

Analyse komplexer Molekülspektren

Diplomarbeit am I. Physikalischen Institut der Universität zu Köln*

Stefan Brackertz

26. August 2016

bei Stephan Schlemmer

*Dies ist nicht die offiziell eingereichte Version, sondern eine leicht überarbeitete Fassung, in der ein Fehler in den Ergebnistabellen zu CD_2H^+ korrigiert ist und wenige Anmerkungen ergänzt wurden.

Inhaltsverzeichnis

1	Einführung	4
1.1	CH ₅ ⁺ : Prototyp eines floppy Moleküls	4
1.2	Was bisher geschah	7
1.2.1	Drei Typen von Kombinationsdifferenzen und ihre Häufigkeit	8
1.2.1.1	Accidental CDs	9
1.2.1.2	Ground state CDs	10
1.2.1.3	Excited state CDs	10
1.2.2	Rekonstruktion der Differenzen zwischen Vibrationsgrundzuständen auf Grund der Häufigkeiten von CDs	11
1.2.3	Rekonstruktion der Vibrationsgrundzustände mit Hilfe temperaturgemischter CDs	12
1.3	Arbeitsprogramm dieser Diplomarbeit	12
2	Testumgebung	15
2.1	Spinstatistik	18
2.2	Was wie parametrisieren?	18
2.3	Zwischenergebnisse	22
2.3.1	Zuverlässigkeit der Auswertung hängt von der CD-Dichte ab	22
2.3.1.1	Variable Bandbreite ist kontraproduktiv	23
2.3.1.2	Auswahlregeln erschweren Auswertung	23
2.3.2	Auswertung mit konstanter Bandbreite ist Kerndichteschätzung	24
2.3.3	Zur Rekonstruktion der Vibrationsgrundzustände mittels temperaturgemischter CDs	25
2.4	Ermittlung statistischer Eigenschaften von Molekül und Messung	26
3	Optimierung des Kerndichteschätzers	27
3.1	Gauß- statt Kastenkernel	27
3.2	Wahl der Bandbreite	29
3.3	Weitere Optimierungsmöglichkeiten	31
4	Peakerkennung	33
4.1	Skalierungsverhalten der CD-Spektren	33
4.2	Rauschunterdrückung mit Hilfe des Skalierungsverhaltens	35
4.3	Weitere Optimierungsmöglichkeiten	36
5	Von Linien über CDs zu Zuständen	39
5.1	CD-Peaks haben verbundene Beinchen	39
5.2	Rekonstruktion der Vibrationsgrundzustände	41
5.2.1	Trennung der Sets	42
5.2.2	Fehlende CD-Peaks und Verbindungen	42
5.2.2.1	Auswahlregeln und Co-CDs	42
5.2.2.2	Rekonstruktion fehlender CD-Peaks und Verbindungen	43

Inhaltsverzeichnis

5.3	Vibrationsangeregte Zustände	43
5.3.1	Check der Set-Aufteilung und der Verbindungen	43
5.3.2	Check der CDs eines CD-Peaks	44
5.3.3	Rekonstruktion der vibrationsangeregten Zustände	45
6	Das Programm	47
6.1	Erstellung und Auswertung des CD-Spektrums	48
6.2	Verbesserungsmöglichkeiten	49
7	Ergebnisse	51
7.1	CD_2H^+	51
7.1.1	A1	53
7.1.2	A2	55
7.1.3	B1	57
7.1.4	B2	59
7.2	CH_5^+	61
7.2.1	Set 1	62
7.2.2	Set 2	64
7.2.3	Set 3	66
7.2.4	Set 4	68
8	Fazit und Ausblick	70
8.1	Was bedeuten die Ergebnisse für die Arbeit an CH_5^+ ?	70
8.2	Potenzial und Weiterentwicklung der modellosen Auswertungsmethode	70
A	Ergänzendes Material	73
A.1	Zuordnung der Linien von CH_5^+ zu Sets	73
A.2	Gesicherte gsCDs von CH_5^+	73
A.3	Scripte	73
	Literaturverzeichnis	74
	Danksagung	76
	Erklärung	77

1 Einführung

Bereits seit den 1920er Jahren sind Moleküle mit Hilfe der Quantenmechanik vollständig beschreibbar. Prinzipiell lässt sich immer ein Hamiltonian aufschreiben, der für jedes Nukleon und jedes Elektron einen kinetischen Term sowie für alle möglichen Elektron-Elektron-, Elektron-Nukleon- und Nukleon-Nukleon-Paare die elektromagnetischen und Spin-Wechselwirkungsterme umfasst. Diese allgemeine Beschreibung eines Moleküls ist allerdings so komplex, dass sich die resultierenden Gleichungen lediglich für Wasserstoff und Helium, nicht aber für größere Atome oder Moleküle direkt lösen lassen.

Üblicherweise nutzt man deshalb die Born-Oppenheimer-Näherung und teilt das Problem in die Beschreibung der Elektronen einerseits und die Beschreibung der Kerne andererseits auf (nachdem man die in der Regel uninteressante Bewegung des Molekül-Schwerpunktes abgespalten hat). Grundlage dieser Aufspaltung ist die Annahme, dass auf Grund der wesentlich höheren Masse der Kerne die Aufenthaltswahrscheinlichkeitsverteilungen der Kerne — im Vergleich zu denen der Elektronen — stark lokalisiert sind und sich nur mit vernachlässigbarer Geschwindigkeit bewegen. In weiteren Näherungsschritten werden dann sowohl die Elektronen- als auch die Kernbeschreibung weiter aufgeteilt. Üblicherweise teilt man die Beschreibung der Kerne in Vibration und Rotation auf. Diese Näherungsmethoden sind gut erforscht, liefern für sehr viele Moleküle gute Ergebnisse und gehören zu den Standardmethoden der Atom- und Molekülphysik. So ermöglichen Programme wie [pgopher] die halbautomatische Bestimmung der Parameter der üblichen Modell-Hamiltonians für die einzelnen Teilbeschreibungen an Hand gemessener Spektren. Dabei werden die gemessenen Linien direkt den nach dem Modell erwarteten Übergängen zugeordnet.

Allerdings funktionieren diese Näherungen bei sogenannten floppy Molekülen, die keine feste Anordnung der Kerne haben, nur schlecht oder gar nicht. Schwierigkeiten bereiten etwa Moleküle, bei denen Teilmoleküle gegeneinander rotieren können oder bei denen sich Rotation und Vibration nicht voneinander trennen lassen. Dies gilt z. B. für CH_5^+ , wie [Schmiedt et al. 2015] gezeigt haben. Für Moleküle, bei denen die Standardmethode nicht funktioniert, gibt es bislang kaum systematische Methoden, weder zur Aufstellung triftiger Modelle, noch zur Auswertung gemessener Spektren.

Die Weiterentwicklung der Mess- und Auswertungsmethoden sowie der Modellbildung für solche Moleküle hat über das Verständnis der Moleküle an sich hinaus Bedeutung, weil sie langfristig dazu beitragen könnte, die chemischen Prozesse selbst zu beobachten und besser zu verstehen, während bisher nur die Edukte und Produkte analysiert werden können. Im Rahmen dieser Diplomarbeit soll am Beispiel von CH_5^+ , einem floppy Molekül mit sehr kompliziertem Spektrum, ein Beitrag zur Weiterentwicklung der Auswertungsmethoden geleistet werden.

1.1 CH_5^+ : Prototyp eines floppy Moleküls

CH_5^+ ist in vielerlei Hinsicht eine Herausforderung: Im Gegensatz zu den meisten anderen Molekülen hat CH_5^+ 120 verschiedene, energetisch äquivalente Anordnungen der Kerne, die nur durch kleine Energiebarrieren voneinander getrennt sind. Deshalb haben die Kerne auch ohne (Vibrations-)anregung keine feste Anordnung, weshalb selbst bei niedrigen Temperaturen ge-

1 Einführung

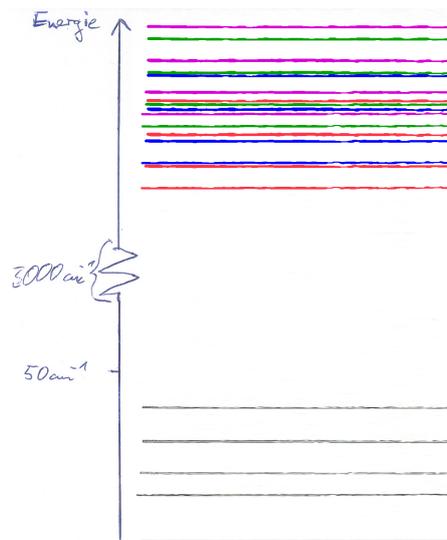


Abbildung 1.1.1: **Schematischer Aufbau eines Energie-Niveau-Schemas** für ein Set bei festem elektronischem Zustand
schwarz: Vibrationsgrundzustände. In je verschiedenen **Farben:** erste ange-regte Zustände von verschiedenen Vibrationsmoden
 Üblicherweise gilt folgende Hierarchie: $E_{Elektronenzustand} \gg E_{Vibration} \gg E_{Rotation}$. Bei CH_5^+ lassen sich Rotation und Vibration nicht wie üblich vonein-ander separieren. Dennoch haben Vibration und Rotation Äquivalente, deren energetische Größenordnungen sich analog verhalten (siehe Text).

messene Spektren sehr viele Linien haben und kaum verstanden sind (siehe [Huang et al. 2006], [Kumar et al. 2006], [Jin et al. 2006]).

Zudem sind die gemessenen Spektren eine Überlagerung von mehreren Teilspektren: Die ver-schiedenen möglichen Kernspin-Konfigurationen führen auf Grund des Pauli-Prinzips dazu, dass das Molekül in drei verschiedenen Species (so wie z. B. Helium in den Species ortho- und para-Helium vorkommt) mit je eigenen Zuständen vorkommt, zwischen denen auf Grund von Auswahl-regeln keine Übergänge möglich sind. Die Zustände dieser Species teilen sich wiederum in je zwei Sets auf, zwischen denen wegen der Paritätserhaltung keine Übergänge stattfinden. Insgesamt ist das CH_5^+ -Spektrum also eine Überlagerung von 6 Einzelspektren. Wie [Asvany et al. 2015] erläutern, beträgt die relative Häufigkeit der Species 1 : 4 : 5 (für die Symmetrien $A_2 : G_2 : H_2$). Dabei gehen [Schmiedt et al. 2016] davon aus, dass die Sets der A_2 -Species quasi unsichtbar seien, weil sie sehr schwache oder sogar verschwindende Übergangsdipolmomente hätten.

Wie [Schmiedt et al. 2016] zeigen, lässt sich die Bewegung der Kerne nicht wie üblich in Rota-tion und Vibration aufteilen. Zwar lässt sich eine Vibration mit 10 verschiedenen Moden von der Gesamtbewegung abspalten, die verbleibende Bewegung ist aber keine einfache Rotation. Dies hat weitreichende Konsequenzen für die Symmetrieeigenschaften und damit die Modellierung von CH_5^+ (und ähnlichen Molekülen), spielt aber im Rahmen dieser Arbeit keine weitere Rolle, weil die entwickelten Auswertungsverfahren keine Modellannahmen über die Moleküle machen.

Spektroskopie von CH_5^+ Bei herkömmlicher Ionenspektroskopie, bei der die Ionen in der Mess-kammer erzeugt werden, werden immer verschiedene Gase und Ionen, die bei der Herstellung des

1 Einführung

eigentlich zu vermessenden Ions entstehen, mit gemessen. Bekannte Linien dieser Fremdgase und -ionen aus dem Spektrum heraus zu rechnen, ist gerade bei dichten Spektren eine große Schwierigkeit. Zusätzlich standen z. B. [White et al. 1999] vor dem Problem, dass sich bei Messungen an CH_5^+ auch CH_4^+ , dessen Spektrum weitgehend unbekannt ist, in der Messkammer befand.

Diese Probleme gibt es bei der Ionenfallenspektroskopie, die Asvany et al. verwendet haben, nicht. Da die Anzahl der Ionen in Ionenfallen recht klein ist, kann die Absorption dabei allerdings nur schlecht direkt beobachtet werden: Der Anteil absorbiertener Photonen ist so gering, dass die Detektoren entweder nicht empfindlich genug sind oder übersteuern. Deshalb muss eine indirekte Detektion verwendet werden, bei der die große Mehrheit der Photonen, die nicht absorbiert wird, keine Rolle spielt. Dazu füllt man beispielsweise bei der LIR-Methode (laser-induced reaction, vergleiche [Schlemmer et al. 2002]) zusätzlich zum zu untersuchenden Ion ein Gas in die Falle, das mit angeregten Ionen reagiert, mit Ionen im (Vibrations-) Grundzustand aber nicht. Führt man das Experiment bei so niedrigen Temperaturen durch, dass die Ionen im (Vibrations-) Grundzustand vorliegen, geschieht diese Reaktion zunächst nicht. Werden dann Ionen durch Absorption angeregt, findet die Reaktion statt, was per Massenspektroskopie des Reaktionsproduktes detektiert werden kann. Andere Methoden wie LIICG (laser-induced inhibition of complex growth, vergleiche [Asvany et al. 2014]) funktionieren insofern ähnlich, als auch hier genutzt wird, dass bestimmte Prozesse (in diesem Fall Komplexbildung) davon abhängig sind, ob das Ion im angeregten Zustand vorliegt oder nicht.

Bei CH_5^+ sind wie bei fast allen anderen Molekülen die Vibrationsenergien wesentlich größer als die Rotationsenergien, sodass mehrere relativ weit auseinander liegende Banden von Zuständen existieren, die jeweils den gleichen Vibrations- aber verschiedene Rotationsquantenzahlen haben. Bei CH_5^+ hat das Band der Vibrationsgrundzustände eine Breite von etwa 60 cm^{-1} , der Abstand dieses Bandes zum ersten vibrationsangeregten Band (bending) beträgt hingegen über 1000 cm^{-1} ([Asvany et al. 2005]). LIR und LIICG sind nicht dafür geeignet, Übergänge zwischen Zuständen desselben Bandes mit nur geringen Energien zu messen, weil sich Ionenzustände mit ähnlichen Energien nicht durch ihre chemischen Reaktionsmöglichkeiten unterscheiden lassen.¹ Damit ist es nicht möglich, die Vibrationsgrundzustände durch direkte Vermessung zu rekonstruieren, was bei einem Ion, das wie CH_5^+ noch nicht gut erforscht ist, eigentlich der naheliegende erste Schritt wäre. Die Vibrationsgrundzustände sind nämlich am übersichtlichsten und kommen in abgewandelter Form auch bei höheren Vibrationsanregungen wieder vor. [Asvany et al. 2015] messen angesichts dessen Übergänge von Vibrationsgrundzuständen zu vibrationsangeregten Zuständen (stretching-Anregung bei 3000 cm^{-1} , siehe Abbildung 1.1.1); die Übergänge zwischen den Vibrationsgrundzuständen lassen sich daraus als Kombinationsdifferenzen rekonstruieren (siehe Abschnitt 1.2.1), was allerdings allein deshalb schon nicht trivial ist, weil CH_5^+ mehrere Vibrationsmoden mit sich überschneidenden Banden hat², sodass es sehr viele mögliche angeregte Zustände gibt und das Spektrum auch bei sehr niedrigen Temperaturen sehr dicht ist.³ Zudem

¹Zusätzlich sind diese Übergänge bei vielen Molekülen auf Grund von Auswahlregeln verboten.

²[Asvany et al. 2005] finden Vibrationsbanden bei 1250 cm^{-1} und zwischen 2200 cm^{-1} und 3400 cm^{-1} . Nach der bisherigen $SO(3)$ -Theorie hat CH_5^+ insgesamt 12 Vibrationsmoden, 5 bending und 7 stretching modes; Asvany et al. interpretieren die Vibrationsmoden bei 1250 cm^{-1} als bending und die zwischen 2200 cm^{-1} und 3400 cm^{-1} als stretching modes. Gemäß [Schmiedt et al. 2016] hat CH_5^+ allerdings $SO(5)$ -Symmetrie und damit insgesamt nur 10 Vibrationsmoden. Im Vergleich zur $SO(3)$ -Theorie ordnen sie 2 bisher zur Vibration gerechnete Freiheitsgrade nun der $SO(5)$ -Rotation zu. Es ist naheliegend, dass diese neu zugeordneten Freiheitsgrade ehemals zu bending modes gehörten, dass also die Gesamtzahl der Längen- und Winkelfreiheitsgrade in beiden Theorien jeweils gleich ist. Damit ist davon auszugehen, dass die Aussagen in [Asvany et al. 2005] über die von [Asvany et al. 2015] vermessenen und in dieser Arbeit untersuchten stretching modes zwischen 2200 cm^{-1} und 3400 cm^{-1} nach wie vor gültig sind.

