinfachung der Berechnung Polkappen verpassen. In Summe sind also lediglich sechs skalare Werte abzuspeichern. Die Seitenflächen der AABB sind immer parallel in Bezug auf das modelleigene Koordinatensystem. Oft existieren in Fasernordnungen bevorzugte globale Richtungen, bspw. wenn alle Fasern parallel zu einer Achse verlaufen. Je parallel ein Segment zu einer Koordinatenrichtung ausgerichtet ist, desto kleiner fällt das Volumen der Box aus.

Im Hinblick auf die beiden Hierarchietypen \mathcal{H}_{T_D} und \mathcal{H}_{T_I} finden wir eine Besonderheit bei der Verwendung von AABBs, der wir durch das folgenden Lemma Ausdruck verleihen,

Lemma 7.3.1 Sei $\mathcal{S} = \{s_1, s_2, \dots, s_n\}$ eine Folge von Segmenten und $\mathcal{M} = \{\text{MEV}_1, \text{MEV}_2, \dots, \text{MEV}_n\}$ mit $\text{MEV}_i = \text{MEV}(s_i)$ die zugehörige Folge von minimal einschließenden Volumina. Dann gilt für beliebige Teilfolgen $s_{i,j} \subseteq \mathcal{S}$ mit $1 \leq i < j \leq n$, dass $\text{MEV}(s_{i,j}) = \text{MEV}(\text{MEV}(s_{i,k}), \text{MEV}(s_{k+1,j}))$, wobei $s_{i,k}$ und $s_{k+1,j}$ beliebige Teilfolgen von $s_{i,j}$ mit $i \leq k \leq j$ sind, falls MEV eine AABB ist.

Das heißt, dass die AABB- \mathcal{H}_{T_D} und die AABB- \mathcal{H}_{T_I} identisch sind, vgl. Abb. (7.6). Der aufwendige Top-down-Konstruktionsprozess der \mathcal{H}_{T_D} kann durch den einfachen Bottom-up-Ansatz ersetzt werden. Der Aufwand für die Konstruktion ist damit linear in der Anzahl der Segmente. Man beachte, dass die versteckten Konstanten sehr klein sind. Eine kombinatorische Beschreibung auf der Basis von AABBs erfordert maximal sechs AABBs.

7.3.7.2 Diskret orientierte Polyeder (k-DOPs)

Während AABBs und Kugeln die einzelnen Segmente der Fasern schlecht approximieren, führt der Einsatz der interessanteren k -DOPS (diskret orientierter Polyeder, engl.: *discrete oriented polytope*) mit $k > 3$ zu enger anliegenden Hüllkörpern und damit zu einer beschleunigten Kollisionserkennung.

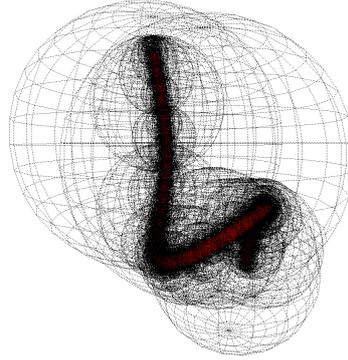
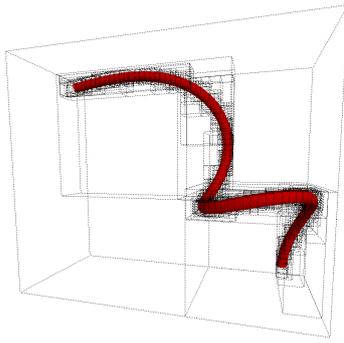


Abbildung 7.6: Hüllkörper der AABB- \mathcal{H}_{T_1} (links) bzw. Kugel- \mathcal{H}_{T_D} (rechts) einer Faser im Vergleich. Man beachte, dass die MEVs bei der Kugel- \mathcal{H}_{T_D} aus den übergeordneten MEVs herausragen.

Ein DOP ist ein konvexes Polyeder und entspricht der Schnittmenge von k Halbräumen, vgl. Abb. (7.5). Jeder Halbraum wird durch eine zugehörige Einheitsnormale $\mathbf{n}_i \in \mathbb{R}^3$ und einen Skalar $d_i \in \mathbb{R}$ definiert,

$$H_i = \{\mathbf{p} \in \mathbb{R}^3 \mid \mathbf{p}^T \mathbf{n}_i \leq d_i\}. \quad (7.5)$$

Gewöhnlich werden paarweise gegenüberliegende Halbraumebenen $\mathbf{n}_{i+k/2} = -\mathbf{n}_i$ gewählt oder so genannte Slabs, so dass wir den k -DOP mathematisch definieren als,

$$\bigcap_{i=1}^k H_i = \{\mathbf{p} \in \mathbb{R}^3 \mid \mathbf{p}^T \mathbf{n}_i \leq d_i\}, \quad i = 1..k. \quad (7.6)$$

Damit sind für den Kollisionstest zwischen zwei k -DOPs $k/2$ Intervall-Überlappungstests nötig. Das Konzept der orientierten Slaps wurde ursprünglich von Kay und Kajiya [Kajiya & Kay, 1989] für Renderingzwecke entwickelt und später von Klosowski et al. [Klosowski et al., 1998] auf das Problem der Kollisionserkennung zwischen nicht deformierbaren polygonalen Modellen angepasst. In ihrem Beitrag konzentrieren sie sich auf die Standardfälle mit $k/2$ Richtungen und $k = \{6, 14, 18, 26\}$.

Im Gegensatz zu existierenden Ansätzen, die sich auf Standardwerte für k beschränken, sind wir an der Frage interessiert, ob für die Kollisionserkennung in Faseranordnungen ein optimales k existiert. Eine sinnvolle Wahl der Richtungen einer variablen Anzahl von Normalen ist kein einfaches Problem und führt uns auf das Gebiet GRASSMANN'Scher Packungsprobleme.

7.3.7.2.1 Parametrische DOPs: Die Bestimmung einer definierten Anzahl von Normalenrichtungen ist eng verwandt mit dem Problem, Punkte auf der Oberfläche einer Kugel so zu verteilen, dass je zwei Punkte sich gegenüberliegen. Ein nahe liegender Ansatz ist die Durchführung einer physikalisch basierten Simulation, welche die Abstände zwischen N Punkten auf der Kugeloberfläche minimiert. Die Bedingungen werden dann bspw. durch Einführen rückstoßender Kräfte erzwungen. Ein ähnliches Problem wurde von [Taosong, 1999] in Verbindung mit dem Quantisierungsproblem im Orientierungsraum bei *QuOSPO*-Hierarchien ² angegangen. Man sucht dabei eine Menge von Normalen, welche die Überschneidungswahrscheinlichkeit der einhüllenden Polyeder in zwei gegebenen Hierarchien direkt nach dem Tumblingprozess³ minimiert. Diese

²Quantized Orientation Slabs with Primary Orientation

³Im Tumblingprozess werden zwei Mengen von Normalen, welche die Polyeder zweier DOP-Hierarchien definieren, aufeinander abgebildet, was normalerweise zu wesentlich größeren Hüllkörpern führt, als eine Neuberechnung.

Wahrscheinlichkeit ist schwer zu berechnen, weil sie sowohl von der zugrunde liegenden Geometrie als auch von deren Position abhängt. Das kann durch Einführung eines geeigneten Ähnlichkeitsmaßes umgangen werden, welches durch die Winkel zwischen den Normalen definiert wird. Die Lösung des Quantisierungsproblems ist eine Menge von N gleichmässig⁴ über die Oberfläche der Einheitskugel verteilten Punkten, vgl. [Rusin, 2001]. Taosong [Taosong, 1999] tesseliert dazu die Kugeloberfläche durch Subdivision eines Oktaeders mit anschließender Projektion der Dreiecksschwerpunkte auf die Kugeloberfläche. Dieses Vorgehen genügt nicht unseren Anforderungen, wonach die Punkte antipodisch sein müssen. Darüber hinaus ist die durch das Subdivisionsverfahren erhaltene Anzahl an Punkten durch $8 \cdot 4^n$ vorgegeben, wobei n die Rekursionstiefe darstellt. Wir möchten eine beliebige Anzahl von antipodischen Punkten generieren, so dass die Strahlen, die durch sie definiert werden, den größtmöglichen Abstand zueinander einnehmen. Ein ähnliche Frage wurde von Conway et al. [Conway et al., 1996] in einer formaleren Art gestellt: Wie müssen N von n -dimensionalen Unterräumen des m -dimensionalen EUKLIDISCHEN Raumes angeordnet werden, so dass jeder die größtmögliche Distanz zu allen anderen Unterräumen einnimmt.

Mathematisch kann das Problem mittels GRASSMANN-Räumen ausgedrückt werden. Der GRASSMANN-Raum $\mathcal{G}(m, n)$ ist die Menge aller n -dimensionalen Unterräume des reellen EUKLIDISCHEN m -dimensionalen Raumes \mathbb{R}^m . Er bildet eine kompakte RIEMANN-Mannigfaltigkeit der Dimension $n(m - n)$. Hier interessieren wir uns insbesondere für den GRASSMANN-Raum $\mathcal{G}(3, 1)$, d.h., die Menge aller Geraden im dreidimensionalen EUKLIDISCHEN Raum. Gemäß Conway et al. [Conway et al., 1996] wird die optimale Packung durch die minimierende chordale Distanz definiert.

Seien $P, Q \in \mathcal{G}(m, n)$ n -Ebenen. Weiterhin seien $\theta_i \in [0, \pi/2]$, $i \in [1, \dots, n]$ die Hauptwinkel zwischen zwei Ebenen, die wie folgt definiert werden können,

$$\cos \theta_i = \max_{u \in P} \max_{v \in Q} uv = u_i v_i, \quad (7.7)$$

mit $i = 1, \dots, n$. Die Vektoren u_i und v_i werden als Hauptvektoren bezeichnet, die zum Paar P und Q gehören. Weiterhin gilt, $uu = vv = 1$, $uu_j = 0$, $vv_j = 0$, ($1 \leq j \leq i - 1$).

Dann ist die *chordale Distanz* gegeben durch,

$$d_c(P, Q) = \left(\sum_{i=1}^n \sin^2 \theta_i \right)^{\frac{1}{2}}. \quad (7.8)$$

Im Gegensatz zu anderen Abstandsmaßen ist das Quadrat der chordalen Distanz überall differenzierbar.

Das führt zu folgender Definition des Packungsproblems: Gegeben seien N n -dimensionale Unterräume des \mathbb{R}^3 , die gepackt werden sollen. Gesucht ist eine Menge von Ebenen $P_1, \dots, P_N \in \mathcal{G}(m, n)$ mit der größtmöglichen chordalen Distanz $\min_{i \neq j} d_c(P_i, P_j)$ zwischen beliebigen Paaren von Ebenen. Das Problem ist wohl definiert, weil $\mathcal{G}(m, n)$ kompakt ist.

Dieses Problem wurde von Conway et al. [Conway et al., 1996] als Optimierungsproblem neu formuliert. Mittels eines Musteroptimierers wurden für einen großen Wertebereich von (m, n)

⁴Der Begriff ist hier intuitiv zu fassen.

mutmaßlich optimale Packungen generiert. Eine große Datenbank ist unter [Sloane, 2005] zu finden. Sie dient als Basis für unseren k -DOP Test. Der Hauptvorteil dieses Optimierungsansatzes besteht darin, dass Qualitätsmetriken, wie bspw. der Winkel zwischen den Unterräumen, direkt als Optimierungskriterium eingeht, wohingegen bei einer physikalisch basierten Simulation die Abstandsbedingungen zwischen den Punkten indirekt über Strafrkräfte erzwungen werden müssen, vgl. [Molino et al., 2003].

7.3.7.2 Teilungsebenen: Bei der Konstruktion einer DOP- \mathcal{H}_R stellt sich die Frage nach einer geeigneten Teilungsebene. Dieses Problem wurde von Klosowski et al. [Klosowski et al., 1998] ausführlich diskutiert. In ihrer Arbeit experimentieren sie mit unterschiedlichen Teilungsstrategien. Jedoch verwenden wir im Folgenden keine dieser Techniken, sondern wählen die Teilungsachsen aus der Menge der $k/2$ Normalen des k -DOPs. Dazu werden die Achsen in Bezug auf die Dicke des jeweiligen Slabs der Größe nach sortiert. Das erfordert die Sortierung von $k/2$ reellen Werten, was in $\mathcal{O}(k \log k)$ geschehen kann. Der Teilungspunkt liegt dann auf halber Distanz zwischen den zwei Ebenen, die den Slab auf gegenüberliegenden Seiten begrenzen. Seine Koordinaten sind durch $1/2 \cdot \mathbf{n}_i(\max_i + \min_i)$ gegeben, mit $0 \leq i < k/2$. Zur Berechnung der Größe des DOPs bzw. der Abstände aller durch die DOP-Normalen definierten Ebenen zum Ursprung in Abhängigkeit vom eingebetteten Zylinder stellen wir eine Distanzfunktion $d_{PC}(P, C)$ auf. Sei C ein Zylinder der Länge r mit Polkappen, den Endpunkten \mathbf{p}_1 bzw. \mathbf{p}_2 und der Zylinderachse $\mathbf{a} = \mathbf{p}_2 - \mathbf{p}_1$. Der minimal mögliche Abstand einer Ebene P zum Zylinder ist r . Liegt die Ebene parallel zum Zylinder ($\mathbf{n}_i^\top \mathbf{a} = 0$), berühren sich beide entlang eines Liniensegmentes, andernfalls haben beide einen Punkt \mathbf{p}_b auf den Polkappen gemeinsam. Der Punkt ist gegeben durch $\mathbf{p}_b = \mathbf{p}_{1|2} + r\mathbf{n}$. Die Wahl des Punktes \mathbf{p}_1 oder \mathbf{p}_2 hängt dabei vom Vorzeichen des Kosinus von Zylinderachse und Ebenennormale \mathbf{n}_i ab,

$$\mathbf{n}_i^\top \mathbf{a} \begin{cases} > 0, & \mathbf{p}_2 \\ = 0, & \mathbf{p}_1 \vee \mathbf{p}_2 \\ < 0, & \mathbf{p}_1. \end{cases} \quad (7.9)$$

Die Projektion des Ortsvektors \mathbf{p}_b auf die Ebenennormale \mathbf{n}_i entspricht genau dem gesuchten Abstand d_i vom Ursprung. Damit haben wir die Normalen der Ebenen und ihre Abstände vom Ursprung für unseren parametrischen DOP-Ansatz festgelegt.

Auch bei DOPs machen wir die schon von den AABBs bekannte Beobachtung, dass DOP- \mathcal{H}_{T_D} und DOP- \mathcal{H}_{T_I} identisch sind. Das Lemma kann auf DOPs und schließlich auf konvexe Polyeder erweitert werden. Wir schließen also daraus, dass, wenn in der Hierarchie konvexe Polyeder mit global festen Richtungen eingesetzt werden, \mathcal{H}_{T_D} und \mathcal{H}_{T_I} identisch sind.

7.3.7.3 Orientierte Quader (OBBs)

Orientierte Quader haben im Bereich der Kollisionserkennung weiten Einsatz gefunden [Barequet et al., 1996, Gottschalk et al., 1996, Lahanas et al., 2000], da sie i.d.R. gute Approximationen der zugrunde liegenden Geometrie bei vergleichsweise moderaten Kosten liefern. Sie sind erstmals von Barequet et al. [Barequet et al., 1996] als Instanzen des allgemeinen Boxtrees⁵ ein-

⁵Bottom-up konstruierter Binärbaum, dessen Blätter die atomaren Einheiten eines Dreiecksnetzes einhüllen.

geführt worden. Gottschalk et al. [Gottschalk et al., 1996] benutzen eine top-down konstruierte OBB- \mathcal{H}_R für die Kollisionserkennung zwischen polygonalen Modellen. Der Hauptnachteil von OBB-Hierarchien ist in den Kollisionserkennungszeiten zu sehen, die selbst bei Einsatz des schnellen *Separating-Axis-Tests* (SAT) um etwa eine Größenordnung höher liegen, als bei der konventionellen AABB- \mathcal{H}_R . Im Gegensatz zu den DOPs existiert eine OBB- \mathcal{H}_{T_D} , weil es keine global festen Richtungen gibt.

Problematisch ist die Berechnung von optimalen OBBs. Der Aufwand dafür ist nach O'Rourke [O'Rourke, 1985] kubisch in der Anzahl der Punkte. Die Idee von O'Rourkes $\mathcal{O}(n^3)$ -Algorithmus besteht in der Anwendung des Rotating-Caliper-Prinzips auf die konvexe Hülle einer Punktmenge im \mathbb{R}^3 . Diese konvexe Hülle formt einen Polyeder. Nach dem Theorem von O'Rourke müssen nun genau zwei Kanten des eingeschriebenen Polyeders in den Ebenen zweier benachbarter Seitenflächen der minimalen einschließenden Box liegen. Es werden alle möglichen Paare von Kanten durchprobiert und für jedes Paar eine dreidimensionale Variante von Toussaints Rotating-Caliper angewendet, um die Box mit dem minimalen Volumen zu finden. Der Vorgang kann vereinfacht werden, wenn man die Koinzidenz mit einer Seitenfläche der Box lediglich für eine Kante des Polyeders fordert. Der Aufwand reduziert sich damit auf $\mathcal{O}(n^2)$.

Barequet und HarPeled [Barequet & Har-Peled, 1999] bieten eine $(1 + \varepsilon)$ -Approximation, die in linearer Zeit $\mathcal{O}(n + 1/\varepsilon^{4.5})$ berechnet werden kann. Bei einer nur 10 %-igen Approximation fällt die Konstante allerdings bereits merklich ins Gewicht.

Lahanas et al. [Lahanas et al., 2000] haben eine Methode basierend auf Powell's Optimierungsverfahren vorgeschlagen. Das Ziel ist dabei eine Menge von optimalen Winkeln $(\theta_1, \theta_2, \theta_3)$ zu finden, welche die Volumenfunktion $V(\theta_1, \theta_2, \theta_3)$ der OBB durch Rotation der Geometrie um die Achsen eines festen Koordinatensystems minimiert. Alternativ kann man die Orientierung auch aus der Berechnung des minimal einschließenden Ellipsoids ableiten.

7.3.7.3.1 Zur Qualität von PCA-Boxen: Ein gängiges Verfahren, von dem auch wir Gebrauch machen werden, ist die Ermittlung einer approximativen OBB auf der Basis einer *Principal-Component-Analysis* (PCA).

Sei $\mathcal{P} = \{\mathbf{r}_1, \dots, \mathbf{r}_n\}$ eine Punktmenge im \mathbb{R}^3 mit dem Schwerpunkt \mathbf{x} . Die Varianz der Punktmenge in der Einheitsrichtung $\mathbf{d} \in \mathbb{R}^3$ ist

$$\text{var}(\mathcal{P}, \mathbf{d}) = \frac{1}{n} \sum_{i=1}^n \langle \mathcal{P} - \mathbf{x}, \mathbf{d} \rangle^2. \quad (7.10)$$

Es existieren orthogonale Hauptrichtungen $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$, in welchen die Varianz maximal wird. Diese Hauptrichtungen sind die Eigenvektoren der Kovarianzmatrix $\mathbf{C} \in \mathbb{R}^{3 \times 3}$. Es gilt also

$$\text{var}(\mathcal{P}, \mathbf{d}) = \langle \mathbf{C} \cdot \mathbf{d}, \mathbf{d} \rangle, \quad (7.11)$$

mit der symmetrischen Kovarianzmatrix

$$\mathbf{C}_{ij} = \frac{1}{n} \sum_{i=1}^n (\mathbf{r}_{ik} - \mathbf{d}_i)(\mathbf{r}_{jk} - \mathbf{d}_j). \quad (7.12)$$

Die Rangfolge der Hauptachsen richtet sich nach der Größe der zugehörigen Eigenwerte $\{\lambda_1, \lambda_2, \lambda_3\}$. Diese lassen sich ganz einfach aus dem charakteristischen Polynom der Kovarianzmatrix bestimmen. Das führt auf ein kubisches Polynom mit bekanntem analytischem Lösungsausdruck, vgl. Abs. (A.3).

Eine Untersuchung über die Qualität von PCA-Boxen im Allgemeinen wurde von Dimitrov et al. [Dimitrov et al., 2007] vorgelegt. Offenbar ist das Verhältnis der Volumina von PCA-Box zu optimal einschließender Box einer gegebenen Punktmenge \mathcal{P} im \mathbb{R}^3 nach unten durch 4, nach oben hingegen durch den Faktor 7,72 beschränkt. Um störende Einflüsse der Punkteverteilung auf die Ergebnisse der PCA zu verhindern, wurden diese Schranken auf der Basis der konvexen Hülle $\text{convh}(\mathcal{P})$ ermittelt. Die PCA-Box liefert also ein mindestens viermal so großes, aber höchstens achtfaches Volumen des optimalen Pendants.

7.3.7.3.2 Aufbau von OBB-Hierarchien: Während man beim Top-Down-Aufbau einer OBB- \mathcal{H}_R bzw. OBB- \mathcal{H}_{T_D} im Wesentlichen dem Ansatz von Gottschalk et al. [Gottschalk et al., 1996] folgen kann, unterscheiden wir bei der OBB- \mathcal{H}_{T_I} drei Ebenen mit unterschiedlichen Konstruktionsansätzen.

Auf der Blattebene folgen die OBBs direkt aus der lokalen Geometrie und Orientierung der Segmente. Auf der zweiten Ebene kombinieren wir die Segmente paarweise entlang der Faser. Das von ihnen aufgespannte Dreieck hat mindestens eine Seite mit dem minimal umschließenden Rechteck gemeinsam, vgl. Theorem von Freeman und Shapira [Freeman & Shapira, 1975]. Die Höhe der Box folgt dann aus der Dicke der Segmente. Terminalsegmente ohne Partner werden mit dem Terminalsegment der nächsten Faser verrechnet.

Die Hüllkörper an den Knoten aller darüber liegenden Ebenen der OBB- \mathcal{H}_{T_I} erhalten wir durch eine PCA auf der Basis der acht Eckpunkte ihrer beiden Kindboxen. Das Volumen beträgt damit das Vier- bis Achtfache der minimalen Box.

Die Basis für die kombinatorische Beschreibung eines MEVs einer OBB- \mathcal{H}_{T_D} besteht aus maximal sechs OBBs.

7.3.7.4 Kugeln

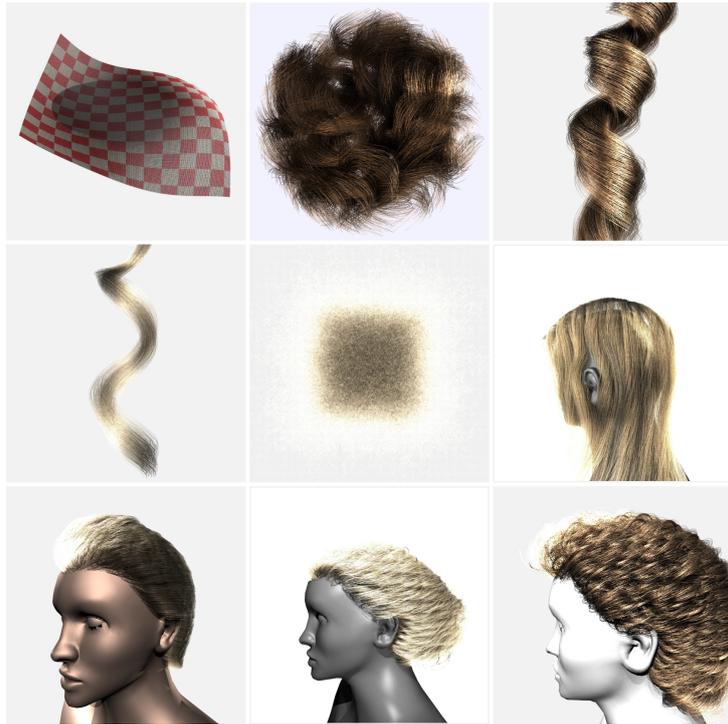
Kugelhierarchien können wegen ihrer Einfachheit als kanonisches Beispiel für Hüllkörper-Hierarchien angesehen werden [Quinlan, 1994, Palmer & Grimsdale, 1995]. Allerdings passen sich Kugeln nicht sehr gut an die Geometrie der Segmente an, was unabhängig von der Orientierung der Segmente zu überflüssigen Überschneidungen führt, vgl. Abb. (7.6).

Das Problem, die minimal umschließende Kugel einer Menge von Kugeln zu berechnen ist nichttrivial und kann auf der Basis des deterministischen Algorithmus von Megiddo

Hüllkörper	\mathcal{H}_R	\mathcal{H}_{T_D}	\mathcal{H}_{T_I}
Achsenparalleler Quader (AABB)	AS	-	AL
Diskret orientierter Polyeder (k -DOP)	DS	-	DL
Orientierter Quader (OBB)	OS	OW	OL
Kugel	SS	SW	SL

Tabelle 7.1: Abkürzungen für die Hierarchietypen.

Abbildung 7.7: Testkandidaten für den Performancetest mit unterschiedlichen Hierarchietypen. Von links nach rechts, die Referenznamen der Anordnungen, die Anzahl der Fasern, die Gesamtzahl der Segmente und die Anzahl impliziter Kollisionen: Tuch (0,5 k; 1,3 M; 17 M); Haarball (0,7 k; 26 k; 1,5 M); Locke (2,5 k; 246 k; 24 k); Tress I+II (0,1 k; 200 k; 1,75 k)/(0,5 k; 1 M; 23 k); Filament Soup (10 k; 1 M; 70 k), im Einheitswürfel generierte Zufallsfasern; Longish (39 k; 2,1 M; 459 k); Charming (49 k; 2,4 M; 542 k); Windy (60 k; 3 M; 1,78 M); Curly (47 k; 4,7 M; 1,89 M).



[Megiddo, 1982] oder dem randomisierten Algorithmus von [Welzl, 1991], angegangen werden, beide mit erwarteter linearer Laufzeit. Eine Kugel- \mathcal{H}_{T_D} lässt sich in $\mathcal{O}(n \log n)$ Zeit aufbauen. Hier ist jedoch zu beachten, dass der Algorithmus von Welzl eine große Konstante der Ordnung $\mathcal{O}(\delta \cdot \delta!)$ birgt, wobei δ gleich $k + 1$ und k die Dimension des Raumes ist, in dem die Hierarchie berechnet werden soll. Für den dreidimensionalen Raum wird die Konstante offensichtlich groß. Die Basis bei der kombinatorischen Beschreibung umfasst maximal vier Kugeln.

7.4 Performancevergleich

Im Rahmen unserer Untersuchungen haben wir alle drei Hierarchievarianten in Kombination mit unterschiedlichen Hüllkörpern implementiert und in unser interaktives Fasersimulationssystem Rapunzel integriert. Im Folgenden benutzen wir für die einzelnen Hierarchietypen nachstehende Abkürzungen, vgl. Tab. (7.1).

Die Laufzeiten wurden für komplexe statische und dynamische Szenen ermittelt. Die Performancetests erfolgten auf einem Pentium IV, 2,4 GHz. Folgende Testfälle wurden dabei im Einzelnen untersucht:

1. *Statischer Faserbündeltest:* Als Grundlage dient eine Faseranordnungen vom Typ Tress I, vgl. Abb. (7.7). Die Zahl der Fasern wird in diesem Test von 50 auf 1000 erhöht, bei 100 Segmenten pro Faser und konstanter Länge.
2. *Dynamischer Faseranordnungstest:* Um die Stabilität der Hierarchien unter dynamischen Bedingungen zu testen, fahren wir zwei Experimente:

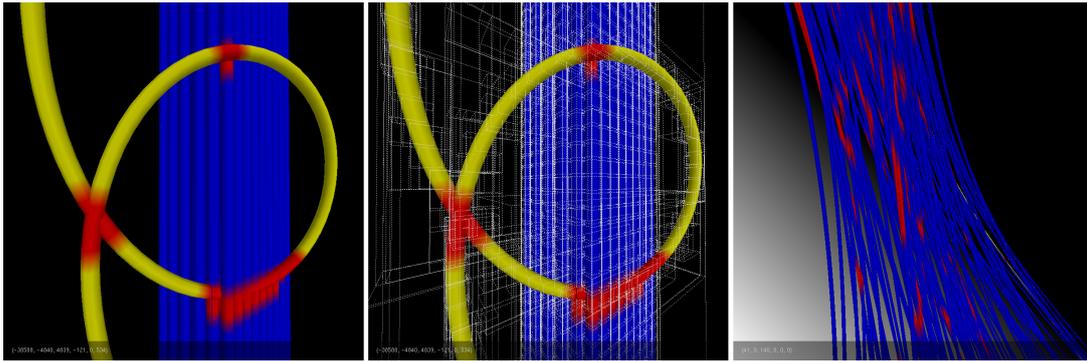


Abbildung 7.8: Ein Schnappschuss des Rapunzel Fasersimulationssystems bei der Durchführung eines dynamischen Faserbündeltests. Links: Interaktive Deformation einer Leitfaser des simulierten Faserbündels. Kollidierende Segmente sind rot markiert. Mitte: Darstellung der Hüllkörper, hier AABBs. Rechts: Simuliertes dynamisches Faserbündel.

- Wir applizieren eine kontrollierte Folge von Kraft-Moment-Paaren am Endpunkt jeder Faser eines Bündels. Wir benutzen dazu vorab aufgezeichnete Kraft-Moment-Trajektorien einer interaktiv deformierten Leitfaser, deren Werte dann mit einer kleinen Zufallsvariation an den Endpunkten der übrigen Fasern appliziert werden. Das Bündel besteht aus 150 Fasern mit je 100 Segmenten. Ausgehend von einer geraden Anfangskonfiguration, d.h., die Fasern sind parallel zur Z-Achse, wird das Bündel dann in 100 Simulationsschritten in eine Schleifenform gebracht, wobei die Zeiten für die Kollisionserkennung und die Aufrechterhaltung der Hierarchien gemessen werden, vgl. Abb. (7.8). Die Berechnung der Faserdeformation basiert auf der numerischen Lösung der statischen COSSERAT-Gleichungen.
 - Im zweiten Experiment untersuchen wir die Entwicklung der Kollisionserkennungszeiten nach Anwendung unterschiedlicher Aktualisierungsstrategien: Refitting einer AS, Neuberechnung des Optimums einer AS und Bottom-up-Aktualisierung einer AL, während der dynamischen Simulation einer Faseranordnung unter Einwirkung der Gravitation.
3. *Statischer Faseranordnungstest:* In diesem Test messen wir die Zeiten für den Aufbau, die Aktualisierung/Aufrechterhaltung sowie die Kollisionserkennung für verschiedene komplexe Faseranordnungen. Die Testkandidaten sind in Abb. (7.7) abgebildet. Weiterhin betrachten wir die Zeiten für das Auffinden der ersten Kollision, gemittelt über 1.000 Versuche. Die Wurzeln der Fasern in den Anordnungen sind mit einer Dichte von 325 Punkten/cm² verteilt. Das entspricht in etwa der gemessenen durchschnittlichen Haardichte auf der menschlichen Kopfhaut.
 4. *Parametrischer DOP-Test:* Auf der Basis unseres parametrischen k -DOP Ansatzes bestimmen wir ein möglichst optimales k für ausgewählte Modelle.

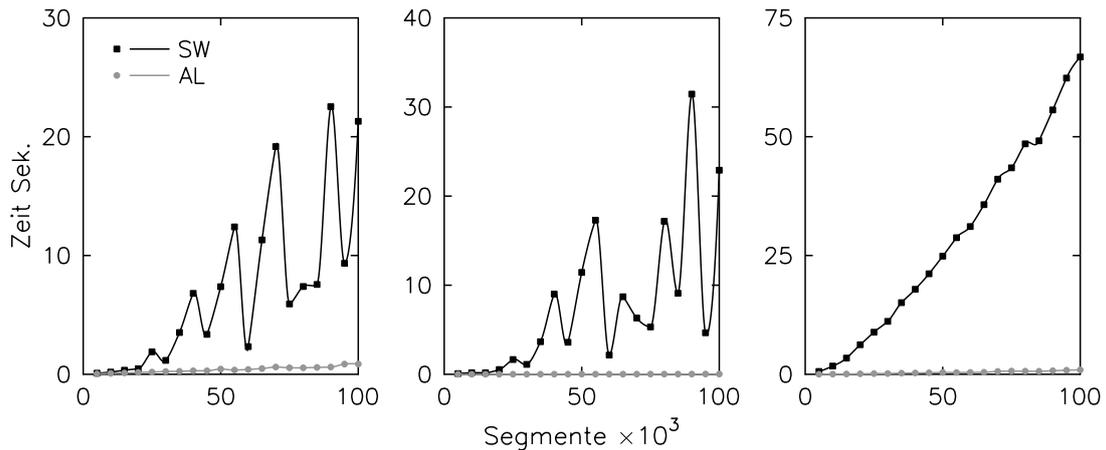


Abbildung 7.9: AL und SW im Vergleich: Laufzeiten für den Aufbau der Hierarchie, die Aktualisierung und die Kollisionserkennung in einer statischen Faseranordnung in Abhängigkeit von der Anzahl der Segmente.

7.4.1 Statischer Faserbündeltest

Im statischen Faserbündeltest wurden die Laufzeiten für den Aufbau, die Aktualisierung und die Kollisionserkennung der AL bzw. der SW miteinander verglichen. Die Messergebnisse der statischen Testanordnung bei veränderlicher Segmentanzahl sind in Abb. (7.9) dargestellt. Die Kosten für den Aufbau und die Aktualisierung der AL hängen linear von der Zahl der Segmente in der Anordnung ab, vgl. Abb. (7.9). Kollisionserkennung auf der Basis der AL ist deutlich schneller, als mit der SW, wie Abb. (7.9) zu entnehmen ist. Der Grund dafür ist die schlechte Approximation der Segmente durch Kugeln. Die Diagramme für die Aktualisierungskosten sowie die Verifikationskaskade der SW zeigen Kostenspitzen für bestimmte Gesamtsegmentzahlen. Für diese extremen Oszillationen muss der randomisierte Algorithmus von Welzl [Welzl, 1991] zur Berechnung der minimal einschließenden Kugeln in der Implementierung von Gärtner [Gärtner & Schönherr, 1997] verantwortlich gemacht werden. Insbesondere benötigt der Algorithmus mehr Zeit für die Ermittlung des MEV von 55 k als von 95 k Segmenten, vgl. Abb. (7.9).

Aktualisierungen der AL sind sehr schnell möglich (vgl. Abb. (7.9)), da die Bottom-up-Technik nur linear von der Anzahl der Segmente abhängt, bei kleiner Konstante. Für die aus 1 k einzelnen Fasern bestehende Haarsträhne aus Abb. (7.10) mit insgesamt 2 M Segmenten nimmt die Aktualisierung lediglich 300 ms in Anspruch. Die Erkennung der über 2 k impliziten Kollisionen in dieser Anordnung benötigt 1,7 Sek. Bei Verdopplung der Segmentanzahl in der Haarsträhne auf 2 k, benötigte die Aktualisierung die erwarteten 600 ms, die Erkennung von 80 k impliziten Kollisionen hingegen 4,3 Sek. mit der AL. Das Auffinden aller 550 k Kollisionen im Modell Charming benötigt 100 Sek., mit der Kugelhierarchie hingegen mehr als 30 Min.

7.4.2 Dynamischer Faseranordnungstest

Im dynamischen Faseranordnungstest auf der Basis der AL, können wir eine starke Korrelation zwischen der Anzahl der Kollisionen, die in jedem Simulationsschritt erkannt wurden, und den

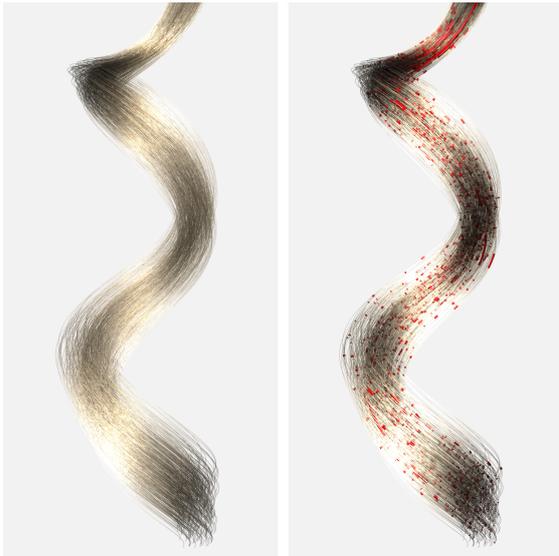


Abbildung 7.10: Komplexe Faseranordnung. Kollisionsbereiche sind rot dargestellt.

korrespondierenden Erkennungszeiten feststellen. Die Ergebnisse des dynamischen Tests sind in Abb. (7.11) dargestellt. Der Erkennungsprozess ist outputsensitiv. Je mehr Kollisionen gefunden werden, desto öfter muss der Kollisionsbaum bis zu den Blättern traversiert werden. Bei sehr schlechter Abdeckung der Segmentgeometrie, wie bspw. durch Kugeln, lassen sich viele Schnitttests zur Feststellung einer Nichtkollision auf der Segment-Segment-Ebene nicht umgehen. Die Aktualisierungszeiten sind linear abhängig von der Segmentzahl und bilden eine konstante Basislinie (Fluktuationen sind durch das Operationssystem verursacht).

Der Vergleich der Entwicklung der Kollisionserkennungszeiten bei der AABB- \mathcal{H}_R nach (a) einer Aktualisierung auf der Basis der Refittingstrategie und (b) einer vollständigen Neuberechnung des Optimums, zeigt, dass sich die Erkennungszeiten im Fall (a) mit fortschreitender Simulationsdauer zunehmend verschlechtern, vgl. Abb. (7.12). Der Unterschied ist bis zum sechzigsten Simulationschritt vernachlässigbar, steigt aber bis zum Ende auf das Vierfache an. Dem gegenüber steht der Aufwand für das Refitting, das mit durchschnittlich 16 ms nur etwa 1/12 des Neuberechnungsaufwandes beträgt. Allerdings kann die Zeiteinsparung durch das Refitting die rapide zunehmenden Kollisionserkennungszeiten ab Schritt 206 nicht mehr abfangen. Die Kombination der Refittingstrategie mit einer gelegentlichen Neuberechnungen sollte hier Abhilfe schaffen.

Auf der anderen Seite steht die AABB- \mathcal{H}_T , mit einem vernachlässigbaren Aufwand von nur 16 ms für die Bottom-up-Aktualisierung und Kollisionserkennungszeiten, die nur etwa 15 ms schlechter sind, als die der optimalen AABB- \mathcal{H}_R . Unter dem Gesichtspunkt der Gesamtkosten schneidet sie etwa um den Faktor 2 bis 2,5 besser ab, als die optimale AABB- \mathcal{H}_R .

7.4.3 Statischer Faseranordnungstest

Die Aktualisierungskosten sind, wie Abb. (7.13) zu entnehmen, für die räumlich angepassten Hierarchien (Aufbau des Baumes ist hier mit eingeschlossen) durchweg größer, als für die topologisch angepassten Varianten. Diese Kosten werden hauptsächlich durch den räumlichen Subdivisionsprozess verursacht. In Bezug auf die Kollisionserkennung hingegen sind die räumlich angepassten

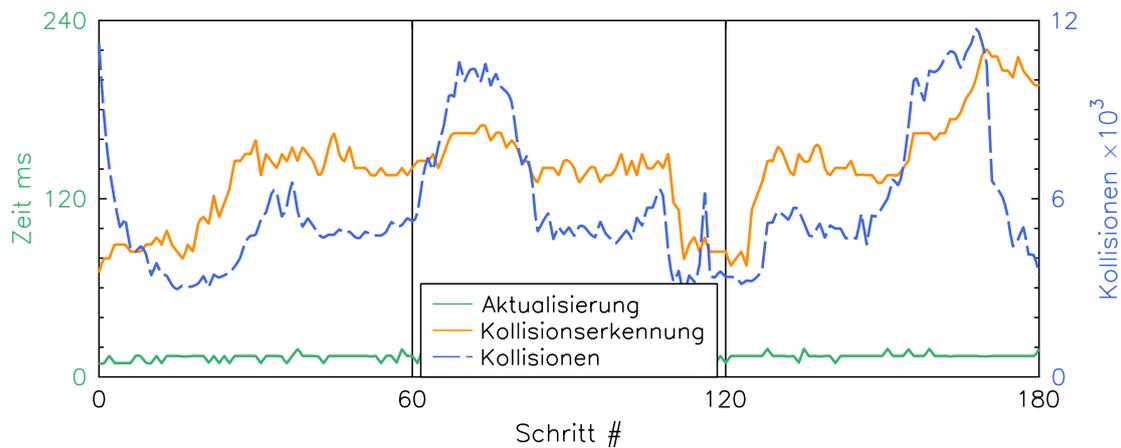


Abbildung 7.11: Laufzeiten für Aktualisierung und Kollisionserkennung mit der AAB- \mathcal{H}_{T_1} bei einer dynamischen Testszene (150 Fasern zu je 100 Segmenten werden aus dem geraden Zustand zu einer Schleife deformiert). Die dritte Kurve zeigt die Anzahl der pro Simulationsschritt erkannten Kollisionen.

Hierarchien in den meisten Fällen um Größenordnungen schneller, vgl. Abb. (7.13). Interessanterweise gibt es auch hier Ausnahmen: Die DL liefert für das Tuch geringfügig bessere Zeiten, als die DS. Im Falle einer topologisch angepassten Hierarchie wird die meiste Zeit in die Überlappungstests investiert. Der Suchpfad zum Auffinden von Kollisionen ist offenbar um einiges länger als bei den räumlich angepassten Hierarchien. Die räumliche Subdivision führt zu wohlseparierten Hüllkörpern bei vergleichsweise hohen Investitionskosten, wohingegen der schnellere Bottom-up-Aufbauprozess entlang der an den Endpunkten verbundenen Filamente deutlich mehr initiale Überlappungen verursacht. Die resultierenden Hüllkörper sind infolgedessen nicht annähernd so gut separiert. Ausschlaggebend ist hierbei die Art der Anordnung. Faseranordnungen ohne erkennbare Ordnung, bei der die Endpunkte der Fasern darüber hinaus beliebig verknüpft werden, führen dann auch zu mehr oder weniger beliebigen bzw. schlecht separierten Hüllkörpern. Einige Beispiele, wie die Filamentsuppe (AS:AL=1:43) oder die Ganzkopfmodelle (\approx 1:10) untermauern diese Aussage. Anders hingegen, wenn eine solche globale Ordnung erkennbar ist, wie bspw. bei Modell Tress I. Dort beträgt das Verhältnis AS:AL nur noch 1:1,7, für DS:DL=1:1.