³Auch Übergänge zwischen den Vibrationsgrundzuständen lassen sich — so sie denn erlaubt sind — mit Hilfe von LIR vermessen, wenn man doppelt indirekt vorgeht: Ist ein Übergang zwischen einem gering besetzten Vibrationsgrundzustand und einem angeregten Vibrationszustand bekannt, kann man einen Laser darauf einstellen.

1 Einführung

ist die Rekonstruktion der Zustände aus den gemessenen Übergängen ohne Modellannahmen über das Molekül keine gut entwickelte Standardmethode.

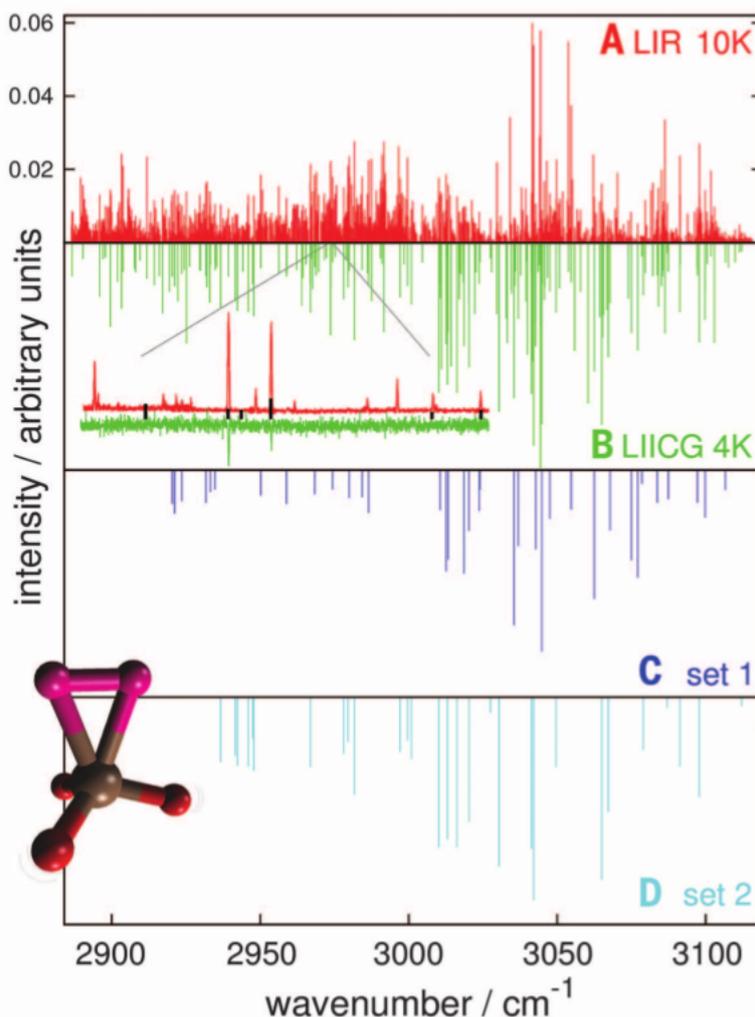
1.2 Was bisher geschah

Konkret knüpft diese Arbeit an die Ergebnisse von [Asvany et al. 2015] an, die hochpräzise Absorptionsspektren von CH_5^+ bei nominal 4 K und 10 K (tatsächlich etwa 12,8 K und 25 K, siehe supplementary material zu [Asvany et al. 2015]) zwischen 2886 cm^{-1} und 3116 cm^{-1} gemessen haben (Abbildung 1.2.1). Mit Hilfe von Kombinationsdifferenzen aus den gemessenen Linien ist es ihnen erstmals gelungen, die Vibrationsgrundzustände von zwei Sets von CH_5^+ teilweise zu rekonstruieren, wozu keine Modellannahmen über CH_5^+ nötig sind. Allerdings kann bisher der bei weitem größte Teil der gemessenen Linien nicht zugeordnet werden und es ist davon auszugehen, dass noch weiter gehende Rekonstruktionen mit Hilfe der vorliegenden Messdaten möglich sind. Deshalb sollen in dieser Arbeit die von Asvany et al. verwendeten Auswertungsverfahren weiter entwickelt werden.

Diese Verfahren arbeiten auf Ebene der Sets und können nicht zwischen verschiedenen Species unterscheiden. Sie setzen weder eine Kenntnis darüber voraus, wie viele Species es gibt, noch wie viele Sets. Allerdings muss bei der Nutzung der im Rahmen dieser Arbeit entwickelten Umgebung zum Test dieser Verfahren (Abschnitt 2) angegeben werden, wie viele Sets simuliert werden sollen und was ihre relative Häufigkeit sein soll.

Induziert man nun mit einem weiteren Laser einen Übergang in den besagten Vibrationsgrundzustand, werden die Ionen von dort aus durch den ersten Laser direkt weiter in den angeregten Vibrationszustand gebracht, wo sie die detektierbare chemische Reaktion bewirken können.

Voraussetzung dafür ist allerdings, dass es zumindest eine These darüber gibt, wo der Übergang, auf den der erste Laser eingestellt wird, liegen könnte. Dies ist bei CH_5^+ bislang nicht gegeben.

Abbildung 1.2.1: CH_5^+ -Spektren

(A) bei nominal 10 K. (B) bei nominal 4 K (C) und (D): Linien, die als zu je einem Set gehörig identifiziert werden konnten.

Quelle: [Asvany et al. 2015]

1.2.1 Drei Typen von Kombinationsdifferenzen und ihre Häufigkeit

In [Asvany et al. 2015] werden zur Rekonstruktion der Vibrationsgrundzustände Kombinationsdifferenzen (combination differences, CDs) berechnet. Dazu werden (nachdem „schlechte“ Messwerte aussortiert wurden) alle möglichen positiven Differenzen von jeweils zwei Vibrationsübergängen gebildet und gezählt, welche Werte dabei wie häufig vorkommen.

Die Linien, aus denen eine Kombinationsdifferenz berechnet ist, können Übergänge zwischen völlig verschiedenen Molekülzuständen sein oder es kann ein Molekülzustand in beide Übergänge involviert sein. Davon abhängig hat die jeweilige Kombinationsdifferenz unterschiedliche Bedeutungen. Zudem hängt davon ab, mit welcher Häufigkeit die Kombinationsdifferenz typischerweise auftaucht. Die in [Asvany et al. 2015] verwendete Methode, die in dieser Arbeit weiter entwickelt

1 Einführung

wird, macht sich diese charakteristischen Häufigkeiten zu Nutze, um Kombinationsdifferenzen zu identifizieren, die Differenzen zwischen Vibrationsgrundzuständen entsprechen.

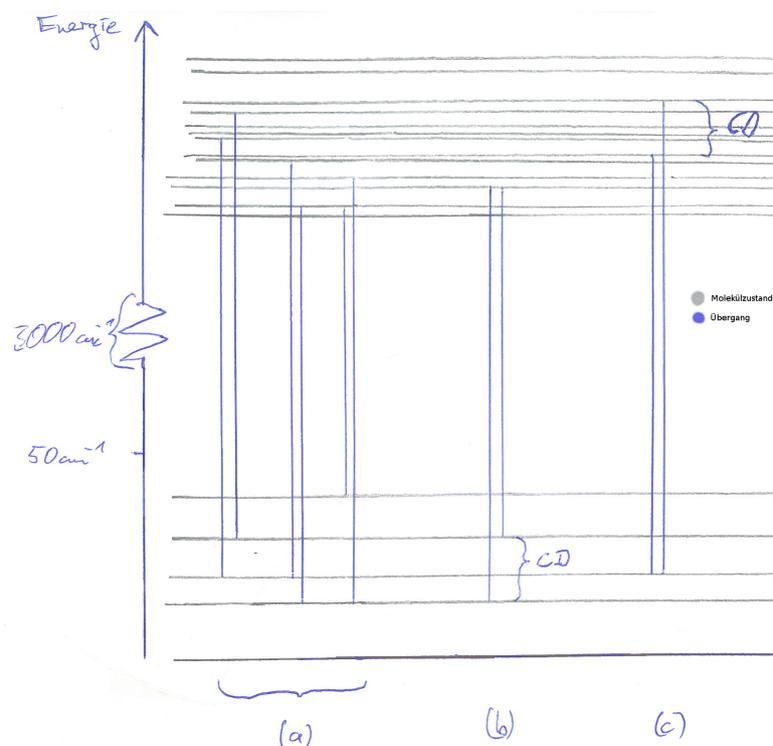


Abbildung 1.2.2: **Verschiedene Typen von CDs (in einem Set)**

(a) unverbundene, accidental CDs, (b) oben verbundene, ground state CDs, (c) unten verbundene, excited state CDs

Die oberen Zustände gehören zu mindestens drei verschiedenen Vibrationsbanden, was im weiteren Verlauf dieser Arbeit aber keine Rolle spielt. Deshalb wurde hier auf eine zur Abbildung 1.1.1 analoge farbliche Unterscheidung verzichtet.

1.2.1.1 Accidental CDs

Der Wert von CDs, die aus Übergängen berechnet sind, die völlig verschiedene Zustände miteinander verbinden (siehe Abbildung 1.2.2, Fall a), ist die Summe (bzw. Differenz) der Energiedifferenz zwischen den beiden involvierten vibrationsangeregten Zuständen und den beiden involvierten Vibrationsgrundzuständen. Diese CDs werden **accidental combination differences**, aCDs, genannt und sind im Allgemeinen uninteressant.

Wenn keine Gruppen von Zuständen mit äquidistanten Energien existieren, ist es sehr unwahrscheinlich, dass mehrere aCDs den gleichen Wert haben. In CD-Spektren machen sich aCDs als Rauschen bemerkbar.

CDs, die aus Linien berechnet sind, die zu verschiedenen Sets von Zuständen gehören, verhalten sich analog.

1.2.1.2 Ground state CDs

Der Wert von CDs, die aus Linien berechnet sind, die denselben angeregten Molekülzustand haben (siehe Abbildung 1.2.2, Fall b), ist der Abstand zwischen zwei Vibrationsgrundzuständen. Diese sog. **ground state combination differences**, gsCDs, werden in [Asvany et al. 2015] und der vorliegenden Arbeit gesucht.

Wenn es keine Auswahlregeln gäbe, sondern alle Differenzen zwischen zwei Zuständen eines Sets gemessen würden, gäbe es für jeden Übergang zwischen zwei Vibrationsgrundzuständen genau so viele CDs mit dem entsprechenden Wert, wie es vibrationsangeregte Zustände gibt (vergleiche Abbildung 1.2.3, Fall a). Dadurch, dass es Auswahlregeln und für die Messungen zu schwache Linien gibt, reduziert sich die Häufigkeit entsprechend.

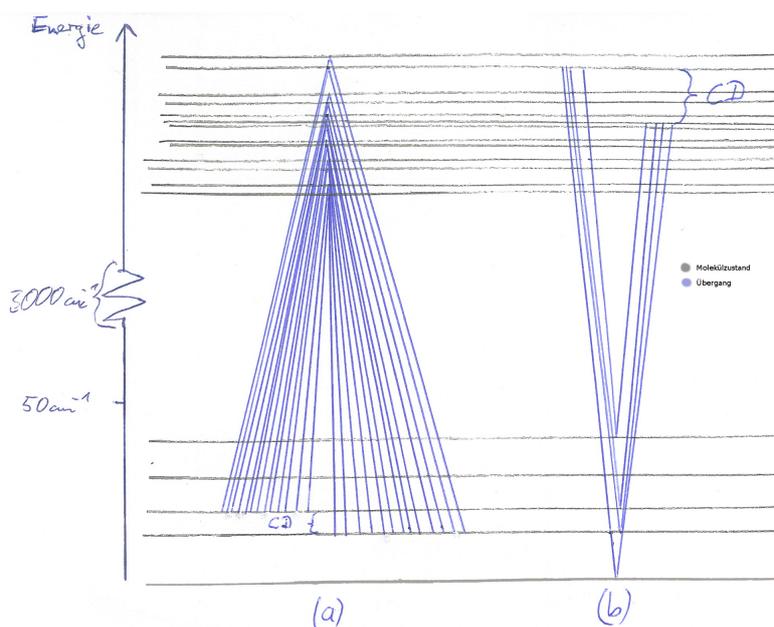


Abbildung 1.2.3: **Häufigkeit von CD-Werten** (ohne Auswahlregeln)

Da CH_5^+ mehrere Vibrationsbanden hat, ist die Anzahl der ersten vibrationsangeregten Zustände (in der Skizze die oberen) ein Vielfaches der Anzahl der Vibrationsgrundzustände (in der Skizze die unteren)

1.2.1.3 Excited state CDs

Im umgekehrten Fall, in dem die Linien einer CD denselben Vibrationsgrundzustand haben (Abbildung 1.2.2, Fall c), entspricht der Wert der CD analog dem Abstand zwischen zwei vibrationsangeregten Zuständen. Diese sog. **excited state combination differences**, esCDs, kommen — wenn man Auswahlregeln nicht berücksichtigt — so oft vor, wie das Set Vibrationsgrundzustände hat (Abbildung 1.2.3, Fall b). Berücksichtigt man Auswahlregeln, sind sie entsprechend seltener.

1.2.2 Rekonstruktion der Differenzen zwischen Vibrationsgrundzuständen auf Grund der Häufigkeiten von CDs

Da CH_5^+ mehrere Vibrationsmoden hat, deren erste angeregte Banden bei 3000 cm^{-1} liegen, ist die Anzahl der Zustände, in die die Ionen bei der Messung durch Absorption angeregt werden, ein Vielfaches der Anzahl der Vibrationsgrundzustände, in denen sich die Ionen vor der Absorption befinden.

Dementsprechend erhält man die Übergänge zwischen den Vibrationsgrundzuständen, die bei der Messtemperatur besetzt sind, indem man die häufigen CD-Energien heraus sucht. Sie machen sich u. a. in einem Histogramm der CD-Werte als Peaks bemerkbar. (Vergleiche Abbildung 1.2.4 (A))

CD-Typ	Bedeutung des CD-Wertes	Häufigkeit der einzelnen Werte
gsCDs	Abstand zwischen Vibrationsgrundzuständen	$\leq \#$ vibrationsangeregte Zustände des Sets
esCDs	Abstand zwischen vibrationsangeregten Zuständen	$\leq \#$ Vibrationsgrundzustände des Sets
aCDs	keine Bedeutung	selten, typischerweise 1, solange Zustände nicht äquidistant sind

Tabelle 1.2.1: Häufigkeiten der verschiedenen CD-Werte

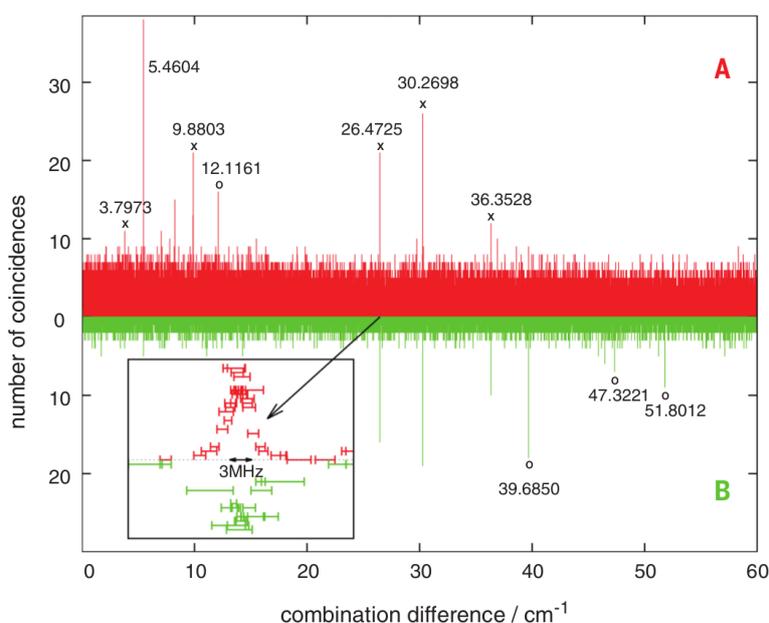


Abbildung 1.2.4: CH_5^+ -CD-Spektren.

(A) bei nominal 10 K. (B) temperaturgemischt, siehe Abschnitt 1.2.3

Quelle: [Asvany et al. 2015]

1.2.3 Rekonstruktion der Vibrationsgrundzustände mit Hilfe temperaturgemischter CDs

Die Ergebnisse des oben erläuterten Verfahrens (Abschnitt 1.2.2) führen zu einer Rekonstruktion der Übergänge zwischen den Vibrationsgrundzuständen. Die Übergänge werden aber weder bestimmten Molekülzuständen zugeordnet, noch wird die Lage der Zustände bestimmt. Um die Lage der Vibrationsgrundzustände zu rekonstruieren, werden deshalb in [Asvany et al. 2015] gemischte CDs gebildet, bei denen immer eine Linie aus dem 10 K-Spektrum von einer Linie aus dem 4 K-Spektrum abgezogen wird (Abbildung 1.2.5). Geht man davon aus, dass bei 4 K nur der niedrigste Vibrationsgrundzustand des jeweiligen Sets besetzt ist, sind die so rekonstruierten Übergänge immer Übergänge von diesem Zustand (dem die Energie 0 cm^{-1} zugewiesen wird) zu irgendeinem anderen Vibrationsgrundzustand des Sets, also die Energien der Vibrationsgrundzustände selbst. Abbildung 1.2.4 (B) zeigt ein CD-Spektrum aus solchen temperaturgemischten CDs.

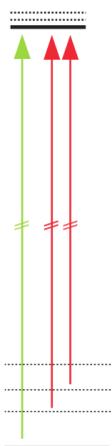


Abbildung 1.2.5: **Temperaturgemischte CDs**

Die Übergänge vom niedrigsten Vibrationsgrundzustand (grün) aus kommen sowohl im 4 K-Spektrum als auch im 10 K-Spektrum vor. Die übrigen Übergänge (rot) kommen nur im 10 K-Spektrum vor. Die gemischten Zustände werden erzeugt, indem immer ein roter von einem grünen Übergang abgezogen wird.

Quelle: [Asvany et al. 2015], Ausschnitt aus der dortigen Abbildung 2

1.3 Arbeitsprogramm dieser Diplomarbeit

Ziel dieser Diplomarbeit ist es, die in [Asvany et al. 2015] verwendeten Methoden weiter zu entwickeln bzw. neue Methoden zu entwickeln, um aus den vorhandenen Messdaten weitere Erkenntnisse zu gewinnen. Nach Möglichkeit sollen die (weiter-) entwickelten Methoden auch zur Auswertung spektroskopischer Daten in anderen Kontexten, bei denen wenig über die untersuchten Moleküle bekannt ist, verwendet werden können.

Bei dem von Asvany et al. verwendeten Auswertungsverfahren gibt es eine prinzipielle Schwierigkeit: Da die Messdaten fehlerbehaftet sind, ist das im Abschnitt 1.2 angesprochene Zählen, wie oft CDs eine bestimmte Energie haben, nicht trivial: Bei der Auswertung von realen Messdaten kommt jeder einzelne CD-Wert nur genau einmal vor, allerdings häufen sich die

1 Einführung

	10 K-CD-Spektrum	kombiniertes 4 K-10 K-CD-Spektrum
CD-Anzahl	4194856 (3815703)	204015(194180)
CD-Fehler (Median)	0,000054 cm ⁻¹ (0,000070 cm ⁻¹)	0,000036 cm ⁻¹ (0,000036 cm ⁻¹)
CD-Abstand (Median)	0,000040 cm ⁻¹ (0,000046 cm ⁻¹)	0,00045 cm ⁻¹ (0,00045 cm ⁻¹)

Tabelle 1.3.1: **Eigenschaften der CD-Spektren**

Wenn wie in [Asvany et al. 2015] nur ausgewählte Linien verwendet werden, gelten die Werte in Klammern

Auf den ersten Blick mag es erstaunen, dass der Median des CD-Fehlers durch Verwerfen „schlechter“ Messungen steigt. Dies liegt daran, dass nur sehr wenige Linien auf Grund eines großen Fehlers aussortiert werden; der Median ist gegenüber solchen Ausreißern unempfindlich; im Gegensatz zum Mittelwert berücksichtigt er nicht, wie weit die Ausreißer vom Rest abweichen, sondern lediglich ihre Anzahl (und auch die nur gering). Wesentlich mehr Linien werden auf Grund zu geringer Intensität aussortiert. Diese Linien haben größtenteils einen unterhalb des Medians liegenden Fehler und haben wegen ihrer größeren Anzahl (ihr genauer Wert spielt wiederum keine Rolle) einen größeren Einfluss auf den Median.

CDs in der Gegend bestimmter Energien. Asvany et al. zählen deshalb für jede CD, wie viele weitere CDs eine ähnliche Energie haben. Zwei CDs werden dabei als nahe genug beieinander liegend gewertet, wenn ihr Abstand kleiner als die Summe der Fehler der beiden CDs ist. (Die Fehler der einzelnen CDs werden aus den Messfehlern der zugehörigen Linien gemäß $\Delta CD = \sqrt{(\Delta Linie1)^2 + (\Delta Linie2)^2}$ ermittelt.) Die Ergebnisse werden in einem CD-Spektrum, einem verallgemeinerten Histogramm der CD-Werte, aufgetragen, in dem jeder Kombinationsdifferenz als horizontale Koordinate ihre Energie und als vertikale Koordinate die ermittelte Anzahl anderer CDs mit ähnlicher Energie zugeordnet wird. Die gesuchten CDs, die häufig vorkommen, sind in so einem Diagramm als Peaks erkennbar.

Mit 185 Linien im 4 K-Spektrum und 2897 Linien im 10 K-Spektrum liegen die Linien sehr dicht. Die Anzahl der CDs, die sich aus n Linien ergeben, ist $\frac{n(n-1)}{2}$. (Für das kombinierte 4 K-10 K-CD-Spektrum lässt sich keine einfache Formel angeben.) Die Eigenschaften der beiden CD-Spektren finden sich in Tabelle 1.3.1 (siehe auch ergänzend Abbildung 3.2.3 weiter unten). Demnach kann die in [Asvany et al. 2015] verwendete Methode gut verwendet werden, um Peaks im kombinierten 4 K-10 K-CD-Spektrum zu finden; wie zuverlässig die Methode für die CDs aus dem 10 K-Spektrum ist, lässt sich allerdings nicht so einfach beurteilen, zumal die CDs zwischen den gesuchten Häufungen nicht gleichmäßig verteilt sind. Im Ergebnis sind eine Reihe von Peaks in den CD-Spektren klar zu identifizieren, an vielen Stellen ist aber unklar, ob tatsächlich ein Peak vorliegt (vergleiche etwa Abbildung 1.2.4). Vorsichtshalber werden diese fraglichen Peaks von Asvany et al. nicht weiter berücksichtigt.

Angesichts dessen ist die vorliegende Diplomarbeit wie folgt aufgebaut:

Testumgebung Zunächst wird eine Testumgebung vorgestellt, mit deren Hilfe es möglich ist, die Zuverlässigkeit der Auswertungsverfahren an Hand von künstlich erzeugten Daten zu testen. Sie dient in einem ersten Schritt dazu, zu beurteilen, inwieweit beim bisherigen Auswertungsverfahren auch den nicht so eindeutigen Ergebnissen vertraut werden kann. Vor allem dient sie aber als direktes feedback bei der systematischen Weiterentwicklung des Auswertungsverfahrens.

Optimierung des Kerndichteschätzers Bei genauerem Hinsehen stellt sich heraus, dass das bisherige Auswertungsverfahren im Wesentlichen einem in der Statistik verbreiteten Verfahren zur

1 Einführung

Schätzung von Wahrscheinlichkeitsverteilungen auf Grund von Messdaten mit Hilfe sog. Kerndichteschätzer entspricht. Dabei entspricht das CD-Spektrum der geschätzten Wahrscheinlichkeitsverteilung für die Energie einer zufällig heraus gegriffenen CD. Diese Erkenntnis und einige grundlegende Eigenschaften von Kerndichteschätzern werden genutzt, um das Auswertungsverfahren einerseits im Detail weiter zu entwickeln und andererseits die Bedeutung der Parameter (die in [Asvany et al. 2015] nur implizit vorkommen) heraus zu arbeiten und Kriterien für eine sinnvolle Wahl zu entwickeln.

Peakerkennung Im Laufe der Erstellung dieser Arbeit hat sich heraus gestellt, dass ein mehrstufiges Verfahren, bei dem Berücksichtigung gesicherter Erkenntnisse weiter Auswertungsdurchläufe, insbesondere eine erneute Kerndichteschätzung auf Grundlage vorselektierter CDs stattfindet, gleichzeitig die besten Ergebnisse verspricht und am robustesten ist. Um die Entscheidung, ob ein Peak vorliegt oder nicht, reproduzierbar und das iterative Verfahren praktisch handhabbar zu machen, wurde eine automatische Peakerkennung entwickelt, die zudem das Skalierungsverhalten der Peaks nutzt, das sich im einfachen CD-Spektrum nicht ablesen lässt.

Rekonstruktion des Energie-Niveau-Schemas Die in [Asvany et al. 2015] verwendete Methode zur Rekonstruktion der Vibrationsgrundzustände mit Hilfe von Messungen bei zwei verschiedenen Temperaturen hat eine Reihe von Nachteilen. Deshalb wurde eine alternative Methode entwickelt, die weitgehend automatisch funktioniert, geringere Ansprüche an die Messungen stellt, die automatische Verifikation der Ergebnisse auf verschiedenen Wegen ermöglicht und auch einen Teil der vibrationsangeregten Zustände rekonstruiert.

Das Programm Auf Grundlage der einzelnen Ergebnisse wurde ein Python-Script entwickelt, das die Auswertung der Messwerte und die Verifikation und ggf. Korrektur der Ergebnisse in einem mehrschrittigen Verfahren erlaubt. Die Funktionsweise dieses Programms wird erläutert.

Ergebnisse CD_2H^+ kann auf herkömmliche Weise modelliert werden. Deshalb wird das entwickelte Programm zunächst auf Messungen an CD_2H^+ angewendet und die Ergebnisse mit denen aus der herkömmlichen Auswertung verglichen. Zuletzt wird das Programm auf die 10 K-Messwerte an CH_5^+ aus [Asvany et al. 2015] angewendet und diskutiert.

2 Testumgebung

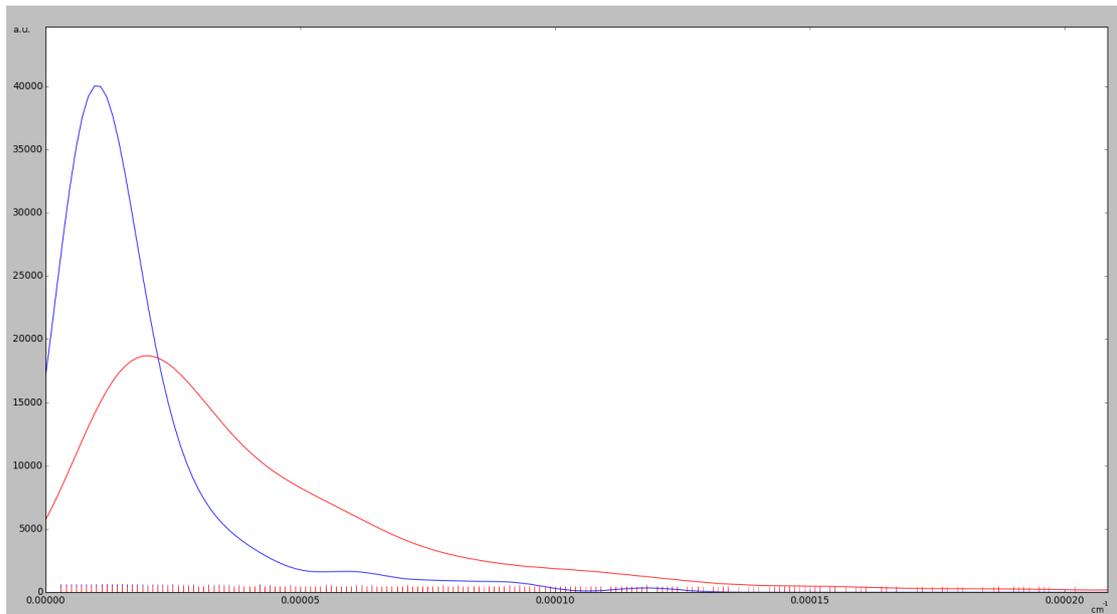


Abbildung 2.0.1: **Verteilung der Messfehler der Übergangsenergien**

aus [Asvany et al. 2015]

blau: 4 K-Messungen. **rot:** 10 K-Messungen.

Berechnet mit einem Kerndichteschätzer; zu Kerndichteschätzern siehe auch Abschnitt 3

Als Testumgebung wurde eine Sammlung von Python-Scripts entwickelt, die ein zufälliges Energie-Niveau-Schema eines Moleküls erzeugt und daraus Messwerte berechnet. Damit lässt sich testen, wie gut die Auswertungsmethoden in der Lage sind, aus den künstlich erzeugten Messwerten das zu Grunde liegende Energie-Niveau-Schema zu rekonstruieren.

Dabei sollen die Auswertungsverfahren möglichst voraussetzungslos aus den gemessenen Energien der Absorptionslinien (die Intensitäten sind nicht zuverlässig vermessen) das Energie-Niveau-Schema rekonstruieren. Dementsprechend muss die Testumgebung die Zuverlässigkeit der Auswertungsverfahren unabhängig davon testen, welche Bedeutung die einzelnen Zustände im Energie-Niveau-Schema haben, was ihre Drehimpuls- oder Spin-Quantenzahlen sind, welche Übergänge erlaubt sind etc. Deshalb wird das einem Test zu Grunde liegende Energie-Niveau-Schema zufällig erzeugt, es werden alle möglichen Übergänge zwischen den erzeugten Zuständen berechnet und ein Teil davon gemäß ebenfalls zufälliger Auswahlregeln verworfen. Um auch Auswertungsverfahren testen zu können, bei denen Messungen bei zwei verschiedenen Temperaturen verwendet werden (wie das von Asvany et al., siehe Abschnitt 1.2.3), werden gleichzeitig Messdaten für zwei verschiedene Temperaturen erzeugt. Im Einzelnen erzeugen die Scripte einen künstlichen Satz

2 Testumgebung

an Messwerten, indem sie die Schritte, die zur Entstehung der tatsächlichen Messdaten führen, nachvollziehen:

Für jedes zu simulierende Set¹ mit relativer Häufigkeit s werden folgende Schritte abgearbeitet:

1. Zufällige Erzeugung einer kleinen Anzahl von Vibrationsgrundzuständen (meist etwa 10 zwischen 0 cm^{-1} und 60 cm^{-1}). Diese Niveaus sollen durch das zu testende Auswertungsverfahren rekonstruiert werden.
2. Zufällige Erzeugung einer großen Anzahl vibrationsangeregter Zustände (meist mehrere 100 zwischen 2800 cm^{-1} und 3200 cm^{-1}).²
3. Berechnung aller möglichen Übergangslinien zwischen jeweils einem vibrationsangeregten Zustand und einem der Vibrationsgrundzustände.
4. Verwerfen eines Teils der Linien gemäß zufälliger Auswahlregeln.
5. Zuweisung einer Übergangswahrscheinlichkeit k zu jeder Linie. k ist eine Zufallszahl, die im Wesentlichen die Bedeutung des unbekanntem Einsteinkoeffizienten für die Absorption eines Photons hat. Da über k keinerlei Informationen vorliegen, wird k gleichverteilt in einem beliebigen, aber festen Intervall gewählt.³

Bis hierhin wird der Aufbau der Moleküle simuliert; die Schritte werden für die einzelnen Sets unabhängig voneinander durchgeführt, die Temperatur spielt noch keine Rolle. In den nächsten Schritten spielt der Zustand der Moleküle eine Rolle, deshalb müssen alle folgenden Schritte für beide Temperaturen (und alle Sets) unabhängig durchlaufen werden.

6. Berechnung der Besetzungswahrscheinlichkeit w sämtlicher Niveaus mit Hilfe der Boltzmann-Verteilung⁴ und der als Parameter vorgegebenen Messtemperatur.⁵
7. Berechnung der Intensitäten der Linien gemäß $\text{Intensität} = s * w * k$. Dabei ist s die relative Häufigkeit des Sets, w die im 6. Schritt berechnete Besetzungswahrscheinlichkeit des unteren Zustandes und k die in Schritt 5 zugewiesene Übergangswahrscheinlichkeit.

Bis hierhin wird die Absorption der Photonen durch die Moleküle der einzelnen Sets simuliert. Deshalb werden diese Schritte auch für jedes Set getrennt durchgeführt. Die weiteren Schritte simulieren den Messprozess, bei dem alle Sets gleichzeitig vermessen werden. Deshalb werden an dieser Stelle die Zwischenergebnisse der einzelnen Sets (für die beiden Temperaturen getrennt) zusammen geführt und die folgenden Schritte für alle Sets gleichzeitig durchlaufen:

8. Verwerfen von Übergängen, die unterhalb einer bestimmten, als Parameter zu wählenden Intensität liegen, denn sie sind faktisch nicht messbar. Da im realen Experiment die Messungen der beiden Temperaturen mit verschiedenen Aufbauten durchgeführt wurden, werden Skalierung und Cutoff für beide Temperaturen separat ermittelt. (Näheres in Abschnitt 2.2)

¹vergleiche Abschnitt 1.1

²Der Energiebereich, in dem die Zustände liegen, muss nicht explizit parametrisiert werden, sondern kann vom Script, das die Testdaten generiert, an Hand der gemessenen Spektren automatisch bestimmt werden. Die Anzahl muss allerdings vorgegeben werden.