Die unterschiedlichen Zeiten bei der Kollisionsauffindung sind damit überwiegend durch die unterschiedliche Anzahl an durchgeführten Überlappungstests begründet, vgl. Abb. (7.14). Die Zahl der tatsächlich auf Überschneidung getesteten Segmente ist bei beiden Varianten fast identisch, während sich die Anzahl der Hüllkörperstests auf dem untersten Niveau um bis zu einer Größenordnung unterscheidet. Die Anzahl überlappender Hüllkörper sind bei den topologisch angepassten Hierarchien ebenfalls größer.

Auch bei der Untersuchung des Verhaltens beim Auffinden der ersten Kollision schneidet die räumlich angepasste Hierarchie besser ab, vgl. Abb. (7.13), unten.

Pro Kollisionserkennungszyklus müssen Kosten sowohl für die Aktualisierung, als auch für die Kollisionserkennung selbst berücksichtigt werden. Unter dem Aspekt der Gesamtkosten also schneiden die topologischen Hierarchien in einigen Fällen interessanterweise sogar besser ab.

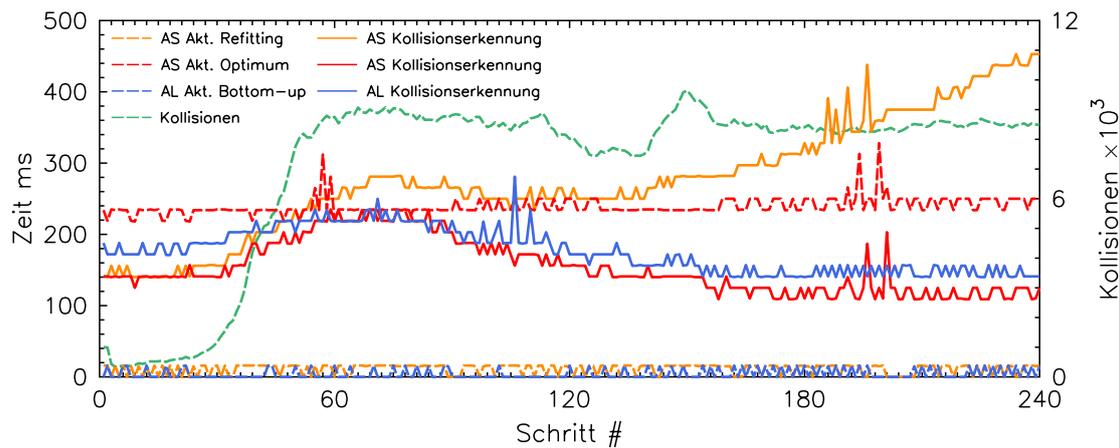


Abbildung 7.12: Entwicklung der Kollisionserkennungszeiten nach Anwendung unterschiedlicher Aktualisierungsstrategien. Betrachtet werden das Refitting und die Neuberechnung des Optimums einer $AABB-\mathcal{H}_R$, sowie die Bottom-up-Aktualisierung einer $AABB-\mathcal{H}_{T_1}$.

Bei Modell Tress I aus Abb. (7.10) bspw. ist die räumlich angepasste der topologischen Variante überlegen, vgl. Abb. (7.14), links. Bei dem komplexeren Beispiel des Ganzkopfmodells wird der zeitliche Aufwand für die Aktualisierung der räumlichen Hierarchie jedoch so groß, dass die topologische Variante trotz einer erheblich höheren Anzahl an Überlappungstests unter dem Aspekt der Gesamtkosten deutlich besser abschneidet. Man macht zudem die Beobachtung, dass sich der Abstand zwischen topologischer und räumlicher Variante um so mehr relativiert, je kleiner die Modelle werden. Bei der AL haben wir mit einem alternativen Aktualisierungsschema experimentiert, welches aber in allen Fällen wegen großer Überlappungen bereits auf den untersten Hierarchieniveaus durchweg schlecht abschneidet, bis zu einer Größenordnung langsamer.

Eine interessante Beobachtung, die wir mit OBB-Hierarchien im Allgemeinen machen konnten, ist, dass ein vorgeschalteter Überlappungstest auf der Basis von SAT immer zeitliche Vorteile bringt, obwohl der direkte Überschneidungstest zwischen zwei Zylindern weniger arithmetische Operationen benötigt und das Volumenverhältnis zwischen einer eng anliegenden OBB und dem Zylinder gerade mal $10/\pi$ beträgt. Diese Erkenntnis kann einige Anhaltspunkte bzgl. der Effizienz von hybriden Hierarchien geben, die unterschiedliche Hüllkörper auf den einzelnen Niveaus einsetzen. Der OBB ist hier auf der Blattebene dem Zylinder der Vorzug zu geben.

Abb. (7.14), rechts, zeigt die durchschnittliche Entwicklung der Hüllkörpervolumina über die Niveaus unterschiedlicher Hierarchietypen, die über Modell Locke aufgebaut wurden. Die Volumina der \mathcal{H}_{T_1} -Hierarchien sind dabei geringfügig größer, als bei den \mathcal{H}_{T_D} - oder \mathcal{H}_R -Hierarchien. Insbesondere die Bottom-up-Konstruktion von orientierten Boxen erweist sich als problematisch, weil die Volumina exponentiell mit dem Niveau anwachsen. Die orientierte Box am Wurzelknoten der Hierarchie unterscheidet sich von seinem Pendant in der Kugelhierarchie um fünf Größenordnungen.

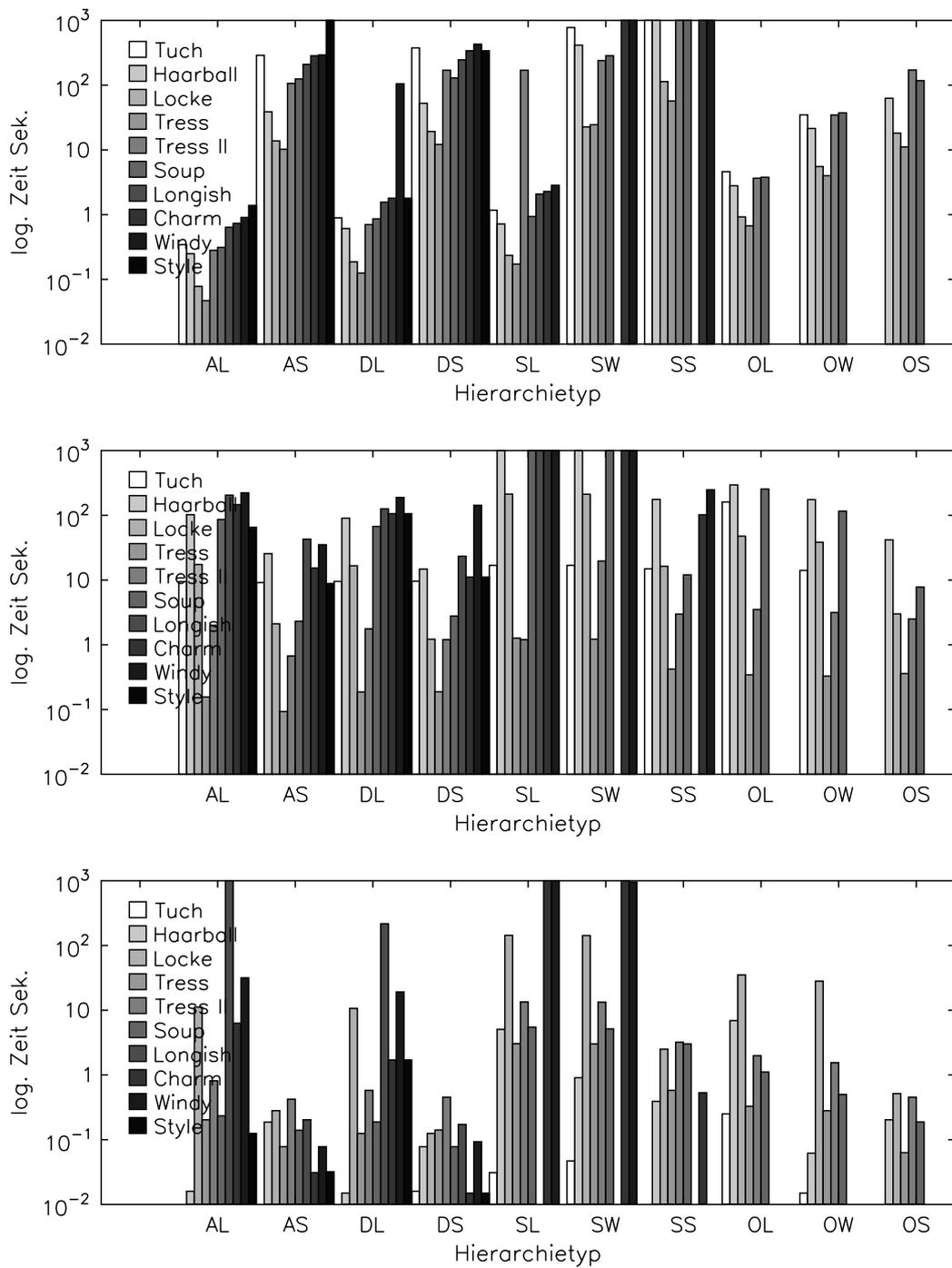


Abbildung 7.13: Zeitmessungen für unterschiedliche Testmodelle und Hierarchietypen. Oben: Gemessene Zeiten für die Aufrechterhaltung bzw. Aktualisierung; Mitte: Zeiten zur Erkennung aller Kollisionen; Unten: Zeiten zur Erkennung der ersten Kollision.

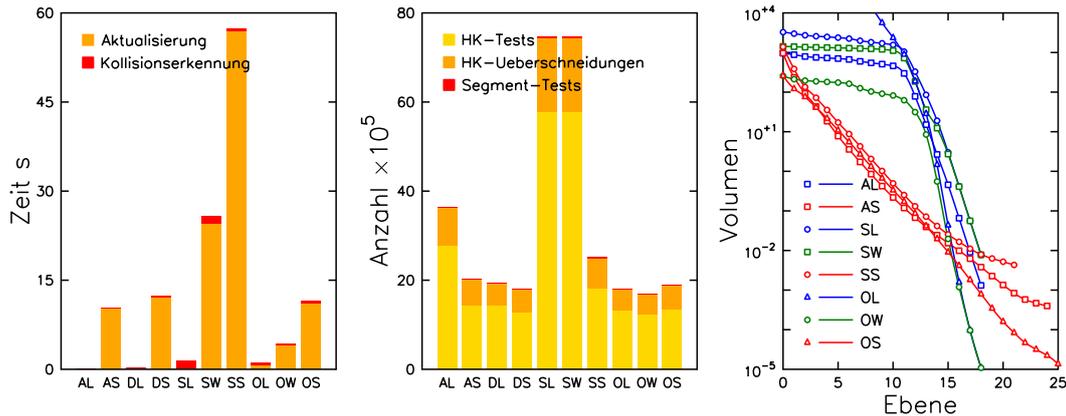


Abbildung 7.14: Links: Aktualisierungs- sowie Erkennungszeiten für Modell Tres I; Mitte: Die korrespondierende Anzahl an Hüllkörperstests, Hüllkörperüberschneidungen, sowie Segmenttests. Rechts: Entwicklung des durchschnittlichen Hüllkörpervolumens über die einzelnen Hierarchieniveaus bei unterschiedlichen Hierarchietypen (Modell Locke).

7.4.4 Parametrischer DOP-Test

Der parametrische DOP-Test wurde mit Modell Locke und einer $DOP-\mathcal{H}_{T_1}$ durchgeführt, wobei die Anzahl der Normalen schrittweise von 6 auf 200 erhöht wurde.

Die Normalenrichtungen wurden dabei nach der Methode aus Abs. (7.3.7.2) ermittelt. Keine der generierten Normalenrichtungen fallen mit den globalen Richtungen zusammen, d.h., sie haben keine zwei Nullkomponenten. Daher sind die Aktualisierungszeiten für den automatisch generierten 6-DOP geringfügig schlechter, als bei der konventionellen $AABB-\mathcal{H}_{T_1}$.

Die Resultate sind in Abb. (7.15) zusammengefasst und zeigen, dass die Anzahl auf Überschneidung untersuchter und die tatsächliche Zahl sich überschneidender Hüllkörper, sowie die der getesteten Primitive mit zunehmendem k abnehmen. Trotzdem liegt das Minimum der Erkennungszeiten offensichtlich im Intervall $[6,200]$ bei $k = 20$. Man beachte, dass die Aktualisierungszeiten monoton mit k anwachsen. In Abb. (7.15) sind die Erkennungszeiten für unterschiedliche Modelle im Intervall $[6,60]$ dargestellt. Das Minimum hängt offenbar von der Geometrie ab. In den meisten Fällen existiert jedoch kein eindeutig identifizierbares Minimum. Jenes ist vielmehr über einen weiten Wertebereich von k „verschmiert“. So ist bspw. die Kurve der Zeiten bei der Filamentsuppe selbst bei $k = 60$ noch abnehmend.

7.5 Diskussion

Im vorangegangenen Kapitel haben wir unterschiedliche Arten von Hüllkörperhierarchien über Modellen unterschiedlicher Komplexität hinsichtlich ihrer Machbarkeit und Performance untersucht. Dabei hat sich gezeigt, dass die klassische, räumlich angepasste $AABB-\mathcal{H}_R$ das optimale Verhältnis von Speicherbedarf und Laufzeitverhalten aufweist. In besonderen Fällen, je nach Art der Geometrie, lässt sich der Einsatz topologisch angepasster $AABB-\mathcal{H}_D$ oder $AABB-\mathcal{H}_{T_1}$ auf der Basis von k -DOPs und $AABB$ s rechtfertigen.

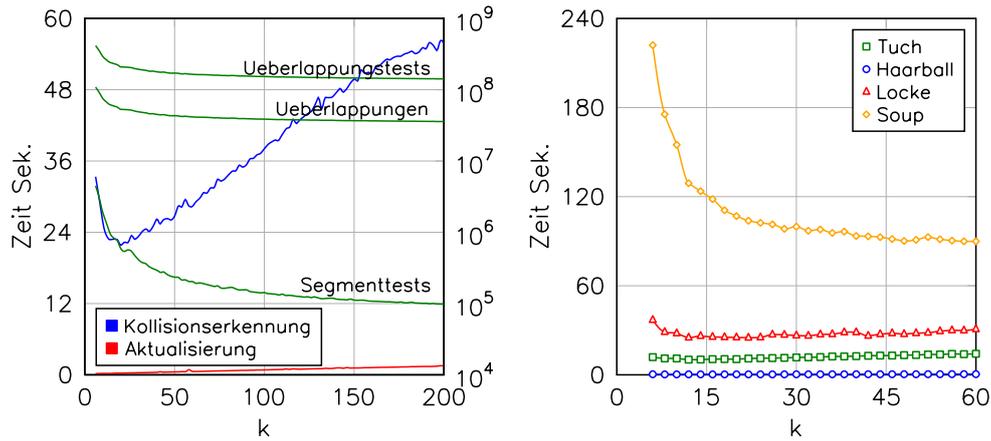


Abbildung 7.15: Links: Ergebnisse des k -DOP-Tests für Modell Locke. Kollisionserkennungs- und Aktualisierungszeiten; Anzahl der Überlappungstests, Hüllkörperüberschneidungen und Segmentstests. Man beachte das Minimum der Erkennungszeiten bei ständig abnehmender Anzahl an Schnitttests. Rechts: Kollisionserkennungszeiten von unterschiedlichen Modellen. Das Minimum hängt offenbar von der Geometrie ab. In den meisten Fällen ist eine eindeutige Identifizierung des Minimums nicht möglich.

OBB-Hierarchien hingegen zeigen zwar eine optimale Packung der Fasergeometrie. Deutliche Laufzeiteinbußen durch aufwendige Schnitttests, die eine Verlangsamung von bis zu einer Größenordnung gegenüber der AABB-Hierarchie bewirken, heben diesen Vorteil wieder auf. Die hier vorgestellte Parametrisierung von k -DOPs auf der Basis von GRASSMANN-Räumen erlaubt uns empirische Tests für veränderliche k zu fahren. Solche Tests liefern nicht nur optimale Werte für k in der momentanen sequentiellen Implementierung, sondern zeigen, dass in potentieller zukünftiger hardwareunterstützter Kollisionserkennung selbst größere Werte für k optimal sein können. k -DOPs haben allerdings den Nachteil, dass die Zahl der Normalen, bei der das Laufzeitoptimum auftritt, bei großen Modellen (hohen Segmentzahlen) zu einem nicht mehr vertretbaren Speichermehraufwand führt. Während in der momentanen sequentiellen Implementierung die Performance von AABBs und k -DOPs sehr ähnlich ist, macht das Potential hinsichtlich Parallelisierung den parametrischen DOP-Ansatz zum vielversprechendsten Kandidaten in dieser Richtung.

Mit den gegebenen Testszenarien konnten wir darüber hinaus zeigen, dass Kollisionserkennung in dicht gepackten Faseranordnungen auf Standard-PCs mit vertretbarem Aufwand selbst bei komplexen Frisuren möglich ist. Damit muss die allgemeine Auffassung, eine faserbasierte Haarsimulation müsse notwendig an der Komplexität der Kollisionserkennung scheitern, zumindest in diesem Bereich des Gesamtprozesses von Simulation, Kollisionsbehandlung und Visualisierung aufgeben werden.

7.6 Kontinuierliche Kollisionserkennung

Die diskreten Kollisionserkennungsverfahren, die wir hier vorgestellt haben, machen lediglich Momentanaufnahmen des Systemzustands zu diskreten Zeitpunkten der Simulation. Das Pro-

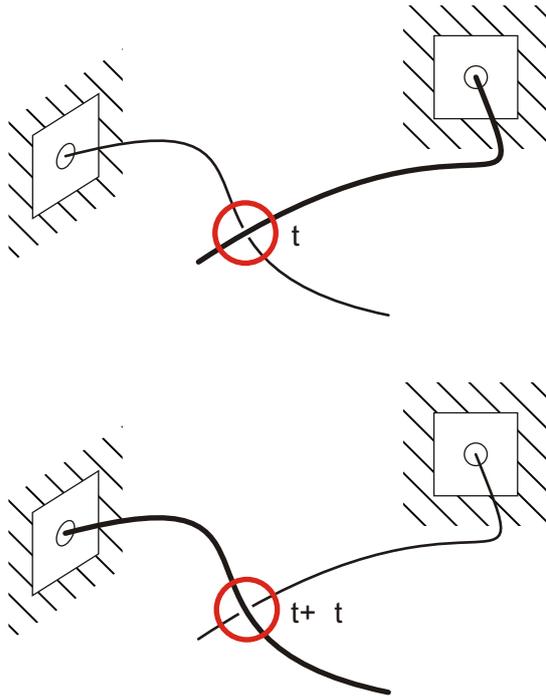


Abbildung 7.16: Sind die beiden Fasern im Verlaufe eines Zeitschritts Δt getunnelt? Die Frage kann bei alleiniger Betrachtung diskreter Zustände nicht beantwortet werden.

blem dabei ist, dass sie keine Informationen darüber liefern, ob das System zwischen diesen Zeitpunkten physikalisch unzulässige Zustände angenommen hat. Wenn sich ein Partikel mit der Geschwindigkeit v auf eine Wand der Dicke δ zu bewegt, dann wird er, sofern der Zeitschritt zwischen zwei Kollisionsbehandlungen zu groß gewählt ist, die Wand passieren. Die Wand stellt in diesem Sinne einen Energiewall dar, der vom Partikel „durchtunnelt“ wird⁶. Physikalisch gesehen, muss der Partikel die Wand, von der er eigentlich hätte abprallen sollen, irgendwann zwischen den beiden Zeitpunkten passiert haben, vgl. Abb. (7.16).

Abhilfe schafft die Betrachtung der Raumzeittrajektorien der Objekte. Unter der Annahme einer konstanten Geschwindigkeit (lineare und Winkelgeschwindigkeit) vollführt ein starrer Körper eine schraubenartige Bewegung, wobei seine Berandung eine komplexe Figur mit gekrümmten Flächen beschreibt. Der Schnitt der Raumzeittrajektorien zweier Körper indiziert eine mögliche Kollision. Da sich die Körper nicht notwendig zum selben Zeitpunkt im Schnittbereich der Trajektorien aufgehalten haben müssen, bringt erst eine Nullstellensuche in den Polynomen, welche die Bewegung zwischen den Zeitschritten charakterisieren, endgültige Sicherheit (Backtracking in der Zeit). Potentielle Kandidaten für solche Untersuchungen findet man, indem die Hüllkörper einer Hierarchie so erweitert, dass sie auch den Körper an seiner neuen Position einschließen, und dann den bekannten hierarchiebasierten Kollisionstest durchführt, vgl. bspw. [Bridson et al., 2002].

Der Aufwand, der mit solchen *kontinuierlichen* Kollisionserkennungsverfahren verbunden ist, sollte nicht unterschätzt werden. Zur Vereinfachung der Schnittkörperbestimmung sind die

⁶Die Wahl des Begriffes des Durchtunnels erfolgt in Anlehnung an den aus der Physik bekannten Tunneleffekt, bei dem Teilchen einen Potentialwall von einem Moment zum Nächsten passieren, etwas, das von der Theorie her unmöglich erscheint.

Raumzeittrajektorien in der Regel erst zu triangulieren [Redon et al., 2004]. Bei mehreren möglicherweise kollidierenden Körpern ist die Nullstellensuche keineswegs trivial. Wir beschränken uns daher in unseren Simulationen zunächst auf den Einsatz diskreter Kollisionserkennungsverfahren in Kombination mit einer sorgfältigen Abstimmung von Schrittweiten, Geschwindigkeiten, sowie Objektradien aufeinander.

7 Kollisionserkennung für 1D-Strukturen

8 Kollisionsantwort

Im Zusammenhang mit der Kollisionserkennung wurde bereits darauf hingewiesen, dass die immanente Eigenschaft der Natur, nach der sich zwei Materiepunkte nicht zum selben Zeitpunkt am selben Ort aufhalten können, auf dem Computer explizit erzwungen werden muss. Deshalb soll die Behandlung oder Auflösung detektierter Kollisionen den Gegenstand des vorliegenden Kapitels bilden. Eine korrekte Kollisionsbehandlung bei der Interaktion einzelner Fasern ist der Schlüssel zur Volumenerhaltung in komplexen Faseranordnungen.

8.1 Verfahren zur Kollisionsbehandlung

Wir geben zunächst eine kurze Übersicht über die im Bereich der Computergraphik etablierte Verfahren und leiten anschließend über zur Klasse der für die Kontaktsimulation in komplexen Faseranordnungen eingesetzten Zeitschrittverfahren. Wir rekapitulieren dazu die grundlegenden Ideen Moreaus zur Behandlung von Systemen mit einseitigen Zwangsbedingungen, orientieren uns dabei aber primär an dem ausgezeichneten Werk von Monteiro Marques [Monteiro Marques, 1993]. Die Effizienz des in diesem Rahmen entwickelten Zeitschrittverfahrens zur Kollisionsbehandlung demonstrieren wir anhand von komplexen Beispielen.

8.1.1 Strafkraften und Anordnungsoptimierung

Unter den Verfahren zur Verhinderung von Durchdringungen haben sich besonders die klassischen Strafkraftverfahren wegen ihrer Einfachheit hervorgetan. Drohende Kollisionen werden durch das Aufbringen einer abstoßenden Kraft in Richtung der Kollisionsnormalen verhindert. Der Betrag folgt irgendeinem Abstandsgesetz. Es wird für die Dauer der Kollision quasi eine temporäre Feder zwischen den kollidierenden Körpern etabliert. Die Federkonstanten sind zur Abwendung von Kollisionen i.d.R. sehr groß zu wählen, was zu den üblichen numerischen Problemen steifer Differentialgleichungen führt, wie kleine Schrittweiten oder fehlende Konvergenz. Strategien, wie das Einbringen einer stabilisierenden künstlichen Dämpfung können dann zu pathologischen Änderungen des Systemverhaltens führen. Da, wo „physikalisch plausible“ Ergebnisse in Bezug auf kleine Durchdringungen vertretbar sind, wie bspw. in Computerspielen, lässt sich dieses Problem durch den Einsatz weicher Federn umgehen. Es bleibt jedoch prinzipiell unklar, wie die Federkonstanten angemessen zu wählen sind. Sie sind abstrakte Begriffe und in diesem Sinne nicht messbar. Was für den einen Zweck angemessen ist, mag im anderen Fall nicht zu den gewünschten Ergebnissen führen. Mirtich [Mirtich, 1996] weist in diesem Zusammenhang darauf hin, dass gerade bei der Verwendung von diskreten Verfahren zur Kollisionserkennung, Körper bereits tief ineinander eingedrungen sind, wenn eine solche Kollision gemeldet wird. Die dann zur Aufhebung dieses Fehlzustandes notwendigen Kräfte sind größer, als beim kontinuierlichen System, was zu numerischen Problemen führen kann. Während die Richtung der Kollisions-

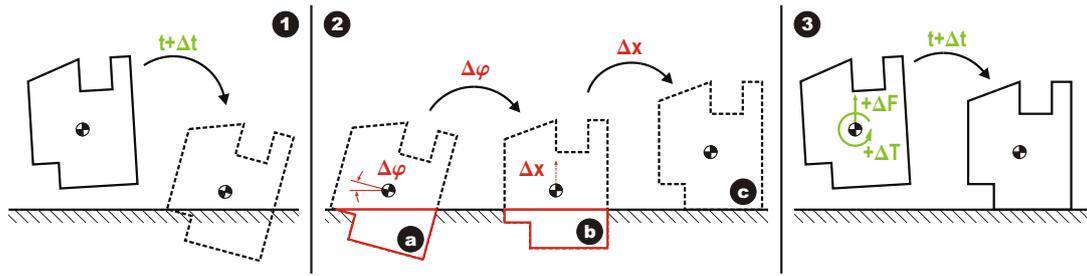


Abbildung 8.1: Anordnungsoptimierung zur Abschätzung von Kollisionsantwortkräften. Das System wird um Δt in der Zeit vorwärts entwickelt. Aus den Durchdringungen berechne man durch Lösen eines geeigneten Programmierungsproblems eine Menge minimaler Verdrehungen bzw. Verrückungen, die notwendig sind, um den Körper aus der misslichen Lage zu befreien. Sie lassen sich dann in äquivalente Kräfte und Drehmomente transformieren und vor der Integration aufbringen, um den Körper durch Integration des Systems in die gewünschte Lage zu bringen.

kraft eindeutig durch die Stoßnormale festgelegt ist, bereitet ihre Skalierung offenbar Probleme. Einer der Hauptgründe dafür ist wohl, dass man das genaue Ziel, also die Frage danach, wo der Körper nach der Abstoßung überhaupt hin soll, nicht berücksichtigt. Damit bleibt die Beschränkung der Antwortkraft nach oben willkürlich. Aus diesem Grund besteht eine einfache Strategie zur Wahl einer geeigneten Skalierung der Kollisionsantwortkraft bzw. zur Überbrückung von Stabilitätsproblemen in der vorausschauenden Behandlung von Systemfehlzuständen. Man entwickelt den Systemzustand um ein definiertes Intervall Δt in der Zeit fort und untersucht auf mögliche Kollisionen. Sind Durchdringungen aufgetreten, sucht man für den aktuellen geometrischen Zustand eine Menge minimaler Verrückungen, sowohl translatorischer, als auch rotatorischer Art, welche, angewandt auf die Systemmitglieder, diesen Zustand gerade auflösen, indem sie das System in eine durchdringungsfreie Konfiguration überführen. Die Verrückungen lassen sich in äquivalente Kräfte transformieren, welche nach der Integration des Systemzustandes genau die durchdringungsfreie Konfiguration liefern, vgl. Abb. (8.1). Verfahren zur Projektion auf zulässige Positionen basieren in der Regel auf der Lösung komplexer Programmierungsprobleme, vgl. [Schmidl, 2002] und [Spillmann & Teschner, 2008]. Im Rahmen der physikalisch basierten Simulation fordert man eine gewisse Kontinuität der Bewegungsabläufe. Bei Verfahren, die primär auf der der Lösung von Programmierungsproblemen basieren, kann eine zeitliche Kohärenz nicht garantiert werden. Das Paradigma der „physikalisch plausiblen Lösung“ kommt hier nur teilweise zum Tragen, da lokal korrektes Verhalten, wie elastisches Abprallen, durch die Projektion kompensiert wird. Offene Fragen, wie die nach der Existenz und Eindeutigkeit von Lösungen unter Einbeziehung von Reibungseffekten lassen sich zudem nicht ohne weiteres beantworten. Unter dem Gesichtspunkt des Berechnungsaufwandes muss das wegen seiner vermeintlichen Einfachheit so beliebte Verfahren der Kollisionsantwortkräfte als ausgesprochen kostenintensiv gelten.

8.1.2 Analytische Methoden

Eine zweite große Klasse von Verfahren sind die analytischen oder auch exakten Methoden. Zwischen der Normalbeschleunigung und der Normalkraft an der Kontaktstelle besteht eine komplementäre Beziehung: Entweder beschleunigen die Körper von einander weg, dann muss die Normalkraft notwendig verschwinden; oder die Körper stehen in Kontakt, dann muss die relative Normalbeschleunigung notwendig verschwinden und eine entsprechende Kontaktkraft wirken. Diese komplementäre Beziehung führt mathematisch ausformuliert auf eine Problemklasse, die auch als *Linear Complementarity Problem (LCP)* bekannt ist. Mit der relativen Normalbeschleunigung $\dot{\mathbf{u}}$, der Normalkraft \mathbf{f} an den Kontaktstellen, sowie der von der momentanen Kontaktkonfiguration abhängigen Matrix \mathbf{M} und dem Vektor \mathbf{b} ist ein LCP mathematisch gegeben durch

$$\text{LCP} := \begin{cases} \dot{\mathbf{u}} = \mathbf{M}\mathbf{f} - \mathbf{b}, \\ \mathbf{0} \leq \dot{\mathbf{u}} \perp \mathbf{f} \geq \mathbf{0}. \end{cases} \quad (8.1)$$

Es sind nur Normalkräfte bzw. relative Normalbeschleunigungen zulässig, welche die Körper voneinander wegbeschleunigen. Ersetzt man die relative Normalbeschleunigung in Gl. (8.1) durch die Lücken- oder Gapfunktion $h \geq 0$, die genau dann von Null verschiedene Werte annimmt, wenn die Körper voneinander wohlsepariert sind, wird die komplementäre Beziehung zwischen Lückenfunktion und Kontaktkraft als SIGNORINI-Bedingung bezeichnet, vgl. [Pfeiffer & Glocker, 1996].

LCPs in der Starrkörpersimulation wurden zuerst von Lötstedt [Lötstedt, 1984], später von Baraff [Baraff, 1989] untersucht. Man kann zeigen, dass für Systeme ohne Reibung Gl. (8.1) immer eine eindeutige Lösung hat. Das LCP ist dann konvex und die Berechnung einer Lösung in erwarteter polynomialer Laufzeit möglich. Im Gegensatz dazu können Existenz und Eindeutigkeit der Lösung bei reibungsbehafteten Systemen lediglich für hinreichend kleine Reibungsbeiwerte garantiert werden [Pang & Trinkle, 1996]. Kontaktprobleme mit Reibung führen auf nicht-konvexe LCPs. Das Auffinden entsprechender Kräfte unter Gleitreibung an den Kontaktpunkten ist NP-Schwer [Baraff, 1991]. Baraff hat durch gesonderte Behandlung einzelner Kontakte mit impulsiven Kräften einen Lösungsansatz auf der Basis des Lemke-Algorithmus vorgestellt.

LCPs haben mittlerweile auch Eingang in den Bereich der Haarsimulation gefunden [Hadap, 2006]. Allerdings beschränkt sich das Modell auf die Simulation ganzer Haarsträhnen. Die Zahl der Kontakte ist in allen Fällen eher als klein zu betrachten, so dass der relativ hohe Berechnungsaufwand für entsprechende Lösungen vertretbar erscheint. Die Zahl potentieller simultaner Kontakte in den hier untersuchten Faseranordnungen hingegen lässt die Behandlung in Form eines LCPs unpraktisch erscheinen. Dabei ist weniger die Dauer des eigentlichen Lösungsprozesses, als vielmehr die algorithmische Unbestimmtheit der Lösung selbst das Problem. Mirtich [Mirtich, 1996] nennt hier den symmetrischen Tisch auf vier Beinen als Beispiel, bei dem unter idealen Verhältnissen die Kontaktkraft an allen vier Auflagepunkten identisch sein muss. Auf der anderen Seite ist die Zahl der Kraftkonfigurationen, welche die Anforderungen von Gl. (8.1) erfüllen, in diesem Fall unendlich groß. Das Problem wird ungleich komplizierter bei Einbezug von Reibung. Hier existieren mehrere Lösungen, die völlig unterschiedliche Beschleunigungen zur Folge haben können. Zeitliche Kohärenz wäre hier also gleichbedeutend mit Kontaktkohärenz, und ihr Einsatz zum beschleunigten Auffinden einer Lösung damit nur eingeschränkt möglich. So elegant die Formulierung als LCP auch auf den ersten Blick erscheinen mag,

so viele Probleme wirft sie doch auf, die sie für die Simulation von komplexen Faseranordnungen mit tausenden von reibungsbehafteten dreidimensionalen Kontakten ungeeignet erscheinen lässt.

8.1.3 Impulsbasierte Techniken

Eine weitere große Klasse von Verfahren zur Kollisionsbehandlung bilden die impuls- oder geschwindigkeitsbasierten Verfahren, vgl. [Hahn, 1988], [Mirtich, 1996]. Da Kollisionskräfte in der realen Welt über einen in Bezug auf die Wirkungsentfaltung von Trägheitskräften vernachlässigbar kurzen Zeitraum wirken und ihre Größe auf der anderen Seite nicht beschränkt ist, kann man das zeitliche Integral über die Kollisionskraft, den so genannten Kraftstoß betrachten. Dieser beschreibt die instantane Impulsänderung infolge einer über ein infinitesimales Zeitintervall wirkenden Kraft beliebiger Größe. Die Geschwindigkeitsänderung des Körpers entspricht gerade dem Kraftstoß geteilt durch seine Masse. Er wirkt in Richtung der Kollisionsnormalen und sein Betrag folgt direkt aus der Relativgeschwindigkeit am Berührungspunkt zweier Körper [Mason, 2001]. Wir werden an anderer Stelle noch näher darauf eingehen.

Die Behandlung von Kontakten auf der Basis von Geschwindigkeiten hat enorme Vorteile gegenüber den kraftbasierten oder analytischen Methoden. Mirtich [Mirtich, 1996] nennt hier u.a. eine fehlende Abhängigkeit von der Kontaktkohärenz, Robustheit, Erweiterbarkeit auf gelenkig verbundene Starrkörper, fehlende Kontaktdegeneriertheit im Gegensatz zu den analytischen Methoden.

8.1.4 Projektive Verfahren für komplexe Kontaktsituationen

Ein Problem, mit dem wir uns bei der hier anstehenden Aufgabe der Simulation von komplexen Faseranordnungen konfrontiert sehen, ist die hohe Zahl von einander abhängiger simultaner Kontakte. Im Gegensatz zu den im Bereich der Computergraphik etablierten Methoden zur Kollisionsbehandlung, die primär auf der Applikation von Strafkraften, der Anordnungsoptimierung, impulsbasierten oder rein analytischen Techniken basieren, liegt hier das Hauptaugenmerk auf der Kombination von projektiven Methoden mit dem Konzept und damit den Vorzügen einer impulsbasierten Simulation. Die Verhinderung von Durchdringungen erfolgt dabei durch Etablierung einer Menge einseitiger Zwangsbedingungen und folgt damit in gewisser Weise den analytischen Methoden. Diese Menge von Zwangsbedingungen spannt den Bereich zulässiger Systemreaktionen auf. Die Ermittlung von Postkollisionszuständen erfolgt durch Projektionen auf solche zulässigen Bereiche. Durch eine mathematisch fundierte Beschreibung des Stoßvorgangs mittels konvexer Analysis lassen sich die Systemreaktionen in den Geschwindigkeitsraum einbetten.

Die mathematische Behandlung von Systemen mit reibungsbehafteten Mehrfachkontakten findet ihr Fundament in der Theorie so genannter *Differentialinklusionen*. Wesentliche Beiträge gehen dabei auf Schatzman [Schatzman, 1973], Moreau [Moreau, 1988] und Monteiro Marques [Monteiro Marques, 1993] zurück. Differentialinklusionen sind – einfach ausgedrückt – Verallgemeinerungen von Differentialgleichungen, mit mengenwertiger Funktion auf der rechten Seite

und im einfachsten Fall von der Form

$$\frac{dx}{dt} \in F(x, t). \quad (8.2)$$

Die Werte von $F(x, t)$ sind geschlossene konvexe Mengen. Man fordert zudem, dass der Graph $\{(x, y) | y \in F(x, t)\}$ von $F(x, t)$ eine geschlossene Menge bildet. Die Verwendung der Form Gl. (8.2) ist dabei essentiell für die Entwicklung von numerischen Lösungsverfahren für Differentialgleichungen mit unstetiger rechter Seite. Herkömmliche Ansätze tun sich schwer mit der Instabilität numerischer Trajektorien im unmittelbaren Bereich der Unstetigkeiten. Die mathematische Theorie solcher Inklusionen soll im Folgenden nur insoweit gestreift werden, als sie für das weitergehende Verständnis der Behandlung von Kollisionen unter Beachtung von Reibung notwendig ist.

Die mathematisch fundierte Beschreibung des Kollisionsprozesses auf der Basis der Theorie der Differentialinklusionen erlaubt Aussagen über die Existenz und Eindeutigkeit von Lösungen für das Problem simultaner Kontakte, solange man sich dabei, unter gewissen vereinfachenden Annahmen, auf Systeme ohne Reibung beschränkt.

Eines der ersten Verfahren, das auf Differentialinklusionen basiert, ist die von Moreau eingeführte *Contact Dynamics Method* [Moreau, 1988], die später zur *Non-Smooth Contact Dynamics Method* [Jean, 1999] erweitert wurde. Diese Verfahren haben sich im Bereich des Ingenieurwesens bei der Simulation hochkomplexer Systeme, mit zehntausenden von reibungsbehafteten simultanen dreidimensionalen Kontakten, wie sie bspw. typisch für die Granulat- oder Schüttgutsimulation sind, bewährt und sollen deshalb die Grundlage für die Kontaktbehandlung in dicht gepackten Faseranordnungen bilden.

8.1.4.1 Zeitschrittverfahren

Beinahe untrennbar verbunden mit dem Begriff der Differentialinklusionen für nicht glatte mechanische Systeme sind die so genannten Zeitschrittverfahren (engl.: *time-stepping schemes*), die in ihrer klassischen Form von Moreau eingeführt worden sind. Die Systemdynamik wird zunächst durch eine Orts- und eine Geschwindigkeitsfunktion auf einem endlichen Zeitintervall $[0, T]$ ausgedrückt. Die Tatsache, dass Körper einander nicht durchdringen können, erlegt dem System gewisse Zwänge auf, die in den klassischen Modellen der Kontaktbehandlung durch eine endliche Menge von Ungleichungen Ausdruck finden. Es sind eben jene Zwangsbedingungen, die für die Unstetigkeit der Systemvariablen verantwortlich zeichnen. Das Zeitschrittverfahren handhabt diese Zwänge auf der Ebene der Geschwindigkeiten. Wenn diese für (fast) alle Zeitpunkte auf dem Intervall $[0, T]$ innerhalb eines bestimmten Kegels liegen, ist die Gültigkeit der Ungleichungen, welche die Systemzwänge ausdrücken, über die Dauer des Zeitintervalls gesichert.

Die Beziehung zwischen der lokalen Kinematik an den Kontaktstellen und den dort auftretenden Kräften wird durch ein Kontaktgesetz beschrieben. Auch das klassische COULOMBSche Gesetz für trockene Reibung lässt sich in einer solchen Form ausdrücken. Im Kontext der durch die Systemzwänge induzierten Unstetigkeiten muss die Systemgeschwindigkeit als Funktion der Zeit mit beschränkter Variation aufgefasst werden. Infolgedessen ist sie nicht mehr differenzierbar im klassischen Sinne, sondern ihre Zeitableitung, die normalerweise der Systembeschleunigung entspricht, wird zum \mathbb{R}^n -wertigen Maß auf dem Intervall $[0, T]$, mit punktuellen Atomen im

Augenblick der Kollisionen. Das eigentliche Zeitschrittverfahren entsteht durch Integration der beiden Systemvariablen der Differentialinklusion auf einem Intervall $[t_i, t_f]$. Das liefert die neue Geschwindigkeit bzw. die neue Position der Systemmitglieder. Das Integral über die Kontaktkraft, der Kraftstoß, das unweigerlich durch den Integrationsvorgang entsteht, bedarf zur seiner Evaluierung einer entsprechenden modifizierten Form des Kontaktgesetzes, das die Beziehung zwischen der relativen Kontaktgeschwindigkeit am Ende des Intervalls und dem Kraftstoß herstellt. Moreau benutzt zur Integration ein implizites Verfahren. Das gewährleistet auf der einen Seite eine gewisse numerische Stabilität, hat aber auf der anderen Seite den gravierenden Nachteil, dass ein hochgradig nichtlineares Gleichungssystem zu lösen ist. Dieses kann z.B. mit dem nicht linearen GAUSS-Seidel-Verfahren in Angriff genommen werden.

Die Klasse der Zeitschrittverfahren bedient sich also der Diskretisierung der Systemdynamik, einschließlich der Kontaktbedingungen, ohne dabei die Idee gewöhnlicher Bewegungsgleichungen, welche die Systembeschleunigungen mit den Kräften in Beziehung setzen, aufzugreifen. Vielmehr erfahren die Differentiale hier eine Umdeutung im Sinne von Maßen. Die Integration des diskreten Gesamtsystems über ein fixes Zeitintervall liefert den neuen Systemzustand in Form von Orts- bzw. Geschwindigkeitsinkrementen der Systemkomponenten. Die Unterscheidung zwischen Kräften impulsiver und nicht impulsiver Natur entfällt, weil lediglich zeitliche Integrale über diese Kräfte betrachtet werden. Der Begriff der Beschleunigungen verliert in diesem Kontext seine Bedeutung, und mit ihm das Problem beliebig großer, stoßinduzierter Kräfte. Zeitschrittverfahren sind damit besonders geeignet für die Vorhersage der Systemdynamik komplexer Systeme mit tausenden von simultanen Kontakten. Der Verzicht auf eine ereignisbasierte Kollisionsbehandlung hat zur Folge, dass ein für das jeweilige Zeitintervall prognostiziertes Systemverhalten maximal approximativ sein kann, und dass Durchdringungen von Körpern bis zu einem bestimmten Grade nicht auszuschließen, ja vielmehr Bestandteil des Verfahrens sind. In gewissem Sinne kommt man damit der Realität so genannter „starrer“ Körper wesentlich näher, weil die Mitte des betrachteten Zeitintervalls der Phase maximaler Kompression während des Stoßvorgangs entspricht. Dieser „Realismus“ spiegelt sich in dem ursprünglich von Moreau [Moreau, 1988] für die Granulatsimulation entwickelten Zeitschrittverfahren wider. Der erste Schritt besteht in der zeitlichen Vorwärtsintegration der Systemdynamik um einen halben Zeitschritt. Die so erhaltenen Geschwindigkeiten und Positionen beschreiben den Zustand des Systems zum Zeitpunkt maximaler Kompression, von dem ausgehend die Systemdynamik über die zweite Hälfte des Intervalls entwickelt wird, und zwar durch Projektion der Geschwindigkeiten auf die erlaubten Zustände, die aus der Menge der unilateralen Zwangsbedingungen geformt werden. Es ist notwendig, die Kollisionen zwischen den Körpern in dieser maximal kompressiven Phase zu detektieren, weil jede Kollision einen einseitigen Zwang induziert, der die Menge erlaubter Zustände weiter einschränkt.

Den Zeitschrittverfahren steht die Klasse der ereignisgesteuerten (engl: *event driven*) Verfahren gegenüber. Das Integrationsgebiet wird als Vereinigung zunächst unabhängiger Teilintervalle aufgefasst, mit stetigen und folglich durch Standardintegratoren für ODE/DAE-Systeme handhabbaren Bewegungsabläufen. Die Teilintervalle werden durch typischerweise kollisionsgetriggerte Ereignisse voneinander separiert. Die ereignisgesteuerten Verfahren sind in diesem Sinne im Bereich der physikalisch basierten Simulation wohletabliert, wenn auch nicht unter diesem Namen bekannt, vgl. Mirtich [Mirtich, 1996]. Ihre Abhängigkeit von individuellen Kollisionsereignissen lässt sie für die Simulation komplexer Systeme mit tausenden von simultanen Kontakten, wie sie

in komplexen Faseranordnungen vorliegen, eher ungeeignet erscheinen.

8.1.4.2 Phase maximaler Kompression

Eine Kollision zwischen zwei Körpern ist ein atomares Ereignis und sollte als solches modelliert werden. Dabei ist das Verhalten von so genannten Starrkörpern vor, während und nach einer Kollision keineswegs starr, wie der Name suggerieren mag. An der Kontaktstelle treten vielmehr lokale, teils elastische, teils plastische, also bleibende Deformationen auf.

Der Stoßvorgang zwischen zwei Körpern teilt sich idealisiert in eine Phase der Kompression, in der eine Umwandlung von kinetischer- in Deformationsenergie stattfindet, sowie die darauf folgende Dekompressionsphase, in der durch Rückverformung die Körper gleichsam wieder voneinander weggestoßen werden. Der Übergang zwischen diesen beiden Phasen ist durch ein Maximum an gespeicherter elastischer Energie gekennzeichnet. Welche Prozesse sich beim Stoßvorgang genau abspielen bzw. für seine dissipative Natur verantwortlich zeichnen, konnte bis heute nicht eindeutig geklärt werden. Neuere Untersuchungen zeigen aber, dass neben lokal plastischen Deformationen, elastische Vibrationen, üblicherweise im Bereich von 100 Hz bis 1 kHz, eine Rolle spielen [Stewart, 2000]. Unter diesen Umständen sind genaue Vorhersagen zukünftiger Kollisionszeitpunkte nur schwer möglich, weil sie in der Größenordnung des Bruchteils solcher Vibrationsdauern liegen müssten. Bei Geschwindigkeiten von bspw. 1 m/s liegen die Genauigkeitsanforderungen dann bereits im Bereich von 0,1 mm.

Es wurde weiter oben bereits angedeutet, dass wir Moreau folgen, wenn wir die Kollisionen zwischen den Fasern dicht gepackter Anordnungen zum Zeitpunkt der maximalen Kompression simultan auflösen. Das impliziert das Auftreten von Durchdringungen zwischen den Fasern nach dem ersten Integrationshalbschritt, die als lokale Deformationen im Sinne des oben Gesagten interpretiert werden müssen. In diesem Zustand sind die Systemgeschwindigkeiten der Körper auf zulässige Zustände zurück zu projizieren. Dem lokalen Deformationsverhalten trägt man an dieser Stelle durch ein adäquates Stoßgesetz Rechnung. Hier haben sich die folgenden zwei Ansätze bewährt. Das Stoßverhalten lässt sich zum Einen durch Einführung eines Restitutionsfaktors (zwischen 0 und 1) regeln, der das Verhältnis zwischen der Geschwindigkeit vor und nach dem Stoß angibt und damit eine Modellierung zwischen voll elastisch und rein plastisch zulässt. Dieser Ansatz ist bekannt als NEWTONSches Stoßgesetz. Man betrachtet hier nur die Geschwindigkeitsanteile in Richtung der Stoßnormalen. Tangentialanteile müssen durch ein entsprechendes Reibungsgesetz gehandhabt werden.

Auf der anderen Seite steht das Stoßgesetz von POISSON, das den Restitutionskoeffizienten als Verhältnis von Normalimpuls in der Dekompressionsphase und dem in der vorausgehenden Kompressionsphase definiert. Der Normalimpuls ist das Integral über die Normalkraft. Nach keinem der beiden Ansätze kann eine Zunahme der Gesamtenergie ausgeschlossen werden, besonders dann nicht, wenn Reibung eine Rolle spielt. Abhilfe schafft die Hypothese von Stronge [Stronge, 1991], die den Koeffizienten mit dem Verhältnis der in den beiden Phasen verrichteten Arbeiten korreliert, also $\varepsilon^2 = -W_{dekomp}/W_{komp}$. Das negative Vorzeichen trägt dem Sachverhalt Rechnung, dass eine der beiden Energien negativ sein muss. Dieser Ansatz garantiert, dass Normal- wie Tangentialkräfte immer dissipativen Charakter haben. Wir werden uns bei der Stoßmodellierung jedoch auf den Ansatz von NEWTON verlassen.

An dieser Stelle sei noch auf den Unterschied zwischen den Begriffen „Kontakt“ und „Kolli-

sion” hingewiesen. Bei einer Kollision liegt der Abstand zwischen den beiden nächsten Punkten auf den Rändern zweier Körper innerhalb eines Toleranzintervalls und die Relativgeschwindigkeit der beiden Punkte, projiziert auf die Stoßnormale, weist in negative Richtung. Bei Kontakten hingegen ist die erste Bedingung erfüllt, aber die Relativgeschwindigkeit in Richtung der Stoßnormalen muss verschwinden.

8.2 Systeme mit unilateralen Zwangsbedingungen

Im Folgenden beziehen wir unsere Betrachtungen auf hinreichend starre Körper $\mathcal{B} := \{\mathbf{q}, \Omega\}$. Sei dabei $\mathcal{M} = SE(3)$ der Konfigurationsraum und $\mathbf{q}(t) \in \mathcal{M}$ die aktuelle Konfiguration des Körpers \mathcal{B} zum Zeitpunkt t . Der Körper schließe einen nicht notwendigerweise konvexen Bereich $\Omega \subset \mathbb{R}^3$ des Raumes ein. Des Weiteren wollen wir den Kontaktpunkt auf der Berandung von \mathcal{B} mit $\mathbf{p} \in \partial\Omega$ bezeichnen.

Wir folgen Monteiro Marques [Monteiro Marques, 1993], wenn wir a) vereinfachend annehmen, dass die Körper durch Punkte im Raum repräsentiert werden und b) für die Herleitung der Theorie der Mehrfachkontakte zunächst die Orientierung der Körper vernachlässigen, indem wir den Konfigurationsraum auf den $\mathcal{M} = \mathbb{R}^3$ beschränken. Der Einbezug der Orientierung bzw. der Drehgrößen wie Drehmoment und Winkelgeschwindigkeit erfolgt dann in einem anschließenden Schritt unter Einführung entsprechender mathematischer Operatoren.

Der Körper \mathcal{B} unterliege einem unilateralen Zwang, einer Undurchdringbarkeitsbedingung, die geometrisch durch eine Ungleichung der Form $h(\mathbf{q}(t)) \geq 0$, beschrieben wird, wobei $h : \mathcal{M} \mapsto \mathbb{R}$ eine reellwertige Funktion ist. Der Zwang ist hier im Sinne einer Abstandsfunktion zu verstehen, die genau dann von null verschiedene Werte annimmt, wenn der Körper von denen durch sie definierten Hyperflächen wohlsepariert ist. Ein Kontakt liegt genau dann vor, wenn $h(\mathbf{q}(t)) = 0$ ist. Es ist zulässig, das „ ≥ 0 “ gegen ein „ ≤ 0 “ einzutauschen. Der Funktion $h(\mathbf{q}(t))$ kommt dann die Bedeutung einer „Lückenfunktion“ (engl.: *gap function*) zu, welche die Eindringtiefe in Richtung der Hyperfläche misst. Für den allgemeinen Fall eines Körpers mit n Zwängen indiziert man die einzelnen Bedingungen,

$$h_\alpha(\mathbf{q}(t)) \geq 0, \quad \alpha = 1..n, \quad (8.3)$$

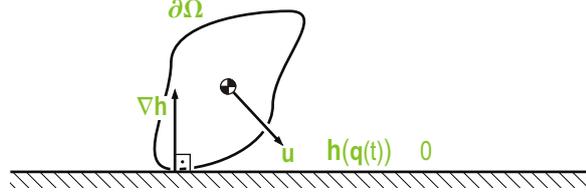
woraus die die Menge aller für den Körper \mathcal{B} zulässigen Systemzustände als

$$\Phi(t) := \{\mathbf{q}(t) \in \mathcal{M} : h_\alpha(\mathbf{q}(t)) \geq 0\} \quad (8.4)$$

folgt.

Anders ausgedrückt beschreibt Gl. (8.4) diejenigen Konfigurationen $\mathbf{q}(t) \in \mathcal{M}$ des Körpers, welche alle unilateralen Zwangsbedingungen erfüllen und den Körper dadurch in einem durchdringungsfreien Zustand halten. Die Zwangsbedingungen definieren an der Stelle $h(\mathbf{q}(t)) = 0$ die Begrenzungshyperflächen der durch Gl. (8.4) eingeschlossenen Region des Konfigurationsraumes. Der hier der Einfachheit halber zum Punkt degenerierte Körper wird durch diese Flächen in dieser Region gleichsam eingeschlossen. Man kann sie in gewissem Sinne als irgendwelchen Körper zugehörig interpretieren, wodurch \mathcal{B} in seiner Bewegungsfreiheit eingeschränkt wird, vgl. Abb. (8.2). Die Zahl solcher Hyperflächen kann sich im Bereich der Granulatsimulation auf die Milliarden belaufen [Moreau, 2003]. Moreau bezeichnet dies als räumliche Unstetigkeit.

Abbildung 8.2: Zur Veranschaulichung des Begriffes unilateraler Zwangsbedingungen.



Wenn man von der Funktion $h(\mathbf{q}(t))$ fordert, dass ihre erste Ableitung zumindest in der Nachbarschaft von $h(\mathbf{q}(t)) = 0$, also an der Kontaktstelle, existiert, steht der Gradient

$$\mathbf{n}_\alpha(\mathbf{q}(t)) = \frac{\nabla h_\alpha(\mathbf{q}(t))}{\|\nabla h_\alpha(\mathbf{q}(t))\|} \quad (8.5)$$

orthogonal auf der durch $h(\mathbf{q}(t)) = 0$ definierten Hyperfläche und bildet damit die Kontaktnormale. Der Wert der Zwangsbedingung alleine lässt allerdings keine Aussagen über die Dauerhaftigkeit eines Kontakts zu. Weist seine Geschwindigkeit weg von der Hyperfläche, muss er als transient gelten, denn er ist im Auflösen inbegriffen, bewegt sich also von der Hyperfläche weg. Die Geschwindigkeit $\mathbf{u} = d\mathbf{q}/dt$ muss eine Komponente antiparallel zur Richtung der Kontaktnormalen besitzen, also in Richtung der Zwangsfläche weisen, damit ein Kontakt als persistent oder aktiv identifiziert werden kann. Liegt die Geschwindigkeit dagegen in der Hyperfläche selbst, bedeutet das nicht notwendigerweise, dass der Kontakt, den wir dann als „glatt“ bezeichnen, dauerhaft ist. Aus algorithmischen Gründen werden wir diesen Fall, dessen Auftreten unter numerischen Gesichtspunkten ohnehin als eher unwahrscheinlich anzusehen ist, trotzdem in die Klasse persistenter Kontakte einordnen. Formal schreiben wir,

$$\text{Kontakt} := \begin{cases} \text{persistent} & \Leftrightarrow h_\alpha(\mathbf{q}(t)) = 0 \wedge \mathbf{n}_\alpha^\top \mathbf{u}_\alpha^- < 0, \\ \text{transient} & \Leftrightarrow h_\alpha(\mathbf{q}(t)) = 0 \wedge \mathbf{n}_\alpha^\top \mathbf{u}_\alpha^- > 0, \\ \text{glatt} & \Leftrightarrow h_\alpha(\mathbf{q}(t)) = 0 \wedge \mathbf{n}_\alpha^\top \mathbf{u}_\alpha^- = 0, \end{cases} \quad (8.6)$$

An dieser Stelle sei darauf hingewiesen, dass \mathbf{u}_α beim starren Körper nicht mit dessen Geschwindigkeit korrespondiert, sondern der Geschwindigkeit am Kontaktpunkt \mathbf{p}_α entspricht. Sie ist bekanntermaßen als

$$\mathbf{v}_\alpha = \dot{\mathbf{p}}_\alpha = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{r}_\alpha \quad (8.7)$$

zu berechnen, wobei \mathbf{r}_α der Richtungsvektor vom Schwerpunkt des Körpers zum Kontaktpunkt \mathbf{p}_α ist. Der Drehanteil in Richtung der Stoßnormalen, $\boldsymbol{\omega}_\alpha$, kann hier vernachlässigt werden.

Weiterhin ist zu beachten, dass die bisher als statisch angenommene Zwangsbedingung auch beweglich sein kann, z.B., wenn sie durch die Berandung eines weiteren Körpers gebildet wird. In diesem Fall entspricht \mathbf{u}_α der Relativgeschwindigkeit der beiden Körper am Kontaktpunkt.

Auf dieser Basis ist die Gesamtmenge der Kontakte gegeben durch

$$\mathcal{U}(t) := \{\mathbf{q}(t) \in \mathcal{M} : h_\alpha(\mathbf{q}(t)) \leq 0\}, \quad (8.8)$$

wohingegen die der aktiven Kontakte durch die folgende Aktivierungsmultifunktion $\mathcal{A}(t)$ beschrieben wird,

$$\mathcal{A}(t) := \{\mathbf{q}(t) \in \mathcal{M} : h_\alpha(\mathbf{q}(t)) \leq 0 \wedge \mathbf{n}_\alpha^\top \mathbf{u}_\alpha^- \leq 0\}. \quad (8.9)$$

Mit Definition Gl. (8.8) werden all jene Fälle eingeschlossen, in denen der Zwang bereits verletzt ist. Das hat pragmatische Gründe, denn die Verletzung einer Bedingung kann nur durch die Mechanismen der Kollisionserkennung detektiert werden und diese ist notwendigerweise diskret. Die Aktivierungsmultifunktion legt die zum Zeitpunkt t aktiven Kontakte über die Persistenzbedingung fest. Man erspart sich dadurch die algorithmisch aufwendige Ermittlung von Postkollisionsgeschwindigkeiten für Zustände, die sich ohnehin in einer der folgenden Zeitschritte auflösen. Die Aktivierung transienter Kontakte hat keine Auswirkungen auf den Systemzustand.

Betrachten wir einen einzigen Kontakt, muss die Geschwindigkeit nach dem Stoß offenbar dem Tangentialhalbraum $\mathcal{T}_\alpha(t)$ zur Menge zulässiger Systemzustände gem. Gl. (8.4) angehören, den wir wie folgt definieren,

$$\mathcal{T}_\alpha(t) := \{\mathbf{u}^* \in T_p\mathcal{M} : (\mathbf{u}^*)^\top \nabla h_\alpha(\mathbf{q}(t)) \geq 0\}. \quad (8.10)$$

Die Geschwindigkeit selbst ist Element des Tangentialraumes $T_p\mathcal{M}$ von \mathcal{M} . Bei Mehrfachkontakten legt jeder Kontakt der Menge $\mathcal{U}(t)$ genau einen solchen Halbraum fest, so dass die Geschwindigkeit nach dem Stoß im Schnitt aller Tangentialhalbräume $\mathcal{T}_\alpha(t)$, dem so genannten Tangentenkegel $\mathcal{T}(\mathbf{q}(t))$ zur Menge zulässiger Systemzustände gem. Gl. (8.4) im Punkt \mathbf{q} liegen muss,

$$\mathcal{T}(\mathbf{q}(t)) := \begin{cases} \bigcap_{\alpha \in \mathcal{U}(t)} \{\mathbf{u}^* \in T_p\mathcal{M} : (\mathbf{u}^*)^\top \nabla h_\alpha(\mathbf{q}(t)) \geq 0\} & , \text{ falls } \mathcal{U}(t) \neq \emptyset \\ \mathcal{M} & , \text{ falls } \mathcal{U}(t) = \emptyset. \end{cases} \quad (8.11)$$

Definition Gl. (8.11) gewährleistet, dass der Tangentenkegel auch bei leerer Kontaktmenge $\mathcal{U}(t) = \emptyset$ durch Zuweisung des gesamten Konfigurationsraumes eine sinnvolle Bedeutung erhält. Der Tangentenkegel ist konvex und bildet die Menge der kinematisch zulässigen Rechts- oder Postkollisionsgeschwindigkeiten. Falls die Geschwindigkeit vor dem Stoß bereits in $\mathcal{T}_\alpha(t)$ liegt, ist sie, solange wir den Fall ohne Reibung betrachten, bereits mit der Geschwindigkeit nach dem Stoß identisch.

Wir wissen jetzt, dass die Menge möglicher Postkollisionsgeschwindigkeiten durch den Tangentenkegel festgelegt wird. Da diese Menge unendlich groß ist, bedarf es einer geeigneten Selektionsregel. Unter Annahme eines rein dissipativen bzw. inelastischen Stoßverhaltens, muss nach Moreau [Moreau, 1988] die Geschwindigkeit auf der Berandung des Tangentenkegels, $\partial\mathcal{T}(\mathbf{q}(t))$ gem. Gl. (8.11), liegen. Sie entspricht dem proximalen Punkt der Linksgeschwindigkeit in der Menge der kinematisch zulässigen Rechtsgeschwindigkeiten. Das drücken wir durch die folgende Projektion aus,

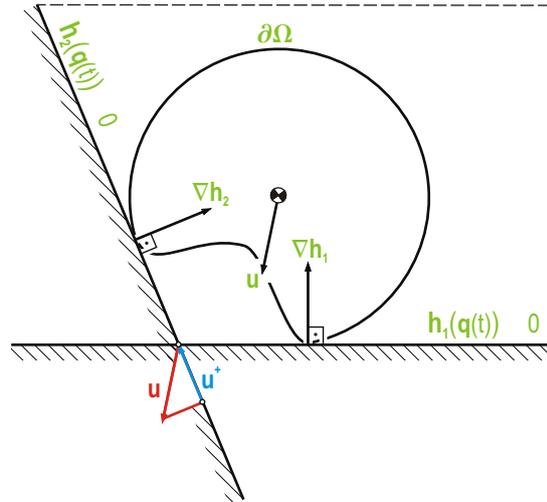
$$\mathbf{u}^+ = \text{proj} \left\{ \mathbf{u}^-, \mathcal{T}(\mathbf{q}(t)) \right\}, \quad \mathbf{u}^+ \in \partial\mathcal{T}_\alpha(t). \quad (8.12)$$

Gesucht ist der proximale Punkt \mathbf{u}^+ zu \mathbf{u}^- auf dem Rand des Tangentenkegels. Die Suche nach dem proximalen Punkt führt mathematisch auf ein konvexes Programmierungsproblem. Das bekannteste Lösungsverfahren ist die *Proximalpunktmethode*, vgl. [Moreau, 1966]. Gl. (8.12) behält ihre Gültigkeit selbst dann, wenn keine Kontakte vorliegen, also $\mathcal{U}(t) = \emptyset$. Der proximale Punkt entspricht dann gerade der Linksgeschwindigkeit.

Aus Gl. (8.12) folgt weiterhin, dass $\|\mathbf{u}^+\| < \|\mathbf{u}^-\|$ ist und damit

$$E^+ = \frac{1}{2} m \|\mathbf{u}^+\|^2 < E^- = \frac{1}{2} m \|\mathbf{u}^-\|^2. \quad (8.13)$$

Abbildung 8.3: Zur Ermittlung der Postkollisionsgeschwindigkeit. Sie entspricht gerade der Projektion auf den Rand des Tangentenkegels. Der Tangentenkegel ist die Schnittmenge der beiden durch die Kollisionsnormalen definierten Halbräume.



Ein Teil der kinetischen Energie geht also durch den Stoßvorgang (Projektion) verloren. Diese Projektion ist maximal dissipativ. Bei einem einzigen Kontakt würde der Stoß den Körper zur Ruhe bringen. Das entspricht dem „Ökonomieprinzip“, nach dem aus der Menge der kinematisch zulässigen Rechtsgeschwindigkeiten die am nächsten liegende gewählt wird. Ein Beispiel soll das verdeutlichen.

Beispiel 8.2.1 Der Körper aus Abb. (8.3) stehe in persistentem Kontakt mit den beiden Hyperflächen $h_1(\mathbf{q}(t)) = 0$ bzw. $h_2(\mathbf{q}(t)) = 0$. D.h., seine Geschwindigkeit besitze eine Komponente, die ihn ohne Gegenwirkung beide Zwänge verletzen lassen würde. Der Tangentenkegel entspricht gerade dem durch die beiden Flächen aufgespannten Kegel (der sichtbare Ausschnitt ist hier gestrichelt umrandet). Wir tragen den Geschwindigkeitsvektor im Ursprung des Tangentenkegels an. Den proximalen Punkt erhält man offensichtlich durch eine Projektion des Vektors auf $h_2(\mathbf{q}(t)) = 0$. Man beachte, dass, weil die Geschwindigkeit im negativen Tangentenkegel liegt, das Vorzeichen der Projektion umzudrehen ist.

8.2.1 Stöße und Reaktionen in Lagrange-Systemen

Die Beziehungen zwischen den Geschwindigkeiten vor und nach einem Stoß wurden im vorangegangenen Abschnitt hergeleitet. Hier wollen wir die Zusammenhänge zwischen den Systemgeschwindigkeiten und den am System operierenden stoßartigen oder impulsiven Kräften näher untersuchen. Als Ausgangsbasis dienen dabei die LAGRANGESchen Bewegungsgleichungen.

Für unser System gelten bei hinreichend glatten Bewegungen, d.h., bei zweimaliger Differenzierbarkeit der generalisierten Ortsfunktion $\mathbf{q}(t)$, die LAGRANGESchen Bewegungsgleichungen,

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} = \mathbf{F}(\mathbf{q}, \dot{\mathbf{q}}, t) + \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} + \boldsymbol{\Psi}(t), \quad (8.14)$$

mit der Massematrix $\mathbf{M}(\mathbf{q})$, den CORIOLIS-Kräften $\mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})$, sowie den kraftbasierten Standardtermen $\mathbf{F}(\mathbf{q}, \dot{\mathbf{q}}, t)$. Auf die rechte Seite von Gl. (8.14) schreiben wir zudem die kollisionsinduzierte Reaktionskraft

$$\boldsymbol{\Psi}(t) = \sum_{\alpha} \mathbf{v}_{\alpha}(t), \quad (8.15)$$

die offenbar im Gleichgewicht mit den Trägheitskräften sowie allen extern am Körper angreifenden Lasten stehen muss. Man beachte, dass die einzelnen Beiträge in Gl. (8.15) beliebig groß sein können. Da $\Psi(t)$ den unstetigen Teil der Gl. (8.14) repräsentiert, sind die Ableitungen nicht im konventionellen Sinne zu verstehen, sondern müssen vielmehr als Distributionen aufgefasst werden [Renouf et al., 2005]. Im Zusammenhang mit dem Begriff der Differentialinklusionen wird näher auf die Verschiebung dieser Bedeutung eingegangen.

In der Perkussionstheorie bemühen wir den Limes des Integrals über die Bewegungsgleichungen Gl. (8.14), und zwar für den Fall, dass die Dauer des Kollisionsprozesses gegen null geht, $t_c \rightarrow 0$. Wir erhalten

$$\lim_{t_c \rightarrow 0} \int_0^{t_c} [\mathbf{M}(\mathbf{q}(t), t) \dot{\mathbf{q}}(t) - \mathbf{C}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) \dot{\mathbf{q}}(t) - \mathbf{F}(\mathbf{q}(t), \dot{\mathbf{q}}(t), t) - \Psi(t)] dt = \mathbf{0}, \quad (8.16)$$

und weiter

$$\begin{aligned} & \mathbf{M}(\mathbf{q}(t), t) (\dot{\mathbf{q}}(t_c) - \dot{\mathbf{q}}(0)) \\ & - \lim_{t_c \rightarrow 0} \int_0^{t_c} [\mathbf{F}(\mathbf{q}(t), \dot{\mathbf{q}}(t), t) + \mathbf{C}(\mathbf{q}(t), \dot{\mathbf{q}}(t)) \dot{\mathbf{q}}(t) + \Psi(t)] dt = \mathbf{0}. \end{aligned} \quad (8.17)$$

Nicht impulsive Anteile unter dem Integral müssen verschwinden, während Position und Massematrix über die infinitesimale Dauer des Stoßes konstant sind. Wir nehmen im Folgenden die Identität zwischen der Konfiguration des Körpers $\mathbf{q}(t)$ und seinen generalisierten Koordinaten $\mathbf{q}(t)$ an und notieren die Zeitableitung $\dot{\mathbf{q}}(t)$ fortan als $\mathbf{u}(t)$.

Für die Linksgeschwindigkeit schreiben wir \mathbf{u}^- , für die Rechtsgeschwindigkeit, also jene genau nach dem Stoß, \mathbf{u}^+ . Wenn wir mit \mathbf{P} den Limes des Integrals über die impulsiven Kräfte $\mathbf{v}(t)$ bzw. den Kraftstoß bezeichnen, ergibt sich die Änderung der Geschwindigkeit eines Körpers infolge des Kraftstoßes in der folgenden kompakten Notation als

$$\mathbf{M}(\mathbf{q}(t)) [\mathbf{u}^+(t) - \mathbf{u}^-(t)] = \sum_{\alpha} \left[\lim_{t_c \rightarrow 0} \int_0^{t_c} \mathbf{v}(t) dt \right] = \sum_{\alpha} \mathbf{P}_{\alpha}. \quad (8.18)$$

Im Falle eines einzigen ($\alpha = 1$) perfekten, reibungslosen Kontaktes, erfährt der Körpers am kontaktierenden Punkt eine Reaktionskraft $\mathbf{v}(t) \in T_p^* \mathcal{M}$ in Richtung der Normalen $\nabla h(\mathbf{q}(t))$, von unbekanntem Betrage,

$$\exists \lambda \in \mathbb{R} : \quad \mathbf{v}(t) = \lambda \nabla h(\mathbf{q}(t)), \quad \lambda \geq 0. \quad (8.19)$$

Notwendige Bedingung für das Auftreten der Reaktionskraft ist die Erfüllung der Persistenzkondition. Setzen wir für $\mathbf{M} = \mathbf{E}$, dann erfüllt \mathbf{P} die Gleichung

$$\mathbf{P} = \mathbf{u}^+(t) - \mathbf{u}^-(t) = c \cdot \nabla h(\mathbf{q}), \quad c > 0. \quad (8.20)$$

Das ist leicht einzusehen, wenn man sich vergegenwärtigt, dass nach Moreau oder dem Ökonomieprinzip der Stoß maximal dissipativ sein und damit der orthogonalen Projektion in den Tangentialhalbraum $\mathcal{T}(\mathbf{q})$ entsprechen muss, vgl. Gl. (8.12). Bei einem Kontakt gilt dann, $\text{proj} \{ \mathbf{u}^-, \mathcal{T}(\mathbf{q}) \} = \mathbf{u}^- - [\mathbf{u}^- \cdot \nabla h(\mathbf{q}) / \|\nabla h(\mathbf{q})\|]$. Auf einer abstrakteren Ebene ist aber für

den Fall $\alpha = 1$ die Normale $\nabla h(\mathbf{q})$ nichts anderes als der nach außen gerichtete Normalenkegel $\mathcal{N}_{\mathcal{T}(\mathbf{q}(t))}(\mathbf{u}^+)$ zum Tangentialraum $\mathcal{T}(\mathbf{q}(t))$ am Punkte $\mathbf{u}^+(t)$. Die Änderung der Geschwindigkeit muss Element dieses Kegels sein und damit auch die Reaktionskraft auf den Stoß, Gl. (8.20), oder formaler

$$-(\mathbf{u}^+(t) - \mathbf{u}^-(t)) \in \mathcal{N}_{\mathcal{T}(\mathbf{q}(t))}(\mathbf{u}^+(t)), \quad \mathbf{u}^+(t) \in \mathcal{T}(\mathbf{q}(t)). \quad (8.21)$$

Für Einzelkontakte haben wir durch Gl. (8.21) die Äquivalenz zwischen der Projektion Gl. (8.12) und der Normalreaktion anschaulich hergestellt. Bei mehreren simultanen Kontakten verteilt sich die Gesamtreaktion über die einzelnen Berührungspunkte. Auch hier gilt dann, dass die Gesamtreaktionskraft auf alle simultanen Stöße bzw. die dadurch bewirkte Geschwindigkeitsänderung ein Element des Normalenkegels zum Tangentialkegel ist, der dem Schnitt der durch die Stoßnormalen definierten Halbräume entspricht. Die Äquivalenz folgt hier nach Anwendung elementarer Aussagen der konvexen Analysis. Formaler schreiben wir die folgende Definition. Der nach außen gerichtete Normalenkegel zum Tangentialkegel am Punkte \mathbf{u}^+ ist die Menge aller Geschwindigkeiten $\mathbf{u}^* \in T_p\mathcal{M}$, die folgende Bedingung erfüllen:

$$\mathcal{N}_{\mathcal{T}(\mathbf{q}(t))}(\mathbf{u}^+) := \left\{ \mathbf{u}^* \in T_p\mathcal{M} \mid \forall \mathbf{v} \in \mathcal{T}(\mathbf{q}(t)) : (\mathbf{v} - \mathbf{u}^+)^T \mathbf{u}^* \leq 0 \right\}. \quad (8.22)$$

Diese Definition ist äußerst unhandlich. Wegen Gl. (8.19) erscheint der folgende Ausdruck angebrachter. Demnach ist der Normalenkegel die konvexe Kombination der Beiträge entlang aller Kontaktnormalen,

$$\mathcal{N}_{\mathcal{T}(\mathbf{q}(t))}(\mathbf{q}(t)) := \left\{ \sum_{\alpha \in \mathcal{A}(t)} \lambda_\alpha \nabla h_\alpha(\mathbf{q}(t)), \lambda_\alpha \geq 0 \right\}. \quad (8.23)$$

Mit Gl. (8.15) und Gl. (8.19) folgt daraus aber

$$\mathcal{N}_{\mathcal{T}(\mathbf{q}(t))}(t) := \sum_{\alpha} \mathbf{v}_\alpha(t) = \mathbf{\Psi}(t). \quad (8.24)$$

Ob man über die Menge aller Kontakte oder lediglich über die der aktiven summiert, ist dabei völlig unbedeutend, weil transiente Kontakte, wie weiter oben bereits angedeutet wurde, den Systemzustand unverändert lassen.

Die elementare Erkenntnis, die sich einstellt, ist die, dass der Normalenkegel gerade der konvexen Summe der Reaktionskräfte entspricht. Anders ausgedrückt, hat sich mit dem Normalenkegel eine Möglichkeit zur Einbettung der Reaktionskräfte in den Raum der Geschwindigkeiten aufgetan. Es ist genau diese Beziehung, die eine Vorhersage der Postkollisionsgeschwindigkeit infolge der stoßartigen am System operierenden Kräfte erlaubt und damit den Schlüssel zu einer rein impulsbasierten Behandlung von Kollisionen zwischen starren Körpern darstellt. Die Geschwindigkeiten sind Elemente des Tangentialraumes $T_p\mathcal{M}$ an den Konfigurationsraum, während die Kräfte dem Kotangentialraum $T_p^*\mathcal{M}$ angehören. Beide sind dual zueinander. Das heißt, dass zwischen ihnen eine bijektive Abbildung existiert, welche die Reaktionskräfte in äquivalente Geschwindigkeiten transformiert und umgekehrt,

$$T_p\mathcal{M} \xrightleftharpoons[\mathbf{M}^{-1}]{\mathbf{M}} T_p^*\mathcal{M} \quad (8.25)$$

Diese Abbildung entspricht nach Gl. (8.20) der Massematrix, weil sie die Geschwindigkeitsänderung und die Kraftstöße in Beziehung zueinander setzt. Im Beispiel zu Gl. (8.20) haben wir sie vereinfachend mit der Identität gleichgesetzt. Ist also die Kraft in den Geschwindigkeitsraum einzubetten, multipliziert man mit der Inversen der Massematrix und umgekehrt.

Weil die Gesamtheit der Reaktionskräfte genau der Summe der Terme in Gl. (8.14) entsprechen muss, dabei $\Psi(t)$ ausgenommen, muss diese im Normalenkegel liegen. Durch Umformung von Gl. (8.14) wird das deutlich,

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} - \mathbf{F}(t, \mathbf{q}, \dot{\mathbf{q}}) - \mathbf{C}(\mathbf{q}, \dot{\mathbf{q}})\dot{\mathbf{q}} \in \mathcal{N}_{\mathcal{T}(\mathbf{q}(t))}. \quad (8.26)$$

Gleichungen dieser Art bezeichnen wir als *Differentialinklusionen*. Die in Erscheinung tretenden Ableitungen erfahren jedoch noch einige Umdeutungen, die wir später besprechen.

8.2.2 Stoßgesetze

Die Gesamtimpulsänderung infolge aller Einzelreaktionen ist durch Gl. (8.21) implizit gegeben. Für weitere Berechnungen ist es oft von Vorteil, wenn man die Beiträge der einzelnen Kontakte zumindest näherungsweise kennt. Dazu ist die Gesamtimpulsänderung auf die einzelnen Stoßnormalen zu projizieren. Sie entspricht dabei gerade der Differenz der Geschwindigkeiten vor und nach dem Stoß multipliziert mit der Massematrix. An den Berührungspunkten erhalten wir damit folgende Einzelbeiträge,

$$\mathbf{v}_\alpha = \left[\mathbf{n}_\alpha^T \mathbf{M}(\mathbf{u}^+ - \mathbf{u}^-) \right] \mathbf{n}_\alpha = \left(\mathbf{n}_\alpha^T \mathbf{M} \Delta \mathbf{u} \right) \mathbf{n}_\alpha. \quad (8.27)$$

Die Reaktion, die durch Gl. (8.21) beschrieben wird, ist maximal dissipativ. D.h., dass der Körper bei nur einem Kontakt nach dem Stoß kein elastisches Verhalten zeigt, weil die Normalkomponente der Reaktion verschwindet. Das entspricht einem Klumpen Ton, der auf einer Tischplatte aufschlägt. Das Verhalten ist völlig plastisch. Es wurde in diesem Zusammenhang eingangs schon darauf hingewiesen, dass dem lokalen Deformationsverhalten durch ein adäquates Stoßgesetz Rechnung getragen wird. Eine geeignete kinetische Normalkomponente nach diesem völlig inelastischen Stoß folgt bspw. aus dem Ansatz von NEWTON. Die Normalkomponente nach dem Stoß entspricht nach diesem empirischen Gesetz ε -mal der negativen Normalkomponente vor dem Stoß,

$$u_N^+ = -\varepsilon u_N^-. \quad (8.28)$$

Der Restitutionskoeffizient ε bestimmt den Grad der Elastizität. Die Geschwindigkeit u_N ist als solche nicht bekannt, weil sie sich bei simultanen Kontakten auf mehrere Stoßnormalen verteilt. Wir kennen allerdings die Normalreaktion bzw. die dazu äquivalente Geschwindigkeitsänderung

$$\Delta \mathbf{u} = \mathbf{u}^+ - \mathbf{u}^-. \quad (8.29)$$

Im Gegensatz zu u_N hat $\Delta \mathbf{u}$ wegen Gl. (8.21) schon das „richtige“ Vorzeichen. Die Geschwindigkeit nach dem Stoß mit elastischem Anteil ist dann wie folgt definiert,

$$\mathbf{u}_e = \mathbf{u}^+ + \varepsilon \Delta \mathbf{u} = (1 + \varepsilon) \mathbf{u}^+ - \varepsilon \mathbf{u}^-. \quad (8.30)$$

An dieser Stelle sei angemerkt, dass der Körper infolge einer unverminderten tangentialen Komponente entlang der Tangentialebene gleitet. Die dazu notwendigen Reibungseffekte wurden bisher noch nicht berücksichtigt.

8.2.3 Differentialinklusionen

Non Smooth Dynamical Systems (NSDS) oder nicht glatte dynamische Systeme lassen sich durch die Diskontinuität der zeitlichen Entwicklung ihrer Bewegungsgleichungen charakterisieren. Für unilaterale Kontaktphänomene ist der so genannte *Moreau Sweeping Prozess (MSP)* erster Ordnung mathematisch gegeben durch die folgenden Gleichungen,

$$d\mathbf{q}(t) - F(t, \mathbf{q}(t))dt = \lambda \in \mathcal{N}_{\Phi(t)}(\mathbf{q}(t)), \quad \mathbf{q} \in \mathbb{R}^n, \lambda \in \mathbb{R}^m, \quad (8.31)$$

mit dem nach außen gerichteten Normalenkegel \mathcal{N} zur konvexen Menge zulässiger Systemzustände $\Phi(t)$ am Punkte $\mathbf{q}(t)$ des HILBERT-Raumes. Ihre Zeitabhängigkeit macht $\Phi(t)$ zu einer „beweglichen“ Menge (engl. *moving set*). Anschaulich beschreibt der zunächst wenig intuitive Begriff des „Sweeping-Prozesses“, wie ein Partikel an der Stelle $\mathbf{q}(t)$ innerhalb der durch die Funktion $\Phi(t)$ beschriebenen konvexen Menge von dieser gleichsam hinweggefegt wird, unter der Bedingung, dass er zu jedem Zeitpunkt ein Element dieser Menge sein soll, $\mathbf{q}(t) \in \Phi(t)$. Liegt $\mathbf{q}(t)$ dabei im Inneren von $\Phi(t)$, ist die Ableitung $\dot{\mathbf{q}}(t) = 0$. Liegt $\mathbf{q}(t)$ dagegen auf dem Rand von $\Phi(t)$, muss die Ableitung $\dot{\mathbf{q}}(t)$ im Normalenkegel $\mathcal{N}_{\Phi(t)}(\mathbf{q}(t))$ zur Menge $\Phi(t)$ am Punkte $\mathbf{q}(t)$ liegen. Falls $\Phi(t)$ unstetig ist, kann in der Folge $\mathbf{q}(t)$ auch unstetig sein. Tritt dieser Fall ein, muss $\mathbf{q}(t^+) - \mathbf{q}(t^-)$ im Normalenkegel $\mathcal{N}_{\Phi(t)}(\mathbf{q}(t))$ liegen.

Der durch Gl. (8.31) definierte Sweeping Prozess erster Ordnung stellt eine so genannte Differentialinklusion dar. Hierbei wird nicht, wie bei Differentialgleichungen, der linken Seite ein bestimmter Punkt zugeordnet, sondern diese soll vielmehr innerhalb einer bestimmten konvexen Menge liegen. Eine Differentialinklusion stellt sich so gesehen als Verallgemeinerung des Konzeptes gewöhnlicher Differentialgleichungen dar. Differentialinklusionen erscheinen typischerweise im Kontext projektiver dynamischer Systeme, die wir in den vorangegangenen Abschnitten schon kennengelernt haben, oder bei Systemen mit COULOMB-Reibung, die noch folgen. In diesen Fällen zeigen die bestimmenden Differentialgleichungen eine diskontinuierliche rechte Seite. Kanonisches Beispiel ist der Block auf der schiefen Ebene. Unter Beachtung der Reibung finden wir im analytischen Ausdruck eine Abhängigkeit vom Vorzeichen der Geschwindigkeit.