³ s und k müssen nicht normiert sein, weil die berechnete Intensität in Schritt 8 so wie so noch einmal skaliert wird.

⁴Zwar sind die Ionen bei den realen Messungen in einer Ionenfalle und damit einem Potenzial ausgesetzt, sodass die Besetzung der Zustände eigentlich nicht Boltzmann-verteilt ist. Allerdings wurde in den Experimenten eine Multipol-Falle mit großem nahezu feldfreien Raum verwendet. Außerdem war die Ionenzahl in der Falle verhältnismäßig gering, sodass die Ionen nicht auf Grund von Coulomb-Abstoßung an den Rand der Falle gedrängt wurden. Deshalb können die Potenzialdifferenzen, denen die Ionen in der Falle ausgesetzt waren, angesichts der Temperatur in der Falle vernachlässigt werden.

⁵Damit die relative Häufigkeit der verschiedenen Sets s , vergleiche Schritt 7, nicht bedeutungslos wird, muss w für jedes Set und jede Temperatur unabhängig normiert werden.

9. Versehen der Energien der verbleibenden Linien mit einem zufälligen Messfehler.⁶

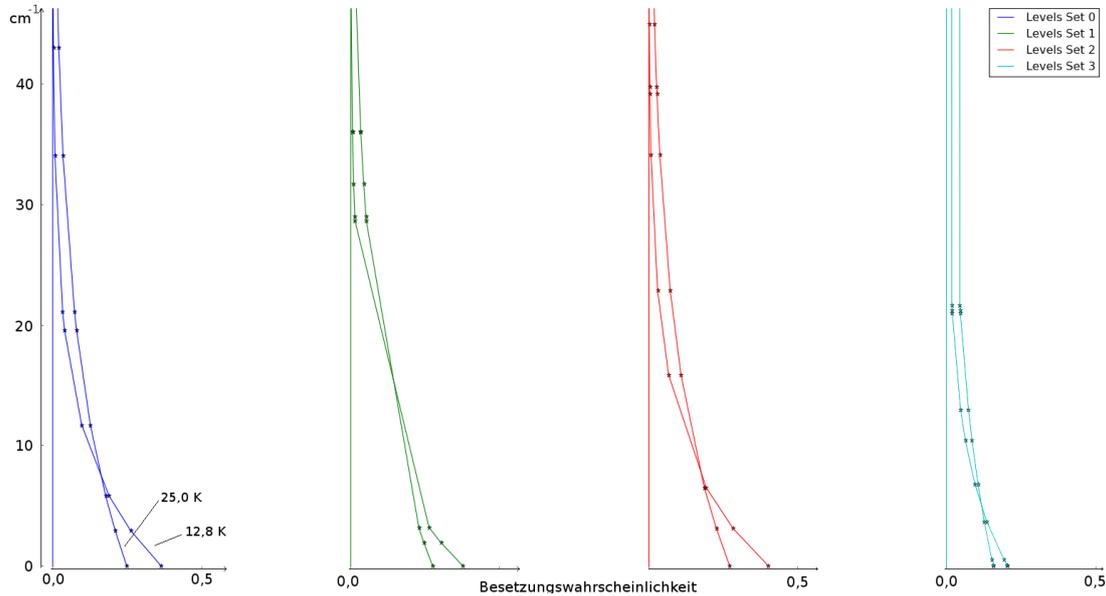


Abbildung 2.0.2: **Zwischenergebnis nach 3. Schritt**
 Zufällig erzeugte Levels für 4 Sets mit Besetzungswahrscheinlichkeiten bei 12,8 K und 25 K

Sämtliche Zwischenergebnisse werden gespeichert, sodass sie zur Bewertung der Auswertungsergebnisse — auch automatisch — heran gezogen werden können. So können etwa im CD-Spektrum eines Auswertungslaufes die Abstände zwischen den im ersten Schritt erzeugten Vibrationsgrundzuständen mit geplottet werden (vergleiche Abbildung 2.0.3).

Die Verwendung von Zufallszahlen an verschiedenen Stellen in diesem Prozess ermöglicht es, einfach und weitgehend automatisch zu testen, ob das Auswertungsverfahren nur zufällig oder zuverlässig funktioniert, indem allein durch mehrfaches Ausführen der Scripte mehrere verschiedene Testdatensätze mit gleichen Parametern erzeugt und getestet werden.

Wie universell das Auswertungsverfahren ist, lässt sich testen, indem mehrere Durchläufe mit veränderten Parametern und Wahrscheinlichkeitsverteilungen für die Zufallszahlen untersucht werden. Dabei macht es in der Regel Sinn, von einem Satz an Parametern und Wahrscheinlichkeitsverteilungen auszugehen, der Daten produziert, die möglichst nah an den tatsächlichen Messdaten liegen (siehe Tabelle 2.0.1).

⁶Da die Messfehler aus mehreren voneinander unabhängigen Fehlerquellen mit endlicher Varianz stammen, können sie als Gauß-verteilt angenommen werden (Zentraler Grenzwertsatz). Allerdings wurden die verschiedenen Linien bei den realen Messungen mit unterschiedlichen Genauigkeiten gemessen (z. B. unterschiedlich lange integriert). Die Messfehler sind zwar nicht gleichverteilt (siehe Abbildung 2.0.1), werden der Einfachheit halber in der Testumgebung aber als gleichverteilt im Intervall $[\min(\text{reale Messfehler}), E(\text{reale Messfehler}) + 3\sigma(\text{reale Messfehler})]$ angenommen. Damit streuen die Messfehler in den Testdaten etwas mehr als in den realen Daten. Stichprobenartige Tests haben ergeben, dass sich dies im Auswertungsprozess nicht bemerkbar macht. Tendenziell werden die Auswertungsverfahren damit auch eher unter etwas erschwerten Bedingungen getestet, sodass ein Verfahren, dass mit Testdaten gut funktioniert, mit den realen Daten jedenfalls nicht schlechter funktionieren sollte.

2 Testumgebung

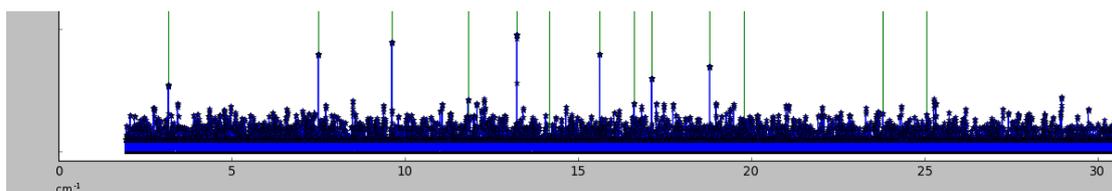


Abbildung 2.0.3: **CD-Spektrum künstlich erzeugter Messwerte**

Die Nutzung der Testumgebung erlaubt die direkte Kontrolle der Auswertungsergebnisse. Die zu rekonstruierenden Abstände zwischen Vibrationsgrundzuständen sind grün geplottet. Wenn die Auswertung gut klappt, sollten an diesen Stellen eindeutig vom Hintergrund unterscheidbare Peaks liegen.

Parameter	Wert	Anmerkungen
# Sets und Gewichtung s	4 Sets mit Gewichtung 4 : 4 : 5 : 5	vergleiche Diskussion in Abschnitt 1.1
Vibrationsgrundzustände	10 zwischen 0 cm^{-1} und 60 cm^{-1}	
vibrationsangeregte Zustände	200 je Set	
Temperaturen	12, 8 K und 25 K	Material zu [Asvany et al. 2015]

Tabelle 2.0.1: Explizit zu setzende Parameter der Testumgebung, die ein den Messdaten ähnliches Testspektrum erzeugen.

2.1 Spinstatistik

Wie bereits erläutert haben die verschiedenen Species und damit die jeweils zwei zugehörigen Sets auf Grund der Spinstatistik verschiedene relative Häufigkeiten, denen bei der Erstellung der Testdaten über den Faktor s Rechnung getragen wird. Innerhalb der einzelnen Sets spielt die Spinstatistik für die Erzeugung von Testdaten dagegen keine Rolle:

Auf Grund des Messverfahrens ist die Messung vollkommen von Ein-Photon-Absorptionsprozessen dominiert. Dementsprechend ändert sich der Gesamtdrehimpuls eines Moleküls bei der Absorption um ± 1 . Da der unterste Zustand einer Vibrationsbande, also hier insbesondere der unterste Zustand jedes Sets, üblicherweise den Gesamtdrehimpuls 0 hat, was bedeutet, dass von da aus nur Absorptionen mit einer Änderung $+1$ möglich sind, könnte man auf die Idee kommen, in der Simulation für Übergänge aus diesem Zustand eine im Schnitt halb so große Übergangswahrscheinlichkeit e (Schritt 7) anzunehmen wie für alle anderen Zustände. Es gibt aber Anhaltspunkte dafür (vergleiche [Schmiedt et al. 2016]), dass bei CH_5^+ bei manchen Sets die jeweils untersten Zustände gar nicht besetzt sind. Damit ist nicht klar, ob überhaupt eins der in Schritt 1 erzeugten Niveaus den Gesamtdrehimpuls 0 hat, und es muss daher von komplett gleichverteilten Übergangswahrscheinlichkeiten ausgegangen werden.

2.2 Was wie parametrisieren?

Die meisten bei der Erstellung der Testdaten notwendigen Parameter lassen sich bei den gemessenen Daten leicht bestimmen und können dann automatisch für die Erzeugung der Testdaten verwendet werden. Die entwickelten Scripte erledigen dies automatisch (erlauben aber auch, die Parameter von Hand zu setzen, um beispielsweise die Abhängigkeit der Zuverlässigkeit der Auswertungsmethoden von den Eigenschaften der Spektren zu untersuchen).

Einige Parameter lassen sich dagegen nicht direkt analog zu den gemessenen Daten setzen. Zu-

2 Testumgebung

dem hängen manche Parameter voneinander ab und es ist zu entscheiden, welchen der Parameter man explizit und welchen implizit setzt.

Auswahlregeln und Gesamtlinienzahl Angesichts des Verfahrens, das das Zustandekommen der Messwerte im realen Experiment nachvollzieht, ist es nahe liegend, für die Auswahlregeln irgendeine Wahrscheinlichkeitsverteilung anzunehmen, was dann automatisch in einer bestimmten Anzahl an Linien resultiert. Ein entscheidender Nachteil dieses Vorgehens ist allerdings, dass die Anzahl der produzierten Linien so nur schlecht zu kontrollieren ist: Die Anzahl der Vibrationsgrundzustände (Schritt 1) ist relativ gering und bei Temperaturen im Bereich der realen Messungen fällt die Boltzmann-Verteilung im Bereich von 0 cm^{-1} bis 60 cm^{-1} stark ab. Deshalb führt schon eine etwas höhere Energie eines der Vibrationsgrundzustände bei diesem Verfahren dazu, dass sehr viel mehr Linien auf Grund zu geringer Intensität aussortiert werden und umgekehrt.

Im Rahmen der Experimente mit verschiedenen Parametern hat sich sehr schnell gezeigt, dass die Dichte der CDs, die quadratisch von der Anzahl der Linien abhängt (also bei festem CutOff noch empfindlicher auf eine etwas andere Lage der Vibrationsgrundzustände reagiert als die Anzahl der Linien), entscheidend für die Qualität der Auswertungsergebnisse ist. (Siehe Abschnitt 2.3.1.) Gleichzeitig erlauben die Messdaten keinerlei direkte Rückschlüsse darauf, wie viele Übergänge durch Auswahlregeln tatsächlich verboten sind, sodass es kaum sinnvolle Anhaltspunkte für die Wahl der Verteilung der Auswahlregeln gibt.

Intensitäts-CutOff Während die Lage der Übergänge und die Auswahlregeln temperaturunabhängig sind, hängen die Intensitäten über die Besetzungswahrscheinlichkeiten der Vibrationsgrundzustände von der Temperatur ab. Deshalb sorgt das Verwerfen der Linien mit niedriger Intensität dafür, dass sich die Spektren bei verschiedenen Temperaturen überhaupt unterscheiden. Die korrekte Wahl des Intensitäts-CutOffs ist daher entscheidend dafür, dass die erzeugten Testdaten mit den Messungen vergleichbar sind.

Die naheliegende Lösung für die Bestimmung des Intensitäts-CutOffs besteht darin, die kleinste in den gemessenen Daten vorkommende Intensität als CutOff zu verwenden. Dies setzt aber voraus, dass die Intensitäten der Testdaten genau so skaliert sind wie die Intensitäten der gemessenen Daten. Die Skalierung der Messdaten ist nicht explizit bekannt. Dennoch lassen sich die Testdaten passend zu den Messdaten skalieren, wenn man die Intensitäten als Resultate eines Zufallsexperimentes ansieht. Die Wahrscheinlichkeitsverteilung dieses Zufallsexperimentes entspricht dann bis auf statistische Schwankungen dem Histogramm der Intensitäten und lässt sich leicht berechnen:

Wie auch in der Testumgebung ist die Intensität i einer Linie, die durch Anregung aus einem Vibrationsgrundzustand mit der Energie e entsteht, Produkt aus einer Exponentialfunktion von e (Boltzmann-Verteilung der Besetzungswahrscheinlichkeiten) und einem Faktor x , in dem im Wesentlichen der Einstein-Koeffizient für die Übergangswahrscheinlichkeit steckt:

$$\begin{aligned} i(e) &= x w(e) & (2.2.1) \\ \text{mit } w(e) &= \exp(-de) \end{aligned}$$

Dabei ist d eine temperaturabhängige Konstante. Zur Herleitung der Wahrscheinlichkeitsdichte der Intensitäten werden die folgenden Definitionen verwendet:

2 Testumgebung

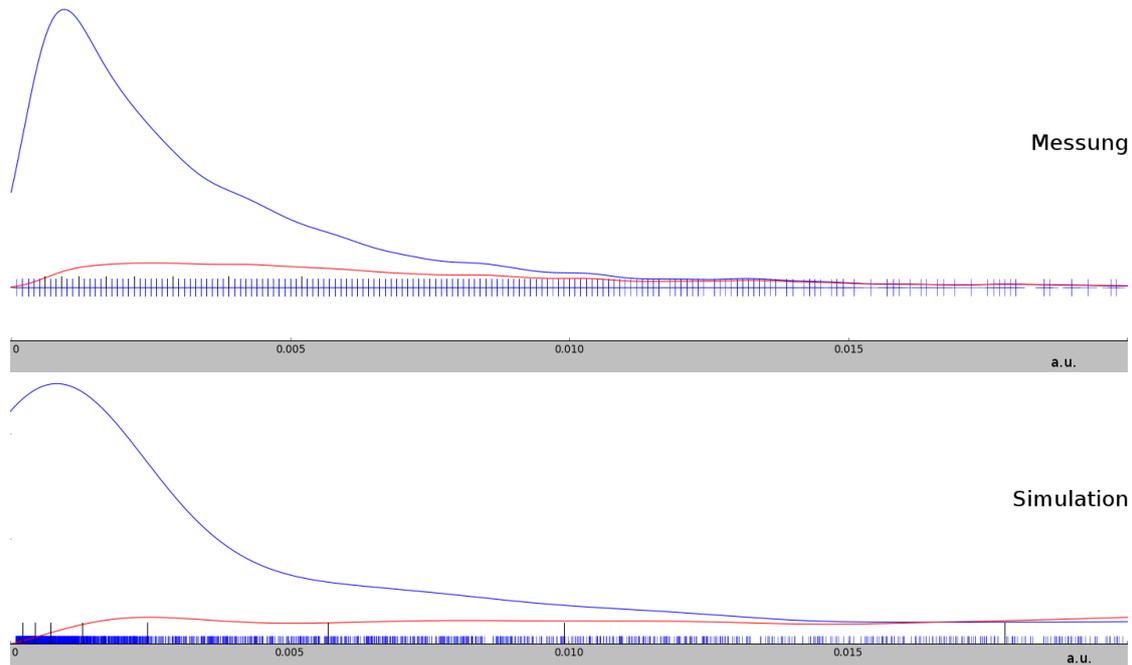


Abbildung 2.2.1: **Intensitätsverteilungen der 10 K-Spektren**

oben: Messdaten. **unten:** Testdaten aus der Simulation, zu den Messdaten passend skaliert.

blau: Intensitätsverteilung $\varrho_I(i)$. **rot:** $i \varrho_I(i)$.