Differentialinklusionen für sich genommen sind zunächst noch kein Allheilmittel für Systeme, die Stößen ausgesetzt sind. Erst in Kombination mit dem Konzept stoßartiger Kräfte führen sie uns auf den Begriff der *Maßdifferentialinklusion*. Der Sweeping-Prozess wurde von Moreau auf LAGRANGE-Systeme mit unilateralen Zwangsbedingungen erweitert. Die Gleichung, die wir mit Gl. (8.21) bereits kennengelernt haben und hier noch einmal in differentieller Form angeben,

$$-d\mathbf{u} \in \mathcal{N}_{\mathcal{T}(\mathbf{q})}(\mathbf{u}(t)), \quad \mathbf{u}(t) \in \mathcal{T}(\mathbf{q}), \quad (8.32)$$

ist eine solche Maßdifferentialinklusion. Wir haben sie als gewöhnliche Differentialinklusion eingeführt, wobei die Ausdeutung zur Maßdifferentialinklusion folgt. Der Ausdruck $d\mathbf{u}$ ist das mit der zweiten Ableitung von $\mathbf{q}(t)$ assoziierte Maß und $\mathcal{N}_{\mathcal{T}(\mathbf{q})}$ der Normalenkegel des Tangentialhalbraumes zur Menge zulässiger Systemzustände. Diese Begriffe müssen wir hier nicht mehr abstrakt fassen, weil sie in den vorangegangenen Abschnitten bereits eine Bedeutung erhalten haben. In einer Maßdifferentialinklusion besteht für die Funktion nicht die Notwendigkeit der Beschränktheit und die Geschwindigkeit muss keine absolut stetige Funktion sein, sondern man fordert lediglich ihre beschränkte Variation auf dem endlichen Intervall $[0, T]$. Das bedeutet, dass

das Supremum

$$\text{var}(\mathbf{u}, [0, T]) = \sup \left\{ \sum_{i=0}^{n-1} \|\mathbf{u}(t_{i+1}) - \mathbf{u}(t_i)\| \right\} < +\infty \quad (8.33)$$

für beliebige n und $0 \leq t_0 \leq t_1 \leq \dots \leq t_n \leq T$ endlich sein muss. In den klassischen LAGRANGESchen Bewegungsgleichungen müssen wir die Differentiale der Ableitung $d^2\mathbf{q}/dt^2 = d\mathbf{u}/dt$ umdeuten,

$$\mathbf{M}(\mathbf{q})d\mathbf{u} - [\mathbf{F}(t, \mathbf{q}, \mathbf{u}) + \mathbf{C}(\mathbf{q}, \mathbf{u})\mathbf{u}(t)]dt = \boldsymbol{\Psi}(t)dt \in \mathcal{N}_{\mathcal{T}(\mathbf{q})}(\mathbf{u}(t)). \quad (8.34)$$

dt ist das LEBESGUE-Maß und $d\mathbf{u}$ das der Geschwindigkeit $\mathbf{u} = \dot{\mathbf{q}}^+(t)$ zugeordnete STIELTJES-Maß, wobei die Geschwindigkeit als *rechtsstetige Funktion lokal beschränkter Variation* in obigem Sinne aufzufassen ist. Auch die LAGRANGE-Multiplikatoren in Gl. (8.23) sind in diesem Sinne Maße. Differentialinklusion Gl. (8.34) muss im Sinne von Differentialmaßen erfüllt sein. Wir führen zunächst ein Vektormaß, hier als Reaktionsmaß, ein

$$d\boldsymbol{\Xi} := \boldsymbol{\Psi}dt = \mathbf{M}(\mathbf{q})d\mathbf{u} - [\mathbf{F}(t, \mathbf{q}, \mathbf{u}) + \mathbf{C}(\mathbf{q}, \mathbf{u})\mathbf{u}(t)]dt. \quad (8.35)$$

Weiterhin sei $d\mu$ ein beliebiges nicht-negatives Maß, bezüglich dessen sowohl dt , als auch $d\mathbf{u}$ absolut stetig sind, so dass $dt = t'_\mu d\mu$ und $d\mathbf{u} = \mathbf{u}'_\mu d\mu$, mit den Dichten $t'_\mu \in L^1([0, T], d\mu, \mathbb{R})$ und $\mathbf{u}'_\mu \in L^1([0, T], d\mu, \mathcal{M})$. Es wird gefordert, dass

$$\frac{d}{d\mu}\boldsymbol{\Xi}(t) = \mathbf{M}(\mathbf{q})\mathbf{u}'_\mu(t) - [\mathbf{F}(t, \mathbf{q}, \mathbf{u}) + \mathbf{C}(\mathbf{q}, \mathbf{u})\mathbf{u}(t)]t'_\mu(t) \in \mathcal{N}_{\mathcal{T}(\mathbf{q})}(\mathbf{u}(t)), \quad (8.36)$$

für $d\mu$ -fast alle $t \in [0, T]$ gilt. Da die rechte Seite ein Kegel ist, muss das nur für ein einziges Maß $d\mu$ in den besagten Bedingungen geprüft werden. Im Falle einer Kollision hat $d\mathbf{u}$ ein Atom an der Stelle t_c . $d\mu$ ist dann identisch mit dem DIRAC-Maß δ_{t_c} , so dass $t'_\mu(t) = 0$ verschwinden muss und $\mathbf{u}'_\mu(t) = d\mathbf{u}/d\delta_{t_c} = \mathbf{u}^+(t) - \mathbf{u}^-(t)$. Daraus folgt sofort die Äquivalenz zwischen Gl. (8.36) und Gl. (8.21). Der zweite Fall, den zu betrachten wir uns anschicken, ist jener, bei dem keine Kollision auftritt und folglich dem System kein Zwang auferlegt wird. Für diesen Fall gilt per Definition Gl. (8.11), dass der Tangentenkegel gleich dem Konfigurationsraum ist und in der Folge der Normalenkegel zum Nullvektor degeneriert. Wir setzen $d\mu = dt$ für die Dauer des kollisionsfreien Intervalls und erhalten für das Reaktionsmaß $d\boldsymbol{\Xi} = 0$, was uns wiederum auf die LAGRANGESchen Bewegungsgleichungen führt.

Die wichtigste Frage, die sich in diesem Zusammenhang stellt, ist jene nach der Existenz einer Lösung und falls eine solche existiert, ob sie eindeutig ist. Unter der Annahme, dass die Gradienten der Zwangsbedingungen definiert sind, also $\nabla h_\alpha(\mathbf{q}(t)) \neq \mathbf{0}$, und die in Gl. (8.34) auftretenden Kraftfunktionen stetig und nach oben beschränkt sind, also keine stoßartigen Anteile enthalten, kann man für den hier betrachteten Multikontaktfall ohne Reibung das folgende zentrale Theorem formulieren, vgl. [Monteiro Marques, 1993]:

Satz 8.2.1 Existenz einer Lösung (Monteiro Marques): Seien $\mathbf{q}_0 \in \mathcal{M}$ die generalisierten Koordinaten eines Punktes und $\mathbf{u}_0 \in \mathcal{T}(\mathbf{q}_0)$ eine zulässige Geschwindigkeit zum Zeitpunkt $t = 0$. Dann existiert eine LIPSCHITZ-stetige Funktion $\mathbf{q} : [0, T] \mapsto \mathcal{M}$, mit $\mathbf{q}(0) = \mathbf{q}_0$, so dass

$$\mathbf{q}(t) = \mathbf{q}_0 + \int_0^t \mathbf{u}(\tau)d\tau, \quad t \in [0, T],$$

gilt, mit der Postkollisionsgeschwindigkeit $\mathbf{u} : [0, T] \mapsto T_p\mathcal{M}$, rechtsstetig und von beschränkter Variation, welche die folgenden Bedingungen im Sinne von Differentialmaßen erfüllt (s.o.):

- (a) $\mathbf{u}(0) = \mathbf{u}_0 = \dot{\mathbf{q}}^+(0)$,
- (b) $\mathbf{u}(t) \in \mathcal{T}(\mathbf{q}), \quad t \in [0, T]$,
- (c) $\mathbf{M}(\mathbf{q})d\mathbf{u} - [\mathbf{F}(t, \mathbf{q}, \mathbf{u}) + \mathbf{C}(\mathbf{q}, \mathbf{u})\mathbf{u}(t)]dt \in \mathcal{N}_{\mathcal{T}(\mathbf{q})}(\mathbf{u}(t))$.

Die etwas ermüdende Beweisführung findet man in besagtem Buch. An dieser Stelle sei noch einmal darauf hingewiesen, dass wir für unsere Betrachtungen den Körper zu einem Punkt degeneriert haben.

Es existiert also eine Lösung für den Multikontaktfall ohne Reibung, allerdings ist sie nicht eindeutig. Für den Fall mit Reibung, den wir im Folgenden besprechen, kann die Existenz einer Lösung nur garantiert werden, wenn man sich auf einen einzigen Kontakt beschränkt. Auch hier ist Eindeutigkeit nicht gegeben.

8.2.4 Diskretisierung

Zur numerischen Behandlung der Kollisionen in komplexen Faseranordnungen ist es notwendig, die Maßdifferentialinklusion Gl. (8.34) zu diskretisieren und auf dem Intervall $\Delta t = (t_i, t_{i+1}]$ numerisch über die Zeit integriert. Unter Anwendung des Differentialmaßes $d\mathbf{u}(\Delta t) = \mathbf{u}(t_{i+1}) - \mathbf{u}(t_i) \approx \mathbf{u}^{i+1} - \mathbf{u}^i$ auf dem Zeitintervall Δt erhält man

$$\mathbf{M}(\mathbf{q}^{i+1})(\mathbf{u}^{i+1} - \mathbf{u}^i) = \int_{t_i}^{t_{i+1}} \mathbf{F}(t, \mathbf{q}, \mathbf{u}) dt + \int_{t_i}^{t_{i+1}} \mathbf{C}(\mathbf{q}, \mathbf{u})\mathbf{u}(t) dt + \Psi^{i+1}, \quad (8.37)$$

$$\mathbf{q}^{i+1} = \mathbf{q}^i + \int_{t_i}^{t_{i+1}} \mathbf{u}(t) dt, \quad (8.38)$$

vgl. [Acary & Brogliato, 2005]. Mittels des Theta-Verfahrens lassen sich die Integrale wie folgt approximieren,

$$\int_{t_i}^{t_{i+1}} \mathbf{F}(t, \mathbf{q}, \mathbf{u}) dt = \Delta t [(1 - \theta)\mathbf{F}(t_{i+1}, \mathbf{q}^{i+1}, \mathbf{u}^{i+1}) + \theta\mathbf{F}(t_i, \mathbf{q}^i, \mathbf{u}^i)], \quad (8.39)$$

$$\int_{t_i}^{t_{i+1}} \mathbf{C}(\mathbf{q}, \mathbf{u})\mathbf{u}(t) dt = \Delta t [(1 - \theta)\mathbf{C}(\mathbf{q}^{i+1}, \mathbf{u}^{i+1})\mathbf{u}^{i+1} + \theta\mathbf{C}(\mathbf{q}^i, \mathbf{u}^i)\mathbf{u}^i], \quad (8.40)$$

$$\mathbf{q}^{i+1} = \mathbf{q}^i + \Delta t [(1 - \theta)\mathbf{u}^{i+1} + \theta\mathbf{u}^i], \quad (8.41)$$

mit $\theta \in [1/2, 1]$. Für $\theta = 1$ entspricht das gerade dem EULER-Verfahren. Für die Geschwindigkeit zum Zeitpunkt $t + \Delta t$ folgt damit

$$\mathbf{u}^{i+1} = \mathbf{u}^i + \mathbf{M}(\mathbf{q}^{i+1})^{-1} \left\{ \Delta t [(1 - \theta)\mathbf{F}(t_{i+1}, \mathbf{q}^{i+1}, \mathbf{u}^{i+1}) + \theta\mathbf{F}(t_i, \mathbf{q}^i, \mathbf{u}^i) + (1 - \theta)\mathbf{C}(\mathbf{q}^{i+1}, \mathbf{u}^{i+1})\mathbf{u}^{i+1} + \theta\mathbf{C}(\mathbf{q}^i, \mathbf{u}^i)\mathbf{u}^i] + \Psi^{i+1} \right\}. \quad (8.42)$$

Die Prä- bzw. Postkollisionsgeschwindigkeiten \mathbf{u}^- und \mathbf{u}^+ , von denen bekannt ist, dass sie sich um die Reaktion Ψ unterscheiden, folgen unmittelbar,

$$\mathbf{u}^- = \mathbf{u}^i + \Delta t \mathbf{M}(\mathbf{q}^{i+1})^{-1} [(1 - \theta)\mathbf{F}^{i+1} + \theta\mathbf{F}^i + (1 - \theta)\mathbf{C}^{i+1}\mathbf{u}^{i+1} + \theta\mathbf{C}^i\mathbf{u}^i], \quad (8.43)$$

$$\mathbf{u}^+ = \mathbf{u}^- + \mathbf{M}(\mathbf{q}^{i+1})^{-1}\Psi^{i+1}. \quad (8.44)$$

Wir benutzen dabei die Approximation $\mathbf{F}^i \approx \mathbf{F}(t_i, \mathbf{q}^i, \mathbf{u}^i)$, für \mathbf{C}^i , \mathbf{F}^{i+1} , \mathbf{C}^{i+1} gilt Entsprechendes. Falls $\mathbf{q} \in \mathcal{SE}(3)$, also nicht nur die Position, sondern auch die Orientierung des Körpers zu berücksichtigen ist, wird der Rotationsanteil \mathbf{R}^i wie in Abs. (3.4.5) beschrieben integriert.

8.3 Ein anisotropes Reibungsmodell für Keratinfasern

Im vorangegangenen Abschnitt wurde auf der Basis der konvexen Analysis die Mechanismen für die Selektion geeigneter Normalreaktionen als Folge der Verletzung von Zwangsbedingungen in Form von Kollisionen entwickelt. Die tangentialen Geschwindigkeitskomponenten, die Gleiten entlang der Tangentialebenen verursachen, blieben bisher unberücksichtigt. Für die Herleitung einer angemessenen Tangentialreaktion nehmen wir an, dass jenes klassische COULOMBSche Reibungsgesetz gilt, das jedem Punkt der Hyperfläche einen generalisierten Reibungskegel $\mathcal{FC}(\mathbf{r}_n)$ zuweist.

Nach COULOMB liegt die Haftreibungskraft eines ruhenden Körpers im Reibungskegel, den man durch eine geometrische Ungleichung der folgenden Art beschreibt.

$$\|\mathbf{u}_t\| = 0 : \|\mathbf{r}_t\| \leq \mu \|\mathbf{r}_n\|. \quad (8.45)$$

Anders ausgedrückt, entspricht die zur Überwindung des Haftzustandes notwendige Kraft mindestens μ mal der Anpresskraft, wobei $\mu \in \mathbb{R}$ ein materialabhängiger Parameter ist. Die Richtung der Haftkraft muss offenbar der Summe der Tangentialkräfte, die am Körper angreifen, entgegenwirken. Ist der Körper dagegen in Bewegung, d.h., ist die Relativgeschwindigkeit an der Kontaktstelle von null verschieden, tritt Gleitreibung auf, dem Betrage nach $\|\mathbf{r}_t\| = \mu \|\mathbf{r}_n\|$ und der Richtung nach der Relativgeschwindigkeit entgegengerichtet, oder formal,

$$\|\mathbf{u}_t\| \neq 0 : \mathbf{r}_t = -\mu \|\mathbf{r}_n\| \frac{\mathbf{u}_t}{\|\mathbf{u}_t\|}. \quad (8.46)$$

Das COULOMBSche Gesetz ist phänomenologischer Natur, weil es die komplexen und weitgehend unverstandenen Wechselwirkungen an der Kontaktfläche approximativ beschreibt. Es zwingt die Kontaktkraft in einen Kegel $\mathcal{FC}(\mathbf{r}_n)$ mit dem Öffnungswinkel $\delta = \arctan \mu$. Wenn der Körper gleitet, liegt die Kontaktkraft auf dem Mantel, beim Haften innerhalb des Kegels. Die Kontaktkraft weicht maximal um den Winkel δ von der Kontaktnormalen ab. Der Reibungskegel wird so gesehen durch die Menge aller Kontaktkräfte aufgespannt, deren tangentialer Anteil die Gleitreibungsbedingung erfüllt, und beschreibt damit das Spektrum an möglichen Reaktionen auf eine bestimmte Kontaktsituation.

Der Übergang vom Haften zur Gleitreibung birgt einige Probleme im Hinblick auf die Stetigkeit der Kraftfunktion. Der COULOMB-Graph (vgl. [Pfeiffer & Glocker, 1996]) weist an der Stelle, an der die relative Kontaktpunktgeschwindigkeit verschwindet, notwendigerweise einen Sprung auf, weil Gleiten erst oberhalb der Haftgrenze auftreten kann. Da die Starrheit von Objekten lediglich eine Approximation ist, kann man COULOMBS Gesetz durch eine entsprechend glatte Funktion ersetzen. Ein Block würde infolgedessen die Rampe mit einer Geschwindigkeit herabkriechen, die in etwa proportional zur Tangentialkraft geteilt durch $\mu \|\mathbf{r}_n\|$ wäre. Bei Experimenten konnte ein solches Kriechverhalten jedoch nicht beobachtet werden [Stewart, 2000]. Der Block verharrt bei

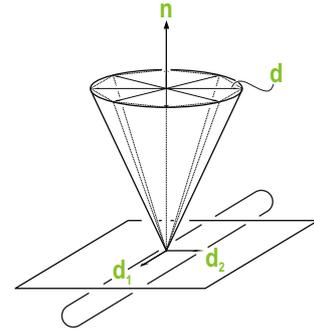


Abbildung 8.4: Reibungskegel und Reibungspyramide. Bei Haftreibung liegt die Kraft im, bei Gleitreibung hingegen auf dem Mantel des Kegels.

einer solchen Anordnung vielmehr in der Ruhelage, bis die Tangentialkraft die durch das COULOMBSche Gesetz vorhergesagte Grenze von $\mu \|\mathbf{r}_n\|$ überschreitet. Neben der fehlenden Deckung mit empirischen Ergebnissen, führen solche Approximationen durch glatte Funktionen auch auf Systeme steifer Differentialgleichungen und die damit verbundenen Probleme.

8.3.1 Der Reibungskegel

Das COULOMBSche Gesetz erzwingt in der dargestellten Form ein isotropes Reibungsverhalten, weil wir den Reibungskoeffizienten implizit für alle Richtungen als gleich angenommen haben. In allgemeinerer Form würde man dem anisotropen Verhalten diverser Materialien durch Wahl unterschiedlicher Reibungsbeiwerte für die beiden Hauptrichtungen der Tangentialebene gerecht. Wir können den Reibungskegel dann durch die Wahl eines geeigneten lokalen Koordinatensystems entsprechend generalisieren. Die Tangentialebene am Kontaktpunkt α werde durch die Richtungsvektoren $\{\mathbf{t}_{1\alpha}, \mathbf{t}_{2\alpha}\}$ aufgespannt, die zusammen mit der Normalen ein rechtshändiges orthonormales Koordinatensystem $\mathbf{R}_\alpha = \{\mathbf{t}_{1\alpha}, \mathbf{t}_{2\alpha}, \mathbf{n}_\alpha\}$ definieren. Der Reibungskegel mit elliptischem Querschnitt wird durch den folgenden analytischen Ausdruck beschrieben,

$$\mathcal{FC}(\mathbf{n}) := \left\{ \mathbf{c} \in \mathbb{R}^3 \mid \mathbf{c} \geq \mathbf{0}, \left(\frac{c_{t_1}}{\mu_{t_1}} \right)^2 + \left(\frac{c_{t_2}}{\mu_{t_2}} \right)^2 \leq c_n^2 \right\}. \quad (8.47)$$

Dieser Ausdruck enthält mit $\mu_{t_1} = \mu_{t_2} = \mu$ isotrope Reibung als Sonderfall. $\mathbf{c} \in \mathbb{R}^3$ ist die Kontaktkraft bzgl. des lokalen Systems \mathbf{R}_α . Der Anteil c_n stellt die Normalreaktion oder im konventionellen COULOMBSchen Sinne die Anpresskraft dar. Die Nichtlinearität, die eine solche Darstellung zwangsläufig induziert, umgeht man üblicherweise durch die Wahl einer polygonalen Approximation des Reibungskegelquerschnitts. Für einen beliebigen, jedoch konvexen Querschnitt, lege man m Richtungsvektoren $\mathbf{d}_\beta \in \mathbb{R}^3$ in der durch \mathbf{t}_1 und \mathbf{t}_2 aufgespannten Ebene fest, welche die auf der Berandung des Querschnitts liegenden Eckpunkte des Polygons definieren, vgl. Abb. (8.4). Der Reibungskegel wird dann durch eine m -seitige Pyramide dargestellt. Besondere Anforderungen, wie bspw. die einer geraden Anzahl oder gleichmäßig verteilter Richtungen wie in [Stewart, 2000], stellen wir nicht. Die Diskretisierung selbst muss notwendigerweise ein anisotropes Verhalten induzieren, weil die innere Approximation des Kegelquerschnitts, die wir hier wählen, durch eine Folge von Sekanten, zwischen den Eckpunkten kleinere Reibungswerte liefert, als durch die ursprüngliche Geometrie festgelegt.

Der skalierbare Querschnitt $\widehat{\mathcal{L}}^c$ des diskreten Reibungskegels ist dann durch die folgende konvexe Menge gegeben,

$$\widehat{\mathcal{L}}^c(\mathbf{n}) := \overline{\text{co}} \{ \mathbf{0} + \mu_\beta \|\mathbf{n}\| \mathbf{d}_\beta, \beta \in \{1, \dots, m\} \mid \mathbf{d}_\beta^\top \mathbf{n} = 0, \|\mathbf{d}_\beta\| = 1, \mu_\beta \geq 0 \}. \quad (8.48)$$

Zusätzlich gilt noch die Konvexitätsbedingung $\mu_i \mathbf{d}_i = w_{i-1} \mu_{i-1} \mathbf{d}_{i-1} + w_{i+1} \mu_{i+1} \mathbf{d}_{i+1}$ mit $w_{i-1} + w_{i+1} \geq 1$, wenn wir das Polygonzentrum in den Ursprung verlegen. Man beachte an dieser Stelle, dass bei anisotropem Verhalten, spezifiziert durch einen nicht kreisrunden Querschnitt, eine Justage desselben auf der Materialoberfläche durch Rotation des lokalen Kontaktkoordinatensystems um die Normale notwendig ist. In der Gliederkette bietet die Richtung der einzelnen Segmente eine Orientierungsmöglichkeit. Wir legen deshalb fest, dass \mathbf{t}_1 mit der Achse des kontaktierende Segmentes in Laufrichtung der Kette zusammenfällt, wenn der aktive Kontakt α auf einem Segment liegt. Die Richtungen \mathbf{d}_β schreibt man als Spaltenvektoren der Generatormatrix \mathbf{D} ,

$$\mathbf{D} := (\mathbf{d}_1, \dots, \mathbf{d}_m), \quad \mathbf{D} \in \mathbb{R}^{3 \times m}. \quad (8.49)$$

Ferner stellt jede Linearkombination der aufspannenden Vektoren des diskreten Kegelquerschnitts

$$\mathbf{f}(w_1, \dots, w_m) = \sum_{\beta=1}^m w_\beta \mathbf{d}_\beta, \quad 0 \leq w_\beta \leq \mu_\beta c_n, \quad (8.50)$$

welche die Bedingung $\sum_{\beta=1}^m w_\beta / \mu_\beta \leq c_n$ erfüllt, eine zulässige Reibungskraft dar. Wenn wir für die Gewichte den Vektor $\mathbf{w} = (w_1, \dots, w_m)^\top \in \mathbb{R}^m$ und für die Reibungskoeffizienten in den unterschiedlichen Richtungen den Vektor $\boldsymbol{\mu} = (\mu_1, \dots, \mu_m)^\top > \mathbf{0} \in \mathbb{R}^m$, ansetzen lässt sich diese Bedingung auch formulieren als

$$\mathbf{f}(\mathbf{w}) = \mathbf{D}\mathbf{w} \in \widehat{\mathcal{L}}^c(c_n \mathbf{n}) \Leftrightarrow \mathbf{e}^\top \mathbf{S}^{-1} \mathbf{w} \leq c_n, \quad (8.51)$$

wobei $\mathbf{S} = \mathbf{E}_m \boldsymbol{\mu}$ die Skalierungsmatrix, $\mathbf{E}_m \in \mathbb{R}^{m \times m}$ die Einheitsmatrix und $\mathbf{e} = (1, \dots, 1)^\top \in \mathbb{R}^m$ der Einheitsvektor ist. Aus diesen Definitionen folgt der diskrete Reibungskegel in der Form

$$\widehat{\mathcal{FC}}(\mathbf{n}) := \overline{\text{co}} \{ \forall \mathbf{w} \in \mathbb{R}^m : c_n \mathbf{n} + \mathbf{D}\mathbf{w} \mid c_n \geq 0, \mathbf{w} \geq \mathbf{0}, \mathbf{e}^\top \mathbf{S}^{-1} \mathbf{w} \leq c_n \}. \quad (8.52)$$

Der Reibungskegel ist eine konvexe, mengenwertige Funktion. Man beachte, dass für $\mu_\beta = \mu$ der Faktor $1/\mu$ aus der Matrix gezogen und auf die linke Seite der Ungleichung gebracht werden kann, womit der klassische Fall isotroper Reibung wieder hergestellt ist.

8.3.2 Maximalleistungsgleichung

Wir haben bisher angenommen, dass bei bekannter Normalkraft diejenige Reibungskraft aus dem Reibungskegel zu wählen ist, deren Tangentialanteil der relativen Kontaktgeschwindigkeit genau entgegen gerichtet ist. Diese zunächst physikalisch nicht zu rechtfertigende Heuristik entspricht einem naiven anisotropen Reibungsmodell, bei dem in Abhängigkeit von der Richtung der Kontaktgeschwindigkeit unterschiedliche Reibungskoeffizienten die Reibungskraft in derselben Richtung begrenzen.

Eine natürliche Grundlage für physikalisch plausible Reibungsmodelle bilden energiebasierte Ansätze. Die Tangentialreaktion sollte so gewählt werden, dass sie der Bewegung den maximal

Normalkräfte bekannt sind und das zugrunde liegende Reibungsgesetz nicht von der Geschwindigkeit abhängt. Dieser Formalismus lässt sich direkt auf den starren Körper übertragen. Die Gesamtreibungskraft und die Geschwindigkeit eines kontaktierenden Körpers erfüllen die Maximalleistungsungleichung auch in Bezug auf eine geschlossene konvexe Limitfläche \mathcal{L}^s . Sie ist in den Raum der Kräfte, $T_p^* \mathcal{M}$, eingebettet und schließt den Ursprung ein. Die Limitfläche folgt aus dem Prinzip der virtuellen Arbeiten, wenn man die Gültigkeit der Maximalleistungsungleichung an allen Kontaktpunkten voraussetzt. Wir erhalten den folgenden Ausdruck,

$$\mathbf{F}^\top \mathbf{U} = \sum_{\alpha} \mathbf{f}_{\alpha} \cdot \mathbf{u}_{\alpha}, \quad \mathbf{F}^\top \mathbf{U}^* = \sum_{\alpha} \mathbf{f}_{\alpha}^* \cdot \mathbf{u}_{\alpha}, \quad (8.56)$$

wobei $\mathbf{F} = (\mathbf{f}, \mathbf{l})$ den generalisierten Kraftvektor der am Körper angreifenden Kräfte und $\mathbf{U} = (\mathbf{v}, \boldsymbol{\omega})$ die Geschwindigkeit des Körpers darstellt. Man beachte, dass die lokale Kontaktkinematik des Körpers durch lediglich drei Freiheitsgerade beschrieben wird, das Gleiten entlang der lokalen Tangentialebene und eine Drehung um die Kontaktnormale. Alle anderen Geschwindigkeitskomponenten müssen notwendig verschwinden, weil sie andernfalls eine Verletzung der unilateralen Zwänge an den Kontaktstellen nach sich ziehen würden. Durch einige Äquivalenzumformungen erhalten wir aus Gl. (8.55) und Gl. (8.56)

$$(\mathbf{F} - \mathbf{F}^*)^\top \cdot \mathbf{U} \geq 0. \quad (8.57)$$

Dabei ist \mathbf{F} die mit der aktuellen Gleitgeschwindigkeit assoziierte Reibungskraft und \mathbf{F}^* ist eine beliebige Kraft, die sich durch Summation möglicher Reibungskräfte \mathbf{f}^* ergibt. Die Limitfläche \mathcal{L}^s begrenzt die Menge möglicher Reibungskräfte infolge einer gegebenen Kontaktnormale. Sie liefert damit eine vollständige Beschreibung der Beziehung zwischen der generalisierten Reibungskraft \mathbf{F} und der Geschwindigkeit \mathbf{U} . In Analogie zur Limitkurve gelten die folgenden Eigenschaften: Falls \mathbf{F} innerhalb von \mathcal{L}^s liegt, muss die Geschwindigkeit notwendig verschwinden. Da, wo die Krümmung der Limitfläche wohldefiniert ist, steht die Richtung der Geschwindigkeit \mathbf{U} senkrecht auf \mathcal{L}^s . Unstetigkeiten im Verlauf der Krümmung sowie Bereiche konstanter Krümmung machen die Beziehung zwischen Reibungskraft und Geschwindigkeit an den entsprechenden Stellen mehrdeutig. Das bedeutet, dass einer Reibungskraft mehrere Geschwindigkeiten zugeordnet sind bzw. für eine Geschwindigkeit mehrere „kompatible“ Reibungskräfte existieren.

Wir wissen jetzt, welche Eigenschaften \mathcal{L}^s hat, insbesondere, dass die Gesamtreibungskraft ein Element des von der Limitfläche eingeschlossenen Raumes sein muss. Wir wissen an dieser Stelle jedoch noch nicht, wie sich die Limitfläche für den kontaktierenden Körper konstruieren lässt. Zunächst ersetzt man dazu die Limitkurven an den einzelnen Kontaktpunkten des Körpers durch entsprechende Limitflächen. Das geschieht durch Einbettung der Limitkurven in den Raum der generalisierten Kräfte. Die Limitfläche eines einzelnen Kontaktpunktes beschreibt die Menge der möglichen generalisierten Reibungskräfte, die von diesem Kontaktpunkt erzeugt werden können. Die Limitfläche \mathcal{L}^s für die am Körper angreifende generalisierte Gesamtreibungskraft ist die konvexe Hülle der MINKOWSKI-Summe der Limitflächen der einzelnen Kontaktpunkte,

$$\mathcal{L}^s := \overline{\text{co}} \bigoplus_{\alpha \in \mathcal{A}(t)} \mathcal{L}_{\alpha}^s. \quad (8.58)$$

Auch hier gilt, dass lediglich aktive Kontakte Beiträge zur Gesamtreibungskraft liefern. Es versteht sich von selbst, dass die MINKOWSKI-Summe nicht explizit berechnet werden muss, sondern in

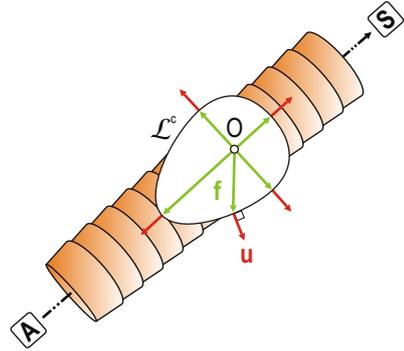


Abbildung 8.6: Zur Veranschaulichung des anisotropen Reibungsmodells für Humanhaar auf der Basis des Limitkurvenkonzeptes.

Form einer Ungleichung in den Projektionsvorgang zur Ermittlung der Postkollisionsgeschwindigkeit eingeht.

An dieser Stelle sei darauf hingewiesen, dass wir für die Kontaktstelle zwischen zwei Fasern vereinfachend punktförmigkeit postulieren. Genau genommen findet aber eine gewisse plastische Verformung statt, die zu einer lokalen Abplattung der Faser führt. Infolgedessen wäre die explizite Berücksichtigung des Momentes bezüglich des lokalen Drehzentrums gerechtfertigt. Die Limitkurve wäre in diesem Fall durch eine Limitfläche mit Ellipsoidform zu ersetzen.

Die Maximalleistungsungleichung und das aus ihr gefolgerte Prinzip der Limitfläche stellt so gesehen eine Verallgemeinerung des COULOMBSchen Reibungsgesetzes hin auf generalisierte Kräfte dar und ermöglicht die einfache Modellierung von anisotropem Reibungsverhalten realer Materialien, vgl. Goyal et al. [Goyal et al., 1991]. In Abb. (8.6) ist die Anwendung des Konzeptes auf Keratinfasern dargestellt. Die durch die Cuticula induzierte Anisotropie findet hier durch eine adäquate Limitkurve Berücksichtigung. Die Limitkurve wird in den Hauptrichtungen durch die Reibungsbeiwerte, die im Kapitel über die mechanischen Eigenschaften angegeben sind, bestimmt.

8.3.3 Tangentialreaktion

Unser primäres Anliegen war, die Tangentialreaktion, die dem Gleiten entlang der Tangentialebene entgegenwirkt, in Form eines äquivalenten Impulses auszudrücken. Über die Beziehung zwischen stoßartigen Kräften und denen durch sie bewirkten instantanen Geschwindigkeitsänderungen wurde weiter oben schon referiert.

Nach Moreau [Moreau, 1988] kann auch das klassische COULOMBSche Reibungsgesetz auf der Basis der Geschwindigkeiten formuliert werden, wenn man den Tangentialimpuls als Differentialinklusion schreibt [Monteiro Marques, 1993],

$$-\mathbf{u}^f \in \mathcal{N}_{\widehat{\mathcal{FC}}(c_{nn})}(\mathbf{u}^+ - \mathbf{u}^-). \quad (8.59)$$

$\mathcal{N}_{\widehat{\mathcal{FC}}(c_{nn})}(\mathbf{u}^+ - \mathbf{u}^-)$ ist der Normalenkegel zur Menge zulässiger Tangentialreaktionen, die den COULOMBSchen Reibungskegel aufspannen. Die Richtung von $\mathbf{u}^+ - \mathbf{u}^-$ fällt bei einem einzelnen Kontakt mit der Stoßnormalen zusammen.

Bei Mehrfachkontakten zwischen starren Körpern ist die Menge zulässiger Tangentialreaktionen zu erweitern, weil anders, als bei Partikeln, auch die durch die lokalen Kontaktkräfte bezüglich des Rotationszentrums induzierten Drehmomente zu berücksichtigen sind. Das entspricht

einer Verallgemeinerung des COULOMBSchen Reibungskegels, wie sie mit dem Konzept der Limitflächen bereits gegeben ist. Der Tangentialimpuls muss bei Mehrfachkontakten offenbar im Schnitt von Tangentenkegel und Normalenkegel der Limitfläche liegen. Das ist leicht einzusehen: Die Postkollisionsgeschwindigkeit liegt Moreaus Selektionsregel zufolge auf dem Rand des Tangentenkegels (Ökonomieprinzip). Wir dürfen infolgedessen nur Geschwindigkeiten dazu addieren, die selbst innerhalb des Tangentenkegels liegen. Da die Limitfläche i.d.R. nicht vollständig innerhalb des Tangentenkegels liegt, sondern diesen vielmehr überlappt, muss der Tangentialimpuls innerhalb der Schnittmenge aus beiden liegen. Diese definieren wir als beschränkten Reibungsimpulsraum

$$\mathcal{R}(\mathbf{q}(t)) = \mathcal{T}(\mathbf{q}(t)) \cap \mathcal{N}_{\mathcal{L}^s(\mathbf{q}(t))}(\mathbf{u}^+ - \mathbf{u}^-). \quad (8.60)$$

Den maximal dissipativen Reibungsimpuls erhalten wir nach Anwendung der Maximalleistungsungleichung Gl. (8.54) als Lösung des folgenden Programmierungsproblems,

$$\mathbf{u}^f = \underset{\mathbf{u}^* \in \mathcal{R}(\mathbf{q}(t))}{\operatorname{argmin}} \left\{ (\mathbf{u}^*)^\top \mathbf{u}^+ \right\}. \quad (8.61)$$

Über entsprechende Nebenbedingungen wird sichergestellt, dass der maximal dissipative Reibungsimpuls im reduzierten Impulsraum \mathcal{R} liegt, so dass das Programmierungsproblem die folgende Form annimmt,

$$\begin{aligned} \mathbf{u}^f &= \underset{\mathbf{u}^* \in T_p \mathcal{M}}{\operatorname{argmin}} \left\{ \mathbf{u}^{*\top} \mathbf{u}^+ \right\}, \\ \text{NB} : \forall \alpha \in \mathcal{A} : \mathbf{n}_\alpha^\top \mathbf{u}_\alpha^* &\geq 0 \wedge \mathbf{M} \mathbf{u}^* \in \widehat{\mathcal{L}}_\alpha^c(\mathbf{v}_\alpha). \end{aligned} \quad (8.62)$$

Der Impuls liegt innerhalb der Limitfläche $\widehat{\mathcal{L}}_\alpha^c$, falls die Projektion auf die aufspannenden Basisvektoren \mathbf{D}_α kleiner als das Produkt aus dem zugeordneten Reibungsbeiwert μ_{β_α} und der Normalkraft $\|\mathbf{v}_\alpha\|$ ist,

$$\mathbf{D}_\alpha^\top \mathbf{M} \mathbf{u} - \|\mathbf{v}_\alpha\| \mu_\alpha \leq \mathbf{0}. \quad (8.63)$$

Damit muss eine Folge von COULOMBSchen Ungleichungen erfüllt sein, die in Richtung der einzelnen Basisvektoren durch die assoziierten Reibungsbeiwerte definiert werden.

Die tangentiale Geschwindigkeitsänderung infolge Reibung \mathbf{u}^f ist zur Geschwindigkeit aus der ersten Projektion zu addieren. Damit erhalten wir die gesuchte Geschwindigkeit für das Mehrfachkontaktproblem als

$$\mathbf{u}^{i+1} = (1 + \varepsilon) \mathbf{u}^+ + \varepsilon \mathbf{u}^- + \mathbf{u}^f. \quad (8.64)$$

8.4 Das Zeitschrittverfahren

In Summe leiten wir den folgenden Algorithmus bzw. das folgende Zeitschrittverfahren zur Berechnung der Geschwindigkeiten und Positionen zum Zeitpunkt $t + \Delta t$ ab.

1. Entwicklung der Orts- bzw. Geschwindigkeitsfunktion um $\Delta t/2$ in die Zukunft (erster Positions- bzw. Geschwindigkeitshalbschritt). Wir erhalten die Testposition $\mathbf{q}^{i+1/2} = \mathbf{q}^i + \Delta t/2 \mathbf{u}^i$ und die Mittelpunktsgeschwindigkeit $\mathbf{u}^{i+1/2}$. Dieser Zustand entspricht der Phase maximaler Kompression.

2. Abschätzung der Kontaktmenge \mathcal{U}_i durch Überprüfung der unilateralen Zwänge

$$h_\alpha(\mathbf{q}^{i+1/2}) \geq 0$$

an der Testposition. Dazu bedient man sich eines Kollisionserkennungsverfahrens. Standardverfahren liefern eine Liste kollidierender Paare mit zusätzlichen Kontaktinformationen, wie der Stoßnormalen \mathbf{n} und der Eindringtiefe d zurück. Diese Informationen sind unter Überprüfung der Persistenzbedingung zunächst in die Menge aktiver Kontakte \mathcal{A}_i pro Körper zu transformieren. Die Projektion der relativen Kontaktpunktgeschwindigkeit auf die Stoßnormale muss außerhalb des durch die Normale definierten Halbraums liegen.

3. Für alle Körper mit aktiver Kontaktmenge $\mathcal{A}_i \neq \emptyset$ berechne man eine zulässige Geschwindigkeiten \mathbf{u}^{i+1} , so dass die durch die Kontakte auferlegten Zwänge nicht weiter verletzt werden. Das geschieht, wie oben beschrieben, durch zwei aufeinander folgende Projektionen: eine für den „Normalimpuls“ und eine für den „Tangentialimpuls“ infolge Reibung. Diese Projektionen werden durch zwei quadratische Programmierungsprobleme ausgedrückt.
4. Falls für einen Körper Kontaktfreiheit besteht, $\mathcal{A}_i = \emptyset$, berechne man die neue Geschwindigkeit \mathbf{u}^{i+1} durch Integration der Geschwindigkeitsfunktion um $t + \Delta t/2$ (zweiter Geschwindigkeitshalbschritt).
5. Berechnung der neuen Positionen \mathbf{q}^{i+1} durch Integration der Geschwindigkeiten \mathbf{u}^{i+1} um $t + \Delta t/2$ (zweiter Positionshalbschritt).

Der Pseudocode für das erweiterte Zeitschrittverfahren von Moreau ist in Alg. (6) dargestellt. Der Algorithmus ist auf jeden Körper \mathcal{B} bzw. auf jedes Segment in der Faseranordnung anzuwenden.

8.4.1 Generalisierung auf $SE(3)$

Während Moreau [Moreau, 1988] und vor allem Monteiro Marques [Monteiro Marques, 1993] ihre Betrachtungen auf Partikel mit linearer Geschwindigkeit beschränken, interessieren wir uns für die Kontaktmodellierung von starren Körpern. Die Orientierung des Körpers hat aber wesentlichen Einfluss auf sein Kontaktverhalten. Die Ortsfunktion \mathbf{q} , vormals Element des \mathbb{R}^3 , gehört durch Ausstattung des Körpers mit einem Koordinatensystem nun dem $SE(3)$ an. Basierend auf den Ideen Masons [Mason, 2001] haben Kaufman et al. [Kaufman et al., 2005] Moreaus Ansatz auf den Konfigurationsraum $SE(3)$ erweitert. D.h., dass neben Kontaktkräften auch Kontaktmomente Berücksichtigung finden. Generalisierte Kräfte, die sich neben dem linearen Anteil auch aus dem Drehmoment zusammensetzen, bezeichnen wir als *Wrench*, generalisierte Geschwindigkeiten hingegen als *Twist* (bei Mirtich findet die Terminologie *spatial forces* und *spatial velocities* Verwendung), hier als $\mathbf{u} = (\mathbf{v}, \boldsymbol{\omega})^\top$. Beides sind also sechsdimensionale Vektoren. Interessanterweise existiert auch für die $SE(3)$ der Begriff des Tangential- bzw. eines dualen Kotangentenraumes. Twists gehören dem ersten an, der auch als $se(3)$ bezeichnet wird. Wrenches dagegen sind Elemente des $se^*(3)$, dem Raum der generalisierten Kräfte.

Um die Generalisierung auf starre Körper bzw. auf die $SE(3)$ als Konfigurationsraum zu berücksichtigen, ist nichts weiter zu tun, als die Kontaktnormale und das innere Produkt neu zu definieren. Die Kontaktnormale enthält nun neben der Richtung normal zur Zwangsfläche auch

Algorithmus 6 : Zeitschrittverfahren zur Berechnung der neuen Positionen und Geschwindigkeiten zum Zeitpunkt $t + \Delta t$.