Unten ist jeweils die Intensität der einzelnen Linien markiert. Das simulierte Spektrum hat bei diesem Schritt noch deutlich mehr Linien als das gemessene, weil noch keine Auswahlregeln angewandt wurden, vergleiche Abschnitt 2.2 im Text.

Zufallsvariable der Energie des Ausgangszustandes: E

$$\text{mit Wahrscheinlichkeitsdichte}^7 : \quad \varrho_E(e) = \begin{cases} \frac{1}{f} & 0 \leq e \leq f \\ 0 & \text{sonst} \end{cases}$$

zu $w(e)$ gehörige Zufallsvariable : $W = w(E)$

mit Wahrscheinlichkeitsdichte : $\varrho_W(w)$

zu x gehörige Zufallsvariable : X

$$\text{mit Wahrscheinlichkeitsdichte}^8 : \quad \varrho_X(x) = \begin{cases} g(x) & 0 \leq x \leq c \\ 0 & \text{sonst} \end{cases}$$

Zufallsvariable der Intensität⁹ : $I = XW$

mit Wahrscheinlichkeitsdichte : $\varrho_I(i)$

2 Testumgebung

Dann ergibt sich mit der Transformationsregel für Wahrscheinlichkeitsdichten:

$$\begin{aligned}
 \varrho_W(w) &= \varrho_E(e(w)) \left| \frac{de}{dw} \right| \\
 &= \dots \\
 &= \frac{1}{dwf} \begin{cases} 1 & w \geq \exp(-df) \\ 0 & \text{sonst} \end{cases} \quad (2.2.2)
 \end{aligned}$$

Da die Zufallsvariable I Produkt der beiden Zufallsvariablen X und W ist, hat I die Wahrscheinlichkeitsdichte $\varrho_I(i) = \int \varrho_X(x) \varrho_W\left(\frac{i}{x}\right) \frac{1}{|x|} dx$. Einsetzen von Gleichung 2.2.2 ergibt:

$$\begin{aligned}
 \varrho_I(i) &= \int \varrho_X(x) \varrho_W\left(\frac{i}{x}\right) \frac{1}{|x|} dx \\
 &= \dots \\
 &= \begin{cases} \frac{1}{dfi} \int_0^{i \exp(df)} g(x) dx & i \leq c \exp(-df) \\ \frac{1}{dfi} & \text{sonst} \end{cases} \quad (2.2.3)
 \end{aligned}$$

Für niedrige Intensitäten i ergibt sich aus Gleichung 2.2.3, dass $i \varrho_I(i) = \frac{1}{df} \int_0^{i \exp(df)} g(x) dx$ gilt. Der Vorfaktor besteht nur aus Konstanten, das Integral wächst monoton, weil $g(x) \geq 0$ gilt. Also wächst auch $i \varrho_I(i)$ monoton. Für hohe Intensitäten i^{10} ist $i \varrho_I(i)$ konstant $\frac{1}{df}$.

Berechnet man also die Histogramme der Intensitäten von gemessenen und erzeugten Daten und multipliziert sie mit den Intensitäten selbst, sollte der resultierende Graph zunächst monoton wachsen und dann konstant sein. Aus dem Vergleich der beiden Konstanten lässt sich berechnen, wie die erzeugten Intensitäten zu reskalieren sind, damit der bei den gemessenen Daten vorgefundene CutOff auf die Testdaten angewendet werden kann. Bei den weniger dichten 4K-Spektren scheint diese Methode nur unzuverlässig zu funktionieren, weil die statistischen Schwankungen zu groß sind. Zumindest für nicht allzu große Intensitäten entsprechen die Ergebnisse bei den dichterem 10K-Spektren aber in etwa den Erwartungen, obwohl die Annahme, dass die Energien der Vibrationsgrundzustände gleichmäßig verteilt liegen ($\varrho_E(e)$), lediglich eine grobe Schätzung ist. (Der Graph fällt allerdings leicht ab, wo er konstant sein sollte, siehe Abbildung 2.2.1.)

Insgesamt funktioniert es am besten, den Übergang zwischen den beiden Abschnitten des Graphen, der sich zuverlässig als lokales Maximum identifizieren lässt, zur Skalierung zu nutzen. (Siehe Abbildung 2.2.1.)

Umsetzung Die Linienzahl bzw. die davon abhängige Dichte der CD-Spektren ist entscheidend für die Auswertungszuverlässigkeit. Deshalb ist die Parameterwahl in den Scripts der Testumgebung darauf ausgerichtet, dass die Anzahl der erzeugten Linien und das Abschneiden der Linien mit zu geringer Intensität möglichst gut abgebildet werden.

In Schritt 8 wird daher zunächst das 10K-Spektrum wie beschrieben zu den Messdaten passend skaliert, sodass die Linien mit zu geringer Intensität analog zu den Messdaten abgeschnitten werden können. Beim 4K-Spektrum, bei dem dies nicht zuverlässig möglich ist, wird ein anderes Verfahren angewendet: Nachdem beim 10K-Spektrum die Linien mit zu geringer Intensität verworfen wurden, werden beim 4K-Spektrum so lange die jeweils schwächsten Linien verworfen, bis

⁷Da nichts über die Lage der Vibrationsgrundzustände bekannt ist, macht es Sinn, sie als gleichverteilt im Intervall $[0, f]$ anzunehmen.

⁸ x ist beschränkt. Deswegen gibt es ein Intervall $[0, c]$, sodass ist $\varrho_X(x)$ außerhalb dieses Intervalls verschwindet. Innerhalb dieses Intervalls ist $\varrho_X(x)$ eine unbekannte nicht-negative Funktion $g(x)$.

⁹vergleiche Gleichung 2.2.1

¹⁰ $i \geq c \exp(-df)$

das Verhältnis der Linienanzahlen von 10 K und 4 K-Spektrum den Messdaten entspricht. Zuletzt werden die Auswahlregeln simuliert und zufällig so viele Linien verworfen, bis die Gesamtzahl der verbleibenden Linien (10 K und 4 K-Spektrum zusammen) der Gesamtzahl der gemessenen Linien entspricht.

Dabei ergibt sich folgende Schwierigkeit: Wie viele Linien auf Grund von Auswahlregeln verworfen werden müssen, ist erst nach dem Intensitäts-CutOff in Schritt 8 klar, die Auswahlregeln sind aber temperaturunabhängig. Deshalb wird dabei folgendermaßen vorgegangen: Statt in Schritt 4 Linien komplett zu verwerfen, wird ihnen lediglich eine Zufallszahl zugeordnet. Nach Schritt 8 werden von den dann noch verbleibenden Linien die verworfen, deren Zufallszahl unterhalb eines bestimmten Schwellwertes liegen. Dieser Schwellwert wird genau so gewählt, dass die richtige Anzahl an Linien übrig bleibt.

2.3 Zwischenergebnisse

Nach Entwicklung der Testumgebung wurde zunächst die Zuverlässigkeit der Auswertungsmethode von [Asvany et al. 2015] untersucht. Dabei wurden recht offensichtliche Hinweise darauf gefunden, wo Verbesserungsmöglichkeiten bestehen, die im weiteren Verlauf dieser Arbeit dann auch erfolgreich ergriffen wurden.

2.3.1 Zuverlässigkeit der Auswertung hängt von der CD-Dichte ab

Wie bereits in Abschnitt 1.2 erläutert, besteht die Hauptschwierigkeit bei der Auswertung der Daten mittels CD-Spektren darin, dass die CDs — gemessen an ihrem Fehler — sehr dicht liegen, sodass nur schlecht entschieden werden kann, ob tatsächlich ein Peak vorliegt oder nur zufällig viele einzelne CDs dicht beieinander liegen. Entscheidend dabei ist, bis zu welchem Abstand zwei CDs zum selben Peak gezählt werden. Dieser Abstand wird im Vorgriff auf den folgenden Abschnitt „Bandbreite“ genannt.

Offensichtlich sollte die Bandbreite b nach Möglichkeit folgendermaßen gewählt werden:

1. Die Bandbreite sollte größer sein als der Fehler der CDs. Sonst werden CDs, die zum selben Peak gehören, auf Grund ihrer Fehler als einzelne (oder zu zwei verschiedenen CD-Peaks gehörige) CDs interpretiert.
2. Die Bandbreite sollte kleiner als der typische Abstand zwischen zwei benachbarten CDs sein.¹¹ Ansonsten werden mit großer Wahrscheinlichkeit auch dort Peaks identifiziert, wo lediglich zufällig CDs nahe beieinander liegen.

In [Asvany et al. 2015] werden — wie schon erwähnt — zwei CDs zum selben Peak gezählt, wenn ihr Abstand kleiner ist als die Summe ihrer Fehler, die sich jeweils gemäß $\Delta CD = \sqrt{(\Delta Linie1)^2 + (\Delta Linie2)^2}$ ergibt. Damit wird sicher gestellt, dass die Bandbreite genau so groß ist wie der Fehler. Kriterium 1 wird damit fast erfüllt (Die Bandbreite sollte eigentlich ein bisschen größer und nicht genau so groß sein wie der Fehler.), Kriterium 2 wird dagegen nicht systematisch berücksichtigt. Da beide Kriterien bei den vorliegenden Daten nicht gleichzeitig erfüllt werden können, wie die Abschätzungen in Abschnitt 1.2 zeigen, wird das 2. Kriterium systematisch nicht erfüllt.

Tests haben gezeigt, dass dies eine sinnvolle Wahl ist, dass also das Kriterium, dass die Bandbreite größer als der Fehler sein soll, wichtiger ist als das Kriterium, dass die Bandbreite kleiner

¹¹Betrachtet man beispielsweise den Median der Abstände benachbarter CDs als typischen Abstand benachbarter CDs, spielen die CDs, die zu einem Peak gehäuft sind, keine Rolle: Es gibt sehr viel mehr aCDs als gsCDs, sodass die gsCDs wie Ausreißer wirken, denen gegenüber der Median unempfindlich ist.

als der typische CD-Abstand sein soll. Dies ist auch plausibel: Eine gemäß Kriterium 1 zu kleine Bandbreite führt tendenziell dazu, dass Peaks zerfallen und im Rauschen untergehen. Dagegen führt eine gemäß Kriterium 2 zu große Bandbreite dazu, dass einzelne CDs, die zufällig nahe beieinander liegen, fälschlicherweise zu einem Peak zusammen gezogen werden. Dabei ist es einerseits sehr unwahrscheinlich, dass dies so viele CDs betrifft, dass dadurch wirklich große Peaks entstehen. Andererseits werden Rauschen und kleine Peaks durch die Glättungswirkung einer großen Bandbreite unterdrückt, was das Problem fälschlicherweise entstehender, in der Regel kleiner Peaks teilweise kompensiert.

In den Tests stellte sich schnell heraus, dass sogar eine noch etwas größere Bandbreite als die von Asvany et al. gewählte, die dann das Kriterium 1 tatsächlich erfüllt, weil sie größer und nicht nur gleich groß wie der Fehler ist, noch bessere Ergebnisse liefert.

2.3.1.1 Variable Bandbreite ist kontraproduktiv

Wird die Bandbreite jeweils von den Fehlern $\Delta CD = \sqrt{(\Delta Linie1)^2 + (\Delta Linie2)^2}$ der beteiligten CDs abhängig gemacht, schwankt sie entlang des CD-Spektrums. Die Auswirkung dieser Schwankung wurde mittels der Testumgebung untersucht mit folgenden Ergebnissen:

- Obwohl die Messfehler der Messungen aus [Asvany et al. 2015] um mehr als eine Größenordnung schwanken¹², hat es keine merklichen Auswirkungen, wenn man die variable Bandbreite durch eine konstante Bandbreite ersetzt. (Die Güte der Ergebnisse hängt allerdings sehr stark davon ab, ob diese konstante Bandbreite sinnvoll gesetzt ist.)
- Untersuchungen mit Testdaten, die eine sehr viel größere Streuung der Messfehler haben, haben ergeben, dass eine wie in [Asvany et al. 2015] variabel gewählte Bandbreite kontraproduktiv ist: CDs mit sehr großen Messfehlern werden bei sehr viel mehr benachbarten CDs mitgezählt als andere CDs. Damit werden Rauschen oder kleine Peaks in der Nähe einer CD mit großem Fehler im CD-Spektrum höher als anderswo und dadurch mitunter fälschlicherweise als (große) Peaks identifiziert. Zudem werden so besonders schlechte Messungen insgesamt besonders stark gewichtet.
- Eine naheliegende Lösung dieses Problems ist eine Reskalierung: Die Peaks werden dabei nicht einfach nur gezählt, sondern dabei mit dem Kehrwert ihres Fehlers gewichtet. Dies führt dazu, dass alle CDs mit gleichem Gesamtgewicht in die Auswertung eingehen. Gleichzeitig ergibt sich dadurch aber ein anderes Problem: Einzelne, besonders genau gemessene Peaks haben, wenn sie denn trotz ihres kleinen Fehlers einmal fälschlicherweise mitgezählt werden (was bei in der Tendenz zu groß gewählter Bandbreite durchaus häufiger vorkommt), in einem sehr schmalen Bereich einen so hohen Wert, dass direkt ein hoher Peak entsteht.

Angesichts dieser Ergebnisse wurde im weiteren Verlauf der Arbeit mit konstanter Bandbreite gearbeitet und die Information über die Messfehler nicht weiter verwendet. So werden zwar Informationen nicht genutzt, aber es ergibt sich dennoch im Vergleich zu vorher eine (kleine) Verbesserung. Vor allem aber wird die Auswertungsmethode dadurch zu einer Kerndichteschätzung, siehe Abschnitt 2.3.2.

2.3.1.2 Auswahlregeln erschweren Auswertung

Während der Tests hat sich gezeigt, dass die Zuverlässigkeit der Auswertungsverfahren aus [Asvany et al. 2015] (und auch aller im Verlauf dieser Arbeit später erarbeiteten Verfahren) sehr

¹²Bis auf wenige Ausreißer liegen die Fehler zwischen $0,00001 \text{ cm}^{-1}$ und $0,00015 \text{ cm}^{-1}$, siehe Abbildung 3.2.3

davon abhängt, wie viele Übergänge auf Grund von Auswahlregeln verboten sind. Bei Testdaten ganz ohne Auswahlregeln liefert das von Asvany et al. verwendete Verfahren auch bei hohen, mit der Dichte der 10 K-Messdaten vergleichbaren Dichten bei richtiger Wahl der Bandbreite fast immer ein richtiges und vollständiges Ergebnis. Sind dagegen z. B. 50 % der Übergänge durch Auswahlregeln verboten, sind bei gleicher CD-Dichte je nach gewählter Bandbreite entweder nur etwa 30 % der Peaks erkennbar oder es werden zwar die richtigen Peaks deutlich, aber zusätzlich etwa doppelt so viele falsche. Bei geringeren CD-Dichten, die mit den 4 K-Messdaten vergleichbar sind, ist das Verfahren auch dann noch zuverlässig, wenn mehr als 50 % der Übergänge auf Grund von Auswahlregeln fehlen.

Diese Ergebnisse sind plausibel: Wenn aCDs fehlen, wird dadurch die Dichte des Rauschens geringer. Wenn gsCDs fehlen, wird die Höhe der Peaks proportional zur Anzahl der fehlenden Übergänge geringer. (Vergleiche Abschnitt 1.2.1.) Da die optimale Bandbreite bei den vorliegenden Daten größer als die Fehler ist (vergleiche Abschnitt 2.3.1), besteht das Problem zerfallender Peaks nicht, sodass auch eine weitere Vergrößerung der Bandbreite, die durch eine geringere Dichte des Rauschens möglich wird, keine Verbesserungen bringt. Die geringere Dichte führt zwar tendenziell auch bei gleicher Bandbreite zu weniger falschen Peaks, in den Tests hat sich allerdings gezeigt, dass dieser Effekt die wesentlich geringere Höhe der richtigen Peaks, die dazu führt, dass manche Peaks gar nicht mehr als solche zu identifizieren sind, nicht kompensieren kann.

Da nur schwer abzuschätzen ist, wie viele Übergänge bei den realen Messdaten auf Grund von Auswahlregeln fehlen, wurden alle Auswertungsverfahren mit Daten getestet, bei denen zwischen 0 % und 75 % der Übergänge fehlten. Die recht hohe Empfindlichkeit gegenüber Auswahlregeln zeigte sich bei allen Verfahren ähnlich.