Data : Körper \mathcal{B} , Massematrix \mathbf{M} , Position \mathbf{q}^i , Geschwindigkeit \mathbf{u}^i , Zeitschritt Δt ,
 Restitutionskoeffizient $\varepsilon \in [0, 1]$
 Result : Position \mathbf{q}^{i+1} , Geschwindigkeit \mathbf{u}^{i+1}

- 1: $\mathbf{u}^- \leftarrow \mathbf{u}^{i+\frac{1}{2}} \leftarrow \text{integrate}(\mathbf{u}^i, \mathbf{F}^{i+\frac{1}{2}}, \mathbf{F}^i, \frac{1}{2}\Delta t, \theta)$
- 2: $\mathbf{q}^{i+\frac{1}{2}} \leftarrow \text{integrate}(\mathbf{q}^i, \mathbf{u}^{i+\frac{1}{2}}, \mathbf{u}^i, \frac{1}{2}\Delta t, \theta)$
- 3: $\mathcal{A} \leftarrow \text{FindPersistentContacts}(\mathcal{B})$
- 4: if $\mathcal{A} \neq \emptyset$ then
 - 5: $\mathbf{u}^+ = \underset{\mathbf{u}^* \in T_p \mathcal{M}}{\text{argmin}} \|\mathbf{u}^* - \mathbf{u}^-\|^2$
 - 6: NB : $\forall \alpha \in \mathcal{A} : \mathbf{n}_\alpha^\top \mathbf{u}^* \geq 0$
 - 7: $\mathbf{v}_\alpha = [\mathbf{n}_\alpha^\top \mathbf{M}(\mathbf{u}^+ - \mathbf{u}^-)] \mathbf{n}_\alpha$
 - 8: $\mathbf{u}^f = \underset{\mathbf{u}^* \in T_p \mathcal{M}}{\text{argmin}} \{ \mathbf{u}^{*\top} \mathbf{u}^+ \}$
 - 9: NB : $\forall \alpha \in \mathcal{A} : \mathbf{n}_\alpha^\top \mathbf{u}_\alpha^* \geq 0 \wedge \mathbf{D}_\alpha^\top \mathbf{M} \mathbf{u} - \|\mathbf{v}_\alpha\| \boldsymbol{\mu}_\alpha \leq \mathbf{0}$
 - 10: $\mathbf{u}^{i+1} = (1 + \varepsilon) \mathbf{u}^+ + \varepsilon \mathbf{u}^- + \mathbf{u}^f$
- 11: else
- 12: $\mathbf{u}^{i+1} \leftarrow \text{integrate}(\mathbf{u}^{i+\frac{1}{2}}, \mathbf{F}^{i+1}, \mathbf{F}^{i+\frac{1}{2}}, \frac{1}{2}\Delta t, \theta)$
- 13: end
- 14: $\mathbf{q}^{i+1} \leftarrow \text{integrate}(\mathbf{q}^{i+\frac{1}{2}}, \mathbf{u}^{i+1}, \mathbf{u}^{i+\frac{1}{2}}, \frac{1}{2}\Delta t, \theta)$

eine Richtung, in der durch den Hebelarm \mathbf{r}_α induzierte Drehimpulse infolge Kontakt aufzubringen sind,

$$\mathbf{n}_\alpha = \frac{(\nabla h_\alpha, \mathbf{r}_\alpha \times \nabla h_\alpha)^\top}{\|(\nabla h_\alpha, \mathbf{r}_\alpha \times \nabla h_\alpha)\|}, \quad (8.65)$$

wobei \mathbf{r}_α der Richtungsvektor vom Schwerpunkt des Körpers zum Kontaktpunkt \mathbf{p}_α ist. Da $SE(3)$ ein „gekrümmter“ Raum ist, muss der klassische Begriff der Orthogonalität fallen gelassen werden. Das „konventionelle“ Skalarprodukt, das die Projektion zweier Vektoren aufeinander beschreibt, verliert seine Gültigkeit. Kaufman et al. [Kaufman et al., 2005] definieren in diesem Zusammenhang eine so genannte *kinetische*, wenn auch beliebige Metrik für die $SE(3)$, aus der das innere Produkt zweier Vektoren durch Multiplikation mit der lokalen Massematrix folgt,

$$\langle \mathbf{a}, \mathbf{b} \rangle := \mathbf{a}^\top \mathbf{M}(\mathbf{q}) \mathbf{b}, \quad \mathbf{a}, \mathbf{b} \in se(3). \quad (8.66)$$

Wenn für zwei Vektoren im $se(3)$ Orthogonalität nicht mehr angenommen werden kann, ist damit auch die Definition für den Reibungskegel obsolet. Diesen hatten wir in seiner klassischen Form weiter oben zwar formal eingeführt, aber in seiner Funktion durch das Konzept der Limitfläche abgelöst.

Neben den genannten Dingen wird auch von Kaufman et al. [Kaufman et al., 2005] eine Einbettung von impulsiven Kräften in den Geschwindigkeitsraum durch Multiplikation mit der Massematrix durchgeführt.

Bisher hatten wir vernachlässigt, dass die vom System induzierten Zwänge nicht notwendigerweise statisch sein müssen, sondern sich selbst „bewegen“ können, also explizit von der Zeit abhängen. Dies mündete in einer Erweiterung des Ansatzes von Kaufman et al. [Kaufman et al., 2005] um eine approximative Behandlung „beweglicher“ Zwangsbedingungen. Solche liegen vor, wenn die Zwangsflächen durch die Berandungen eines oder mehrerer sich bewegender Kollisionspartner gebildet werden. Die Normalgeschwindigkeit des Kontaktpunktes muss nach dem Stoß mindestens so groß sein, wie die Normalgeschwindigkeit der Zwangsbedingung selbst. Das gewährleistet, dass sich Kontaktpunkt und Zwangsbedingung nach dem Stoß mindestens mit derselben Geschwindigkeit in Richtung der Zwangsnormalen bewegen, so dass der Kontaktpunkt von der Zwangsfläche gleichsam vor sich hergeschoben wird. Im Prinzip wird damit ein einfaches Stoßgesetz etabliert. Im Sinne der Impulserhaltung sollten dann zur Berechnung der erforderlichen Mindestnormalgeschwindigkeit die Massen der beiden Körper berücksichtigt werden. Damit ist dann sichergestellt, dass bspw. eine sich bewegende Zwangsfläche einen sehr trägen Körper nicht etwa aus der Ruhelage in Bewegung versetzen kann.

Kaufman et al. [Kaufman et al., 2005] schlagen vor, die notwendige Normalkomponente v_N , die der Körper nach dem Stoß mindestens haben muss, im Zustand der maximalen Kompression zu berechnen. Der Vorteil besteht darin, dass die Normalkomponente der relativen Kontaktpunktgeschwindigkeit in dieser Phase verschwindet, weil die beiden Kontaktpunktgeschwindigkeiten in Richtung der Normalen gleich groß sind, vgl. Mirtich [Mirtich, 1996]. Zieht man hingegen die Geschwindigkeiten beider Körper nach dem Stoß heran, muss man mit Größen rechnen, die noch nicht bekannt sind. Die Projektionen müssten dann für alle Körper simultan ausgeführt werden. Das betrifft insbesondere die Geschwindigkeit des zweiten Körpers nach dem Stoß.

Die gesuchte Normalgeschwindigkeit am Kontaktpunkt erhält man durch eine Mittelung der

beiden Geschwindigkeiten vor dem Stoß mit den Massen der Körper,

$$v_N = \left[\frac{m_{\alpha_i}(\mathbf{v}_i + \boldsymbol{\omega}_i \times \mathbf{r}_{\alpha_i}) + m_{\beta_j}(\mathbf{v}_j + \boldsymbol{\omega}_j \times \mathbf{r}_{\beta_j})}{m_{\alpha_i} + m_{\beta_j}} \right]^T \frac{\nabla \mathbf{h}_{\alpha}}{\|\nabla \mathbf{h}_{\alpha}\|}. \quad (8.67)$$

Die Berechnung der Massen kann im Detail in Baraff und Witkin [Baraff & Witkin, 2003] nachgelesen werden und führt auf folgende Beziehung,

$$m_{\alpha} = \frac{1}{\frac{1}{m} + \frac{1}{\|\nabla \mathbf{h}_{\alpha}\|} (\mathbf{r}_{\alpha} \times \nabla \mathbf{h}_{\alpha})^T \mathbf{I}^{-1} (\mathbf{r}_{\alpha} \times \nabla \mathbf{h}_{\alpha})}. \quad (8.68)$$

Damit ist die Bedingung für die Projektion auf den Tangentenkegel in Alg. (6) wie folgt zu erweitern,

$$\forall \alpha \in \mathcal{A} : \mathbf{n}_{\alpha}^T \mathbf{u}^* \geq v_N. \quad (8.69)$$

8.5 Kollisionsantwort bei kinematischen Ketten

Im folgenden Abschnitt stellen wir ein iteratives Verfahren zur Kombination der impulsbasierten Kollisionsantwort mit der Simulation kinematischer Ketten vor.

Die kontaktierende kinematische Kette kann unter Einführung entsprechender uni- bzw. bilateraler Zwangsbedingungen vom Typ $h(\mathbf{q}(t)) \geq 0$ bzw. $g(\mathbf{q}(t)) = 0$ als LAGRANGE-System formuliert werden,

$$\begin{aligned} \mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} &= \mathbf{F}(\mathbf{q}, \dot{\mathbf{q}}, t) + \boldsymbol{\Psi}(\mathbf{q}) + \sum \lambda \nabla g(\mathbf{q}(t)), \\ h(\mathbf{q}(t)) &\geq 0, \quad h \in \mathbb{R} \\ g(\mathbf{q}(t)) &= 0, \quad g \in \mathbb{R}. \end{aligned} \quad (8.70)$$

mit der Massematrix $\mathbf{M}(\mathbf{q})$, den externen Lasten $\mathbf{F}(\mathbf{q}, \dot{\mathbf{q}}, t)$, den impulsiven Kräften $\boldsymbol{\Psi}(\mathbf{q})$ und den bilateralen Zwängen $\sum_{i=1}^m \lambda \nabla g(\mathbf{q}(t))$. Gavrea et al. [Gavrea et al., 2008] nach besitzen solche Systeme keine Lösung im klassischen Sinne. Eine approximative Lösung soll uns hier aber reichen.

Die Idee besteht darin, dass wir das System Gl. (8.70) von den Gelenkzwängen bereinigen und für jedes Segment der kinematischen Kette unter Berücksichtigung aller Kontakte lösen. Damit folgen wir dem im vorangegangenen Abschnitt beschriebenen Zeitschrittverfahren. In einem zweiten Schritt werden durch einen noch zu beschreibenden Ansatz die Gelenkbedingungen erzwungen. Da beide Prozess von einander entkoppelt sind, kann es zu partiell widersprechenden Systemzuständen kommen. Zur hinreichenden Befriedigung von Kontakt- und Gelenkbedingungen iteriert man über beide Teilaspekte. Im Folgenden bleibt zu klären, wie die Gelenkzwänge selbst zu erfüllen sind. Es ist nahe liegend, die auf dem Geschwindigkeits-Impuls-Level operierende Kontaktbehandlung um ein geeignetes Impuls-basiertes Schema zur Befriedigung der Gelenkzwänge zu erweitern. Mirtich [Mirtich, 1996] hat in seiner Dissertation die Idee einer impulsbasierten Gelenkdarstellung verworfen. Die Zwangsbedingung müsste in so einem Fall durch die Kollisionsbehandlung zwischen den Gelenkkomponenten aufrechterhalten werden. Ein solches Vorgehen sei wegen des vergleichsweise hohen Aufwandes zu langsam. Diese

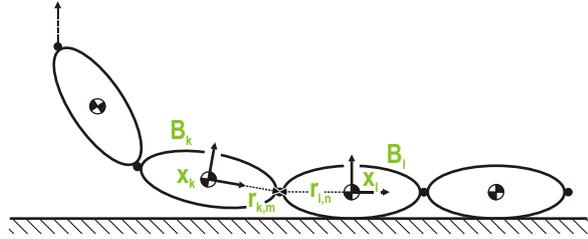


Abbildung 8.7: Zur Notation bei Gliederketten.

Idee lässt sich aber durchaus in gering modifizierter Form umsetzen. Man berechnet dazu die notwendigen Geschwindigkeitskorrekturen bzw. die äquivalenten Impulse oder Kraftstöße, die vor der Integration der Systemvariablen an jedem Körper zu applizieren sind, damit die Gelenkzwänge nach der Integration erfüllt sind. Das wird sich in den wenigsten Fällen in analytischer Form umsetzen lassen, sondern vielmehr iterativ zu erfolgen haben. Bei der Entwicklung eines Impulskorrekturverfahrens können wir auf Vorarbeiten von Bender [Bender, 2007] und Weinstein et al. [Weinstein et al., 2006] zurückgreifen.

8.5.1 Verfahren I: Impulskorrektur

Die instantane Änderung der Geschwindigkeit eines Körpers infolge des Kraftstoßes \mathbf{P} ist nach Gl. (8.18) durch

$$\Delta \mathbf{u} = \mathbf{M}(\mathbf{q})^{-1} \mathbf{P} \quad (8.71)$$

gegeben. Dabei setzt sich der Vektor der Geschwindigkeitsänderungen $\Delta \mathbf{u} = (\Delta \mathbf{v}, \Delta \boldsymbol{\omega})^T$ aus der Änderung der linearen bzw. der Winkelgeschwindigkeit des Körpers zusammen. Der Vektor \mathbf{P} ist als räumlicher Drehimpuls aufzufassen, mit den Komponenten

$$\mathbf{P} = \begin{pmatrix} \mathbf{p} \\ \mathbf{L} \end{pmatrix} = \begin{pmatrix} \mathbf{p} \\ \mathbf{L}^* + \mathbf{r} \times \mathbf{p} \end{pmatrix} = \begin{pmatrix} \mathbf{E}_3 & \mathbf{0} \\ \tilde{\mathbf{r}} & \mathbf{E}_3 \end{pmatrix} \begin{pmatrix} \mathbf{p} \\ \mathbf{L}^* \end{pmatrix}. \quad (8.72)$$

Hier gilt es zu beachten, dass neben dem reinen Drehimpuls \mathbf{L}^* auch der lineare Impuls \mathbf{p} über den Hebelarm des Angriffspunktes einen Anteil zum Gesamtdrehimpuls \mathbf{L} beiträgt.

Für das Erzwingen der Gelenkbedingungen $g(\mathbf{q}) = 0$ ist ein komplementäres Kraftstoßpaar $\pm \mathbf{P}$ von unbekannter Größe und Richtung an den beiden Körpern \mathcal{B}_k und \mathcal{B}_l links und rechts vom Gelenk aufzubringen, vgl. Abb. (8.7). Die daraus resultierende Geschwindigkeitsänderung $\Delta \mathbf{u}$ nach Gl. (8.71) wird auf die aktuellen Geschwindigkeiten addiert,

$$\widehat{\mathbf{u}}_k^i = \mathbf{u}_k^i + \Delta \mathbf{u}_k^i = \mathbf{u}_k^i + (\mathbf{M}_k^i)^{-1} \mathbf{P}_k^i. \quad (8.73)$$

Wir setzen den Konfigurationsraum $\mathcal{M} = \mathcal{SO}(3)$ und expandieren Gl. (8.73) für $\mathbf{u} \in se(3)$ als

$$\begin{pmatrix} \widehat{\mathbf{v}}_k^i \\ \widehat{\boldsymbol{\omega}}_k^i \end{pmatrix} = \begin{pmatrix} \mathbf{v}_k^i \\ \boldsymbol{\omega}_k^i \end{pmatrix} + \begin{pmatrix} \frac{1}{m_k} \mathbf{E}_3 & \mathbf{0} \\ \mathbf{0} & (\mathbf{I}_k^i)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{p}_k^i \\ \mathbf{L}_k^i \end{pmatrix}, \quad (8.74)$$

Das System wird nach Applikation des Kraftstoßes \mathbf{P} probatorisch um Δt in der Zeit vorwärts integriert und die Gelenkbedingung überprüft. Dieser Prozess ist dann gegebenenfalls für eine Folge von Kraftstößen $\mathbf{P}_k^0, \dots, \mathbf{P}_k^v$ zu wiederholen, dabei immer ausgehend vom Systemzustand

zum Zeitpunkt t_i , bis die Zwänge hinreichend genau erfüllt sind. Der auf diese Weise ermittelte finale Korrekturkraftstoß wird auf die Systemgeschwindigkeiten zum Zeitpunkt t_i addiert und die „eigentliche“ Integration des Systems vorgenommen. Das impliziert eine Konservierung des Systemzustandes zum Zeitpunkt t_i , da er für den iterativen Prozess immer wieder gebraucht wird. Aus dieser Beschreibung wird sofort klar, dass es keinen kumulativen Fehler, wie bei anderen Verfahren, bspw. der Gelenkdarstellung mittels LAGRANGE-Multiplikatoren, geben kann. Jede Verletzung eines Zwanges hat einen adäquaten Korrekturkraftstoß im nächsten Zeitschritt zur Folge. Fehler aus einer vorangegangenen unzureichenden Rückprojektion auf den zulässigen Bereich können in nachfolgenden Integrationsschritten korrigiert werden.

Ausgehend von den Geschwindigkeiten erhält man die neuen Positionen \mathbf{q}^{i+1} durch Integration. Eine Überprüfung der Qualität des Korrekturkraftstoßes erfolgt durch Einsetzen der Positionen in die Funktion $g(\mathbf{q}_k^{i+1}, \mathbf{q}_l^{i+1})$. Solange diese nicht hinreichend kleine Werte annimmt, ist eine weitere Modifikation des Korrekturkraftstoßes notwendig.

8.5.1.1 Gelenkbedingung

Die Funktion $g(\mathbf{q})$ ist nichtlinear und entspricht in ihrer elementarsten Form einer Punkt-zu-Punkt-Bedingung, welche die Koinzidenz zweier Punkte, i.d.R. auf unterschiedlichen Körpern gelegen, fordert. Das ist die einfachste Form eines Gelenkes, mit drei rotatorischen Freiheitsgraden. Für andere Gelenktypen ist die bilaterale Zwangsbedingung entsprechend anzupassen. Wir verwenden für das Erzwingen des Zusammenhalts zwischen den Segmenten einer Faser ein solches Punktgelenk.

Sei der m -te Anker- oder Gelenkpunkt des Körpers \mathcal{B}_k durch den lokalen Vektor $\mathbf{a}_{k,m}$ definiert, vgl. Abb. (8.7). Der globale Richtungsvektor $\mathbf{r}_{k,m}$, der vom Schwerpunkt \mathbf{x}_k zum Ankerpunkt zeigt, folgt aus der Transformation

$$\mathbf{r}_{k,m} = \mathbf{R}_k \mathbf{a}_{k,m}. \quad (8.75)$$

Die Richtung zwischen korrespondierenden Ankerpunkten auf den Körpern \mathcal{B}_k und \mathcal{B}_l zum Zeitpunkt t_{i+1} ist folglich

$$\begin{aligned} \mathbf{d}_{k,m,l,n}(\mathbf{q}_k^{i+1}, \mathbf{q}_l^{i+1}) &= \left[\mathbf{R}_k^{i+1} \mathbf{a}_{k,m}^{i+1} + \mathbf{x}_k^{i+1} - (\mathbf{R}_l^{i+1} \mathbf{a}_{l,n}^{i+1} + \mathbf{x}_l^{i+1}) \right] \\ &= \left[\mathbf{r}_{k,m}^{i+1} - \mathbf{r}_{l,n}^{i+1} + \mathbf{x}_k^{i+1} - \mathbf{x}_l^{i+1} \right], \end{aligned} \quad (8.76)$$

mit den Schwerpunktvektoren \mathbf{x}_k^{i+1} bzw. \mathbf{x}_l^{i+1} und der Länge

$$\begin{aligned} g_{k,m,l,n}(\mathbf{q}_k^{i+1}, \mathbf{q}_l^{i+1}) &= \|\mathbf{d}_{k,m,l,n}(\mathbf{q}_k^{i+1}, \mathbf{q}_l^{i+1})\| \\ &= \sqrt{(\mathbf{d}_{x_{k,l}})^2 + (\mathbf{d}_{y_{k,l}})^2 + (\mathbf{d}_{z_{k,l}})^2}. \end{aligned} \quad (8.77)$$

Unter Hinzugabe eines geeigneten Kraftstoßes soll die Distanz verschwinden, bzw. die Zwangsbedingung, wie gefordert, den Wert Null annehmen. Dazu sind die Integrationsprozesse, welche die notwendigen neuen Positionen und Orientierungen $\mathbf{q}^{i+1} = \{\mathbf{x}^{i+1}, \mathbf{R}^{i+1}\}$ liefern, noch genauer zu spezifizieren.

8.5.1.2 Integration

Die Änderung des Schwerpunktes von Körper \mathcal{B}_k über das Zeitintervall Δt erhalten wir durch Integration der Geschwindigkeit, etwa in der Art

$$\mathbf{x}_k^{i+1} = \mathbf{x}_k^i + \Delta t \left(\mathbf{v}_k^i + \frac{\Delta t}{2m_k} \mathbf{f}_k^i \right). \quad (8.78)$$

Hier eignet sich grundsätzlich jedes Verfahren, auch ein implizites. Allerdings ist für die Kraftstoßermittlung nur eine relativ grobe Abschätzung der neuen Konfiguration des Körpers notwendig. Der Aufwand sollte deshalb so gering wie möglich ausfallen.

Die Richtungsvektoren vom Schwerpunkt zu den Gelenkpunkten ändern sich mit der Orientierung des Körpers. Der neue Richtungsvektor zum Zeitpunkt $t + \Delta t$ ist gegeben durch

$$\mathbf{r}_{k,m}^{i+1} = \Delta \mathbf{R}_k^i \mathbf{r}_{k,m}^i = \Delta \mathbf{R}_k^i \mathbf{R}_k^i \mathbf{a}_{k,m} = \mathbf{R}_k^{i+1} \mathbf{a}_{k,m}. \quad (8.79)$$

Dabei ist $\Delta \mathbf{R}$ die inkrementelle Rotation, welche die Form eines Matrixexponentialen annimmt. Wir nähern die Orientierung eines Körpers zum Zeitpunkt $t + \Delta t$ durch den Ausdruck (vgl. [Krysl & Endres, 2005])

$$\mathbf{R}_k^{i+1} = \Delta \mathbf{R}_k^i \mathbf{R}_k^i = \exp \left[\Delta t \left(\tilde{\boldsymbol{\omega}}_k^i + \frac{\Delta t}{2} \tilde{\boldsymbol{\alpha}}_k^i \right) \right] \mathbf{R}_k^i \quad (8.80)$$

an, wobei sich die Winkelbeschleunigung $\boldsymbol{\alpha}_k^i = (\mathbf{I}_k^i)^{-1} \mathbf{I}_k^i$ direkt aus den externen Momenten \mathbf{I} und dem inversen Trägheitstensor ergibt. Das Matrixexponential approximiert man durch die bestens bekannte RODRIGUES-Formel, vgl. Gl. (3.162).

An dieser Stelle sind wir soweit, dass wir die durch den Kraftstoß bewirkten Geschwindigkeitsänderungen direkt aufaddieren können. Man substituiert nun in Gl. (8.78) und Gl. (8.80) die Geschwindigkeiten durch ihre modifizierten Versionen $\hat{\mathbf{v}}$ und $\hat{\boldsymbol{\omega}}$ aus Gl. (8.74), setzt in Gl. (8.76) ein und erhält den folgenden Ausdruck,

$$\begin{aligned} \mathbf{f}_{k,l}(\mathbf{P}^*) = & \left\{ \exp \left[\Delta t \left(\tilde{\boldsymbol{\omega}}_k^i + (\mathbf{I}_k^i)^{-1} \mathbf{L}_k^i + \frac{\Delta t}{2} \tilde{\boldsymbol{\alpha}}_k^i \right) \right] \mathbf{R}_k^i \mathbf{a}_{k,m} - \right. \\ & \exp \left[\Delta t \left(\tilde{\boldsymbol{\omega}}_l^i - (\mathbf{I}_l^i)^{-1} \mathbf{L}_l^i + \frac{\Delta t}{2} \tilde{\boldsymbol{\alpha}}_l^i \right) \right] \mathbf{R}_l^i \mathbf{a}_{l,n} + \\ & \mathbf{x}_k^i + \Delta t \left[\mathbf{v}_k^i + \frac{1}{m_k} \left(\mathbf{p}_k^i + \frac{\Delta t}{2} \mathbf{f}_k^i \right) \right] - \\ & \left. \mathbf{x}_l^i - \Delta t \left[\mathbf{v}_l^i - \frac{1}{m_l} \left(\mathbf{p}_l^i - \frac{\Delta t}{2} \mathbf{f}_l^i \right) \right] \right\} = \mathbf{0}, \end{aligned} \quad (8.81)$$

wobei $\mathbf{p}_l^i = \mathbf{p}_k^i$ und $\mathbf{L}_l^i = \mathbf{L}_k^i$ ist. Entgegengesetzte Vorzeichen, die der Komplementarität der Impulse entspringen, sind schon in den Gleichungen berücksichtigt. Man beachte zudem, dass der Drehimpuls \mathbf{L} die Summe aus reinem Drehimpuls \mathbf{L}^* und dem durch den linearen Impuls bezüglich des Schwerpunktes induzierten Anteil ist, also

$$\mathbf{L}_k = \mathbf{L}_k^* + \mathbf{r}_{k,m} \times \mathbf{p}_k, \quad (8.82)$$

Neben der Koinzidenz der Gelenkpunkte macht es Sinn, die Körper rechts und links vom Gelenk in eine bestimmte gegenseitige Lage zu zwingen. Nach EULERS Theorem kann jede Drehung durch eine Drehachse $\mathbf{\Omega}$ und einen korrespondierenden Winkel θ ausgedrückt werden. Das entspricht gerade der Inversen der Exponential-Map-Darstellung oder der Log-Funktion im $\mathcal{SO}(3)$. Sie ist gegeben durch [Murray et al., 1994]

$$2 \cos \theta = \text{trace}(\mathbf{R}) - 1, \quad \tilde{\mathbf{\Omega}} = \frac{1}{2 \sin \theta} (\mathbf{R} - \mathbf{R}^\top), \quad \mathbf{R} \neq \mathbf{E}_3. \quad (8.83)$$

Sei $\mathbf{R}_{k,l}^{i+1} = (\mathbf{R}_k^{i+1})^\top \mathbf{R}_l^{i+1}$ die relative Lage der beiden Systeme zueinander und $\mathbf{R}_{k,l}^*$ die stationäre Zielorientierung, beide ausgedrückt im System von Körper \mathcal{B}_k . Beide Orientierungen fallen zusammen, wenn der Drehwinkel zur Rotation $(\mathbf{R}_{k,l}^{i+1})^\top \mathbf{R}_{k,l}^*$ verschwindet. Das ist genau dann der Fall, wenn gilt,

$$\theta = 0 \Leftrightarrow (\mathbf{R}_{k,l}^{i+1})^\top \mathbf{R}_{k,l}^* = \mathbf{E}_3. \quad (8.84)$$

Zusammen mit Gl. (8.83) ergibt sich daraus die zweite Funktion zur Bestimmung des gesuchten Kraftstoßes,

$$f_{k,l}^\circ(\mathbf{P}^*) = \text{trace} \left\{ \left\{ \left\{ \exp \left[\Delta t \left(\tilde{\boldsymbol{\omega}}_k^i + (\mathbf{I}_k^i)^{-1} \mathbf{L}_k^i + \frac{\Delta t}{2} \tilde{\boldsymbol{\alpha}}_k^i \right) \right] \mathbf{R}_k^i \right\}^\top \right. \right. \\ \left. \left. \exp \left[\Delta t \left(\tilde{\boldsymbol{\omega}}_l^i - (\mathbf{I}_l^i)^{-1} \mathbf{L}_l^i + \frac{\Delta t}{2} \tilde{\boldsymbol{\alpha}}_l^i \right) \right] \mathbf{R}_l^i \right\}^\top \mathbf{R}_{k,l}^* \right\} - 3 = 0. \quad (8.85)$$

Die Gelenkpunkte beschreiben nicht lineare Trajektorien. Mit $\mathbf{f}(\mathbf{P}^*) : \mathbb{R}^6 \mapsto \mathbb{R}^3$ und $f^\circ(\mathbf{P}^*) : \mathbb{R}^6 \mapsto \mathbb{R}$ liegen zwei nicht lineare Funktionen vor. Das wird insbesondere ersichtlich, wenn man die auftretenden Matrixexponentiale durch die RODRIGUES-Formel substituiert.

Gesucht ist eine gemeinsame Nullstelle $\mathbf{P}^* = (\mathbf{p}, \mathbf{L}^*)^\top$ beider Funktionen. Zu ihrer Bestimmung kollabiert man die beiden Funktionen in eine einzige,

$$F_{k,l}(\mathbf{P}^*) = \|\mathbf{f}_{k,l}(\mathbf{P}^*)\|_2 + |f_{k,l}^\circ(\mathbf{P}^*)| = 0. \quad (8.86)$$

Lösungsstrategien für nichtlineare Gleichungssysteme wurden bereits im Abschnitt über die COSERAT-Gleichungen besprochen, vgl. Abs. (3.4.7). Das NEWTON-Verfahren scheint ein geeigneter Kandidat. Die unumgängliche Berechnung der JACOBI-Matrix führt man zweckmässigerweise in symbolischer Form mit einem Computeralgebraprogramm durch, z.B. Maple.

8.5.1.3 Gleichungskopplung

Bisher haben wir unsere Betrachtungen auf paarweise gelenkig verbundene Starrkörper beschränkt. Am Körper wirken lediglich die Impulse des einen gemeinsamen Gelenkes. Wir erhalten ein System von Gleichungen, die nicht gekoppelt sind. Jeder Korrekturimpuls infolge des einen Gelenkes kann die Zwangsbedingung, die durch ein weiteres gestellt wird, verletzen. So gesehen besteht die Notwendigkeit, über den Lösungsvektor zu iterieren, bis alle Zwänge befriedigt sind. Der Ansatz ist damit sehr ähnlich zu dem von Weinstein et al. [Weinstein et al., 2006]. Wir können die fehlende Kopplung der Gleichungen vornehmen, indem wir an jedem Körper die

Auswirkung aller dort durch die Gelenke verursachten Korrekturimpulse betrachten. Dazu ist lediglich die Geschwindigkeitsänderung infolge der Kraftstöße neu zu definieren. Der Impuls am Körper \mathcal{B}_k ist die Summe aller Korrekturimpulse. Wir nehmen o.B.d.A. an, dass jedes Glied der kinematischen Kette durch ein inwärts gerichtetes Gelenk mit seinem Vorgänger, und über ein auswärts gerichtetes Gelenk mit seinem Nachfolger verbunden ist. Wir machen das im Folgenden durch die Verwendung der Indizes \mathbf{I} bzw. \mathbf{A} kenntlich. Daraus resultiert für die Änderung der Geschwindigkeit des Gelenkpunktes \mathbf{r}_{A_k} infolge der an beiden Gelenkpunkten \mathbf{r}_{I_k} und \mathbf{r}_{A_k} applizierten Korrekturimpulse die folgende Beziehung,

$$\begin{aligned} \mathbf{f}_{k,l}(\mathbf{P}_{I_k}^*, \mathbf{P}_{A_k, I_l}^*, \mathbf{P}_{A_l}^*) = & \left\{ \exp \left[\Delta t \left(\tilde{\omega}_k^i + (\mathbf{I}_k^i)^{-1} (\mathbf{L}_{A_k}^i - \mathbf{L}_{I_k}^i) \frac{\Delta t}{2} \tilde{\alpha}_k^i \right) \right] \mathbf{R}_k^i \mathbf{a}_{k,m} - \right. \\ & \exp \left[\Delta t \left(\tilde{\omega}_l^i - (\mathbf{I}_l^i)^{-1} (\mathbf{L}_{A_l}^i - \mathbf{L}_{I_l}^i) + \frac{\Delta t}{2} \tilde{\alpha}_l^i \right) \right] \mathbf{R}_l^i \mathbf{a}_{l,n} + \\ & \mathbf{x}_k^i + \Delta t \left[\mathbf{v}_k^i + \frac{1}{m_k} \left((\mathbf{P}_{A_k}^i - \mathbf{p}_{I_k}^i) + \frac{\Delta t}{2} \mathbf{f}_k^i \right) \right] - \\ & \left. \mathbf{x}_l^i - \Delta t \left[\mathbf{v}_l^i - \frac{1}{m_l} \left((\mathbf{P}_{A_l}^i - \mathbf{p}_{I_l}^i) - \frac{\Delta t}{2} \mathbf{f}_l^i \right) \right] \right\} = \mathbf{0}. \end{aligned} \quad (8.87)$$

Entsprechende Änderungen sind an den Gleichungen für die Koinzidenz von aktueller und Zielorientierung durchzuführen, was in folgender Beziehung resultiert,

$$\begin{aligned} f_{k,l}^o(\mathbf{P}_{I_k}^*, \mathbf{P}_{A_k, I_l}^*, \mathbf{P}_{A_l}^*) = \text{trace} & \left\{ \left\{ \exp \left[\Delta t \left(\tilde{\omega}_k^i + (\mathbf{I}_k^i)^{-1} (\mathbf{L}_{A_k}^i - \mathbf{L}_{I_k}^i) + \frac{\Delta t}{2} \tilde{\alpha}_k^i \right) \right] \mathbf{R}_k^i \right\}^T \right. \\ & \left. \exp \left[\Delta t \left(\tilde{\omega}_l^i - (\mathbf{I}_l^i)^{-1} (\mathbf{L}_{A_l}^i - \mathbf{L}_{I_l}^i) + \frac{\Delta t}{2} \tilde{\alpha}_l^i \right) \right] \mathbf{R}_l^i \right\}^T \mathbf{R}_{k,l}^* \Big\} - 3 = 0. \end{aligned} \quad (8.88)$$

Man kann dieses für die m Gelenke eines Körpers \mathcal{B}_k fortsetzen, indem die Summe $\sum \mathbf{P}_A - \sum \mathbf{P}_I$ der Impulse in die Gleichungen eingesetzt wird. Hierbei sind lediglich die Vorzeichen zu beachten, da die Impulse in komplementärer Form auf die jeweiligen Opponenten einwirken. Als Konvention legen wir fest, dass Impulse infolge inwärts, also zur Wurzel des Gelenkbaumes hin, gerichteter Gelenke ein negatives und jene an den auswärts gerichteten Gelenken ein positives Vorzeichen erhalten.

Man schreibt die Gleichungen für die m Gelenke der kinematischen Kette als Gesamtgleichungssystem der Form

$$\mathbf{F}(\widehat{\mathbf{P}}^*) = \left\{ \begin{array}{c} F_1(\mathbf{0}, \mathbf{P}_1^*, \mathbf{P}_2^*) \\ F_2(\mathbf{P}_1^*, \mathbf{P}_2^*, \mathbf{P}_3^*) \\ \vdots \\ F_{m-1}(\mathbf{P}_{m-2}^*, \mathbf{P}_{m-1}^*, \mathbf{P}_m^*) \\ F_m(\mathbf{P}_{m-1}^*, \mathbf{P}_m^*, \mathbf{0}) \end{array} \right\} = \mathbf{0}, \quad (8.89)$$

und sieht sofort die Kopplung, die jetzt zwischen den Gleichungen durch die Einführung der gemeinsamen an den inwärts bzw. auswärts gerichteten Gelenken applizierten Impulse besteht.

Dabei ist

$$\widehat{\mathbf{P}}^* := (\mathbf{P}_1^*, \dots, \mathbf{P}_m^*)^T \in \mathbb{R}^{6 \times m} \quad (8.90)$$

der globale Impulsvektor, der sämtliche bilateralen Zwänge erfüllt. Die Funktion $\mathbf{F}(\widehat{\mathbf{P}}^*) : \mathbb{R}^{6 \times m} \mapsto \mathbb{R}^m$ bildet in den \mathbb{R}^m ab.

8.5.2 Verfahren II: Linearisierter Ansatz

Die Formulierung über die relative Kontaktpunktgeschwindigkeit bietet einen alternativen Zugang zum Problem der Gelenkdarstellung. Bei der Kollision zweier Körper \mathcal{B}_k bzw. \mathcal{B}_l wirkt am Kontaktpunkt eine beliebig große Kraft über eine infinitesimale Dauer. Das Integral über diese impulsive Kraft, der Kraftstoß, wirkt mit demselben Betrag, aber entgegengesetzten Vorzeichen auf beide Körper ein. Zur Befriedigung bilateraler Zwangsbedingung sucht man einen adäquaten Kraftstoß. Hier ist dann der Kraftstoß nicht etwa die Folge eines Kontaktes, sondern es verhält sich vielmehr umgekehrt. Der Kontaktpunkt ist das Ziel eines solchen gesteuerten Kraftstoßes.

Die lineare Geschwindigkeit eines beliebigen Punktes eines Körpers \mathcal{B}_k , dessen Koordinaten wir relativ zum Schwerpunkt desselben angeben, ändert sich mit

$$\Delta \mathbf{v}_{k_\alpha} = \Delta \mathbf{v}_k + \Delta \boldsymbol{\omega}_k \times \mathbf{r}_{k_\alpha}. \quad (8.91)$$

Zur Erinnerung: $\Delta \mathbf{v}_k$ und $\Delta \boldsymbol{\omega}_k$ (ohne Index α) sind die lineare- bzw. die Winkelgeschwindigkeit des Körpers.

Man kann in Analogie dazu einen Ausdruck für die Änderung der Winkelgeschwindigkeit am selben Punkt des Körpers angeben. Es kann offensichtlich nur der Anteil der Winkelgeschwindigkeit in Richtung des Vektors vom Schwerpunkt zum Punkt, \mathbf{r}_{k_α} , einen Beitrag leisten, also

$$\Delta \boldsymbol{\omega}_{k_\alpha} = \frac{1}{\|\mathbf{r}_{k_\alpha}\|^2} (\Delta \boldsymbol{\omega}_k \mathbf{r}_{k_\alpha}) \mathbf{r}_{k_\alpha}. \quad (8.92)$$

Dieser Anteil mag bei der Darstellung von Kontakten keine Rolle spielen, bei der Modellierung von Gelenken hingegen schon, weil er die Verdrehung um die Kontaktpunktachse beschreibt. Nehmen wir als Beispiel zwei Zylinder, deren Achsen kollinear sind und die eine Änderung der Winkelgeschwindigkeit, mit der sie um dieselbe rotieren, erfahren. Konzentriert man seine Betrachtungen lediglich auf die Änderung des linearen Geschwindigkeitsanteils, so wird man keine Änderung bzgl. eines auf der Drehachse gelegenen Punktes feststellen, obwohl die Körper in Bewegung sind. Die Beschreibung von Torsion ist ohne diesen Anteil offensichtlich nicht möglich.

Man schreibt die Änderung der Geschwindigkeit in Matrixform als

$$\Delta \mathbf{u}_{k_\alpha} = \begin{pmatrix} \Delta \mathbf{v}_{k_\alpha} \\ \Delta \boldsymbol{\omega}_{k_\alpha} \end{pmatrix} = \begin{pmatrix} \Delta \mathbf{v}_k + \Delta \boldsymbol{\omega}_k \times \mathbf{r}_{k_\alpha} \\ \frac{1}{\|\mathbf{r}_{k_\alpha}\|^2} (\Delta \boldsymbol{\omega}_k \mathbf{r}_{k_\alpha}) \mathbf{r}_{k_\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{E}_3 & -\tilde{\mathbf{r}}_{k_\alpha} \\ \mathbf{0} & \frac{1}{\|\mathbf{r}_{k_\alpha}\|^2} \mathbf{r}_{k_\alpha} \mathbf{r}_{k_\alpha}^T \end{pmatrix} \begin{pmatrix} \Delta \mathbf{v}_k \\ \Delta \boldsymbol{\omega}_k \end{pmatrix}. \quad (8.93)$$

Wir schreiben die Projektion der Änderung der Winkelgeschwindigkeit auf die Drehpunktachse als dyadisches Produkt $\mathbf{r} \mathbf{r}^T \Delta \boldsymbol{\omega}$. Durch Einsetzen von Gl. (8.72) in Gl. (8.18) und anschließend in

Gl. (8.93) erhält man schließlich für den linearen Anteil

$$\begin{aligned}
\Delta \mathbf{v}_{k_\alpha}(t) &= (-1)^{k-1} (\Delta \mathbf{v}_k + \Delta \boldsymbol{\omega}_k \times \mathbf{r}_{k_\alpha}) \\
&= (-1)^{k-1} \left(\frac{1}{m_k} \mathbf{p}(t) + (\mathbf{I}_k^{-1} \mathbf{L}(t)) \times \mathbf{r}_{k_\alpha} \right) \\
&= (-1)^{k-1} \left(\frac{1}{m_k} \mathbf{p}(t) + \tilde{\mathbf{r}}_{k_\alpha} \mathbf{I}_k^{-1} \mathbf{L}(t) \right) \\
&= (-1)^{k-1} \left(\frac{1}{m_k} \mathbf{E}_3, \tilde{\mathbf{r}}_{k_\alpha} \mathbf{I}_k^{-1} \right) \mathbf{p}(t), \quad k = \{1, 2\}.
\end{aligned} \tag{8.94}$$

Auf der anderen Seite haben wir für die Änderung der Winkelgeschwindigkeit

$$\begin{aligned}
\Delta \boldsymbol{\omega}_{k_\alpha}(t) &= (-1)^{k-1} \left(\frac{1}{\|\mathbf{r}_{k_\alpha}\|^2} (\Delta \boldsymbol{\omega}_k \mathbf{r}_{k_\alpha}) \mathbf{r}_{k_\alpha} \right) \\
&= (-1)^{k-1} \left(\frac{1}{\|\mathbf{r}_{k_\alpha}\|^2} ((\mathbf{I}_k^{-1} \mathbf{L}(t)) \mathbf{r}_{k_\alpha}) \mathbf{r}_{k_\alpha} \right) \\
&= (-1)^{k-1} \left(\frac{1}{\|\mathbf{r}_{k_\alpha}\|^2} \mathbf{r}_{k_\alpha} \mathbf{r}_{k_\alpha}^\top \mathbf{I}_k^{-1} \mathbf{L}(t) \right) \\
&= (-1)^{k-1} \left(\frac{1}{\|\mathbf{r}_{k_\alpha}\|^2} \mathbf{r}_{k_\alpha} \mathbf{r}_{k_\alpha}^\top \mathbf{I}_k^{-1} \right) \mathbf{L}(t), \quad k = \{1, 2\}.
\end{aligned} \tag{8.95}$$

Die relative Kontaktpunktgeschwindigkeit ist die Differenz der beiden Änderungen $\Delta \mathbf{u}_{k_\alpha} - \Delta \mathbf{u}_{l_\beta}$ oder in Matrixform

$$\begin{aligned}
\Delta \mathbf{u}(t) &= \begin{pmatrix} \left(\frac{1}{m_k} + \frac{1}{m_l} \right) \mathbf{E}_3 & \left(\tilde{\mathbf{r}}_{k_\alpha} \mathbf{I}_k^{-1} + \tilde{\mathbf{r}}_{l_\beta} \mathbf{I}_l^{-1} \right) \\ \mathbf{0} & \left(\frac{1}{\|\mathbf{r}_{k_\alpha}\|^2} \mathbf{r}_{k_\alpha} \mathbf{r}_{k_\alpha}^\top \mathbf{I}_k^{-1} + \frac{1}{\|\mathbf{r}_{l_\beta}\|^2} \mathbf{r}_{l_\beta} \mathbf{r}_{l_\beta}^\top \mathbf{I}_l^{-1} \right) \end{pmatrix} \begin{pmatrix} \mathbf{p}(t) \\ \mathbf{L}(t) \end{pmatrix} \\
&= \mathbf{K} \mathbf{P}(t).
\end{aligned} \tag{8.96}$$

\mathbf{K} ist die „Kollisionsmatrix“, konstant, nicht-singulär, positiv definit und nicht-symmetrisch, vgl. [Mirtich, 1996]. In unserem Fall ist sie aufgrund der Berücksichtigung des Einflusses von Drehmomenten notwendigerweise anders definiert, aber die genannten Eigenschaften behalten bis auf die fehlende Symmetrie ihre Gültigkeit, wie man in Analogie zum Beweis in [Mirtich, 1996] leicht zeigen kann.