In [Asvany et al. 2015] wurden bei der Auswertung Linien, die nur mit niedriger Intensität oder mit großem Fehler gemessen wurden, weggelassen. Dies wirkt faktisch wie zusätzliche Auswahlregeln. Deshalb wurden diese Linien in dieser Arbeit nicht aussortiert. Die schlechte Qualität dieser Messdaten hat — wie sich mit entsprechend konstruierten Messdaten gezeigt hat — (bei konstanter Bandbreite) keinen negativen Einfluss: Die Intensität wird so wie so nicht berücksichtigt und die Linienfehler selbst der am schlechtesten gemessenen Linien liegen immer noch innerhalb der Bandbreite, die die besten Ergebnisse liefert. Umgekehrt sind allein durch diese Maßnahme (wenige) neue Peaks aufgetaucht, die im weiteren Verlauf der Arbeit bestätigt werden konnten.¹³

2.3.2 Auswertung mit konstanter Bandbreite ist Kerndichteschätzung

Mit konstanter Bandbreite ist die von Asvany et al. angewandte Methode zur Erstellung der CD-Spektren im Wesentlichen die Anwendung eines Kerndichteschätzers mit Kastenkerneln.¹⁴ Kerndich-

¹³Dies legt auch der Satz von Nadaraja (vergleiche z. B. [Wertz 1978], Corollary 4.4 auf Seite 38) nahe, der besagt, dass sich die Qualität einer Schätzung unter bestimmten Voraussetzungen, die hier gegeben sind, beliebig verbessern lässt, indem man immer mehr Daten dazu nimmt. Allerdings bezieht sich der Satz auf Einzelereignisse, in diesem Fall also Absorptionsprozesse einzelner Photonen, auf die ein Kerndichteschätzer (vergleiche Abschnitt 2.3.2) angewendet wird. Hier wird aber auf der Ebene nach Integration bereits bestimmter Linienpositionen gearbeitet, also mit vorverarbeiteten Daten, sodass dieser Satz nicht direkt anwendbar ist.

¹⁴Asvany et al. zählen für jede CD die benachbarten CDs im Bandbreitenfenster, eine Kerndichteschätzung zählt die CDs im Bandbreitenfenster. Das ergibt an der Stelle einer CD (und woanders berechnen Asvany et al. ja gar nichts) immer einen um 1 erhöhten Wert, der von der CD an der Stelle stammt, an der gerade gezählt wird. Deshalb ist das Ergebnis einer Kernschätzung einfach überall um 1 gegenüber dem Ergebnis von Asvany et al. nach oben verschoben. Zudem ist eine klassische Kerndichteschätzung gemäß der Wahrscheinlichkeitsinterpretation so normiert, dass $\int \varrho_{\text{geschätzt}}(E) de = 1$ gilt. Dies wird durch $\int K_b(x) dx = 1$ und den Vorfaktor $\frac{1}{n}$ sichergestellt. Bei Asvany et al. ist das Ergebnis willkürlich normiert. Der Unterschied ist aber lediglich eine Multiplikation des Gesamtergebnisses mit einem einfach zu ermittelnden konstanten Faktor.

teschätzer sind eine Weiterentwicklung von Histogrammen für Daten, die nicht nur diskrete, sondern alle möglichen Werte annehmen können.¹⁵ Sie sind ein in der Statistik seit den 1950er Jahren verwendetes, gut dokumentiertes Werkzeug zur Schätzung von Wahrscheinlichkeitsverteilungen an Hand von Zufallsexperimenten¹⁶ und werden auch in zahlreichen anderen Zusammenhängen, z. B. in der Signalverarbeitung zur Rauschunterdrückung verwendet.¹⁷

Dies legt folgende (Re-)Interpretation der CD-Spektren nahe: Das CD-Spektrum gibt die auf Grund der Messdaten geschätzte Wahrscheinlichkeit an, dass eine CD, die aus den Messwerten zweier zufällig heraus gegriffener Absorptionsprozesse berechnet wird, eine bestimmte Energie hat.

In [Asvany et al. 2015] werden die vertikalen Koordinaten im CD-Spektrum, die im Wesentlichen diesen geschätzten Wahrscheinlichkeiten entsprechen, nur für die Energien berechnet, bei denen im Experiment tatsächlich eine CD gemessen wurde. Selbstverständlich lässt sich aber für beliebige Energien schätzen, wie wahrscheinlich es ist, dass eine zufällig heraus gegriffene CD diese Energie hat. Vor allem für die schnelle visuelle Interpretation der CD-Spektren hat es sich im Verlauf dieser Arbeit bewährt, die Wahrscheinlichkeiten nicht nur für die Energien, bei denen CDs gemessen wurden, zu berechnen, sondern z. B. zusätzlich an 3 Zwischenpunkten.¹⁸

2.3.3 Zur Rekonstruktion der Vibrationsgrundzustände mittels temperaturgemischter CDs

Die Zuverlässigkeit der in Abschnitt 1.2.3 beschriebenen Methode von Asvany et al. zur Rekonstruktion nicht nur der Übergänge zwischen den Vibrationsgrundzuständen, sondern der Vibrationsgrundzustände selbst hängt stark davon ab, ob die kältere Messung bei einer so niedrigen Temperatur stattgefunden hat, dass tatsächlich in jedem Set nur der Zustand mit der jeweils niedrigsten Energie (und nicht mehrere Vibrationsgrundzustände) merklich besetzt ist. Versuche mit Testdaten haben gezeigt, dass dies meist der Fall ist, wenn die niedrigere Temperatur unter etwa 6 K liegt. Bei den Daten aus [Asvany et al. 2015] entspricht die Nominaltemperatur von 4 K zwar dieser Anforderung, die tatsächliche Temperatur von etwa 12,8 K ist aber zu hoch. Dies hat zur Folge, dass Peaks, die Differenzen zwischen Vibrationsgrundzuständen entsprechen, bei denen aber nicht der unterste Vibrationsgrundzustand involviert ist, also z. B. zwischen 2. unterstem und 5. unterstem Vibrationsgrundzustand, fälschlicherweise für Vibrationsgrundzustände gehalten werden.

Asvany et al. sind diesem Problem begegnet, indem sie die Ergebnisse zusätzlich einem Plausibilitätscheck auf Grund der in den CDs steckenden Linien unterzogen und falsche Ergebnisse verworfen haben. Im Laufe dieser Arbeit wurden diese Plausibilitätschecks zu einem Verfahren weiter entwickelt, das es allein auf Grund kombinatorischer Überlegungen und ohne eine zusätzliche Messung bei niedriger Temperatur ermöglicht, die Niveaus direkt zu rekonstruieren. Siehe

¹⁵Zum Zusammenhang zwischen Histogrammen und Kerndichteschätzungen siehe z. B. Abschnitt 3.1.2 von [Wolfgang et al. 2004].

¹⁶Lesenswerte Einführungen sind z. B. der erste Abschnitt von [Wertz 1978] oder die ersten beiden Kapitel von [Silverman 1986]. Bei Wirtz findet sich eine (inzwischen nicht mehr ganz vollständige) systematische Erläuterung der Anwendungsgebiete, Silvermann bringt stattdessen eine Reihe sehr konkreter Beispiele aus Natur-, Ingenieurs- und Wirtschaftswissenschaften sowie Medizin.

¹⁷Bekanntestes Beispiel ist der Gaußsche Weichzeichner in der Bildbearbeitung. Er bestimmt den Wert eines Pixels des weichgezeichneten Bildes als mit einem Gauß-Kern gewichteten Mittelwert aller Pixel des Originalbildes. Eine breite Sammlung moderner Anwendungen von Kernschätzungen in der Bild- und Signalverarbeitung findet sich im Buch von [Camps-Valls et al. 2007].

¹⁸Die naheliegende Lösung, die Wahrscheinlichkeiten für äquidistante Energien zu berechnen, hat sich nicht bewährt: Entweder sind die Abstände so groß, dass schmale Peaks mitunter übersprungen werden und nicht sichtbar sind, oder es müssen so viele Werte berechnet werden, dass Rechenaufwand und Speicherverbrauch unpraktisch werden.

Abschnitt 5.

2.4 Ermittlung statistischer Eigenschaften von Molekül und Messung

Sind die Parameter der Testumgebung so angepasst, dass sie ein Spektrum liefert, das zu den gemessenen Daten passt, können aus den Parametern Rückschlüsse über Molekül und Messvorgang gezogen werden:

- Gelingt der CutOff der Linien mit niedriger Intensität in Schritt 8 der Testdatenerzeugung für beide Temperaturen zuverlässig, hängt das resultierende Verhältnis der Linienanzahlen im Wesentlichen nur noch von den simulierten Temperaturen ab: Bis zum 5. Schritt der Erzeugung des Testspektrums werden beide Spektren gemeinsam temperaturunabhängig erzeugt. Nichts, was bis dahin eingeht, kann auf das Verhältnis der Linienanzahlen der beiden Spektren einen Einfluss haben. In den Schritten 6 und 7 werden die Linienintensitäten berechnet, die die Grundlage für den CutOff in Schritt 8 sind, durch den der Anzahlunterschied zu Stande kommt. Dabei gehen als einzige Parameter die jeweiligen Temperaturen ein. Dementsprechend ist ein Verhältnis der Linienanzahlen der beiden Spektren, das nicht zu den Messdaten passt, ein Anzeichen falsch gewählter Simulationstemperaturen. Da die Skalierung für die 4K-Spektren wie in Abschnitt 2.2 beschrieben nicht zuverlässig funktioniert und deshalb umgekehrt die Linienanzahl für den CutOff beim 4K-Spektrum heran gezogen wird (vergleiche Abschnitt 2.2), funktioniert dies derzeit aber nicht.
- Die Analyse des Graphen von $i \varrho_I(i)$ der realen Messdaten ermöglicht die Rekonstruktion der Wahrscheinlichkeitsverteilung der Übergangswahrscheinlichkeiten (vergleiche Gleichung 2.2.3): Die Steigung des Graphen ist bis auf den Vorfaktor $\frac{1}{df}$ die gesuchte Wahrscheinlichkeitsverteilung $g(x)$. $\frac{1}{df}$ lässt sich wie beschrieben an Hand des Verlaufs des Graphen für große Intensitäten ermitteln.

3 Optimierung des Kerndichteschätzers

Auch wenn das in [Asvany et al. 2015] verwendete Verfahren bei Verwendung einer konstanten Bandbreite die Anwendung eines Kerndichteschätzers ist, lassen sich viele der bekannten Erkenntnisse über Kerndichteschätzer nicht direkt anwenden. Die meisten dieser Erkenntnisse beziehen sich — soweit dem Autor bekannt — auf Situationen, in denen Kerndichteschätzer direkt auf die Daten von Zufallsexperimenten angewendet werden. Bei den vorliegenden Messungen sind diese Zufallsexperimente die Absorptionen einzelner Photonen. Diese Daten liegen aber nicht vor. Vielmehr wurden die Daten über die Dauer der Messung integriert und anschließend die Lage der einzelnen Linien bestimmt. Dem Autor sind keine Ergebnisse darüber bekannt, wie sich die nicht direkt anwendbaren Erkenntnisse über Kerndichteschätzer sinnvoll auf derart vorverarbeitete Daten übertragen lassen. Dies selbst zu erarbeiten, sprengt aber den Rahmen dieser Arbeit.

Angesichts dessen werden hier lediglich sehr grundlegende Eigenschaften von Kerndichteschätzern genutzt, die aber bereits eine weitreichende Optimierung der Kerndichteschätzung ermöglichen.

3.1 Gauß- statt Kastenkernel

Die Berechnung der geschätzten Wahrscheinlichkeiten im CD-Spektrum mit Hilfe eines Kerndichteschätzers funktioniert gemäß folgender Formel:¹

$$\varrho_{\text{geschätzt}}(E) = \frac{1}{n} \sum_i K_b(E - CD_i) \quad (3.1.1)$$

Dabei sind CD_i die gemessenen CDs, n die Anzahl der CDs und K_b eine Kernfunktion mit der Bandbreite b ; E ist die Energie, für die die Wahrscheinlichkeitsdichte geschätzt wird. Als Kern sind beliebige Funktionen mit $K_b : \mathbb{R} \rightarrow [0, \infty)$ und $\int K_b(x) dx = 1$ verwendbar. Eine Übersicht verbreiteter Kerne zeigt Abbildung 3.1.1.

Kerne unterscheiden sich vor allem hinsichtlich dreier Kriterien: ihrer Glattheit, ihrer Symmetrie und ihrer Effizienz. Die Effizienz ist ein Maß für das Verhältnis von verwendeter Datenmenge und der aufintegrierten quadratischen Differenz zwischen Schätzung und tatsächlicher Wahrscheinlichkeitsverteilung (*AMISE*). Wenn keine weiteren Informationen über die zu schätzende Verteilung vorliegen, ist der Epanechnikov-Kernel am effektivsten, die verschiedenen Kerne unterscheiden sich aber nur geringfügig.² Vor allem am Rand des Datensatzes neigen Kerndichteschätzer zu einem Bias, also einer systematischen Verfälschung der Werte. Um diesen Effekt so gering wie möglich zu halten, verwendet man Kerne mit möglichst guten Symmetrieeigenschaften.³ Solche Verfälschungen bedeuteten im vorliegenden Anwendungsfall leicht verschobene CD-Peaks. Die Größenordnungen eines solchen Bias liegt aber deutlich unter der Bandbreite und ist damit im vorliegenden Fall ebenfalls nicht entscheidend. Die Glattheit des Kerns spielt vor

¹Zur Skalierung siehe auch Fußnote 14 in Abschnitt 2.3.2.

²Vergleiche Abschnitt 3.1.2 in [Simonoff 1996].

³Vergleiche die Diskussion in Abschnitt 2.8 von [Wand et al. 1995] und speziell zur Frage des *boundary bias* Abschnitt 3.2.1 in [Simonoff 1996].

3 Optimierung des Kerndichteschätzers

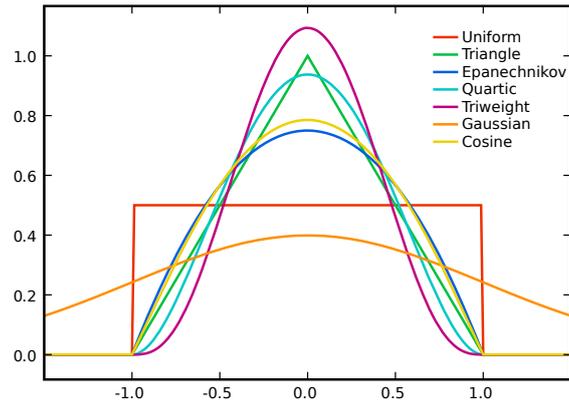


Abbildung 3.1.1: **Verschiedene Kernfunktionen**

Der Kastenkernel ist hier mit „Uniform“ bezeichnet.

Quelle: [Wikipedia 2008]

allein für die Robustheit der Schätzung eine große Rolle: Kerne mit geringer Glattheit führen insbesondere, wenn die zu schätzende Verteilung ebenfalls nicht besonders glatt ist und die Anzahl der Datenpunkte innerhalb des Bandbreitenfensters so gering ist, dass sie großen statistischen Schwankungen unterliegt, dazu, dass die geschätzte Verteilung sehr empfindlich von einzelnen Parametern abhängt.⁴ Dies ist auch bei der von [Asvany et al. 2015] verwendeten Methode zu beobachten, die mit einem Kastenkernel, dem am wenigsten glatten Kern unter den sinnvoll einsetzbaren, arbeitet. Zudem besteht die zu rekonstruierende Verteilung aus Peaks, ist also auch nicht besonders glatt.

Unter diesen Bedingungen entscheiden schon minimale Änderungen an der Bandbreite oder das Hinzufügen / Weglassen einzelner CDs darüber, ob bestimmte CD-Peaks deutlich hervor stechen oder im Rauschen untergehen. (Vergleiche die Diskussion der CD-Peaks von CH_5^+ bei 7.04935 cm^{-1} , 36.91995 cm^{-1} und 46.46478 cm^{-1} im Abschnitt 7.2.) Dies kann man sich folgendermaßen verplausibilisieren: Für einen Kasten-Kern wird Gleichung 3.1.1 zu:

$$\varrho_{\text{geschätzt}}(E) = \frac{1}{n} \sum_i \frac{1}{b} \mathbf{1}_{|E - CD_i| < \frac{b}{2}}$$

Das heißt, es wird (bis auf Vorfaktoren) für jede Energie E gezählt, wie viele CDs in einem Fenster der Breite b um E liegen. Liegt nun eine CD in der Nähe des Fensterrandes, reichen kleine Änderungen der Parameter dafür aus, dass sie einmal mitgezählt wird und einmal nicht, obwohl eigentlich nur ein sehr geringer Unterschied besteht.