Die Eigenschaften der \mathbf{K} -Matrix garantieren die Existenz ihrer Inversen, so dass bei bekannter Relativgeschwindigkeit der beiden Gelenkpunkte ein äquivalenter Kraftstoß in der Form $\mathbf{P}(t) = \mathbf{K}^{-1} \Delta \mathbf{u}(t)$ direkt aus Gl. (8.96) folgt. Auch hier führen wir eine Kopplung der Gleichungen durch und stellen damit ein globales Gleichungssystem der Form

$$\widehat{\mathbf{K}} \widehat{\mathbf{P}} = \Delta \widehat{\mathbf{u}}, \tag{8.97}$$

für die gesamte kinematische Kette auf, mit der Gesamtkollisionsmatrix $\widehat{\mathbf{K}}$,

$$\widehat{\mathbf{K}} := \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{K}_{21} & \mathbf{K}_{22} & \mathbf{K}_{23} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{K}_{32} & \mathbf{K}_{33} & \mathbf{K}_{34} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{m-1,m-2} & \mathbf{K}_{m-1,m-1} & \mathbf{K}_{m-1,m} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}_{m,m-1} & \mathbf{K}_{m,m} \end{pmatrix}, \quad (8.98)$$

mit einer Haupt- und zwei Nebendiagonalen und den Vektoren

$$\widehat{\mathbf{P}} := (\mathbf{P}_1, \dots, \mathbf{P}_m)^\top \in \mathbb{R}^{6 \times m}, \quad (8.99)$$

$$\Delta \widehat{\mathbf{u}} := (\Delta \mathbf{u}_1, \dots, \Delta \mathbf{u}_m)^\top \in \mathbb{R}^{6 \times m}. \quad (8.100)$$

Die Kollisionsmatrizen \mathbf{K}_{ii} auf der Hauptdiagonalen folgen der Definition Gl. (8.96). Wir verzichten auf eine Herleitung der \mathbf{K}_{ij} mit $i \neq j$. Eine Orientierungshilfe diesbezüglich ist in dem ähnlichen gearteten Ansatz von Bender [Bender, 2007] zu finden.

Eine Lösung von Gl. (8.97) ist nur möglich, wenn $\Delta \mathbf{u}$ bekannt ist. Die Geschwindigkeitsänderung $\Delta \mathbf{u}$ infolge des unbekanntes Kraftstoßes kann approximativ ermittelt werden, indem man die Abstandsfunktion aus Gl. (8.76) durch den Zeitschritt Δt teilt,

$$\Delta \mathbf{v}_{k,l} \approx \frac{1}{\Delta t} \mathbf{d}_{k,m,l,n}(\mathbf{q}_k^{i+1}, \mathbf{q}_l^{i+1}). \quad (8.101)$$

Der Kraftstoß soll also genau so groß sein, dass die durch ihn bewirkte Geschwindigkeitsänderung die Lücke zwischen den Gelenkpunkten schließt. $\Delta \omega_{k,l}$ hängt hingegen von der relativen Orientierung der beiden Gelenkpartner ab und kann bspw. über den Winkel zwischen den Achsen beider Systeme abgeschätzt werden.

8.5.3 Integration elastischen Verhaltens

Es hat sich gezeigt, dass ein Erzwingen der relativen Lage beider Gelenkpartner zu einem unangemessen starren Verhalten führt. Wir verwerfen deshalb den zweiten Term in Gl. (8.86). Ein biege- und torsionssteifes Verhalten der Faser wird durch den Einbau entsprechender Federn erzwungen. Die Änderungen $\Delta \kappa_1$ bzw. $\Delta \kappa_2$ der lokalen Biegungen κ_1 und κ_2 werden wie folgt über einen Zeitschritt Δt approximiert,

$$\Delta \kappa_1 = \frac{\Delta t}{2} \left[(\boldsymbol{\omega}_l^{i+1} + \boldsymbol{\omega}_l^i)^\top \mathbf{d}_{1_l}^i - (\boldsymbol{\omega}_k^{i+1} + \boldsymbol{\omega}_k^i)^\top \mathbf{d}_{1_l}^i \right], \quad (8.102)$$

$$\Delta \kappa_2 = \frac{\Delta t}{2} \left[(\boldsymbol{\omega}_l^{i+1} + \boldsymbol{\omega}_l^i)^\top \mathbf{d}_{2_l}^i - (\boldsymbol{\omega}_k^{i+1} + \boldsymbol{\omega}_k^i)^\top \mathbf{d}_{2_l}^i \right]. \quad (8.103)$$

Die Änderung $\Delta \kappa_3$ der relativen Verdrehung entlang der Faserachse oder Torsion am Gelenkpunkt (m, n) pro Zeitschritt sei

$$\Delta \kappa_3 = \frac{\Delta t}{2} \left[(\boldsymbol{\omega}_l^{i+1} + \boldsymbol{\omega}_l^i)^\top \mathbf{d}_{3_l}^i - (\boldsymbol{\omega}_k^{i+1} + \boldsymbol{\omega}_k^i)^\top \mathbf{d}_{3_k}^i \right]. \quad (8.104)$$

Daraus folgt die lokale Biegung als $\boldsymbol{\kappa}^{i+1} = \boldsymbol{\kappa}^i + \Delta \boldsymbol{\kappa} - \boldsymbol{\kappa}^\circ$, mit der Anfangskrümmung $\boldsymbol{\kappa}^\circ$. An jedem Gelenk wird ein zusätzliches komplementäres Drehimpulspaar aufgebracht, das der lokalen Verdrehung $\Delta \boldsymbol{\kappa}$ entgegenwirkt. Die Berechnung erfolgt dabei in Analogie zu Gl. (3.57).

8.5.4 Algorithmus

Es hat sich in Experimenten gezeigt, dass der erste Ansatz (Korrekturimpuls) unter numerischen Gesichtspunkten einfacher zu handhaben ist, so dass wir bei der Darstellung kinematischer Ketten ausschließlich davon Gebrauch machen werden. Der Algorithmus zur Integration des Kontaktmodells mit dem Ansatz für kinematische Ketten sieht wie folgt aus, vgl. Alg. (7). In der

Algorithmus 7: Harmonisierung der Zwänge. Kombiniertes Zeitschrittverfahren zur Berechnung der neuen Positionen und Geschwindigkeiten in einer Anordnung kinematischer Ketten zum Zeitpunkt $t + \Delta t$ unter Berücksichtigung etwaiger Kollisionen.

```

Data : Anordnung  $\mathcal{C}^i$ ,  $\Delta t$ ,  $\epsilon$ 
Result : Anordnung  $\mathcal{C}^{i+1}$ 
1: for  $\mathcal{C}_k^i \in \mathcal{C}^i$  do
2:    $\widehat{\mathbf{P}}_k^* \leftarrow \mathbf{0}$ 
3:    $\widehat{\mathbf{P}}_k' \leftarrow \mathbf{0}$ 
4:    $\mathcal{Q}_k^* \leftarrow \text{integrate}(\mathcal{C}_k^i, \Delta t)$ 
5:   while  $g(\mathcal{Q}_k^*) > \epsilon$  do
6:      $\mathbf{F} \leftarrow \text{buildNLEQS}(\mathcal{Q}_k^*)$ 
7:      $\widehat{\mathbf{P}}_k^* \leftarrow \text{solveImpulse}\{\mathbf{F}(\widehat{\mathbf{P}}_k^*) = \mathbf{0}\}$ 
8:      $\widehat{\mathbf{P}}_k' \leftarrow \text{compElasticity}(\mathcal{Q}_k^*)$ 
9:      $\mathcal{Q}_k^* \leftarrow \text{applyImpulse}(\mathcal{C}_k^i, \widehat{\mathbf{P}}_k^* + \widehat{\mathbf{P}}_k')$ 
10:     $\mathcal{Q}_k^* \leftarrow \text{integrate}(\mathcal{Q}_k^*, \Delta t)$ 
11:   end
12:    $\mathcal{C}_k^i \leftarrow \text{applyImpulse}(\mathcal{C}_k^i, \widehat{\mathbf{P}}_k^* + \widehat{\mathbf{P}}_k')$ 
13: end
14: for  $\mathcal{S}_k^i \in \mathcal{C}^i$  do
15:    $\mathcal{S}_k^{i+1}\{\mathbf{q}^{i+1}, \mathbf{u}^{i+1}\} \leftarrow \text{resolveContacts}(\mathcal{S}_k^i, \Delta t)$ ; // Siehe Alg. (6)
16: end

```

ersten Schleife werden adäquate Korrekturimpulse zur Befriedigung der Gelenkbedingungen bei der nachfolgenden Integration berechnet. Diese müssen innerhalb der zulässigen Bereiche (Tangentenkegel) liegen, sofern die durch das jeweilige Gelenk gekoppelten Segmente mit anderen Körpern kollidieren. Das Zeitschrittverfahren zur Kollisionsauflösung fungiert in dieser Hinsicht als nachgeschalteter Filter, der die Korrekturimpulskomponenten, die außerhalb des zulässigen Bereiches liegen, entfernt. Die dadurch notwendigerweise bewirkte Verfälschung der Korrekturwirkung führt dann zu einer erneuten, wenn auch geringen, Verletzung der Gelenkbedingung. Solche Fehlzustände lassen sich wegen der gegenseitigen Abhängigkeit nicht durch die isolierte Betrachtung einzelner Segmente beheben, sondern nur dadurch, dass beide Aspekte kombiniert werden. Eine direkte Umsetzung mündet allerdings in einem hochdimensionalen Programmierungsproblem, dessen Dimension linear in der Anzahl der Gelenkbedingungen ist. Jeder Kontakt induziert zusätzlich eine nichtlineare Nebenbedingung. Gesucht ist der Vektor der Korrekturimpulse, der alle Gelenkbedingungen befriedigt. Bei kontaktierenden Segmenten muss er innerhalb des Tangentenkegels liegen, ist allerdings nicht mehr notwendig maximal dissipativ. Zusätzlich

wirkt eine dem Tangentialanteil entgegengerichtete Reibungskomponente. Wir verzichten an dieser Stelle bewusst auf eine tiefer gehende Betrachtung. Eine direkte Kombination des Kontaktansatzes mit dem COSSERAT-Modell, die allerdings nicht mehr Gegenstand dieser Arbeit ist, scheint hier sinnvoller und soll im Folgenden kurz skizziert werden.

8.5.5 Integration mit Cosserat-Modellen

Die oben vorgestellte spezielle Theorie der COSSERAT-Stäbe beschreibt die Faser als Kontinuum. Es ist deshalb nahe liegend, die Kollisionsantwort direkt mit diesem Modell zu kombinieren. Die einzige „Schnittstelle nach außen“ in den Bewegungsgleichungen stellen die extern angreifenden Kräfte und Momente dar. Die grundlegenden Probleme einer kraftbasierten Kollisionsantwort wurden zu Beginn dieses Kapitels ausführlich diskutiert und sollen an dieser Stelle nicht wiederholt werden. Erste Experimente bestätigen die Vorbehalte in wesentlichen Punkten. Interessanter wäre in diesem Zusammenhang die Behandlung auf der Basis von Kraftstößen. In diesem Sinne lässt sich das vorgestellte Zeitschrittverfahren für die Verwendung mit COSSERAT-Modellen anpassen. Man integriert zunächst einen Halbschritt in die Zukunft, prüft auf Kollisionen und berechnet für kontaktierende Segmente die zulässige Geschwindigkeit nach dem Stoß. Bei der nachfolgenden Integration um einen weiteren Halbschritt, ist dann ein Randwertproblem mit intermediären Randbedingungen für die Geschwindigkeiten der kollidierenden Segmente zu lösen. Die Entwicklung eines geeigneten numerischen Verfahrens steht noch aus.

Eleganter scheint uns jedoch die Erweiterung der COSSERAT-Gleichungen um Terme für stoßartige Kräfte. Durch eine Grenzwertbetrachtung (s.o.) erhält man Funktionen, welche die instantanen Geschwindigkeitsänderungen entlang der Faser infolge eines Kraftstoßes an beliebiger Stelle beschreiben. Nach Superposition dieser Änderungen mit dem aktuellen Zustand kann das Gleichungssystem in bekannter Weise gelöst werden. Für den Einbezug von Reibung sind die COSSERAT-Gleichungen allerdings im Kontext der Theorie der Differentialinklusionen zu erweitern. Über Existenz und Eindeutigkeit der Lösung ist nichts bekannt. Erste, wenn auch primär theoretische Ausführungen diesbezüglich sind bei Schuricht [Schuricht, 1998] zu finden. Jedoch beschränken sich die Betrachtungen auf die zweidimensionalen, statischen COSSERAT-Gleichungen. Die dort gewonnenen Erkenntnisse lassen sich also nicht notwendig auf das vorliegende Problem übertragen.

8.6 Numerische Beispiele

Alle in Rahmen dieser Arbeit vorgestellten Algorithmen wurden in das Fasersimulationssystem Rapunzel (Universität Bonn) implementiert. Dabei handelt es sich um eine über Konfigurationsdateien gesteuerte prototypische Simulationsumgebung zur Berechnung des mechanischen Verhaltens komplexer Faseranordnungen. Rapunzel ist mit einer Benutzerschnittstelle und einfachen Möglichkeiten für das visuelle Debugging ausgestattet. Darüber hinaus verfügt die Simulationsumgebung über Schnittstellen zu Modellierungswerkzeugen wie LighWaveTM, POV-RayTM, sowie den hauseigenen Hochleistungsfaserverrenderer LightSimulator.

Im Folgenden präsentieren wir einige Fallbeispiele, die wir durch Anwendung unseres Multi-kontaktalgorithmus mittlerweile effizient behandeln können. Dabei legen wir den Schwerpunkt auf die Auswirkungen von Reibung und die Erhaltung des Haarvolumens. Die Simulationen bzw.



Abbildung 8.8: Die Bedeutung der Reibung bei der Simulation von Humanhaar. Diese Bilder zeigen die Dynamik einer Faseranordnung unter Einwirkung von Gravitation. Die Fasern kollidieren mit der Oberfläche einer Kugel, auf der sie liegen bleiben, wenn Reibungseffekte berücksichtigt werden (obere Reihe). Ohne Reibung teilt sich die Anordnung in zwei Portionen unterschiedlicher Größe, die je rechts und links der Kugel herabfallen (untere Reihe). Dieser Spliteffekt von Haarsträhnen kann mit leithaarbasierten Modellen und anschließender Interpolation von Fasern nicht dargestellt werden.

Messungen wurden in der Regel auf einem Rechner mit Intel Pentium IV und 2 GB RAM ausgeführt. Für die Fasern wurde durchgängig ein Querschnitt von $70 \mu\text{m}$ angenommen. Wir möchten an dieser Stelle betonen, dass die Berechnungszeiten je nach Grad der Komplexität der simulierten Szenen recht hoch ausfallen können. Die Beschleunigung der Ausführung bspw. unter Ausnutzung von paralleler Hardware (z.B. NVIDIA Tesla™) stellt jedoch unter algorithmischen Gesichtspunkten kein Problem dar. Die Ausführungszeiten spielen aber momentan eine eher untergeordnete Rolle, weil das Hauptaugenmerk hier auf einer Demonstration der Machbarkeit einer physikalisch faserbasierten Haarsimulation liegt.

8.6.1 Horizontale Anordnung auf Kugel

Als erstes Beispiel betrachten wir eine horizontale, einseitig fixierte Anordnung, bestehend aus 500 einzelnen Fasern der Länge 30 cm, mit je 40 Segmenten. Die Anordnung bewegt sich unter Einwirkung der Gravitation auf die Oberfläche einer Kugel zu, die unter der Mittellinie der Faseranordnung zentriert ist, vgl. Abb. (8.8). Wir betrachten die Fälle mit und ohne Reibung.

Die Simulationsergebnisse stehen nicht im Widerspruch zur Erfahrung oder Intuition. Die Berechnung der Reibung verhindert ein Abrutschen der Fasern von der Kugeloberfläche. Das Ausblenden von Reibungseffekten führt zu einer dynamischen Teilung der Anordnung an der Kugeloberfläche, so dass ein Teil links, der andere rechts zur ursprünglichen Faserachse ausweicht. Die Fähigkeit zur physikalisch basierten Darstellung dieses Spliteffektes zeichnet die expliziten Modelle gegenüber impliziten Ansätzen aus.

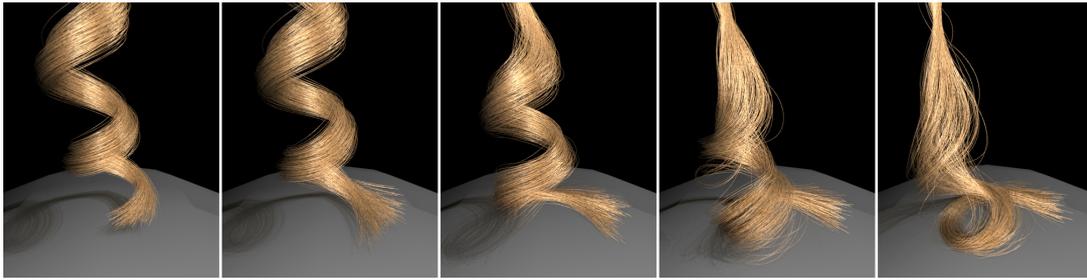


Abbildung 8.9: Diese Bildsequenz zeigt die Entwicklung der Dynamik einer Haarsträhne, die auf der Basis unseres Ansatzes berechnet wurde. Es findet eine Abwicklung der Anordnung unter Gravitation statt, die mit einer Annäherung an die Kugeloberfläche einhergeht. Die bei der Berührung mit der Oberfläche und zwischen den Fasern auftretenden Kollisionen werden detektiert und anschließend durch Aufbringen eines reibungsbehafteten Antwortimpulses aufgelöst. Die korrekte Behandlung der Kollisionen zwischen den Fasern ist ein Garant dafür, dass das Volumen der auf der Oberfläche liegenden Anordnung erhalten bleibt, weil es ein „Durchtunneln“ der Fasern verhindert. Die Volumenkonstanz ist eines der wichtigsten Eigenschaften für den globalen Realismus in der Haarsimulation. Typische Effekte, wie das Absteigen einzelner Fasern tragen zu einer weiteren Verbesserung des visuellen Erscheinungsbildes bei. Die Anordnung besteht aus 355 Einzelfasern bzw. 34.789 Segmenten.

8.6.2 Vertikale Anordnung auf Kugel

Als weiteres Beispiel, das auch unter ästhetischen Gesichtspunkten interessant ist, betrachten wir eine vertikale, in der Ausgangskonfiguration helikale, Anordnung bestehend aus 355 einzelnen Fasern der Länge 30 cm, mit je 100 Segmenten, vgl. Abb. (8.9) und Abb. (8.10), im Folgenden als Referenzszene bezeichnet. Die Fasern fallen unter Einwirkung der Gravitation auf die Oberfläche einer unterhalb der Anordnung positionierten Kugel.

Unter Einfluss der Gravitation findet eine Abwicklung der Anordnung statt. Die unteren Enden der Fasern legen sich dabei sanft auf die Oberfläche der Kugel, wobei die Reibung ein weiteres Abrutschen verhindert. Dieses Beispiel demonstriert eindrucksvoll, wie wichtig eine reibungsbehaftete Kollisionauflösung zwischen den einzelnen Fasern für den globalen Realismus bei der Simulation solcher Anordnungen ist. Die Fasern bleiben, bedingt durch die Reibung, aufeinander liegen und formen ein stabiles Volumen aus. Die Volumenkonstanz belegt andererseits, dass der Effekt des Durchtunnels bei der gewählten Schrittweite von 1/1.000 Sek. weitestgehend eliminiert ist. Abb. (8.10) zeigt eine Nahaufnahme der Anordnung in ihrer Endlage nach wenigen Sekunden Simulationszeit. Kompakte Bereiche aufeinander ruhender Fasern sind hier sehr gut auszumachen.

8.6.3 Vertikale Anordnung auf Terrain

Als drittes Beispiel betrachten wir die vertikale Anordnung aus dem vorhergehenden Beispiel. Unter Einwirkung der Gravitation fallen die Fasern auf konkave Bereiche eines darunter platzierten Geländemodells, vgl. Abb. (8.11). Auch hier werden Kollisionen zwischen den Fasern und mit der Peripherie aufgelöst. Durch den Einsatz von Distanzfeldern sind auch in den konkaven Bereichen des Geländemodells keine Durchdringungen auszumachen, vgl. Abb. (8.15).



Abbildung 8.10: Dieses Bild zeigt eine Nahaufnahme der Anordnung in ihrer Endlage nach wenigen Sekunden Simulationszeit. Der Effekt der konsequenten Kollisionauflösung ist sehr gut zu sehen. Die Fasern ruhen infolge der reibungsbehafteten und teilweise inelastischen Antwortimpulse aufeinander und formen ein statisches Volumen aus.

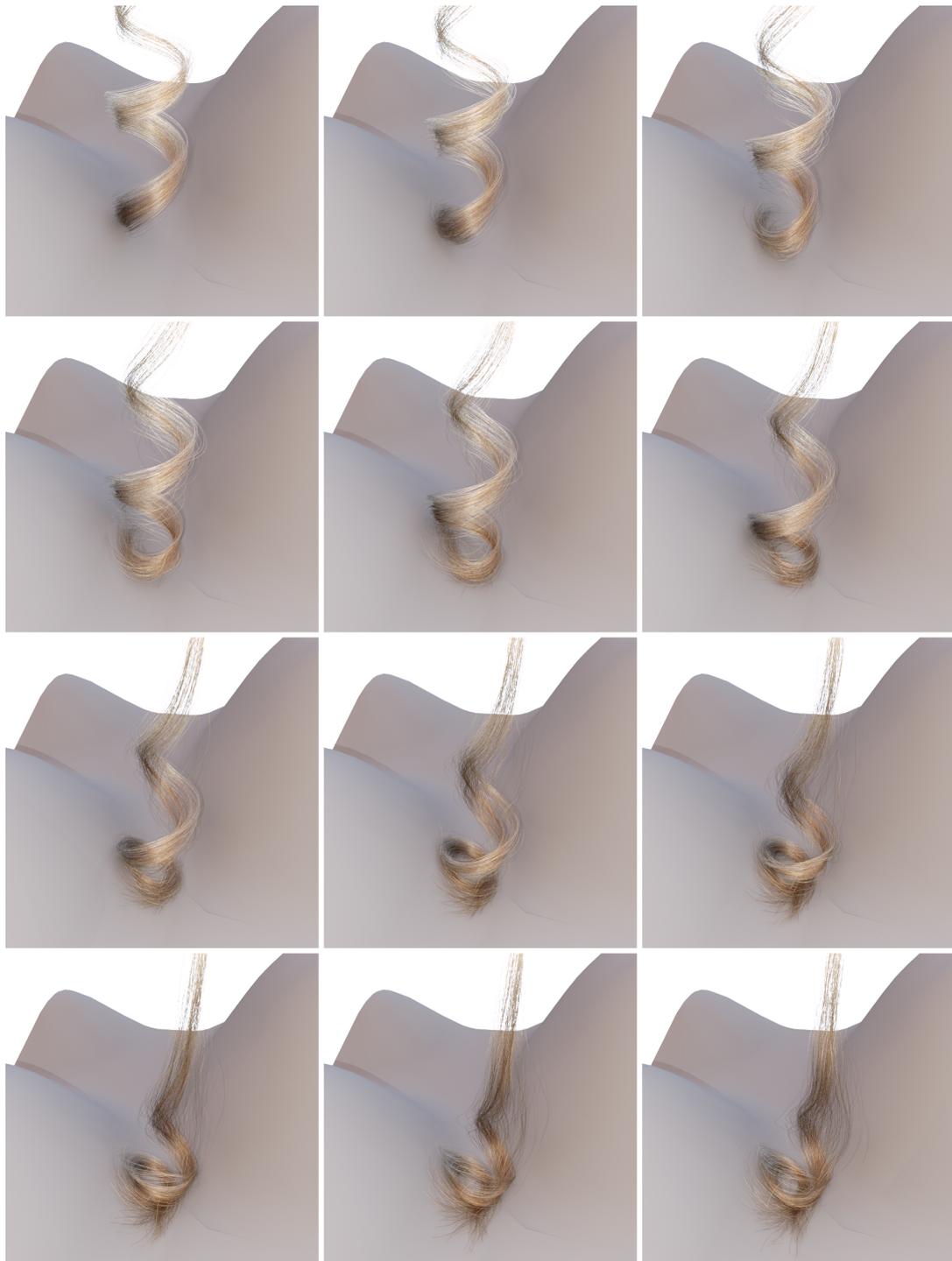


Abbildung 8.11: *Dynamisches Verhalten einer Faseranordnung unter Einwirkung der Gravitation. Die Fasern fallen auf konkave Bereiche eines darunter platzierten Geländemodells. Kollisionen zwischen den Fasern und mit der Peripherie werden aufgelöst.*

8.6.4 Horizontale Anordnung auf schiefer Ebene

Als viertes Beispiel betrachten wir eine horizontale, helikale Anordnung, bestehend aus 250 Einzelfasern der Länge 30 cm, zu je 100 Segmenten, die sich unter Einwirkung der Gravitation auf eine unter ihr befindliche schiefe Ebene (Neigung $22,5^\circ$) zu bewegt, vgl. Abb. (8.12) und Abb. (8.13).

Die helikale Anordnung setzt mit den Extrempunkten auf der Ebene auf. An diesen Stützstellen findet eine Auffächerung der Fasern statt. Der Reibungskoeffizient ist so gewählt, dass ein weitergehendes Abrutschen der Fasern verhindert wird.

8.6.5 Diskussion

Die Effizienz des Kollisionsantwortverfahrens lässt sich am einfachsten prüfen, wenn wir Nahaufnahmen der Resultate einer visuellen Inspektion unterziehen, vgl. Abb. (8.14). Die erste Nahaufnahme zeigt die Anordnung im Bereich der fixierten Enden. Im Verlaufe der Simulation straffen sich die Fasern der initial helikalen Anordnung im Bereich der Aufhängung, wodurch es zu einer Kompaktion kommt. Diese Kompaktion, die ein mehr statisches Aufeinanderruhen der Fasern bewirkt, muss durch entsprechende reibungsbehaftete Kollisionsantwortimpulse, die der Tendenz zu gegenseitigen Durchdringung entgegenwirken, aufrechterhalten werden. Das Bild zeigt deutlich, dass dies hier besonders gut gelingt, weil die Fasern sich aufgrund des Schwungs in den unteren Bereichen leicht umeinanderwickeln. Die Verwicklung erkennt man gut. Sie muss sich also über mehrere Zeitschritte gebildet haben. Wie folgern also, dass die Auflösung der Kollision ordnungsgemäß funktioniert.

Das zweite Bild, das in etwa dem mittleren Teil der Anordnung entnommen ist, zeigt klar erkennbare Faserschichten mit globalen Hauptrichtungen. Die Krümmungen kommen hier durch das Aufliegen der unteren Bereiche auf der Kugeloberfläche und der damit verbundenen Entlastung in den mittleren Bereichen zustande. Die globale Ordnung zeigt eindrucksvoll, dass die Kollisionsantwort den Prozess über die gesamte Simulationsdauer in eine physikalisch akzeptable Richtung gelenkt hat.

Besonders interessant ist das Verhalten an der Kreuzungsstelle, an der sich der noch in der Abwicklung befindliche Faserstrang über den bereits auf der Kugeloberfläche ruhenden legt. Die oberen Fasern bilden ein klar differenziertes Faserbündel, das auf den unteren Fasern ruht. Es sind keine Unterwanderungen der unteren Schichten durch einzelne Fasern auszumachen, wie sie etwa durch Einsacken infolge einer unzureichenden Kollisionsbehandlungsroutine zu erwarten wären. Auch das dürfen wir als weiteren Indikator dafür werten, dass unser Zeitschrittverfahren für die Behandlung vieler simultaner Kontakte, wie sie typisch für solche Anordnungen sind, angemessen ist.

Das untere Ende der Anordnung zeigt klar geordnete und infolge der Fasersteifigkeit nach außen gerichtete Spitzen. Querschläger infolge schlechter Kollisionsauflösung sind auch hier nicht auszumachen.

Die Behandlung von Kollisionen mit der Peripherie auf der Basis von Distanzfeldern erweist sich ebenfalls als äußerst robust. Durchdringungen der Oberfläche oder andere Artefakte treten nicht auf. Da der Hauptaufwand auf das Preprocessing verlegt wurde und Schnitttests zwischen geometrischen Primitiven nicht notwendig sind, können die Abstände zu einer großen Anzahl an Punkten, mit denen wir die kinematischen Ketten abtasten, mit zeitlich vernachlässigbarem Auf-



Abbildung 8.12: Eine horizontale, helikale Faseranordnung fällt unter Einwirkung der Gravitation auf eine darunter platzierte schiefe Ebene (Teil I). Kollisionen zwischen den Fasern und zwischen den Fasern und der Peripherie werden aufgelöst.



Abbildung 8.13: Eine horizontale, helikale Faseranordnung fällt unter Einwirkung der Gravitation auf eine darunter platzierte schiefe Ebene (Teil II). Die Anordnung setzt sich mit den Extremstellen auf der Ebene auf. Deutliche Abplattungen oder Auffächerungen sind zu erkennen.



Abbildung 8.14: Eine genaue Kollisionsbehandlung ist der Schlüssel zu der so wichtigen Volumenerhaltung bei der Haarsimulation. Dieser Bilder zeigen Nahaufnahmen an unterschiedlichen Stellen derselben Anordnung, die mit unserem Ansatz berechnet wurden. Die Szenen sind sehr komplex, mit vielen Kollisionen, die perfekt aufgelöst wurden. Die Fasern gleiten übereinander hinweg oder ruhen aufeinander. Die untere Reihe zeigt Scans von realen Haaren zum Vergleich.

wand bestimmt werden, im Gegensatz zur Erkennung von Kollisionen zwischen den Fasern. Bei der hier gewählten Segmentlänge erweist sich eine Überprüfung der Endpunkte auf Eindringen als ausreichend, um Kollisionen sicher abzuwenden. Eine Abtastung zwischen den Endpunkten der Segmente ist nicht erforderlich.

Ein weiterer Vorteil bei der Verwendung von Distanzfeldern besteht in der Möglichkeit den Level vorzugeben. Die Fasern liegen dann nicht auf der Oberfläche, sondern auf einer beliebigen zu spezifizierenden Niveaufläche. Damit lässt sich ein positiver Toleranzbereich festlegen, weil es gerade im Bereich der Nullniveaufläche wegen des dort anstehenden Vorzeichenwechsel zu numerischen Problemen kommen kann und bereits geringe numerische Abweichungen in den Koordinaten das Eintauchen in das Volumen der Peripherie zur Folge haben können. Solche Artefakte machen sich dann bei Nahaufnahmen störend bemerkbar. Wie zuverlässig sich Kollisionen auch in konkaven Bereichen der Peripherie detektieren lassen, sei anhand von Abb. (8.15) verdeutlicht. Die Fasern fallen durch Abwicklung unter Gravitation in eine stetige Kavität und bleiben selbst im Bereich der hier stark gekrümmten Oberfläche ohne Durchdringungen auf derselben liegen.

8.6.5.1 Inelastische Stöße

Allgemein scheint die Applikation unelastischer Antwortimpulse bei der Simulation besonders angebracht, da ein elastisches Aufeinanderfedern auch bei realen Haaren nicht zu beobachten ist. Wir setzen daher den Restitutionskoeffizienten in allen Fällen auf $\varepsilon = 0$. Die teilweise fatalen Auswirkungen einer teilelastischen Kollisionsantwort ist in Abb. (8.16) zu sehen. Nach über 20.000 Simulationsschritten und mehr als zwei Wochen Berechnungszeit, sprechen die Abbildungen eine deutliche Sprache. Das System strebt keinem erkennbaren minimalen Energiezustand entgegen, kommt nicht zur Ruhe. Die kinetische Energie des Systems scheint vielmehr auf einem konstanten Niveau zu stagnieren, ein Phänomen, das sich in einem unaufhaltsamen Fließen der bereits in einen völlig planaren Zustand übergegangenen Faseranordnung manifestiert. Dieses Beispiel könnte in dramatisch pragmatischer Weise die Grenzen des NEWTONSchen Stoßgesetzes, dessen wir uns hier bedienen, aufzeigen. Danach ist ein Ansteigen der Systemenergie nicht grundsätzlich auszuschließen. Die Ergründung der genauen Ursachen für das katastrophale Systemverhalten sei aber zukünftigen Forschungsbemühungen vorbehalten.

Ein weiteres, fast schon groteskes Beispiel der Auswirkungen einer stagnierenden Systemenergie ist in Abb. (8.17) zu sehen. Die Fasern fallen zunächst unter dem Einfluss der Gravitation, wie erwartet, auf die Oberfläche der Terrains, um sich dort leicht zu entspannen. Über einen bestimmten Zeitpunkt hinaus geht ein energetischer Ruck durch das System, infolgedessen sich die Fasern im weiteren Verlauf der Simulation beständig bergauf bewegen, um teilweise über den Rand des Terrains hinauszuwandern und in die Tiefe zu stürzen.

Kommen wir an dieser Stelle noch einmal auf Abb. (8.16) zurück. Mit zunehmender Simulationsdauer treten hier Durchdringungen auf, weil durch die elastische Komponente die dünnen Fasern selbst innerhalb der mit $1/1.000$ Sekunde recht klein bemessenen Schrittweite einander durchtunneln. Ein solches Tunneln lässt sich durch eine diskrete Kollisionserkennung, wie sie für die Berechnung der hier vorgestellten Beispiele Verwendung fand, nicht feststellen. Hier sind vielmehr die Mechanismen eines kontinuierlichen Ansatzes notwendig. In derselben Weise dringen Fasern teilweise in die Peripherie ein, was in einigen Bildern deutlich zu erkennen ist. In



Abbildung 8.15: Der Einsatz von Distanzfeldern bei der Kollisionserkennung zwischen den Fasern einer Anordnung und der Peripherie erweist sich als äußerst robust. Die Fasern bleiben auf der gekrümmten Oberfläche des Terrains liegen. Durchdringungen sind nicht zu beobachten.



Abbildung 8.16: Die Bilder zeigen die Probleme, die mit der Applikation teilelastischer Antwortimpulse verbunden sind. Die kinetische Energie des Systems stagniert auf einem bestimmten Niveau. Die Fasern durchtunneln einander und dringen teilweise in die Ebene ein. Eine stabile Konfiguration kann nicht ausgebildet werden. Der zeitliche Abstand zwischen den Bildern beträgt 3,3 Sek. (100 Frames oder 3.333 Simulationsschritte, von links nach rechts und oben nach unten).



Abbildung 8.17: Das Bild zeigt die Probleme, die mit der Applikation teilelastischer Antwortimpulse verbunden sind. Die kinetische Energie des Systems stagniert auf einem bestimmten Niveau. In der Folge wandern die Fasern teilweise bergauf.

der Folge kommt es dann partiell sogar zum Kettenabbriss, ein sicherer Indikator für ein lokales numerisches Desaster.

Tunneln lässt sich bei der hier eingesetzten diskreten Kollisionserkennung nur durch eine sehr vorsichtige Wahl der Schrittweite verhindern. Es ist ein komplexes Zusammenspiel von Geschwindigkeiten, „Objektdurchmessern“ und der Schrittweite selbst, das hier durch eine geeignete Wahl der Letzteren zu harmonisieren ist. Unsere Experimente haben gezeigt, dass man bei einem Faserdurchmesser von $70\ \mu\text{m}$ mit $1/1.000$ Sekunde in allen vorgestellten Fällen, in denen wir auf elastische Stöße verzichten, auf der sicheren Seite liegt.

Der allgemeinen Auffassung, dass der Hauptteil des Berechnungsaufwandes bei der physikalisch basierten Simulation in die Kollisionserkennung fließt, muss an dieser Stelle entschieden widersprochen werden. Vielmehr hat sich gezeigt, dass der mit dem Auffinden von Kollisionen verbundene zeitliche Aufwand in den hier betrachteten Fällen eher vernachlässigbar ist. Die meiste Zeit beansprucht die eigentliche numerische Berechnung von Gelenkkorrektur- und Antwortimpulsen, die i.d.R. iterativ erfolgt. Hinzu kommt, dass dieses Verfahren für jede Faser der Anordnung zu wiederholen ist. Dieser Befund bleibt selbst dann aktuell, wenn man die diskrete durch eine kontinuierliche Kollisionserkennung austauschen würde, was aus den o.a. Gründen als durchaus gerechtfertigt erscheint. Die Systemzeit müsste in solchen Fällen bis zum Zeitpunkt des Auftretens der ersten Kollision zurückgesetzt und die zeitaufwändigen numerischen Berechnungen u.U. mehrfach wiederholt werden.

Für die Berechnung eines einzigen Zeitschrittes hat man, je nach Anzahl der notwendigen Iterationen, um die Gelenkkorrektur mit der Impulsantwort zu harmonisieren, von 0,5 bis zu drei Minuten für die Referenzszene („Strähne auf Kugel“) bei rein serieller Abarbeitung (eine Faser nach der anderen) auf der oben genannten Hardware zu veranschlagen. Die Tatsache, dass die numerische Integration der Fasern weitgehend unabhängig von einander abläuft, birgt jedoch ein großes Beschleunigungspotential, bspw. durch Parallelisierung. Bei serieller Bearbeitung ist für eine Sequenz von fünf Sekunden Dauer bereits eine Woche Berechnungszeit zu veranschlagen, wenn man von zwei Minuten pro Zeitschritt ausgeht. Diese Zeitangaben sind starken Schwankungen unterworfen, weil die Dauer für die Harmonisierung stark von der Zahl der Kollisionen abhängt, die bedingt durch den physikalisch basierten Kompaktionsprozess zu Beginn der Simulation stark ansteigt, um dann in Abhängigkeit von der Systemdynamik zwischen dem Gleichgewichtszustand mehr oder weniger stark zu schwanken. Des Weiteren sollte man beachten, dass oftmals ein mehrmaliger Neustart der Simulation unumgänglich ist, um das notwendige manuelle Parameterfitting vorzunehmen. Die von einer auf eine andere Szene achtlos übertragenen Reibungswerte bspw., zeitigen dort selten denselben Effekt, können vielmehr zu numerischen Problemen führen, welche die Berechnung extrem verlangsamen. Das hier vorgestellte Modell enthält noch zu viele ad-hoc Parameter, die nicht klar erkennen lassen, wie das System skaliert.

Ein solcher ad-hoc-Ansatz ist das hier eingeführte Biegefedermodell, das mit den typischen Problemen bei der Wahl passender Steifigkeitsparameter aufwartet. Dass es aber durchaus seinen Dienst zu vollster Zufriedenheit erfüllt, davon kann man sich anhand von Abb. (8.18) überzeugen. Die Biegefedern verleihen den Fasern eine gewisse Steifigkeit. In der ersten Szene wird der Einfluss der Federn auf die Fasern berücksichtigt, die infolgedessen stabile Schleifen ausformen. Diese Schleifen verhindern, dass die Anordnungen einer planaren Endkonfiguration entgegenstrebt. Des Weiteren stellen sie einen überwiegend glatten Krümmungsverlauf der Fasern sicher. In der zweiten Szene sind die Einflüsse dieser Federn ausgeblendet und die Fasern sinken unter



Abbildung 8.18: Die Bilder zeigen die Simulation einer Faseranordnung mit aktivem bzw. inaktivem Biegefedermodell. Der Unterschied ist offensichtlich: Ohne Aktivierung der Biegefedern nehmen die Fasern eine zackige Form an. Die überwiegende Anzahl von Faserschlaufen existiert nur für einen kurzen Augenblick, bevor sie kollabieren. Bei aktivierten Biegefedern hingegen ist der Krümmungsverlauf entlang der Fasern glatt. Schlaufen bleiben stabil, bis sie durch die Wucht kollidierender Segmente niedergerissen werden.

ihrem Eigengewicht in eine völlig unregelmäßige Form zusammen.

8.6.5.2 Reibungsanisotropie

Ein in den gängigen Modellen zur Simulation von Haaren in der Computergraphik eher stiefmütterlich behandelter Aspekt ist der Einfluss der Reibung. Wie weiter oben bereits berichtet wurde, zeigt sich bei den typischen Kämm- und Bürstversuchen der kosmetischen Industrie, dass u.a. der Einfluss der Reibung auf das Haarvolumen ganz erheblich ist. Hier gilt die Heuristik, dass, je größer die Reibung zwischen den Fasern ausfällt, desto größer das Volumen der Anordnung ist. Neben diesen Faktoren spielen auch noch andere Einflüsse, wie bspw. die Steifigkeit der Fasern eine Rolle, die wir an dieser Stelle nicht mit in die Untersuchung einbeziehen wollen. In diesem

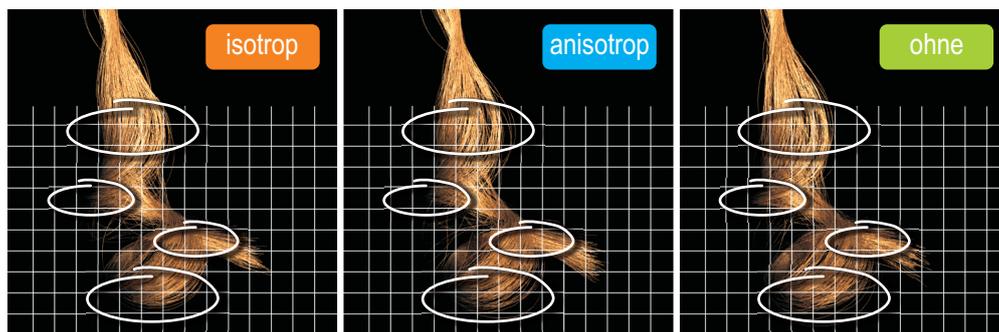


Abbildung 8.19: Die Anordnung aus dem Referenzbeispiel zeigt frappante Unterschiede in Abhängigkeit vom eingesetzten Reibungsmodell (isotrop, anisotrop und ohne Reibung). Markante Stellen sind durch ellipsoide Markierungen hervorgehoben.

Zusammenhang sind wir an der Frage interessiert, ob sich die seit langem bekannten Befunde der kosmetischen Industrie durch eine numerische Simulation bestätigen lassen. Abb. (8.19) zeigt die Ergebnisse einer solchen Simulation mit dem Referenzbeispiel. Alle drei Bilder entstammen der Simulation nach ca. 0,7 Sekunden. Diese wurden unter denselben Bedingungen gefahren, jedoch mit unterschiedlichen Reibungsmodellen. Im ersten Fall wählen wir ein isotropes Modell, d.h., wir behandeln die Faser als Zylinder mit gleicher Oberflächenrauigkeit in allen Richtungen und ordnen ihr einen Reibungsbeiwert von $\mu=0,47$ zu. Im zweiten Fall wechseln wir zu einem anisotropen Reibungsmodell. Dabei ordnen wir den Richtungen parallel, antiparallel, sowie orthogonal zur Faserachse unterschiedliche Reibungsbeiwerte zu. Das geschieht unter Einbezug des oben vorgestellten Konzepts der Limitkurven. Im dritten Fall verzichten wir auf die Reibung, d.h., die Berechnung eines adäquaten Reibungsimpulses unterbleibt.

Die Unterschiede fallen deutlich aus. Der anisotrope Fall zeigt weniger Auffächerung im Bereich der Kreuzungsstelle gegenüber dem isotropen Fall. Darüber hinaus scheint das Volumen etwas kleiner. Einzelne Fasern separieren sich klar vom Rest der Anordnung. Der Fall ohne Reibung zeigt das kleinste Volumen. Durch die fehlende Reibung gleiten die auf der Kugeloberfläche liegenden Fasern übereinander hinweg und nehmen in ihrer Endposition weniger Raum ein, als in den reibungsbehafteten Fällen. Diese Befunde lassen sich auch quantifizieren, wenn wir das Volumen des einschließenden achsenparallelen Quaders betrachten, vgl. Tab. (8.1).

Reibungsmodell	Länge [cm]	Breite [cm]	Höhe [cm]	Volumen [cm ³]	Verh. isotr. [%]
Isotrop	8,69	14,09	8,35	1022,39	0,99
Anisotrop	8,53	14,09	8,59	1032,41	1,00
Keine Reibung	8,12	13,98	8,24	935,38	0,91

Tabelle 8.1: Volumen des einschließenden achsenparallelen Quaders bei Verwendung von unterschiedlichen Reibungsmodellen. Die Befunde bestätigen, was in der kosmetischen Industrie seit langem bekannt ist: Reibung beeinflusst das Volumen.

8.6.5.3 Harmonisierung der Zwangsbedingungen

Bei der Harmonisierung der Zwangsbedingungen sind zwei sich einander partiell widersprechende Projektionen iterativ in Übereinstimmung zu bringen.

Das lokale Kontaktverhalten hängt davon ab, welche der beiden Projektion man zuerst ausführt. Beide haben Vor- bzw. Nachteile. Erfolgt die Behebung der Durchdringungen vor der Projektion zur Einhaltung der Gelenkzwänge, kommt es in der Folge zu einer erneuten Verletzung der Kontaktbedingungen. Bei zwei aufeinander liegenden Fasern manifestiert sich diese Verletzung in einem sukzessiven Durchsacken der oberen durch die untere Faser. Behebt man dagegen die Durchdringungen erst nach der Erfüllung der Gelenkzwänge, driften die Segmente der Fasern an den Gelenkpunkten wieder auseinander. Dieser Fehlzustand wird in der nachfolgenden Iteration durch einen entsprechenden Korrekturimpuls behoben, der wiederum die Durchdringung vergrößert, usw. Aus Sicht der Volumenerhaltung ist ein Durchsacken nicht akzeptabel, weil stabile Kontakte einen der Stützpfiler der Volumenerhaltung bei der faserbasierten Haarsimulation darstellen. Aus diesem Grunde folgt unser Algorithmus bei der Harmonisierung der Zwangsbedingungen stets Variante zwei.

In der Praxis hat sich zur Eingrenzung der durch Variante zwei verursachten Nebeneffekte die Wahl hinreichend kleiner Schrittweiten bewährt. Mit $\Delta t = 1/1.000$ Sekunde liegt man hier auf der sicheren Seite. Es hat sich zudem gezeigt, dass in allen hier vorgestellten numerischen Beispielen bereits eine bis drei Iteration zur Harmonisierung ausreichend sind.

8.7 Ausblicke

In Rahmen dieser Arbeit haben wir ein parametrisches Modell zur Simulation von Faseranordnungen auf der Basis einzelner Fasern entwickelt. Für die Simulation des mechanischen Verhaltens der Einzelfaser wurden dabei unterschiedliche diskrete und kontinuumsmechanische Ansätze, hier insbesondere die spezielle Theorie der COSSERAT-Stäbe, besprochen und numerische Lösungsverfahren entwickelt. Die Interaktionsmechanismen zwischen den Fasern werden durch ein geeignetes Zeitschrittverfahren auf der Basis der Theorie der Differentialinklusionen abgebildet. Die Effizienz unseres Ansatzes wurde an zahlreichen komplexen Beispielen demonstriert. Insbesondere konnte gezeigt werden, dass eine konsequente Behandlung der Faser-Faser-Interaktionen die Hauptkomponente zur Volumenerhaltung in solchen Anordnungen bildet. Erst die physikalisch korrekte, faserbasierte Geometriesimulation in Verbindung mit geeigneten Renderingverfahren ermöglicht einen bis dato im Bereich der Haarsimulation einmaligen Realismus.

Was bleibt also zu tun?

1. Bisher wurde die Kollisionsantwort lediglich in Kombination mit der impulsbasierten Darstellung von kinematischen Ketten betrachtet. Der Nachteil liegt auf der Hand. Partiiell sich widersprechende Zwangsbedingungen uni- und bilateraler Natur, die durch Kontakte und Gelenke induziert werden, erfordern eine iterative Harmonisierung, die bspw. bei Fasern, die unter hohen Zugspannungen stehen, nicht gelingen will. Zur Projektion des globalen Lösungsvektors auf die Schnittmenge der durch die Zwangsbedingungen aufgespannten zulässigen Teilräume bedarf es hingegen geeigneter Lösungsstrategien für hochdimensionale Optimierungsprobleme mit einer großen Anzahl an nicht linearen Nebenbedingungen. Zur Beseitigung dieser Probleme bietet sich der Einsatz eines implizit beschränkten Ansatzes, wie der speziellen Theorie der COSSERAT-Stäbe an, bei der die „bilateralen Zwänge“ bereits durch entsprechende Differentialgleichungen in kontinuierlicher Form Eingang in das Gesamtsystem gefunden haben. Wie man mit den daraus notwendigerweise resultierenden Steifigkeitsproblemen fertig wird, haben wir im Rahmen dieser Arbeit ausführlich diskutiert. Im Hinblick auf die Einflechtung stoßartiger Kräfte in die COSSERAT-Theorie, kann, wie weiter oben bereits angedeutet, vom Konzept der Differentialinklusionen Gebrauch gemacht werden. Die Ausarbeitung entsprechender Gleichungen würde allerdings den Rahmen dieser Arbeit sprengen und bleibt damit der Zukunft vorbehalten.
2. Betrachtet man die klassische COULOMBSche Reibung, die hier auf den anisotropen Fall generalisiert wurde, als Teilaspekt des Überbegriffes der Adhäsion, ist die Erweiterung auf andere Faser-Faser „Wechselwirkungen“ makroskopischer Natur, wie bspw. das Zusammenkleben infolge von auf der Oberfläche haftenden Substanzen naheliegend. Während die generalisierte impulsbasierte COULOMBSche Reibung lediglich der Tangentialgeschwindigkeit entgegenwirkt, beeinflusst die Adhäsion im klassischen Sinne auch die Geschwindigkeit normal dazu. Eine entsprechende Erweiterung des projektiven Ansatzes scheint generell möglich.
3. Die Beispiele, die zur Demonstration unseres Ansatzes vorgestellt wurden, dürfen mit Recht als komplex bezeichnet werden. Zur Berechnung vollständiger Frisuren ist eine Beschleunigung bestimmter Teilaspekte der Prozesskette jedoch sinnvoll. Die Integration der Differentialgleichungen im Rahmen des Zeitschrittverfahrens ist für jede Faser einer Anordnung durchzuführen und kann wegen der Unabhängigkeit dieses Teilaspektes parallelisiert werden. In diesem Kontext bietet sich der Einsatz entsprechender Graphikhardware an (z.B. NVIDIA TeslaTM).

Die systematische Erweiterung des Ansatzes im oben vorgeschlagenen Rahmen ermöglicht die simultane Betrachtung von Faser-Faser-Interaktionen, sowohl attraktiver (Wasser, Lipide, kosmetische Produkte, etc.), als auch repulsiver (elektrostatische Kräfte, Kollisionen) Natur. Unter Berücksichtigung der o.a. Aspekte scheint eine akkurate Berechnung des mechanischen Verhaltens komplexer Frisuren auf der Basis einzelner Fasern mit „vertretbarem“ Aufwand bereits in greifbare Nähe gerückt.

A Anhang

A.1 Notation

Die Notation ist in den Tabellen Tab. (A.1), Tab. (A.2) und Tab. (A.3) zusammengefasst. Auf die Angabe von Funktionsparametern oder Indices wurde dabei verzichtet.

Symbol	Bedeutung
g	Bilateraler Zwang im \mathbb{R}^3
h	Unilateraler Zwang im \mathbb{R}^3
∇h	Gradient des unilateralen Zwanges im \mathbb{R}^3
i, j, k, l	Laufvariablen
k	Modul
m	Masse (pro Einheitslänge)
m_3	Projektion internes Moment auf Tangente \mathbf{d}_3
m_z	Projektion internes Moment auf globale Z-Achse
n_z	Projektion interne Kraft auf globale Z-Achse
t	Zeit
Δt	Zeitschrittweite
s	Kurvenparameter
Δs	Räumliche Schrittweite
L, l	Länge, LAGRANGE-Funktion
z	Substitution für $\cos \theta$
z_1, z_2, z_3	Reelle Wurzeln des kubischen Polynoms
A	Querschnittsfläche
E	E-Modul, YOUNGSche-Zahl, Energie
G	Torsionsmodul, Schubmodul
\mathcal{H}	HAMILTON-Funktion
T	Kinetischen Energie
U	Potentielle Energie
W	Energiefunktion (elastische Energie einer Faser)
I_1, I_2, I_3	Hauptflächenträgheitsmomente pro Einheitslänge
ε	Restitutionskoeffizient
$\alpha_s, \alpha_t, \beta_s, \beta_t, \gamma_s, \gamma_t$	Beiwerte des generalisierten α -Verfahrens
λ	Eigenwert, LAGRANGE-Multiplikator
ϕ, θ, ψ	EULER-Winkel
ρ	Spektralradius
ρ	Lineare Dichte
μ	Reibungsbeiwert, Relaxationsparameter, polares Flächenträgheitsmoment pro Einheitslänge
Γ	Logarithmisches Dekrement

Tabelle A.1: Notation. Skalare.

A.2 Der Lagrange-Formalismus

Über die Anwendung des LAGRANGE-Formalismus kommt man schnell zu den Bewegungsgleichungen selbst komplexer Systeme. Voraussetzung dafür ist die Einführung von verallgemeiner-

Symbol	Bedeutung
\mathbf{a}	Beschleunigung, lokaler Vektor
\mathbf{d}_k	Koordinatenachsen
\mathbf{e}	Einheitsvektor
\mathbf{f}	Externe Kraft
\mathbf{g}	Gravitationsvektor
\mathbf{l}	Externes Moment
\mathbf{m}	Internes Moment
\mathbf{n}	Normale, interne Kraft
\mathbf{p}	Impuls
\mathbf{q}	Transformation $\{\mathbf{x}, \mathbf{R}\}$, Quaternion
$\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}$	Generalisierte Koordinaten, Geschwindigkeiten, Beschleunigungen
\mathbf{r}	Ortsvektor
\mathbf{u}	Räumliche Geschwindigkeit $(\{\mathbf{v}, \boldsymbol{\omega}\})$
\mathbf{v}	Dehnung (Scherung und Extension)
\mathbf{w}	Gewichtungsvektor
\mathbf{x}	Kontaktpunkt, Zustandsvektor
$\boldsymbol{\kappa}$	DARBOUX-Vektor, Biegung und Torsion
$\boldsymbol{\kappa}^\circ$	Initiale Biegung und Torsion
$\boldsymbol{\omega}$	Winkelgeschwindigkeit, Twist-Vektor
$\boldsymbol{\nu}$	Reaktionskraft
$\boldsymbol{\Psi}$	Gesamtreaktionskraft
\mathbf{Y}	Dissipationspotential
$\boldsymbol{\Lambda}$	Kraftterme in Strukturgleichung
$\boldsymbol{\Omega}$	Drehachse
\mathbf{A}	Dämpfungsmatrix (Stabilitätsanalyse)
\mathbf{C}	CORIOLIS-Matrix
\mathbf{D}	Dämpfungsmatrix, Generatormatrix
\mathbf{E}_3	Einheitsmatrix $\mathbb{R}^{3 \times 3}$
\mathbf{I}	Trägheitstensor
\mathbf{J}	Jacobimatrix
\mathbf{K}	Steifigkeitsmatrix
\mathbf{L}	Drehimpuls
\mathbf{M}	Massematrix
\mathbf{P}	Räumlicher Impuls $(\{\mathbf{p}, \mathbf{L}\})$
\mathbf{Q}	Generalisierte Kräfte in LAGRANGE-Gleichung
\mathbf{R}	Rotationsmatrix
\mathbf{T}	Tangente

Tabelle A.2: Notation. Vektoren und Matrizen.

ten Koordinaten $\mathbf{q}(t) = (q_1(t), \dots, q_n(t))$, die den Systemzustand zu jedem Zeitpunkt t eindeutig charakterisieren. Die generalisierten Koordinaten spannen den Konfigurationsraum des Systems auf. Man definiert die so genannte LAGRANGE-Funktion als

$$L(\mathbf{q}, \dot{\mathbf{q}}, t) = T(\mathbf{q}, \dot{\mathbf{q}}, t) - U(\mathbf{q}, t). \quad (\text{A.1})$$

Hierin sind T und U die kinetische bzw. die potentielle Energie des Systems, ausgedrückt durch seine verallgemeinerten Koordinaten. Die Bewegungs- oder LAGRANGE-Gleichungen zweiter Art nehmen die folgende Form an,

$$\frac{d}{dt} \frac{\partial}{\partial \dot{q}_i} L(\mathbf{q}, \dot{\mathbf{q}}, t) - \frac{\partial}{\partial q_i} L(\mathbf{q}, \dot{\mathbf{q}}, t) = 0, \quad (\text{A.2})$$

mit den verallgemeinerten Geschwindigkeiten $\dot{\mathbf{q}} = (\dot{q}_1, \dots, \dot{q}_n)$. Nach Herleitung der Energiefunktionen T und U ergeben sich durch Einsetzen und Differenzieren die Bewegungsgleichungen als System gewöhnlicher Differentialgleichungen zweiter Ordnung, die durch Integration zu

Symbol	Bedeutung
\mathcal{A}	Menge aktiver Kontakte
\mathcal{B}	Starrkörper
\mathcal{C}	Konfiguration einer Faser
\mathcal{M}	Konfigurationsraum
\mathcal{N}	Normalenkegel
\mathcal{L}^c	Limitkurve
\mathcal{L}^s	Limitfläche
\mathcal{T}	Zusammengesetzter Tangentenkegel
\mathcal{U}	Kontaktmenge
\mathcal{FC}	Reibungskegel
$T_p\mathcal{M}$	Tangentialraum zum Konfigurationsraum
$T_p^*\mathcal{M}$	Kotangentialraum zum Konfigurationsraum
Φ	Menge zulässiger Systemzustände, Distanzfeld
sn, cn, dn	JACOBIsche elliptische Funktionen
$F(\pi/2, k)$	unvollständiges elliptisches Integral der ersten Art
$E(\phi, k)$	unvollständiges elliptisches Integral der zweiten Art
$\Pi(\phi, n, k)$	unvollständiges elliptisches Integral der dritten Art
$K(k)$	vollständiges elliptisches Integral der ersten Art
∂_t	Partielle Ableitung nach der Zeit
∂_s	Partielle Ableitung nach dem Kurvenparameter s
\square	Zeitableitung einer Größe
\square'	Ableitung einer Größe nach dem Kurvenparameter s
\square^{-1}	Inverse
\square^T	Transponierte
$\tilde{\square}$	Schief-symmetrische Matrixdarstellung eines Vektors

Tabelle A.3: Notation. Mengen, elliptische Integrale und andere.

lösen sind. In Ausnahmefällen, u.a. bei sehr einfachen mechanischen Systemen, ist eine analytischen Integration der Ausdrücke möglich.

Die LAGRANGE-Gleichungen selber lassen sich aus der Forderung rechtfertigen, dass das Wirkungsintegral $S = \int_{t_1}^{t_2} L(\mathbf{q}, \dot{\mathbf{q}}, t) dt$ der LAGRANGE-Funktion stationär werden muss. Das entspricht dem HAMILTONSchen Prinzip oder dem Prinzip der kleinsten Wirkung. Danach ist jedes mechanische System durch die LAGRANGE-Funktion bzw. durch seine generalisierten Koordinaten und Geschwindigkeiten eindeutig bestimmt. Die Bewegung des System zwischen zwei Zuständen zu den Zeitpunkten t_1 und t_2 verläuft so, dass das Wirkungsintegral einen extremen Wert annimmt.

Der LAGRANGE-Formalismus kann auf nicht konservative Systeme erweitert werden. Das sind Systeme, in denen nicht ausschließlich konservative Kräfte wirken. Damit sind Potentialkräfte gemeint, wie bspw. die Gewichtskraft eines Körpers infolge Gravitation. In diesem Fall werden die generalisierten Kräfte Q_i , die am System angreifen, auf die rechte Seite der LAGRANGE-Gleichungen geschrieben. Auch eine Erweiterung auf Nebenbedingungen der Form $\sum_i b_{k_i}(\mathbf{q}_i, \dot{\mathbf{q}}_i, t) \delta q_i$ ist durch Einführung von LAGRANGE-Multiplikatoren möglich. Auf der rechten Seite erscheint ein weiterer Term der Form $\sum_i \lambda_i b_{k_i}$, dessen Multiplikatoren dann bei der Integration zu bestimmen sind.

Beispiel A.2.1 Wir geben im Folgenden ein einfaches Beispiel zum besseren Verständnis an: den Körper auf der schiefen Ebene mit der Neigung α . Als generalisierte Koordinate wählt man die Weglänge q , die der Körper auf der schiefen Ebene zurücklegt. Wir führen zunächst eine Variablentrans-

formation durch,

$$\mathbf{x}(q) = \{q \cos \alpha, 0, q \sin \alpha\}.$$

Die kinetische und potentielle Energie des Systems bzw. die korrespondierende LAGRANGE-Funktion sind dann

$$T(\mathbf{x}, \dot{\mathbf{x}}) = \frac{1}{2} m \dot{\mathbf{x}}^2, \quad U(\mathbf{x}) = m \mathbf{g} x_z,$$

$$L(q, \dot{q}) = T(q, \dot{q}) - U(q) = \frac{1}{2} m \dot{q}^2 - m \mathbf{g} q \sin \alpha.$$

Durch Anwenden des LAGRANGE-Formalismus erhält man die folgende gewöhnliche Differentialgleichung erster Ordnung,

$$\frac{d}{dt} \dot{q}(t) = -g \sin \alpha,$$

mit der bekannten analytischen Lösung

$$q(t) = -\frac{1}{2} g t^2 \sin \alpha + \dot{q}_0 t + q_0.$$

Auch die Reibung zwischen Körper und Unterlage kann berücksichtigt werden. Dazu braucht man die Normalkraft ($\mathbf{N} = m \mathbf{g} \cos \alpha$), die aber infolge der generalisierten Koordinatendarstellung nicht direkt angegeben werden kann. Wir wenden das Prinzip der virtuellen Verrückungen zur Ermittlung der generalisierten Kräfte $Q_{q_i} = \delta W / \delta q_i$ an,

$$\begin{aligned} \delta W &= [-\operatorname{sgn}(q) \mu \mathbf{N}] \delta q \\ &= [-\operatorname{sgn}(q) \mu m \mathbf{g} \cos \alpha] \delta q \\ Q_{q_i} &= \delta W / \delta q_i \\ &= -\operatorname{sgn}(q) \mu m \mathbf{g} \cos \alpha. \end{aligned}$$

Beispiel A.2.2 Als weiteres Beispiel sollen die Bewegungsgleichungen für den symmetrischen Drehkreisel hergeleitet werden. Alle Größen werden in Bezug auf ein körperfestes Koordinatensystem angegeben. Als generalisierte Koordinaten wählt man die drei EULER-Winkel $q_i = \{\phi, \theta, \psi\}$. Die lokale Winkelgeschwindigkeit $\boldsymbol{\omega}$ in Abhängigkeit von den EULER-Winkeln ist

$$\boldsymbol{\omega} = \begin{pmatrix} \sin \theta \sin \psi & \cos \psi & 0 \\ \sin \theta \cos \psi & -\sin \psi & 0 \\ \cos \theta & 0 & 1 \end{pmatrix} \begin{pmatrix} \dot{\phi} \\ \dot{\theta} \\ \dot{\psi} \end{pmatrix}.$$

Mit dem Trägheitstensor $\mathcal{I} = \operatorname{diag}(I_1, I_2, I_3)$ und der Symmetriebedingung $I_1 = I_2$ kann die kinetische Energie des Kreisels durch die EULER-Winkel bzw. ihre Zeitableitungen $\dot{q}_i = \{\dot{\phi}, \dot{\theta}, \dot{\psi}\}$ ausgedrückt werden.

$$\begin{aligned} T &= \frac{1}{2} \boldsymbol{\omega}^T \mathcal{I} \boldsymbol{\omega} \\ &= \frac{1}{2} \left[I_1 (\dot{\phi} \sin \theta \sin \psi + \dot{\theta} \cos \psi)^2 + I_2 (\dot{\phi} \sin \theta \cos \psi - \dot{\theta} \sin \psi)^2 + \right. \\ &\quad \left. I_3 (\dot{\phi} \cos \theta + \dot{\psi})^2 \right] \\ &= \frac{1}{2} \left[I_1 (\dot{\phi}^2 \sin^2 \theta + \dot{\theta}^2) + I_3 (\dot{\phi} \cos \theta + \dot{\psi})^2 \right]. \end{aligned}$$

Die Potentielle Energie ist $U = U(\phi, \theta, \psi) = m\mathbf{g} \cos \theta$, so dass sich schließlich mit der LAGRANGE-Funktion $L(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) - U(\mathbf{q})$ die folgenden Bewegungsgleichungen ergeben,

$$\begin{aligned} \frac{d}{dt} [I_1 \dot{\phi} \sin^2 \theta + I_3 \cos \theta (\dot{\phi} \cos \theta + \dot{\psi})] + \frac{\partial}{\partial \phi} U &= 0, \\ I_1 (\ddot{\theta} - \sin \theta \cos \theta \dot{\phi}^2) + I_3 \dot{\phi} \sin \theta (\dot{\phi} \cos \theta + \dot{\psi}) + \frac{\partial}{\partial \theta} U &= 0, \\ I_3 \frac{d}{dt} (\dot{\phi} \cos \theta + \dot{\psi}) + \frac{\partial}{\partial \psi} U &= 0. \end{aligned}$$

Die Standardform der LAGRANGE-Funktion lässt sich sehr elegant auf dissipative Kräfte erweitern, die von den generalisierten Geschwindigkeiten abhängen. Im ersten Beispiel hatten wir lediglich Gleitreibung betrachtet, die in den Term der generalisierten Kräfte aufgenommen wird. Eine besondere Form der geschwindigkeitsabhängigen Reibung wird durch die RAYLEIGH-Dissipationsfunktion beschrieben,

$$\Upsilon(\mathbf{q}, \dot{\mathbf{q}}, t) = \frac{1}{2} \dot{\mathbf{q}}^T \mathbf{D}(\mathbf{q}, t) \dot{\mathbf{q}} = \frac{1}{2} \sum_i \sum_j \mathbf{D}_{ij}(\mathbf{q}, t) \dot{q}_i \dot{q}_j. \quad (\text{A.3})$$

Man kann zeigen, dass die Ableitung der Dissipationsfunktion nach den generalisierten Geschwindigkeiten folgende Form annimmt,

$$\frac{\partial}{\partial \dot{q}_i} \Upsilon(\mathbf{q}, \dot{\mathbf{q}}, t) = \sum_j \mathbf{D}_{ij}(\mathbf{q}, t) \dot{q}_j = -Q_i. \quad (\text{A.4})$$

Dissipative Kräfte werden in einem separaten Term auf die linke Seite der LAGRANGE-Gleichungen geschrieben,

$$\frac{d}{dt} \frac{\partial}{\partial \dot{q}_i} L(\mathbf{q}, \dot{\mathbf{q}}, t) - \frac{\partial}{\partial q_i} L(\mathbf{q}, \dot{\mathbf{q}}, t) + \frac{\partial}{\partial \dot{q}_i} \Upsilon(\mathbf{q}, \dot{\mathbf{q}}, t) = Q_i, \quad (\text{A.5})$$

wobei Q_i jetzt ausschließlich externe Kräfte, z.B. infolge Reibung oder Kollisionen beschreibt.

A.3 Berechnung der Wurzeln

Wir betrachten die rechte Seite der Gleichung

$$\left(\frac{dz}{ds} \right)^2 = 2n(z - z_1)(z_2 - z)(z_3 - z). \quad (\text{A.6})$$

Diese kann als kubisches Polynom der Form

$$P = c_1 z^3 + c_2 z^2 + c_3 z + c_4 \quad (\text{A.7})$$

ausgedrückt werden. Die Berechnung der Wurzeln $\{z_1, z_2, z_3\}$ von P führt auf,

$$z_1 = -2\sqrt{Q} \cos\left(\frac{\zeta}{3}\right) - \frac{c_1}{3}, \quad (\text{A.8})$$

$$z_2 = -2\sqrt{Q} \cos\left(\frac{\zeta + 2\pi}{3}\right) - \frac{c_1}{3}, \quad (\text{A.9})$$

$$z_3 = -2\sqrt{Q} \cos\left(\frac{\zeta - 2\pi}{3}\right) - \frac{c_1}{3}, \quad (\text{A.10})$$

mit

$$Q = \frac{c_1^2 - 3c_2}{9}, \quad (\text{A.11})$$

$$R = \frac{2c_1^3 - 9c_1c_2 + 27c_3}{54}, \quad (\text{A.12})$$

$$\zeta = \arccos \frac{R}{\sqrt{Q^3}}. \quad (\text{A.13})$$

A.4 Analytische Jacobi-Matrizen für die α -Methode

Im Folgenden geben wir die JACOBI-Untermatrizen $\partial\mathbf{F}/\partial\mathbf{x}_j^i$ an, die für das NEWTON-Verfahren zur Lösung der dynamischen COSSERAT-Gleichungen gebraucht werden. Auf die Angabe entsprechender Matrizen für $\partial\mathbf{F}/\partial\mathbf{x}_{j-1}^i$ verzichten wir an dieser Stelle aus Platzgründen.

$$\frac{\partial\mathbf{F}}{\partial v_j^i} = \begin{bmatrix} -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} & (1-\beta_t)(1-\beta_s)(\kappa_z)_j^i & -(1-\beta_t)(1-\beta_s)(\kappa_y)_j^i \\ -(1-\beta_t)(1-\beta_s)(\kappa_z)_j^i & -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} & (1-\beta_t)(1-\beta_s)(\kappa_x)_j^i \\ (1-\beta_t)(1-\beta_s)(\kappa_y)_j^i & -(1-\beta_t)(1-\beta_s)(\kappa_x)_j^i & -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ \frac{m(1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} & -(1-\beta_t)(1-\beta_s)m(\omega_z)_j^i & (1-\beta_t)(1-\beta_s)m(\omega_y)_j^i \\ (1-\beta_t)(1-\beta_s)m(\omega_z)_j^i & \frac{m(1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} & -(1-\beta_t)(1-\beta_s)m(\omega_x)_j^i \\ -(1-\beta_t)(1-\beta_s)m(\omega_y)_j^i & (1-\beta_t)(1-\beta_s)m(\omega_x)_j^i & \frac{m(1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \end{bmatrix} \quad (\text{A.14})$$

$$\frac{\partial \mathbf{F}}{\partial \omega_x^i} = \begin{bmatrix} 0 \\ (1-\beta_t)(-1+\beta_s) \\ 0 \\ -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \\ -(1-\beta_t)(1-\beta_s)(\kappa_z)_j^i \\ (1-\beta_t)(1-\beta_s)(\kappa_y)_j^i \\ 1/4 \frac{\rho_c \pi r^4 (1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \\ -1/4 (1-\beta_t)(1-\beta_s) \rho_c \pi r^4 (\omega_z)_j^i \\ 0 \\ 0 \\ -(1-\beta_t)(1-\beta_s) m(v_z)_j^i \\ (1-\beta_t)(1-\beta_s) m(v_y)_j^i \end{bmatrix} \quad (\text{A.15})$$

$$\frac{\partial \mathbf{F}}{\partial \omega_y^i} = \begin{bmatrix} (1-\beta_t)(1-\beta_s) \\ 0 \\ 0 \\ (1-\beta_t)(1-\beta_s)(\kappa_z)_j^i \\ -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \\ -(1-\beta_t)(1-\beta_s)(\kappa_x)_j^i \\ 1/4 (1-\beta_t)(1-\beta_s) \rho_c \pi r^4 (\omega_z)_j^i \\ 1/4 \frac{\rho_c \pi r^4 (1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \\ 0 \\ (1-\beta_t)(1-\beta_s) m(v_z)_j^i \\ 0 \\ -(1-\beta_t)(1-\beta_s) m(v_x)_j^i \end{bmatrix} \quad (\text{A.16})$$

$$\frac{\partial \mathbf{F}}{\partial \omega_z^i} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -(1-\beta_t)(1-\beta_s)(\kappa_y)_j^i \\ (1-\beta_t)(1-\beta_s)(\kappa_x)_j^i \\ -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \\ 1/4 (1-\beta_t)(1-\beta_s) (\omega_y)_j^i \rho_c \pi r^4 \\ -1/4 (1-\beta_t)(1-\beta_s) \rho_c \pi r^4 (\omega_x)_j^i \\ 1/2 \frac{\rho_c \pi r^4 (1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \\ -(1-\beta_t)(1-\beta_s) m(v_y)_j^i \\ (1-\beta_t)(1-\beta_s) m(v_x)_j^i \\ 0 \end{bmatrix} \quad (\text{A.17})$$

$$\frac{\partial \mathbf{F}}{\partial \kappa_x^i} = \begin{bmatrix} 0 \\ (1-\beta_t)(1-\beta_s)(v_z)_j^i \\ -(1-\beta_t)(1-\beta_s)(v_y)_j^i \\ \frac{(1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \\ (1-\beta_t)(1-\beta_s)(\omega_z)_j^i \\ -(1-\beta_t)(1-\beta_s)(\omega_y)_j^i \\ -\frac{EI_1(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \\ (1-\beta_t)(1-\beta_s) \left(-(\kappa_z)_j^i EI_1 + G\mu \left((\kappa_z)_j^i - (\kappa_z^o)_j \right) \right) \\ (1-\beta_t)(1-\beta_s) \left(-EI_2 \left((\kappa_y)_j^i - (\kappa_y^o)_j \right) + (\kappa_y)_j^i EI_1 \right) \\ 0 \\ (1-\beta_t)(1-\beta_s)(n_z)_j^i \\ -(1-\beta_t)(1-\beta_s)(n_y)_j^i \end{bmatrix} \quad (\text{A.18})$$

$$\frac{\partial \mathbf{F}}{\partial \kappa_y^i} = \begin{bmatrix} -(1-\beta_t)(1-\beta_s)(v_z)_j^i \\ 0 \\ (1-\beta_t)(1-\beta_s)(v_x)_j^i \\ -(1-\beta_t)(1-\beta_s)(\omega_z)_j^i \\ \frac{(1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \\ (1-\beta_t)(1-\beta_s)(\omega_x)_j^i \\ (1-\beta_t)(1-\beta_s) \left(-G\mu \left((\kappa_z)_j^i - (\kappa_z^o)_j \right) + (\kappa_z)_j^i EI_2 \right) \\ -\frac{EI_2(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \\ (1-\beta_t)(1-\beta_s) \left(-(\kappa_x)_j^i EI_2 + EI_1 \left((\kappa_x)_j^i - (\kappa_x^o)_j \right) \right) \\ -(1-\beta_t)(1-\beta_s)(n_z)_j^i \\ 0 \\ (1-\beta_t)(1-\beta_s)(n_x)_j^i \end{bmatrix} \quad (\text{A.19})$$

$$\frac{\partial \mathbf{F}}{\partial \kappa_z^i} = \begin{bmatrix} (1-\beta_t)(1-\beta_s)(v_y)_j^i \\ -(1-\beta_t)(1-\beta_s)(v_x)_j^i \\ 0 \\ (1-\beta_t)(1-\beta_s)(\omega_y)_j^i \\ -(1-\beta_t)(1-\beta_s)(\omega_x)_j^i \\ \frac{(1-\alpha_t)(1-\alpha_s)}{\gamma_t \Delta t} \\ (1-\beta_t)(1-\beta_s) \left(-(\kappa_y)_j^i G\mu + EI_2 \left((\kappa_y)_j^i - (\kappa_y^o)_j \right) \right) \\ (1-\beta_t)(1-\beta_s) \left(-EI_1 \left((\kappa_z)_j^i - (\kappa_z^o)_j \right) + (\kappa_x)_j^i G\mu \right) \\ -\frac{G\mu(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \\ (1-\beta_t)(1-\beta_s)(n_y)_j^i \\ -(1-\beta_t)(1-\beta_s)(n_x)_j^i \\ 0 \end{bmatrix} \quad (\text{A.20})$$

$$\frac{\partial F}{\partial n_j^i} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & (1-\beta_t)(1-\beta_s) & 0 \\ (1-\beta_t)(-1+\beta_s) & 0 & 0 \\ 0 & 0 & 0 \\ -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} & (1-\beta_t)(1-\beta_s)(\kappa_z)_j^i & -(1-\beta_t)(1-\beta_s)(\kappa_y)_j^i \\ -(1-\beta_t)(1-\beta_s)(\kappa_z)_j^i & -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} & (1-\beta_t)(1-\beta_s)(\kappa_x)_j^i \\ (1-\beta_t)(1-\beta_s)(\kappa_y)_j^i & -(1-\beta_t)(1-\beta_s)(\kappa_x)_j^i & -\frac{(1-\beta_t)(1-\beta_s)}{\gamma_s \Delta s} \end{bmatrix} \quad (\text{A.21})$$

A.5 Elliptische Integrale und Jacobische elliptische Funktionen

Ausgehend von der Definition für das unvollständige elliptische Integral der ersten Art,

$$F(\phi, k) = \int_0^\phi \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}}, \quad (\text{A.22})$$

mit dem Modul $k \in [0, 1]$ und der Amplitude ϕ ,

$$\phi = \text{am } F(\phi, k), \quad (\text{A.23})$$

definiert man die JACOBISCHEN elliptischen Funktionen sn , cn und dn wie folgt,

$$\text{sn } F(\phi, k) = \sin \phi, \quad (\text{A.24})$$

$$\text{cn } F(\phi, k) = \cos \phi, \quad (\text{A.25})$$

$$\text{dn } F(\phi, k) = \sqrt{1 - k^2 \sin^2 \phi}. \quad (\text{A.26})$$

Für $k = 0$ gilt,

$$\text{sn } F(\phi, 0) = \sin F(\phi, 0), \quad \text{cn } F(\phi, 0) = \cos F(\phi, 0), \quad \text{dn } F(\phi, 0) = 1, \quad (\text{A.27})$$

für $k = 1$ hingegen,

$$\text{sn } F(\phi, 1) = \tanh F(\phi, 1), \quad \text{cn } F(\phi, 1) = \text{sech } F(\phi, 1), \quad \text{dn } F(\phi, 1) = \text{sech } F(\phi, 1). \quad (\text{A.28})$$

Für die Ableitungen nach dem Parameter ϕ gelten die folgenden Beziehungen,

$$\frac{d}{d\phi} \text{sn } F(\phi, k) = \text{cn } F(\phi, k) \text{dn } F(\phi, k), \quad (\text{A.29})$$

$$\frac{d}{d\phi} \text{cn } F(\phi, k) = -\text{sn } F(\phi, k) \text{dn } F(\phi, k), \quad (\text{A.30})$$

$$\frac{d}{d\phi} \text{dn } F(\phi, k) = -k^2 \text{sn } F(\phi, k) \text{cn } F(\phi, k). \quad (\text{A.31})$$

Die elliptischen Integrale der zweiten und dritten Art folgen den Definitionen

$$E(\phi, k) = \int_0^\phi \sqrt{1 - k^2 \sin^2 \varphi} \, d\varphi, \quad (\text{A.32})$$

$$\Pi(\phi, n, k) = \int_0^\phi \frac{d\varphi}{(1 - n \sin^2 \varphi) \sqrt{1 - k^2 \sin^2 \varphi}}. \quad (\text{A.33})$$

Auf die vollständigen Integrale der ersten, zweiten und dritten Art kommt man durch Setzen von $\phi = \pi/2$. Für $F(\pi/2, k)$ ist die Abkürzung $K(k)$ üblich. Oftmals überführt man die elliptischen Integrale durch die Substitution $u = \sin \phi$ oder $s = F(\phi, k)$ in die vom Sinus der Amplitude oder dem Kurvenparameter s abhängigen Formen. Wir verzichten an dieser Stelle auf die Angabe der substituierten Formen und verweisen auf die einschlägige Literatur, vgl. z.B. [Abramowitz & Stegun, 1970].

A.6 Definition der Räume

A.6.1 Spezielle Orthogonale Gruppe, $\mathcal{SO}(3)$

Die spezielle orthogonale Gruppe $\mathcal{SO}(3)$ oder Drehgruppe im \mathbb{R}^3 ist die Menge aller linearen Transformationen, welche die Länge eines Vektors erhalten und die Eigenschaft erfüllen,

$$\mathcal{SO}(3) := \{ \mathbf{R} : \mathbf{R} \in \mathbb{R}^{3 \times 3}, \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det \mathbf{R} = +1 \}. \quad (\text{A.34})$$

Dies entspricht der Menge aller Rotationen um den Ursprung des \mathbb{R}^3 . Die LIE-Algebra der $\mathcal{SO}(3)$ wird als $so(3)$ bezeichnet und ist durch die Menge der schief-symmetrischen Matrizen der Form

$$\tilde{\boldsymbol{\omega}} := \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}, \quad (\text{A.35})$$

gegeben. Sie bildet einen Vektor $\boldsymbol{\omega} \in \mathbb{R}^3$ auf einer Matrix $\tilde{\boldsymbol{\omega}} \in so(3)$ ab. Ferner gilt

$$[\tilde{\boldsymbol{\omega}}_1, \tilde{\boldsymbol{\omega}}_2] = \tilde{\boldsymbol{\omega}}_1 \tilde{\boldsymbol{\omega}}_2 - \tilde{\boldsymbol{\omega}}_2 \tilde{\boldsymbol{\omega}}_1. \quad (\text{A.36})$$

Dies entspricht gerade der Kreuzproduktoperation im \mathbb{R}^3 , $\boldsymbol{\omega}_1 \times \boldsymbol{\omega}_2$, mit $\boldsymbol{\omega}_1, \boldsymbol{\omega}_2 \in \mathbb{R}^3$. Für beliebige Vektoren $\mathbf{p} \in \mathbb{R}^3$ ist die Multiplikation mit der schief-symmetrischen Matrixdarstellung von $\boldsymbol{\omega}$ identisch mit der Kreuzproduktoperation im \mathbb{R}^3 , $\tilde{\boldsymbol{\omega}} \mathbf{p} \equiv \boldsymbol{\omega} \times \mathbf{p}$.

A.6.2 Spezielle Euklidische Gruppe, $\mathcal{SE}(3)$

Die Spezielle Euklidische Gruppe $\mathcal{SE}(3)$ ist die Menge aller homogenen Starrkörpertransformationen im dreidimensionalen EUKLIDischen Raum \mathbb{R}^3 ,

$$\mathcal{SE}(3) := \left\{ A : A = \begin{bmatrix} \mathbf{R} & \mathbf{d} \\ 0 & 1 \end{bmatrix}, \mathbf{R} \in \mathbb{R}^{3 \times 3}, \mathbf{d} \in \mathbb{R}^3, \mathbf{R}^T \mathbf{R} = \mathbf{I}, \det \mathbf{R} = 1 \right\}. \quad (\text{A.37})$$

Man kann zeigen, dass $\mathcal{SE}(3)$ eine Mannigfaltigkeit und bei Matrixmultiplikation wie eine Gruppe zu behandeln ist. $\mathcal{SE}(3)$ ist eine LIE-Gruppe der Dimension sechs. Der Tangentenraum einer

LIE-Gruppe hat die Struktur einer LIE-Algebra. Die LIE-Algebra der $\mathcal{SE}(3)$ wird als $se(3)$ bezeichnet und ist wie folgt definiert,

$$se(3) := \left\{ \begin{bmatrix} \tilde{\boldsymbol{\omega}} & \boldsymbol{v} \\ 0 & 1 \end{bmatrix} : \tilde{\boldsymbol{\omega}} \in \mathbb{R}^{3 \times 3}, \boldsymbol{v} \in \mathbb{R}^3, \tilde{\boldsymbol{\omega}}^T = -\tilde{\boldsymbol{\omega}} \right\}. \quad (\text{A.38})$$

$\tilde{\boldsymbol{\omega}} \in so(3)$ ist die schief-symmetrische Matrixdarstellung des Vektors $\boldsymbol{\omega} \in \mathbb{R}^3$.

Literaturverzeichnis

- [Abramowitz & Stegun, 1970] Abramowitz, M. & Stegun, I. (1970). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. Dover, New York, NY.
- [Acary & Brogliato, 2005] Acary, V. & Brogliato, B. (2005). Numerical time integration of higher order dynamical systems with state constraints. In *ENOC 2005, Fifth Euromech Nonlinear Dynamics conference*.
- [Adams et al., 1990] Adams, M. J., Briscoe, B. J., & Wee, T. K. (1990). The differential friction effect of Keratin fibres. *J. Phys. D: Appl. Phys.*, 23, 406–414.
- [Anjyo et al., 1992] Anjyo, K., Usami, Y., & Kurihara, T. (1992). A simple method for extracting the natural beauty of hair. In *ACM SIGGRAPH Computer Graphics*, volume 26(2) (pp. 111–120).
- [Antman, 1995] Antman, S. S. (1995). *Nonlinear Problems of Elasticity*, volume 107 of *Appl. Math. Sci.* Berlin and New York: Springer-Verlag.
- [Bando et al., 2003] Bando, Y., Chen, B.-Y., & Nishita, T. (2003). Animating hair with loosely connected particles. In *Computer Graphics Forum (Proceedings of Eurographics 2003)*, volume 22(3).
- [Baraff, 1989] Baraff, D. (1989). Analytical methods for dynamic simulation of non-penetrating rigid bodies. *ACM SIGGRAPH Computer Graphics*, 23(3), 223–232.
- [Baraff, 1991] Baraff, D. (1991). Coping with friction for non-penetrating rigid body simulation. *ACM SIGGRAPH Computer Graphics*, 25(4), 31–41.
- [Baraff & Witkin, 2003] Baraff, D. & Witkin, A. (2003). Physically based modeling. In *SIGGRAPH 2003 Course Notes*.
- [Barequet et al., 1996] Barequet, G., Chazelle, B., Guibas, L. J., Mitchell, J. S. B., & Tal, A. (1996). BOXTREE: A hierarchical representation for surfaces in 3D. *Computer Graphics Forum*, 15(3), 387–396.
- [Barequet & Har-Peled, 1999] Barequet, G. & Har-Peled, S. (1999). Efficiently approximating the minimum-volume bounding box of a point set in three dimensions. In *SODA: ACM-SIAM Symposium on Discrete Algorithms (A Conference on Theoretical and Experimental Analysis of Discrete Algorithms)*.
- [Barnes & Hut, 1986] Barnes, J. & Hut, P. (1986). A hierarchical $O(N \log N)$ force-calculation algorithm. *Nature*, 324, 446–449.
- [Bender, 2007] Bender, J. (2007). *Impulsbasierte Dynamiksimulation von Mehrkörpersystemen in der virtuellen Realität*. PhD-thesis in computer science, Universität Fridericiana zu Karlsruhe (TH).
- [Bereg, 2004] Bereg, S. (2004). Cylindrical hierarchy for deforming necklaces. *International Journal of Computational Geometry and Applications*, 14(1–2), 3–18.
- [Bernoulli, 1728] Bernoulli, D. (1728). Methodus universalis determinandae curvaturae filii a potentiis quamcunque legem inter se, observantibus extensi, una cum solutione problematum quorundam novorum eo pertinentium. *Comm. Acad. Sci. Petrop.*, 3, 62–69.
- [Bernoulli, 1694] Bernoulli, J. (1694). Curvatura laminae elasticae. *Opera*, 1, 576–600.
- [Bertails et al., 2006] Bertails, F., Audoly, B., Cani, M.-P., Querleux, B., Leroy, F., & Lévêque, J.-L. (2006). Super-helices for predicting the dynamics of natural hair. *ACM Trans. Graph.*, 25(3), 1180–1187.
- [Bertails et al., 2003] Bertails, F., Kim, T.-Y., Cani, M.-P., & Neumann, U. (2003). Adaptive wisp tree - a multiresolution control structure for simulating dynamic clustering in hair motion. In *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (pp. 207–213, 377).
- [Blanz & Vetter, 1999] Blanz, V. & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *SIGGRAPH 1999, Computer Graphics Proceedings* (pp. 187–194).
- [Bogaty, 1967] Bogaty, H. (1967). Torsional Properties of Hair in Relation to Permanent Waving and Setting. *J. Soc. Cosmet. Chem.*, 18, 575–589.