Um zu einer höheren Robustheit der Ergebnisse zu kommen, bietet es sich deshalb an, mit einem glatteren Kern zu arbeiten. Der Gaußkern ist maximal glatt⁵ und symmetrisch. Seine Effizienz liegt zwischen (und sehr dicht bei) den Effizienzen von Epanechnikov- und Kasten-Kern.⁶ Deshalb wird im Folgenden ein Gaußkern verwendet, was die Ergebnisse gegenüber dem von Asvany et al. verwendeten Kastenkernel ein wenig verbessert, sie aber vor allem robuster macht.

⁴[Simonoff 1996] diskutiert diesen Effekt in Abschnitt 2.4 sehr instruktiv für Histogramme und einfache (Kastenkernel-)Schätzer.

⁵Der maximal effiziente Epanechnikov-Kern ist zwar auch hoch symmetrisch, aber nicht so glatt.

⁶Vergleiche Tabelle 3.1 in [Simonoff 1996].

3.2 Wahl der Bandbreite

Die Wahl der Bandbreite ist entscheidend für die Qualität einer Kerndichteschätzung. Typische Auswirkungen einer zu kleinen, passenden und zu großen Bandbreite zeigen die Abbildungen 3.2.1 und 3.2.2. Zur Wahl der Bandbreite von Kerndichteschätzern gibt es umfangreiche Literatur⁷ aber kein einfaches Rezept. (Es sei denn, es gibt eine offensichtlich sinnvolle Wahl, wie es der Fall wäre, wenn sich beide in Abschnitt 2.3.1 beschriebenen Kriterien gleichzeitig erfüllen ließen.)

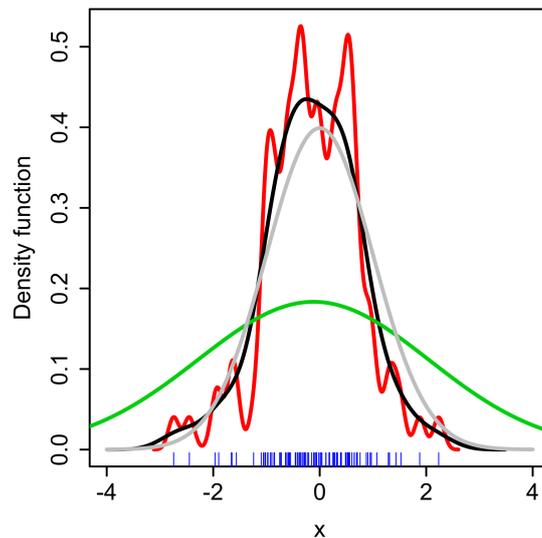


Abbildung 3.2.1: **Wahl der Kernbandbreite**

grau: zu schätzende Wahrscheinlichkeitsdichte. **blau:** Ergebnisse eines Zufallsexperimentes, das dieser Wahrscheinlichkeitsverteilung unterliegt. **rot:** Kerndichteschätzung mit zu kleiner Bandbreite: Das Ergebnis ist stark verrauscht und der Peak zerfällt in mehrere Einzelpeaks. **grün:** Kerndichteschätzung mit zu großer Bandbreite: Die geschätzte Wahrscheinlichkeitsverteilung ist zu stark geglättet, Details gehen verloren. **schwarz:** Schätzung mit passender Bandbreite.

Quelle: [Wikipedia 2010]

Prinzipiell gibt es zwei verschiedene Ansätze:

1. Unter bestimmten Voraussetzungen gilt für die optimale Bandbreite eines Kerndichteschätzers $\frac{c}{n^a}$, wobei n die Anzahl der Datensätze ist und c und a Zahlen, die von der zu schätzenden Dichteverteilung abhängen.⁸ Es existieren automatische Verfahren, die in mehreren Schätzdurchläufen auf Grund von Heuristiken c und a iterativ immer besser wählen. Sie sind in den meisten Statistik-Software-Paketen implementiert.⁹
2. Wahl der Bandbreite auf Grund von Tests mit Daten gleichen Typs, für die das Ergebnis bekannt ist.

⁷Einen aktuellen Überblick findet man etwa bei [Zambom et al. 2012] in Kapitel 4.

⁸Für eine Begründung dieses Ansatzes zur Bandbreitenwahl siehe beispielsweise die Erläuterungen zu Theorem 4.1 in [Wertz 1978].

⁹[Guidoum 2014] dokumentiert z. B. die Implementierung in GNU R.

3 Optimierung des Kerndichteschätzers

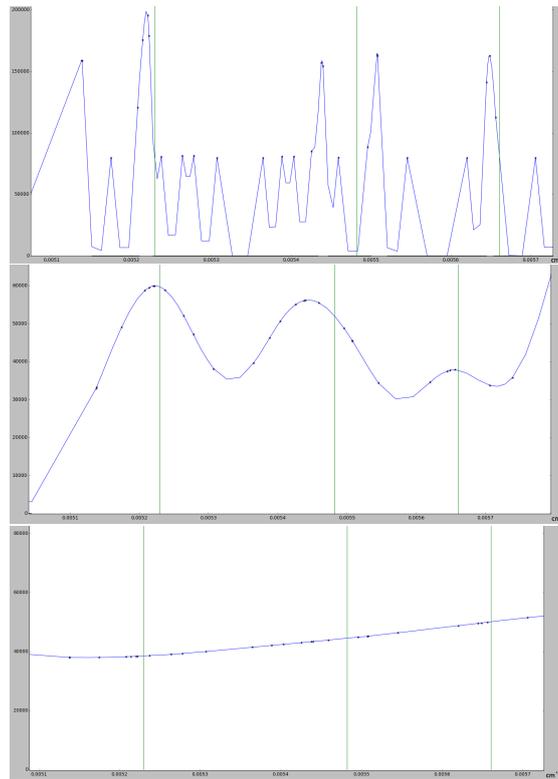


Abbildung 3.2.2: **Kerndichteschätzungen mit Testdaten**
oben: zu kleine Bandbreite ($0,000005 \text{ cm}^{-1}$). **Mitte:** geeignete Bandbreite ($0,00005 \text{ cm}^{-1}$). **unten:** zu große Bandbreite ($0,00025 \text{ cm}^{-1}$). **grün:** Lage der zu rekonstruierenden Peaks.

Beim ersten Ansatz werden fast immer Kenntnisse über Differenzierbarkeitseigenschaften der zu schätzenden Wahrscheinlichkeitsdichten bzw. zu rekonstruierenden Signale vorausgesetzt. Diese Verfahren funktionieren in der Regel umso besser, je glatter die zu schätzende Verteilung bzw. das zu rekonstruierende Signal ist. Im vorliegenden Fall ist die zu schätzende Wahrscheinlichkeitsverteilung eine Summe von Gaußfunktionen.¹⁰ Damit ist sie eigentlich für die automatischen Verfahren zur Bandbreitenbestimmung ideal geeignet. Allerdings sind die meisten dieser Verfahren für die Verwendung mit Ergebnissen einzelner Zufallsexperimente ausgelegt, und nicht für — wie oben erläutert — vorverarbeitete Messdaten.

Im Gegensatz dazu kann der zweite Ansatz sehr einfach angewendet werden, weil mit Hilfe der Testumgebung einfach Daten gleichen Typs erzeugt werden können. Die bei dieser Methode oft gegebene Gefahr der Überanpassung an die Testdaten¹¹ besteht dabei nicht, weil die Wahl der

¹⁰Die zu rekonstruierende Wahrscheinlichkeitsdichte gehört zur Summe zweier Zufallsvariablen. Eine der beiden Zufallsvariablen ist die Energie der CDs des Moleküls selbst und ihre Wahrscheinlichkeitsverteilung ist auf Grund der Energie-Quantelung eine Summe von Delta-Distributionen. Die andere Zufallsvariable ist der Fehler des Messprozesses. Geht man wiederum davon aus, dass der Fehler des Messprozesses sich aus mehreren Fehlerquellen mit endlicher Varianz zusammensetzt, ist die Fehler-Zufallsvariable Gauß-verteilt (Zentraler Grenzwertsatz). Die Gesamtwahrscheinlichkeitsdichte ergibt sich daraus als Summe von Gaußfunktionen.

¹¹Wird ein Auswertungsverfahren an Testdaten optimiert, versteht man unter Überanpassung eine Wahl der Parameter, die zwar die Fehler des Verfahrens bei Anwendung auf den Testdatensatz reduziert, das gesamte

3 Optimierung des Kerndichteschätzers

Bandbreite nicht auf Grund eines einzigen Datensatzes geschieht, sondern mit der Testumgebung immer neue, gleichartige aber verschiedene Datensätze erzeugt werden können.

Bewährt hat sich, die Bandbreite möglichst so zu wählen, dass sie gerade oberhalb des niedrigsten Drittels der CD-Abstände liegt, aber auf keinen Fall kleiner ist als zwei Drittel der CD-Messfehler. Im Zweifel hat letzteres Kriterium Vorrang.¹² Gemäß dieser Methode wählt das entwickelte Programm automatisch eine passende Bandbreite an Hand der zu verarbeitenden Daten. Dies zeigt exemplarisch Abbildung 3.2.3.

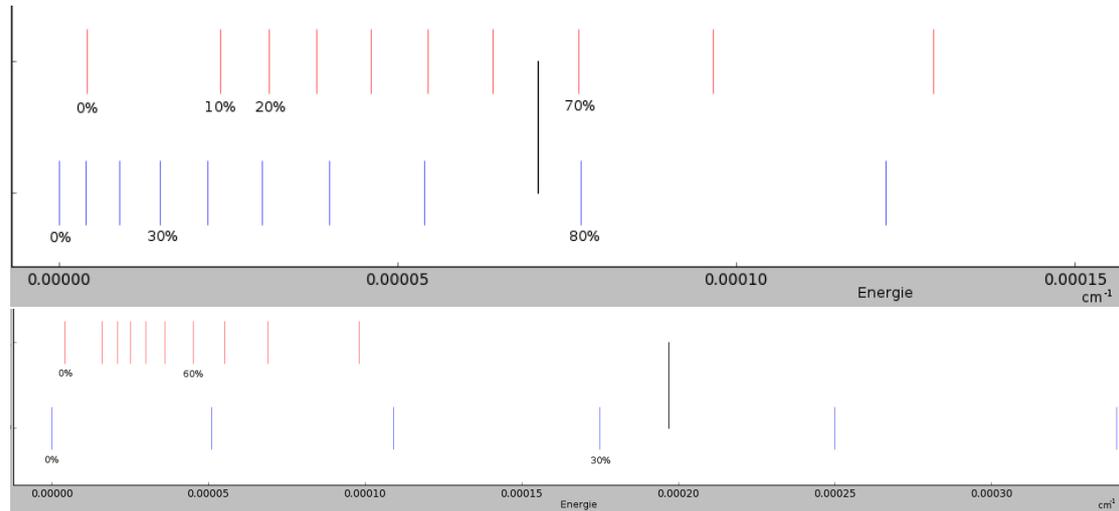


Abbildung 3.2.3: **Verteilung von CD-Messfehlern und CD-Abständen**
rot: Fehler. **blau:** CD-Abstände. **schwarz:** automatisch gewählte Bandbreite
oben: 10 K-Spektrum. **unten:** kombiniertes 4 K-10 K-Spektrum.

3.3 Weitere Optimierungsmöglichkeiten

Im Wesentlichen kann die Nutzung von Kerndichteschätzern durch geschickte Wahl von Kern und Bandbreite optimiert werden. Die in dieser Arbeit verwendeten Kombinationen haben sich insofern bewährt, als sich an den Testdaten gezeigt hat, dass kleinere Änderungen der Bandbreite oder die Verwendung anderer symmetrischer und relativ glatter Kerne die Ergebnisse nur wenig verändern. Deshalb wurde der Kernschätzer im Rahmen dieser Arbeit nicht weiter optimiert, zumal — wie zu Beginn dieses Teils der Arbeit erläutert — existierende Kenntnisse über Kernschätzer nicht direkt angewendet werden können.

Will man den Kernschätzer dennoch weiter optimieren, gibt es prinzipiell vier Ansätze:

Verfahren dafür aber so sehr auf den spezifischen Testdatensatz zuschneidet, dass es im Allgemeinen schlechter wird. Die Wahl der Bandbreite ist ein Kompromiss zwischen geringem Bias (genauer Schätzung), aber hoher Varianz (starker Schwankung) bei niedriger Bandbreite und großem Bias, aber geringer Varianz bei hoher Bandbreite (*bias-variance tradeoff*). Eine Überanpassung bestünde in diesem Fall darin, dass die gewählte Bandbreite nicht dem im Allgemeinen günstigsten Kompromiss entspräche, sondern z. B. größer wäre, dabei aber bei einem speziellen Testdatensatz, an Hand dessen sie gewählt wurde, zufällig ein sehr gutes Ergebnis liefert, weil die übermäßig geglättete Kurve zufällig genau die Datenpunkte trifft.

¹²Dies gilt für den hier verwendeten Gauß-Kern. Für andere Kernel müsste die Bandbreite gemäß Tabelle 3.2 in [Simonoff 1996] angepasst werden.

3 Optimierung des Kerndichteschätzers

- Verwendung einer variablen Bandbreite die die stark schwankende lokale Dichte des CD-Spektrums berücksichtigt. Einen Überblick über Ansätze für die variable Wahl der Bandbreite von Kernschätzern findet sich in Abschnitt 4.4.1 von [Zambom et al. 2012].
- Nutzung von *Significance Zero Crossings*, einer modernen Weiterentwicklung der Kerndichteschätzung, die darauf beruht, mehrere Kerndichteschätzungen bei verschiedenen Bandbreiten durchzuführen und die signifikanten Eigenschaften der Daten aus den verschiedenen Schätzungen zusammen zu sammeln. Für eine kurze Erläuterung der Methode und weitergehende Literaturhinweise siehe [Zambom et al. 2012], Kapitel 5.
- Übertragung der Kenntnisse über Kernschätzer auf vorverarbeitete Daten, also in diesem Fall Lage und Fehler der Absorptionslinien, die aus den Einzelphotonen-Absorptionen ermittelt wurden. Dies wäre auch ein vielversprechender Weg, um die Informationen über die Messfehler (und Intensitäten der Linien, wenn sie denn verlässlicher gemessen wären) sinnvoll nutzen zu können, anstatt sie wie in dieser Arbeit zu ignorieren (vergleiche die Diskussion um konstante versus variable Bandbreite in Abschnitt 2.3.1.1).
- Wäre bei der Messung die Integrationszeit für jede vermessene Energie gleich gewesen oder protokolliert worden und wären die Intensitätsmessungen zuverlässig, könnte prinzipiell auch mit den Rohdaten gearbeitet werden. Die für die einzelnen Energien gemessenen Intensitäten wären dann (ggf. nach Gewichtung mit den jeweiligen Integrationszeiten und Empfindlichkeiten der Messapparatur im jeweiligen Wellenlängenbereich) proportional zur Anzahl der Photonenabsorptionen mit der entsprechenden Energie. Dann könnten sämtliche über Kernschätzer bekannten Ergebnisse direkt angewendet werden. Dabei wären tendenziell auch insgesamt bessere Ergebnisse zu erwarten, weil durch die Vorverarbeitung schon Informationen verloren gehen. Es ergäbe sich allerdings ein entscheidendes praktisches Problem: Die Anzahl dieser Ereignisse wäre um Größenordnungen höher als die Anzahl der jetzt bearbeiteten Linien. Da die Anzahl der CDs quadratisch von der Linienanzahl abhängt, wäre der Rechenaufwand immens. (Die im Rahmen dieser Arbeit verwendeten Scripte sind sicher nicht besonders gut auf Laufzeit optimiert; dennoch sind bereits für die Arbeit mit Linien (statt Einzelphotonen-Ereignissen) zahlreiche Optimierungen eingearbeitet worden, um auf einem Standard-PC eine Rechenzeit, die im Bereich einiger Minuten statt eines Tages liegt, realisieren zu können.)

4 Peakerkennung

Wie bereits im Arbeitsprogramm erläutert (Abschnitt 1.3) wurde im Rahmen dieser Diplomarbeit eine automatische Peakerkennung entwickelt. Die einfachste Form einer Peakerkennung besteht darin, Stellen, an denen das CD-Spektrum eine bestimmte Schwellenwert-Höhe überschreitet, als Peak zu identifizieren. Da die optimale Bandbreite des Kernschätzers wie in Abschnitt 2.3.1 diskutiert größer als der mittlere Abstand zweier benachbarter CDs ist, besteht dabei die Gefahr falscher Peaks. Nutzt man allein eine Peakerkennung auf Grund der Höhe im CD-Spektrum, muss man deshalb die Schwellenwert-Höhe sicherheitshalber verhältnismäßig groß wählen, sodass auch viele richtige Peaks nicht erkannt werden.