- [Bouaricha & Schnabel, 1994] Bouaricha, A. & Schnabel, R. (1994). TENSOLVE: A software package for solving systems of nonlinear equations and nonlinear least squares problems using tensor methods. Technical Report CRPC-TR94586, Rice University, Houston, Texas.
- [Breen et al., 2001] Breen, D. E., Mauch, S., Whitaker, R. T., & Mao, J. (2001). 3D Metamorphosis between different types of geometric models. *Eurographics 2001 Proceedings*, 20(3), 36–48.
- [Breuer, 1972] Breuer, M. (1972). The binding of small molecules to hair - I: The hydration of hair and the effect of water on the mechanical properties of hair. *J. Soc. Cosmet. Chem.*, 23, 447–470.
- [Bridson et al., 2002] Bridson, R., Fedkiw, R., & Anderson, J. (2002). Robust treatment of collisions, contact and friction for cloth animation. In *SIGGRAPH 2002, ACM TOG 21* (pp. 594–603).
- [Brown et al., 2003] Brown, J., Latombe, J.-C., & Montgomery, K. (2003). Real-time knot tying simulation. *The Visual Computer*, 20(2), 165–179.
- [Bustard & Smith, 1990] Bustard, H. & Smith, R. (1990). Studies of factors affecting light scattering by individual human hair fibres. *Int. J. Cosmet. Sci.*, 12, 121–133.
- [Cao et al., 2006] Cao, D. Q., Liu, D., & Wang, C. H.-T. (2006). Three-dimensional nonlinear dynamics of slender structures: Cosserat rod element approach. *International Journal of Solids and Structures*, 43(3–4), 760–783.
- [Chang et al., 2002] Chang, J. T., Jin, J., & Yu, Y. (2002). A practical model for hair mutual interactions. In *ACM SIGGRAPH Symposium on Computer Animation, San Antonio* (pp. 73–80).
- [Chapman, 1973] Chapman, B. (1973). Bending Stress Relaxation and Recovery of Wool, Nylon 66, and Terylene Fibers. *J. App. Polymer Sci.*, 17, 1693–1713.
- [Chapman, 1975] Chapman, B. (1975). The rheological behaviour of keratin during the aging process. *Rheol. Acta*, 14, 466–470.
- [Choe et al., 2005] Choe, B., Choi, M. G., & Ko, H.-S. (2005). Simulating complex hair with robust collision handling. In *Proceedings of the 2005 ACM SIGGRAPH* (pp. 153–160).
- [Choe & Ko, 2005] Choe, B. & Ko, H.-S. (2005). A statistical wisp model and pseudophysical approaches for interactive hairstyle generation. *IEEE Transactions on Visualization and Computer Graphics*, 11(2), 160–170.
- [Chung & Hulbert, 1993] Chung, J. & Hulbert, G. (1993). A time integration algorithm for structural dynamics with improved numerical dissipation - the generalized-alpha method. *Journal of Applied Mechanics*, 60(6), 371–375.
- [Clebsch, 1862] Clebsch, A. (1862). *Theorie der Elasticität fester Körper*. Teubner.
- [Cohen et al., 1995] Cohen, J. D., Lin, M. C., Manocha, D., & Ponamgi, M. (1995). I-COLLIDE: an interactive and exact collision detection system for large-scale environments. In *SI3D '95: Proceedings of the 1995 symposium on Interactive 3D graphics* (pp. 189–198).
- [Coming & Staadt, 2006] Coming, D. S. & Staadt, O. G. (2006). Kinetic sweep and prune for multi-body continuous motion. *Computers & Graphics*, 30(3), 439–449.
- [Conway et al., 1996] Conway, J., Hardin, R., & Sloane, N. (1996). Packing lines, planes, etc.: Packings in Grassmannian space. *Experimental Mathematics*, 5, 139–159.
- [Corcuff et al., 1993] Corcuff, P., Gremillet, P., Jourlin, M., Duvault, Y., Leroy, F., & Leveque, J.-L. (1993). 3d Reconstruction of human hair by confocal microscopy. *J. Soc. Cosmet. Chem.*, 44, 1–12.
- [Cosserat & Cosserat, 1907a] Cosserat, E. & Cosserat, F. (1907a). Sur la mécanique générale. *Comptes Rendus de l'Academie des Sciences, Paris*, 145, 1139–1142.
- [Cosserat & Cosserat, 1907b] Cosserat, E. & Cosserat, F. (1907b). Sur la statique de la ligne déformable. *Comptes Rendus de l'Academie des Sciences, Paris*, 145, 1409–1412.
- [Cosserat & Cosserat, 1908] Cosserat, E. & Cosserat, F. (1908). Sur la théorie des corps minces. *Comptes Rendus de l'Academie des Sciences, Paris*, 146, 169–172.
- [Daldegan et al., 1993] Daldegan, A., Thalmann, N. M., Kurihara, T., & Thalmann, D. (1993). An integrated system for modeling, animating and rendering hair. *Computer Graphics Forum*, 12(3), 211–221.

- [Darden et al., 1993] Darden, T. A., York, D. M., & Pedersen, L. G. (1993). The effect of long-range electrostatic interactions in simulations of macromolecular crystals: A comparison of the Ewald and truncated list methods. *The Journal of Chemical Physics*, 99(10), 8345–8348.
- [Deem & Rieger, 1968] Deem, D. E. & Rieger, M. M. (1968). Mechanical Hysteresis of Chemically Modified Hair. *J. Soc. Cosmet. Chem.*, 19, 395–410.
- [Dichmann, 1994] Dichmann, D. J. (1994). *Hamiltonian Dynamics of a Spatial Elastica and the Stability of Solitary Waves*. PhD-thesis in mathematics, University of Maryland College Park.
- [Dichmann & Maddocks, 1996] Dichmann, D. J. & Maddocks, J. H. (1996). An impetus-striction simulation of the dynamics of an elastica. *Journal of Nonlinear Science*, 6(3), 271–292.
- [Dimitrov et al., 2007] Dimitrov, D., Knauer, C., Kriegel, K., & Rote, G. (2007). New upper bounds on the quality of the pca bounding boxes in \mathbb{R}^2 and \mathbb{R}^3 . In *Proceedings of the 23rd ACM Symposium on Computational Geometry, Gyeongju, South Korea* (pp. 275–283).
- [Erdmann, 1994] Erdmann, M. (1994). On a representation of friction in configuration space. *The International Journal of Robotics Research*, 13(3), 240–271.
- [Ericson, 2005] Ericson, C. (2005). *Realtime Collision Detection*. Morgan Kaufmann.
- [Euler, 1771a] Euler, L. (1771a). De oscillationibus annulorum elasticorum. *Opera Omnia II*, 11, 378–382.
- [Euler, 1771b] Euler, L. (1771b). Genuina principia doctrinae de statu aequilibrum et motu corporum tam perfecte flexibilium quam elasticorum. *Opera Omnia II*, 11, 37–61.
- [Euler, 1774] Euler, L. (1774). De genuina methodo tam aequilibrium quam motum corporum flexibilium determinandi et utriusque egregio consensu. *Opera Omnia II*, 11, 180–193.
- [Feugelmann & Curiskis, 1983] Feugelmann, M. & Curiskis, J. (1983). Finite Element Analysis of the Composite Fiber, Alpha Keratin. *Text. Res. J.*, 53, 271–274.
- [Feughelman, 1959] Feughelman, M. (1959). A Two-Phase Structure for Keratin Fibres. *Text. Res. J.*, 29, 223–228.
- [Feughelman, 1973] Feughelman, M. (1973). Mechanical Hysteresis in Wool Keratin Fibers. *J. Macromol. Sci. Phys.*, B7(3), 569–582.
- [Feughelman, 1982] Feughelman, M. (1982). The physical properties of alpha-keratin fibers. *J. Soc. Cosmet. Chem.*, 33, 385–406.
- [Feughelman, 1997] Feughelman, M. (1997). *Mechanical Properties and Structure of Alpha-Keratin-Fibres.: Wool, Human Hair and Related Fibres*. Sydney: University of New South Wales Press.
- [Fisher & Lin, 2001] Fisher, S. & Lin, M. C. (2001). Deformed distance fields for simulation of non-penetrating flexible bodies. In *Proceedings of the Eurographic workshop on Computer animation and simulation* (pp. 99–111).
- [Floater, 1997] Floater, M. S. (1997). Parametrization and smooth approximation of surface triangulations. *Computer Aided Geometric Design*, 14(4), 231–250.
- [Freeman & Shapira, 1975] Freeman, H. & Shapira, R. (1975). Determining the minimum-area encasing rectangle for an arbitrary closed curve. *Communications of the ACM*, 18(7), 409–413.
- [Fung, 1993] Fung, Y. (1993). *Biomechanics. Mechanical Properties of Living Tissues*. New York: Springer-Verlag, 2. edition.
- [Gamez-Garcia, 1993] Gamez-Garcia, M. (1993). Effect of some oils, emulsions, and other aqueous systems on the mechanical properties of hair at small deformations. *J. Soc. Cosmet. Chem.*, 44, 69–87.
- [Gärtner & Schönherr, 1997] Gärtner, B. & Schönherr, S. (1997). Smallest enclosing ellipses - fast and exact. Report B 97-03.
- [Gavrea et al., 2008] Gavrea, B., Anitescu, M., & Potra, F. (2008). Convergence of a class of semi-implicit time-stepping schemes for nonsmooth rigid multibody dynamics. *SIAM Journal on Optimization*, 19(2), 969–1001.
- [Gobat & Grosenbaugh, 2001] Gobat, J. & Grosenbaugh, M. (2001). Application of the generalized-alpha method to the time integration of the cable dynamics equations. *Computational Methods in Applied Mechanics and Engineering*, 190(37), 4817–4829.

- [Gobat et al., 2002] Gobat, J., Grosenbaugh, M., & Triantofyllou, M. (2002). Generalized-alpha time integration solutions for hanging chain dynamics. *Journal of Engineering Mechanics*, 128(6), 677–687.
- [Gobat, 2000] Gobat, J. I. (2000). *The Dynamics of Geometrically Compliant Mooring Systems*. PhD-thesis in oceanographic engineering, Massachusetts Institute of Technology / Woods Hole Oceanographic Institution.
- [Goldsmith & Baden, 1970] Goldsmith, L. A. & Baden, H. P. (1970). The Mechanical Properties of Hair. I. The Dynamic Sonic Modulus. *J. Invest. Derm.*, 55(4), 256–259.
- [Goldsmith & Baden, 1971] Goldsmith, L. A. & Baden, H. P. (1971). The Mechanical Properties of Hair. II. Chemical Modifications and Pathological Hairs. *J. Invest. Derm.*, 56(3), 200–204.
- [Gottschalk et al., 1996] Gottschalk, S., Lin, M. C., & Manocha, D. (1996). OBBTree: A hierarchical structure for rapid interference detection. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (pp. 171–180).
- [Goyal, 2006] Goyal, S. (2006). *Dynamic Rod Model to Simulate Mechanics of Cables and DNA*. PhD-thesis in mechanical engineering and scientific computing, University of Michigan.
- [Goyal et al., 2003] Goyal, S., Perkins, N. C., & Lee, C. L. (2003). Torsional buckling and writhing dynamics of elastic cables and DNA. In *Proceedings of ASME 2003 Design Engineering and Technical Conference, Chicago, USA, September 2nd, 6th*.
- [Goyal et al., 2005] Goyal, S., Perkins, N. C., & Lee, C. L. (2005). Nonlinear dynamics and loop formation in Kirchhoff rods with implications to the mechanics of DNA and cables. *Journal of Computational Physics*, 209(1), 371–389.
- [Goyal et al., 1991] Goyal, S., Ruina, A., & Papadopoulos, J. M. (1991). Planar sliding with dry friction. Part 1: Limit surface and moment function. *WEAR*, 143(2), 307–330.
- [Greengard & Rokhlin, 1987] Greengard, L. F. & Rokhlin, V. (1987). A fast algorithm for particle simulation. *Journal of Computational Physics*, 73(325), 325–348.
- [Grosan et al., 2006] Grosan, C., Abraham, A., & Gelbukh, A. (2006). Evolutionary method for nonlinear systems of equations. In *MICAI 2006: Advances in Artificial Intelligence*, volume 4293 of *Lecture Notes in Computer Science* (pp. 283–293).: Springer, Berlin / Heidelberg.
- [Guibas et al., 2002] Guibas, L., Nguyen, A., Russel, D., & Zhang, L. (2002). Collision detection for deforming necklaces. In *Proceedings of the 18th annual symposium on Computational geometry* (pp. 33–42).
- [Guibas et al., 2003] Guibas, L. J., Nguyen, A., & Zhang, L. (2003). Zonotopes as bounding volumes. In *Proceedings of the 14th annual ACM-SIAM symposium on discrete algorithms* (pp. 803–812).
- [Guioulet et al., 1987] Guioulet, A., Garson, J., & Leveque, J. (1987). Study of the optical properties of human hair. *Int. J. Cosmet. Sci.*, 9, 111–124.
- [Hadap, 2006] Hadap, S. (2006). Oriented strands - dynamics of stiff multi-body system. In *ACM SIGGRAPH Symposium on Computer Animation (2006)* (pp. 91–100).
- [Hadap & Magnenat-Thalmann, 2000] Hadap, S. & Magnenat-Thalmann, N. (2000). Interactive hair styler based on fluid flow. In *Computer Animation and Simulation* (pp. 87–100).
- [Hadap & Magnenat-Thalmann, 2001] Hadap, S. & Magnenat-Thalmann, N. (2001). Modeling dynamic hair as a continuum. In *Computer Graphics Forum*, volume 20(3).
- [Hahn, 1988] Hahn, J. K. (1988). Realistic animation of rigid bodies. In *Computer Graphics*, volume 22(4) (pp. 299–308).
- [Haider & Kaneko, 2000] Haider, A. & Kaneko, T. (2000). Hair shape modeling from video captured images and CT data. In *The Proceedings of ICAT2000* (pp. 52–57).
- [Halperin & Overmars, 1994] Halperin, D. & Overmars, M. H. (1994). Spheres, molecules, and hidden surface removal. In *Symposium on Computational Geometry* (pp. 113–122).
- [Healey & Mehta, 2003] Healey, T. J. & Mehta, P. G. (2003). Straightforward computation of spatial equilibria of geometrically exact Cosserat rods. <http://tam.cornell.edu/Healey.html>.
- [Hilber et al., 1977] Hilber, H. M., Hughes, T. J. R., & Taylor, R. L. (1977). Improved numerical dissipation for time integration algorithms in structural dynamics. *Earthquake Engineering and Structural Dynamics*, 5, 283–292.

- [Hollfelder et al., 1995] Hollfelder, B., G.Blankenburger, Wolfram, L., & Höcker, H. (1995). Chemical and physical properties of pigmented and non-pigmented hair ('grey hair'). *Int. J. Cosmet. Sci.*, 17, 87–89. Letter to the Editor.
- [Huck & Baddiel, 1971] Huck, P. & Baddiel, C. (1971). The mechanical properties of virgin and treated human hair fibres; a study by means of the oscillating beam method. *J. Soc. Cosmet. Chem.*, 22, 401–410.
- [Jachowicz & Yao, 1996] Jachowicz, J. & Yao, K. (1996). Dynamic hairspray analysis. I. Instrumentation and preliminary results. *J. Soc. Cosmet. Chem.*, 47, 73–84.
- [James & Pai, 2004] James, D. L. & Pai, D. K. (2004). BD-Tree: Output-sensitive collision detection for reduced deformable models. *ACM Transactions on Graphics (SIGGRAPH 2004)*, 23(3), 393–398.
- [Jean, 1999] Jean, M. (1999). The non smooth contact dynamics method. *Computer Methods in Applied mechanics and Engineering*, 177, 235–257. Special issue on computational modeling of contact and friction.
- [Johri & Jatar, 1982] Johri, M. & Jatar, D. (1982). Young's Modulus in the Identification of Human Scalp Hair. *Med. Sci. Law*, 22(1), 63–67.
- [Jollès et al., 1997] Jollès, P., Zahn, H., & Höcker, H. (1997). *Formation and Structure of Human Hair*. Basel: Birkhäuser-Verlag.
- [Jones & Satherley, 2001] Jones, M. W. & Satherley, R. (2001). Using distance fields for object representation and rendering. In *Proc. 19th Ann. Conf. of Eurographics (UK Chapter)* (pp. 37–44). London.
- [Kajiya & Kay, 1989] Kajiya, J. T. & Kay, T. L. (1989). Rendering fur with three dimensional textures. In *ACM SIGGRAPH Computer Graphics*, volume 23(3) (pp. 271–280).
- [Kassenbeck, 1999] Kassenbeck, P. (1999). *Das Haar und Seine Struktur*. Darmstadt: Wella AG.
- [Kaufman et al., 2005] Kaufman, D. M., Edmunds, T., & Pai, D. K. (2005). Fast frictional dynamics for rigid bodies. *ACM Trans. Graph.*, 23(3), 946–956.
- [Kehrbaum & Maddocks, 1997] Kehrbaum, S. & Maddocks, J. H. (1997). Elastic rods, rigid bodies, quaternions and the least quadrature. *Phil. Trans. R. Soc. Lond. A*, 355(1732), 2117–2136.
- [Kim & Neumann, 2002] Kim, T.-Y. & Neumann, U. (2002). Interactive multiresolution hair modeling and editing. In *ACM Transactions on Graphics (TOG)*, volume 21(3) (pp. 620–629).
- [Kirchhoff, 1859] Kirchhoff, G. R. (1859). Über das Gleichgewicht und die Bewegung eines unendlich dünnen elastischen stabes. *Journal für die reine und angewandte Mathematik*, 56, 285–313.
- [Klosowski et al., 1998] Klosowski, J., Held, M., Mitchell, J., Sowizral, H., & Zikan, K. (1998). Efficient collision detection using bounding volume hierarchies of k-DOPs. In *IEEE Transactions on Visualization and Computer Graphics*, volume 4(1).
- [Krishnan et al., 1998a] Krishnan, S., Gopi, M., Lin, M. C., Manocha, D., & Pattekar, A. (1998a). Rapid and accurate contact determination between spline models using shelltrees. In *Proceedings of Eurographics'98* (pp. 803–812).
- [Krishnan et al., 1998b] Krishnan, S., Pattekar, A., Lin, M. C., & Manocha, D. (1998b). Spherical shell: A higher order bounding volume for fast proximity queries. In *Proceedings of the Workshop on the Algorithmic Foundations of Robotics WAFR'98* (pp. 177–190).
- [Krysl & Endres, 2005] Krysl, P. & Endres, L. (2005). Explicit Newmark/Verlet algorithm for time integration of the rotational dynamics of rigid bodies. *International Journal for Numerical Methods in Engineering*, 62(15), 2154–2177.
- [Kurihara et al., 1994] Kurihara, T., Anjyo, K., & Thalmann, D. (1994). Hair animation with collision detection. In N. Magnenat Thalmann & D. Thalmann (Eds.), *Models and Techniques in Computer Animation* (pp. 128–138). Tokyo: Springer.
- [Lahanas et al., 2000] Lahanas, M., Kemmerer, T., Milickovic, N., Karouzakis, K., Baltas, D., & Zamboglou, N. (2000). Optimized bounding boxes for three-dimensional treatment planning in brachytherapy. *Medical Physics*, 27(10), 2333–2342.
- [Larsson & Akenine-Möller, 2001] Larsson, T. & Akenine-Möller, T. (2001). Collision detection for continuously deforming bodies. In *Proceedings of Eurographics, short presentations* (pp. 325–333).
- [Lee et al., 2002] Lee, C.-Y., Chen, W.-R., Leu, E., & Ouhyoung, M. (2002). A rotor platform assisted system for 3d hairstyles. In *Proceedings of the 10th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2002 (WSCG 2002)*, volume 10(1–3) (pp. 271–278).

- [Levoy, 1999] Levoy, M. (1999). The Digital Michelangelo Project. Stanford University.
- [Lokovic & Veach, 2000] Lokovic, T. & Veach, E. (2000). Deep shadow maps. In *Proceedings of the 27th annual conference on computer graphics and interactive techniques* (pp. 385–392).
- [Lotan et al., 2002] Lotan, I., Schwarzer, F., Halperin, D., & Latombe, J.-C. (2002). Efficient maintenance and self-collision testing for kinematic chains. In *Proceedings of the 18th annual symposium on Computational geometry, Barcelona, Spain* (pp. 43–52).
- [Lötstedt, 1984] Lötstedt, P. (1984). Numerical simulation of time-dependent contact and friction problems in rigid body mechanics. *SIAM Journal on Scientific and Statistical Computing*, 5(2), 370–393.
- [Love, 1893] Love, A. E. H. (1893). *A Treatise on the Mathematical Theory of Elasticity*. Dover, 1 edition.
- [Marschner et al., 2003] Marschner, S. R., Jensen, H. W., Cammarano, M., Worley, S., & Hanrahan, P. (2003). Light scattering from human hair fibers. In *ACM Transactions on Graphics (TOG)*, volume 22(3) (pp. 780–791).
- [Mason, 2001] Mason, M. T. (2001). *Mechanics of Robotic Manipulation*. MIT Press.
- [Mauch, 2003] Mauch, S. P. (2003). *Efficient Algorithms for Solving Static Hamilton-Jacobi Equations*. PhD-thesis in computer science, California Institute of Technology, Pasadena, California, USA.
- [Megiddo, 1982] Megiddo, N. (1982). Linear-time algorithms for linear programming in \mathbb{R}^3 and related problems. In *Proceeding of the 23rd IEEE Symposium on Foundations of Computer Science* (pp. 329–338).
- [Menkart et al., 1966] Menkart, J., Wolfram, L., & Mao, I. (1966). Caucasian Hair, Negro Hair, and Wool: Similarities and Differences. *J. Soc. Cosmet. Chem.*, 17, 769–787.
- [Miller, 1988] Miller, G. S. P. (1988). The motion dynamics of snakes and worms. In *Proceedings of the 15th annual conference on Computer graphics and interactive techniques* (pp. 169–173).
- [Mirtich, 1996] Mirtich, B. V. (1996). *Impulse-based Dynamic Simulation of Rigid Body Systems*. Phd-thesis in computer science, University of California at Berkeley.
- [Molino et al., 2003] Molino, N., Bridson, R., Teran, J., & Fedkiw, R. (2003). A crystalline, red green strategy for meshing highly deformable objects with tetrahedra. In *Proceedings of the 12th International Meshing Roundtable* (pp. 103–114).
- [Monteiro Marques, 1993] Monteiro Marques, M. D. (1993). *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction*. Birkhäuser, Basel, Boston, Berlin.
- [Moon & Marschner, 2006] Moon, J. T. & Marschner, S. R. (2006). Simulating multiple scattering in hair using a photon mapping approach. In *ACM SIGGRAPH conference proceedings* (pp. 1067–1074).
- [Moon et al., 2008] Moon, J. T., Walter, B., & Marschner, S. R. (2008). Efficient multiple scattering in hair using spherical harmonics. In *ACM SIGGRAPH conference proceedings*.
- [Moré et al., 1980] Moré, J. J., Garbow, B. S., & Hillstom, K. E. (1980). *User Guide for MINPACK-1*. Technical Report ANL-80-74, Argonne National Laboratory.
- [Moreau, 1966] Moreau, J. (1966). Quadratic programming in mechanics: Dynamics of one-sided constraints. *SIAM Journal on control and optimization*, 4(1), 153–158.
- [Moreau, 1988] Moreau, J. (1988). *Unilateral contact and dry friction in finite freedom dynamics*, volume 302, (pp. 1–82). Springer-Verlag, Wien, New York.
- [Moreau, 1994] Moreau, J. (1994). Some numerical methods in multibody dynamics: application to granular materials. *European Journal of mechanics / A*, 13(4), 93–114.
- [Moreau, 2003] Moreau, J. (2003). An introduction to unilateral dynamics. In *Novel Approaches in Civil Engineering. Proceedings of the December 2000 meeting of Laboratoire Lagrange* (pp. 1–37).: Springer-Verlag.
- [Morton & Hearle, 1997] Morton, W. & Hearle, J. (1997). *Physical Properties of Textile Fibres*. Bath, UK: The Textile Institute, third edition.
- [Murray et al., 1994] Murray, R. M., Li, Z., & Sastry, S. S. (1994). *A Mathematical Introduction to Robotic Manipulation*. CRC Press.
- [Nagase et al., 2000] Nagase, S., Ohshika, M., Ueda, S., Satoh, N., & Tsujii, K. (2000). A Universal Structural Model for Human Hair to Understand the Physical Properties 1. *Bull. Chem. Soc. Jpn.*, 573, 2161–2167.

- [Naruse & Fujita, 1971] Naruse, N. & Fujita, T. (1971). Changes in the Physical Properties of Human Hair with Age. *Journal of the American Geriatrics Society*, 19(4), 308–314.
- [Nelson & Cox, 2000] Nelson, D. L. & Cox, M. M. (2000). *Lehninger Principles of Biochemistry*. Worth Publishing, 3. edition.
- [Nikiforidis et al., 1992] Nikiforidis, G., Balas, C., & Tsambaos, D. (1992). Mechanical parameters of human hair: possible application in the diagnosis and follow-up of hair disorders. *Clin. Phys. Physiol. Meas.*, 13(3), 281–290.
- [Nikiforidis et al., 1993] Nikiforidis, G., Tsambaos, D., Balas, C., & Bezerianos, A. (1993). A Method for the Determination of Viscoelastic Parameters of Human Hair in Relation to Its Structure. *Skin Pharmacol*, 6(1), 32–37.
- [Nizette & Goriely, 1999] Nizette, M. & Goriely, A. (1999). Towards a classification of Euler-Kirchhoff filaments. *Journal of Mathematical Physics*, 40, 2830–2866.
- [O'Connor et al., 1995] O'Connor, S. D., Komisarek, K. L., & Baldeschwieler, J. D. (1995). Atomic Force Microscopy of Human Hair Cuticles: A Microscopic Study of Environmental Effects on Hair Morphology. *J. Invest. Derm.*, 105(1), 96–99.
- [Ogle & Fox, 1999] Ogle, R. R. & Fox, M. J. (1999). *Atlas of Human Hair. Microscopic Characteristics*. Boca Raton, London, New York, Washington D.C.: CRC Press.
- [Orfanos & Happle, 1990] Orfanos, C. & Happle, R. (1990). *Hair and Hair Diseases*. Berlin, Heidelberg, New York: Springer-Verlag.
- [Orfanos & L.Brückmann, 1982] Orfanos, C. & L.Brückmann (1982). Messungen der Haardichte an normaler menschlicher Kopfhaut mit und ohne kosmetische Vorbehandlung. *Akt. Dermatol.*, 8, 222–224.
- [Orfanos & Ruska, 1968a] Orfanos, C. & Ruska, H. (1968a). Die Feinstruktur des menschlichen Haares. I. Die Haar-Cuticula. *Arch. klin. exp. Derm.*, 231, 97–110.
- [Orfanos & Ruska, 1968b] Orfanos, C. & Ruska, H. (1968b). Die Feinstruktur des menschlichen Haares. II. Der Haar-Cortex. *Arch. klin. exp. Derm.*, 231, 264–278.
- [O'Rourke, 1985] O'Rourke, J. (1985). Finding minimal enclosing boxes. *International Journal of Parallel Programming*, 14(3), 183–199.
- [Pai, 2002] Pai, D. K. (2002). STRANDS: Interactive simulation of thin solids using Cosserat models. In *Computer Graphics Forum*, volume 21(3) (pp. 347–352).
- [Palmer & Grimsdale, 1995] Palmer, I. & Grimsdale, R. (1995). Collision detection for animation using sphere-trees. In *Computer Graphics Forum*, volume 14(2) (pp. 105–116).
- [Pang & Trinkle, 1996] Pang, J.-S. & Trinkle, J. C. (1996). Complementarity formulations and existence of solutions of dynamic multi-rigid-body contact problems with coulomb friction. *Mathematical Programming: Series A and B*, 73(2), 199–226.
- [Paris et al., 2004] Paris, S., no, H. B., & Sillion, F. (2004). Capture of hair geometry from multiple images. *ACM Transactions on Graphics (Proceedings of the SIGGRAPH conference)*.
- [Patel, 1988] Patel, C. (1988). Hair conditioning composition and method. United States Patent, Patent Number US04719104. <http://210.69.13.146/search/PNUS04719104>.
- [Pecoraro & Astore, 1990] Pecoraro, V. & Astore, I. (1990). Measurements of Hair Growth Under Physiological Conditions. In C. Orfanos & R. Happle (Eds.), *Hair and Hair Diseases* (pp. 237–254). Berlin, Heidelberg, New York: Springer-Verlag.
- [Pfeiffer & Glocker, 1996] Pfeiffer, F. & Glocker, C. (1996). *Multibody Dynamics with Unilateral Contacts*. Wiley Series in Nonlinear Science. John Wiley and Sons.
- [Phillips, 1987] Phillips, D. (1987). Effects of Humidity, Ageing, Annealing, and Tensile Loads on the Torsional Damping of Wool Fibers. *Text. Res. J.*, 57, 42–49.
- [Plante et al., 2001] Plante, E., Cani, M.-P., & Poulin, P. (2001). A layered wisp model for simulating interactions inside long hair. In *Computer Animation and Simulation*.
- [Plante et al., 2002] Plante, E., Cani, M.-P., & Poulin, P. (2002). Capturing the complexity of hair motion. *Graphical Models*, 64(1), 40–58.

- [Press et al., 2002] Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2002). *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press; 2nd Edition, February.
- [Quinlan, 1994] Quinlan, S. (1994). Efficient distance computation between non-convex objects. In *IEEE International Conference on Robotics and Automation* (pp. 3324–3329).
- [Rankin, 2002] Rankin, W. T. (2002). DPMTA - A distributed implementation of the parallel multipole tree algorithm - Version 3.1. Department of Electrical Engineering, Duke University. <http://www.ee.duke.edu/wrankin/Dpmta/>.
- [Redon et al., 2004] Redon, S., Kim, Y. J., Lin, M. C., Manocha, D., & Templeman, J. (2004). Interactive and continuous collision detection for avatars in virtual environments. In *Computer Graphics* (pp. 117–124).
- [Reich & Robbins, 1993] Reich, C. & Robbins, C. R. (1993). Light scattering and shine measurement of human hair: A sensitive probe of the hair surface. *J. Soc. Cosmet. Chem.*, 44, 221–234.
- [Renouf et al., 2005] Renouf, M., Acary, V., & Dumont, G. (2005). 3D frictional contact and impact multibody dynamics. A comparison of algorithms suitable for real-time applications. In *ECCOMAS Thematic Conference Multibody Dynamics, Madrid, Spain, 2005*.
- [Rigby et al., 1974] Rigby, H., Mitchell, T., & Robinson, M. (1974). Changes with Time in Mechanical Properties of Wool Fibers. *J. Macromol. Sci. Phys.*, B10(2), 255–269.
- [Robbins, 2002] Robbins, C. (2002). *Chemical and Physical Behavior of Human Hair*. New York: Springer-Verlag, 4. edition.
- [Robbins & Reich, 1986] Robbins, C. & Reich, C. (1986). Prediction of hair assembly characteristics from single fiber properties. Part II. The relationship of fiber curvature, friction, stiffness, and diameter to combing behavior. *J. Soc. Cosmet. Chem.*, 37, 141–158.
- [Rosenblum et al., 1991] Rosenblum, R. E., Carlson, W. E., & III, E. T. (1991). Simulating the structure and dynamics of human hair: Modelling, rendering and animation. *The Journal of Visualization and Computer Animation*, 2(4), 141–148.
- [Ruetsch & Weigmann, 1996] Ruetsch, S. B. & Weigmann, H.-D. (1996). Mechanism of tensile stress release in the keratin fiber cuticle: I. *J. Soc. Cosmet. Chem.*, 47, 13–26.
- [Rusin, 2001] Rusin, D. (2001). Frequently-asked questions about spheres. <http://www.math.niu.edu/~rusin/known-math/index/spheres.html>.
- [Sakai et al., 2000] Sakai, M., Nagase, S., Okada, T., Satoh, N., & Tsujii, K. (2000). A Universal Structural Model for Human Hair to Understand the Physical Properties 2. Mechanical and Permeation Behaviors. *Bull. Chem. Soc. Jpn.*, 73, 2169–2177.
- [Schatzman, 1973] Schatzman, M. (1973). Sur une classe de problèmes hyperboliques non linéaires. *Comptes Rendus de l'Academie des Sciences, Paris, Sér. A-B*, 277, A671–A674.
- [Schmidl, 2002] Schmidl, H. J. (2002). *Optimization-based animation*. PhD-thesis in computer science, University of Miami, Coral Gables, FL, USA.
- [Schuricht, 1998] Schuricht, F. (1998). Regularity for shearable nonlinearly elastic rods on obstacle problems. *Archive for Rational Mechanics Analysis*, 145, 23–49.
- [Schwartz & Knowles, 1963] Schwartz, A. M. & Knowles, D. (1963). Frictional Effects in Human Hair. *J. Soc. Cosmet. Chem.*, 14, 455–463.
- [Scott & Robbins, 1978] Scott, G. & Robbins, C. (1978). Stiffness of human hair fibers. *J. Soc. Cosmet. Chem.*, 29, 469–485.
- [Sethian, 1999] Sethian, J. (1999). *Level Set Methods and Fast Marching Methods*. Cambridge, UK: Cambridge University Press.
- [Simo & Vu-Quoc, 1988] Simo, J. & Vu-Quoc, L. (1988). On the dynamics in space of rods undergoing large motions - a geometrically exact approach. *Computer Methods in Applied Mechanics and Engineering*, 66, 125–161.
- [Sloane, 2005] Sloane, N. J. A. (2005). Home page. <http://www.research.att.com/~njas/index.html>.
- [Spillmann & Teschner, 2007] Spillmann, J. & Teschner, M. (2007). CoRdE: Cosserat rod elements for the dynamic simulation of one-dimensional elastic objects. In *Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on Computer animation* (pp. 63–72).

- [Spillmann & Teschner, 2008] Spillmann, J. & Teschner, M. (2008). An adaptive contact model for robust simulation of knots. In *Computer Graphics Forum*, volume 27(2) (pp. 497–506).
- [Stamm et al., 1977a] Stamm, R. F., Garcia, M. L., & Fuchs, J. J. (1977a). The optical properties of human hair I. Fundamental considerations and goniophotometer curves. *J. Soc. Cosmet. Chem.*, 28, 571–599.
- [Stamm et al., 1977b] Stamm, R. F., Garcia, M. L., & Fuchs, J. J. (1977b). The optical properties of human hair II. The luster of hair fibers. *J. Soc. Cosmet. Chem.*, 28, 601–609.
- [Stewart, 2000] Stewart, D. E. (2000). Rigid-body dynamics with friction and impact. *SIAM Review*, 42(1), 3–39.
- [Stronge, 1991] Stronge, W. J. (1991). Unraveling paradoxical theories for rigid body collisions. *Journal of Applied Mechanics*, 58(4), 1049–1055.
- [Swanbeck & Nyren, 1970] Swanbeck, G. & Nyren, J. (1970). The Mechanical Properties of Hairs from Patients with Different Types of Hair Diseases. *J. Invest. Derm.*, 54(3), 248–251.
- [Swift, 1995] Swift, J. (1995). Some simple theoretical considerations on the bending stiffness of human hair. *International Journal of Cosmetic Science*, 17, 245–253.
- [Swift, 1997] Swift, J. A. (1997). Mechanism of split-end formation of human hair. *J. Soc. Cosmet. Chem.*, 48, 123–126.
- [Tao & Postle, 1989a] Tao, X. & Postle, R. (1989a). A Viscoelastic Analysis of the Keratin Composite. Part I. Longitudinal and Transverse Mechanical Properties. *Text. Res. J.*, 59, 123–138.
- [Tao & Postle, 1989b] Tao, X. & Postle, R. (1989b). A Viscoelastic Analysis of the Keratin Composite. Part II: Thermal und Hygral Expansion. *Text. Res. J.*, 59, 300–306.
- [Taosong, 1999] Taosong, H. (1999). Fast collision detection using QuOSPO trees. In *Proceedings of the Symposium on Interactive 3D Graphics* (pp. 55–62).
- [Theetten et al., 2007] Theetten, A., Grisoni, L., Andriot, C., & Barsky, B. (2007). Geometrically exact dynamic splines. *Computer-Aided Design*, 40(1), 35–48.
- [Timoshenko, 1921] Timoshenko, S. P. (1921). On the correction for shear of the differential equations for transverse vibrations of prismatic bars. *Phil. Mag.*, 41(6), 744–746.
- [van den Bergen, 1997] van den Bergen, G. (1997). Efficient collision detection of complex deformable models using AABB trees. *Journal of Graphics Tools*, 2(4), 1–14.
- [Vernall, 1961] Vernall, D. (1961). A study of the size and shape of cross-sections of hair from four races of men. *Am. J. Phys. Anthropol.*, 19, 345–350.
- [Wakamatsu & Hirai, 2004] Wakamatsu, H. & Hirai, S. (2004). Static modeling of linear object deformation based on differential geometry. *The International Journal of Robotics Research*, 23(3), 293–311.
- [Ward et al., 2002] Ward, K., Fisher, S., & Lin, M. C. (2002). *Simplified Representations for Modeling Hair*. Technical Report TR02-020, University of North Carolina at Chapel Hill.
- [Ward et al., 2004] Ward, K., Galoppo, N., & Lin, M. C. (2004). Modeling hair influenced by water and styling products. In *Proceedings of the Computer Animation and Social Agents 2004*.
- [Ward et al., 2003] Ward, K., Lin, M. C., Lee, J., Fisher, S., & Macri, D. (2003). Modeling hair using level-of-detail representations. In *Proc. of Computer Animation and Social Agents*.
- [Wei et al., 2005] Wei, Y., Ofek, E., Quan, L., & Shum, H.-Y. (2005). Modeling hair from multiple views. In *ACM Transactions on Graphics (TOG), 2004*, volume 24(3) (pp. 816–820).
- [Weinstein et al., 2006] Weinstein, R., Teran, J., & Fedkiw, R. (2006). Dynamic simulation of articulated rigid bodies with contact and collision. *IEEE Transactions on Visualization and Computer Graphics*, 12(3), 365–374.
- [Welzl, 1991] Welzl, E. (1991). Smallest enclosing disks (balls and ellipsoids). In H. Maurer (Ed.), *New Results and New Trends in Computer Science*, number 555 in LNCS (pp. 359–370). Springer.
- [Wickett et al., 1992] Wickett, R. R., Sramek, J. A., & Throbaugh, C. M. (1992). Measurement of the adhesive strength of hair-hairspray junctions. *J. Soc. Cosmet. Chem.*, 43, 169–178.
- [Wolfram & Albrecht, 1985] Wolfram, L. J. & Albrecht, L. (1985). Torsional behavior of human hair. *J. Soc. Cosmet. Chem.*, 36, 87–99.

- [Wong et al., 1994] Wong, M., Wis-Surel, G., & Epps, J. (1994). Mechanism of hair straightening. *J. Soc. Cosmet. Chem.*, 45, 347–352.
- [Wood et al., 1981] Wood, W. L., Bossak, M., & Zienkiewicz, O. C. (1981). An alpha modification of Newmarks method. *International Journal for Numerical Methods in Engineering*, 15, 1562–1566.
- [Xu & Yang, 2001] Xu, Z. & Yang, X. D. (2001). V-HairStudio: An interactive tool for hair design. *IEEE Computer Graphics and Application*, 21(3)(3), 36–43.
- [Yin et al., 1976] Yin, N. E., Kissinger, R. H., Tolgyesi, W. S., & Cottington, E. M. (1976). The Effect of fiber diameter on the cosmetic aspects of hair. *J. Soc. Cosmet. Chem.*, 28, 139–150.
- [Yoshizawa et al., 2004] Yoshizawa, S., Belyaev, A. G., & Seidel, H.-P. (2004). A fast and simple stretch-minimizing mesh parameterization. In *International Conference on Shape Modeling and Applications. June 7-11, 2004, Genova, Italy* (pp. 200–208).
- [Yu, 2001] Yu, Y. (2001). Modeling realistic virtual hairstyles. In *Proceedings of Pacific Graphics* (pp. 295–301).
- [Zahn, 1998] Zahn, H. (1998). Das Menschenhaar und seine innere Struktur. *PZ Dermo Pharmazie*, 11, 10–14.
- [Zinke, 2008] Zinke, A. (2008). *Photo-Realistic Rendering of Fiber Assemblies*. Phd-thesis in computer science, University of Bonn.
- [Zinke & Weber, 2006] Zinke, A. & Weber, A. (2006). Global illumination for fiber based geometries. In *Electronic proceedings of the Ibero American Symposium on Computer Graphics (SIACG 2006)*.
- [Zinke & Weber, 2007] Zinke, A. & Weber, A. (2007). Light scattering from filaments. *IEEE Transactions on Visualization and Computer Graphics*, 13(2), 342–356.
- [Zinke et al., 2008] Zinke, A., Yuksel, C., Weber, A., & Keyser, J. (2008). Dual scattering approximation for fast multiple scattering in hair. In *ACM SIGGRAPH conference proceedings*.