Eine naheliegende Alternative ist, die Form der Peaks zu untersuchen: Ein echter Peak sollte in etwa die Form einer Gauß-Funktion haben.¹ Ein falscher Peak, der in Wirklichkeit von mehreren, nahe beieinander liegenden esCDs oder aCDs herrührt, hat dagegen die Form einer Summe von mehreren, nahe beieinander liegenden Gauß-Funktionen. Allerdings gibt es bei der Analyse der Peakform mindestens drei Schwierigkeiten:

1. Die fraglichen (und deshalb interessanten) Peaks bestehen aus einer verhältnismäßig geringen Anzahl von CDs, sodass die Form relativ großen statistischen Schwankungen unterliegt.
2. Wie in Abschnitt 2.3.1 erläutert macht es Sinn, die Bandbreite des Kernschätzers verhältnismäßig groß zu wählen. Dadurch spielt zwar Punkt 1 eine nicht ganz so große Rolle, gleichzeitig werden aber Abweichungen von der Gauß-Form der CD-Peaks, also das Erkennungsmerkmal falscher Peaks, weitgehend weg geglättet.
3. Eine robuste automatische Bewertung der Peakform ist nicht einfach zu programmieren.

Angesichts dessen wurde die Analyse der Peakform nicht weiter verfolgt und stattdessen das Skalierungsverhalten der Peaks bei Veränderung der Anzahl einbezogener CDs genutzt. Dazu wird vor der herkömmlichen Peakerkennung über einen Höhen-Schwellenwert die in Abschnitt 4.2 erläuterte Methode zur Rauschunterdrückung angewendet.

4.1 Skalierungsverhalten der CD-Spektren

Normiert man den Kernschätzer nur mit seiner Bandbreite, nicht aber mit der Anzahl verwendeter Datenpunkte n , so weisen die Peakhöhen ein charakteristisches Skalierungsverhalten auf, wenn man die Anzahl der verwendeten CDs n im Datensatz variiert. (Siehe Gleichung 3.1.1.) Diese Variation wird im entwickelten Auswertungsscript dadurch erreicht, dass vor der Kernschätzung ein bestimmter Anteil der CDs zufällig aussortiert wird. Dabei ergibt sich folgendes Verhalten² (vergleiche Abbildung 4.1.1):

¹Dass die Peaks überhaupt eine Breite haben, liegt im Wesentlichen an den Messfehlern. Da im betrachteten Experiment die Messfehler aus mehreren einzelnen Fehlerquellen (mit endlicher Varianz) stammen, ist ihre Summe, also der Gesamtmeßfehler, normalverteilt (Zentraler Grenzwertsatz).

²Die Punkte 1 und 3 lassen sich heuristisch an Hand von Gleichung 3.1.1 begründen. Vor allem aber wurde dieses Verhalten an Hand von Testdaten untersucht.

4 Peakerkennung

1. **Echte Peaks** Gemäß Gleichung 3.1.1 ist die Höhe des Peaks proportional zur Anzahl berücksichtigter CDs.
2. **Zusammengesetzte Peaks** Besteht ein großer Peak in Wirklichkeit aus mehreren kleineren Peaks, gilt dies im Schnitt auch, allerdings mit einer anderen Proportionalitätskonstanten. Da sich die CDs quasi auf die Einzelpeaks aufteilen, ist die Geschwindigkeit, mit der die Peakhöhe bei steigender Anzahl berücksichtigter CDs wächst, geringer. (Im Wesentlichen ist diese Geschwindigkeit proportional zum Kehrwert der Anzahl der im zusammengesetzten Peak steckenden echten Peaks.) Umgekehrt zerfallen zusammengesetzte Peaks in die einzelnen Peaks, wenn man die Anzahl der verwendeten CDs zunehmend reduziert.
3. **Rauschen** Variiert man die CD-Anzahl, so ändert sich die Höhe des Rauschens nicht. Stattdessen variiert seine Dichte, weil zusätzliche CDs in aller Regel in eine Lücke zwischen andere aCDs, von denen das Rauschen ja kommt, fallen.

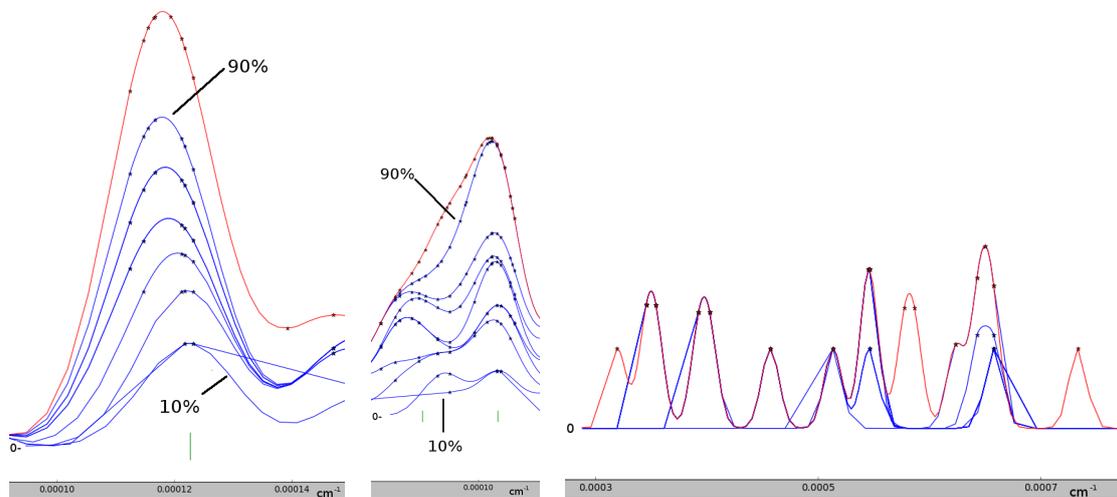


Abbildung 4.1.1: **Skalierungsverhalten des CD-Spektrums**

rot: Schätzung mit vollem Datensatz. **blau:** Schätzungen mit 10%, 20%, ... der CDs. **grün:** Lage echter Peaks.

links: echter Peak

Mitte: aus zwei echten Peaks zusammengesetzter Peak

rechts: Rauschen

Um das Skalierungsverhalten zur Peakerkennung zu nutzen, wird eine Kernschätzung mit vollständigem Datensatz mit Schätzungen mit reduziertem Datensatz verglichen. Damit der Effekt groß genug ist, müssen zur Erzeugung der reduzierten Datensätze bis zu 90% der CDs zufällig verworfen werden. Dies hat zur Folge, dass nur noch wenige CDs pro Peak übrig bleiben, sodass statistische Schwankungen eine große Rolle spielen und das Ergebnis stark davon abhängt, welche CDs genau verworfen wurden. Deshalb ist es notwendig, mit 5–40 reduzierten Datensätzen zu arbeiten. Die Ergebnisse der verschiedenen Durchläufe dürfen dabei nicht arithmetisch gemittelt werden, denn sonst ergäbe sich im Wesentlichen wieder das Ergebnis des Durchlaufs mit komplettem Datensatz.

Da es um das Skalierungsverhalten geht, wird zur graphischen Analyse für die jeweils untersuchte Stelle des CD-Spektrums jeweils der Quotient aus Schätzung mit vollem und einer der

4 Peakerkennung

Schätzungen mit reduziertem Datensatz gebildet:

$$\text{Quotient}_i = \frac{\text{Kernschätzung mit vollem Datensatz}}{\text{Kernschätzung mit reduziertem Datensatz}_i} \quad (4.1.1)$$

Der obere Teil von Abbildung 4.1.2 zeigt für zwei idealtypische Stellen eines CD-Plots jeweils das Histogramm dieser Quotienten. Im unteren Teil der Abbildung sind zahlreiche solcher Histogramme in einem Diagramm zusammen gefasst.

Rauschen macht sich dadurch bemerkbar, dass die Kernschätzungen mit reduziertem Datensatz mit großer Wahrscheinlichkeit den Wert 0 haben und der Quotient damit extrem groß wird. Extrem große Quotienten (hier > 20) wurden im Histogramm bei -1 einsortiert.

Auch wenn das Skalierungsverhalten in den Histogrammen zunächst offensichtlich zu sein scheint, ist das zuverlässige (automatische) auffinden von Peaks mit dieser Methode nicht trivial.

4.2 Rauschunterdrückung mit Hilfe des Skalierungsverhaltens

Relativ einfach lässt sich das Skalierungsverhalten des Rauschens nutzen, um vor der eigentlichen Peakerkennung einen Großteil des Rauschens, das durch die aCDs entsteht, zu eliminieren, so dass die eigentliche Peakerkennung, die die Peakhöhe nutzt, wesentlich empfindlicher eingestellt werden kann: Da die aCDs nebeneinander liegen, ist das Rauschen bei Schätzerläufen mit reduzierten Daten gleich hoch, aber weniger dicht. Es fallen einzelne Stachel des Rauschens weg, die verbleibenden behalten ihre Höhe (vergleiche Abbildung 4.1.1 rechts). Bei verschiedenen Durchläufen, bei denen jeweils unterschiedliche CDs weggelassen werden, fallen unterschiedliche Stachel des Rauschens weg. Bildet man die Enveloppe von genügend vielen solcher Durchläufe, erhält man deshalb wieder alle Stachel, also das komplette Rauschen zurück.

Im Gegensatz dazu unterschieden sich die Ergebnisse verschiedener Durchläufe mit reduzierten Datensätzen bei Peaks kaum voneinander und liegen wesentlich unter dem Ergebnis mit komplettem Datensatz, sodass dies auch für ihre Enveloppe gilt (Abbildung 4.1.1 links).

Parameter	10 K-CD-Spektrum	kombiniertes 4 K-10 K-CD-Spektrum
Rauschunterdrückung		
# Durchläufe	10	10
Erwartungswert der Sprungweite	10	5
Erwartungswert der Clustergröße	7	2,5
für Enveloppe verwendete Percentile	90	80
Peaksuche durch Höhenanalyse		
Höhen-Schwellenwert	1,8/Bandbreite bis 3,0/Bandbreite	1,5/Bandbreite

Tabelle 4.2.1: Parameter der Peakerkennung

Dies führt dazu, dass das Rauschen weitgehend verschwindet, wenn vom Ergebnis mit vollem Datensatz die Enveloppe mehrerer Durchläufe mit reduzierten Datensätzen abgezogen wird, wohingegen die Peaks nur minimal flacher werden, wie Abbildung 4.2.1 zeigt.

Damit das Verfahren auch bei dichten Spektren zuverlässig funktioniert, sind zwei Details wichtig:

4 Peakerkennung

- Bei der Berechnung der Enveloppe wird an den einzelnen Stellen des CD-Spektrums jeweils nur ein bestimmtes unteres Percentil verwendet, sodass Ausreißer nach oben nicht berücksichtigt werden. So wird das Verfahren robuster gegenüber einer übermäßigen Abflachung der Pekas auf Grund einzelner statistischer Ausreißer. (Man kann statt zufälligem Verwerfen auch Verwerfen nach bestimmten Mustern verwenden, die Ergebnisse sind dabei aber nicht so gut.)
- Bei optimaler Bandbreite des Kernschätzers liegen auch die aCDs nicht einzeln nebeneinander, sondern verschmelzen. Deshalb wird beim Verwerfen der CDs folgendermaßen vorgegangen: Es wird immer abwechselnd eine bestimmte Anzahl an CDs übersprungen und ein Cluster mit einer bestimmten Anzahl an CDs behalten. Die Größe sowohl der Sprünge als auch der Cluster wird dabei jedes Mal zufällig zu gegebenem Erwartungswert festgelegt.³ Die bei diesem Verfahren verwendeten Parameter finden sich in Tabelle 4.2.1.

4.3 Weitere Optimierungsmöglichkeiten

Im Wesentlichen gibt es bei der Peak-Erkennung drei Verbesserungsmöglichkeiten:

- (Halb-)automatische / systematische Bestimmung der Parameter, die bislang nur durch Ausprobieren bestimmt wurden
- Verwendung des Skalierungsverhaltens über die Rauschunterdrückung hinaus
- Verwendung der Peakform

³gleichverteilt zwischen 0 und dem Doppelten des Erwartungswertes

4 Peakerkennung

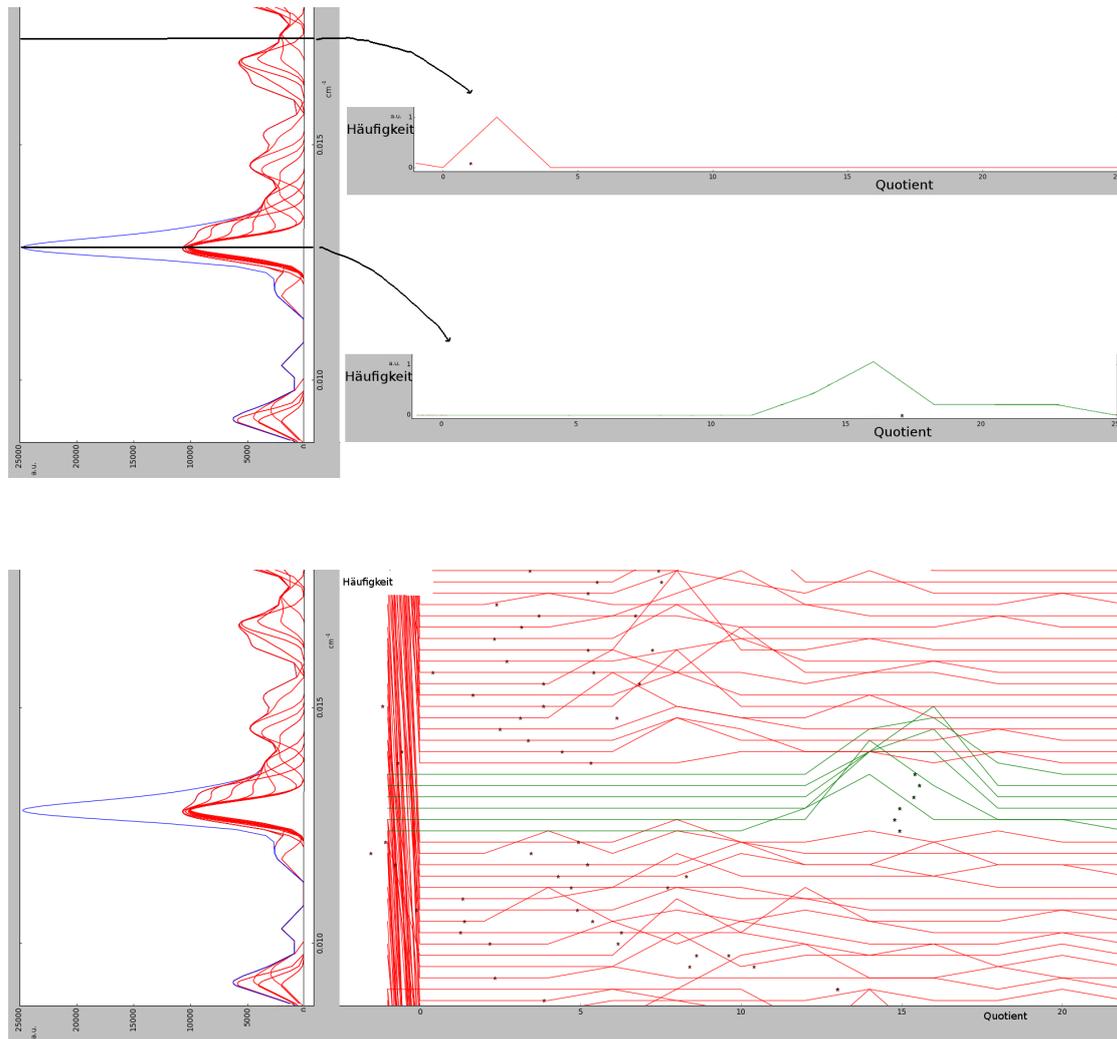


Abbildung 4.1.2: **Histogramm-Analyse des Skalierungsverhaltens der Kernschätzungen**

jeweils links hochkant: CD-Spektrum. In blau Schätzung mit vollständigem Datensatz, in rot Schätzungen mit jeweils 10% der Daten.

jeweils links: Histogramme der Quotienten (Gleichung 4.1.1) aus den verschiedenen Durchläufen. Die Schwerpunkte der Histogramme (einmal mit und einmal ohne -1 -Bin) sind jeweils mit einem Stern gekennzeichnet.

oben: Histogramme der Quotienten von Schätzung mit vollem Datensatz und Schätzungen mit reduziertem Datensatz an einer Stelle des CD-Spektrums mit Rauschen (rot) und einer Stelle des CD-Spektrums mit einem Peak (grün).

unten: zahlreiche solcher Diagramme für verschiedene Stellen des CD-Spektrums. Sie sind vertikal so verschoben, dass ihre jeweilige Nulllinie auf gleicher Höhe ist wie die zugehörige Stelle im Hochkant-CD-Spektrum links daneben.

4 Peakerkennung

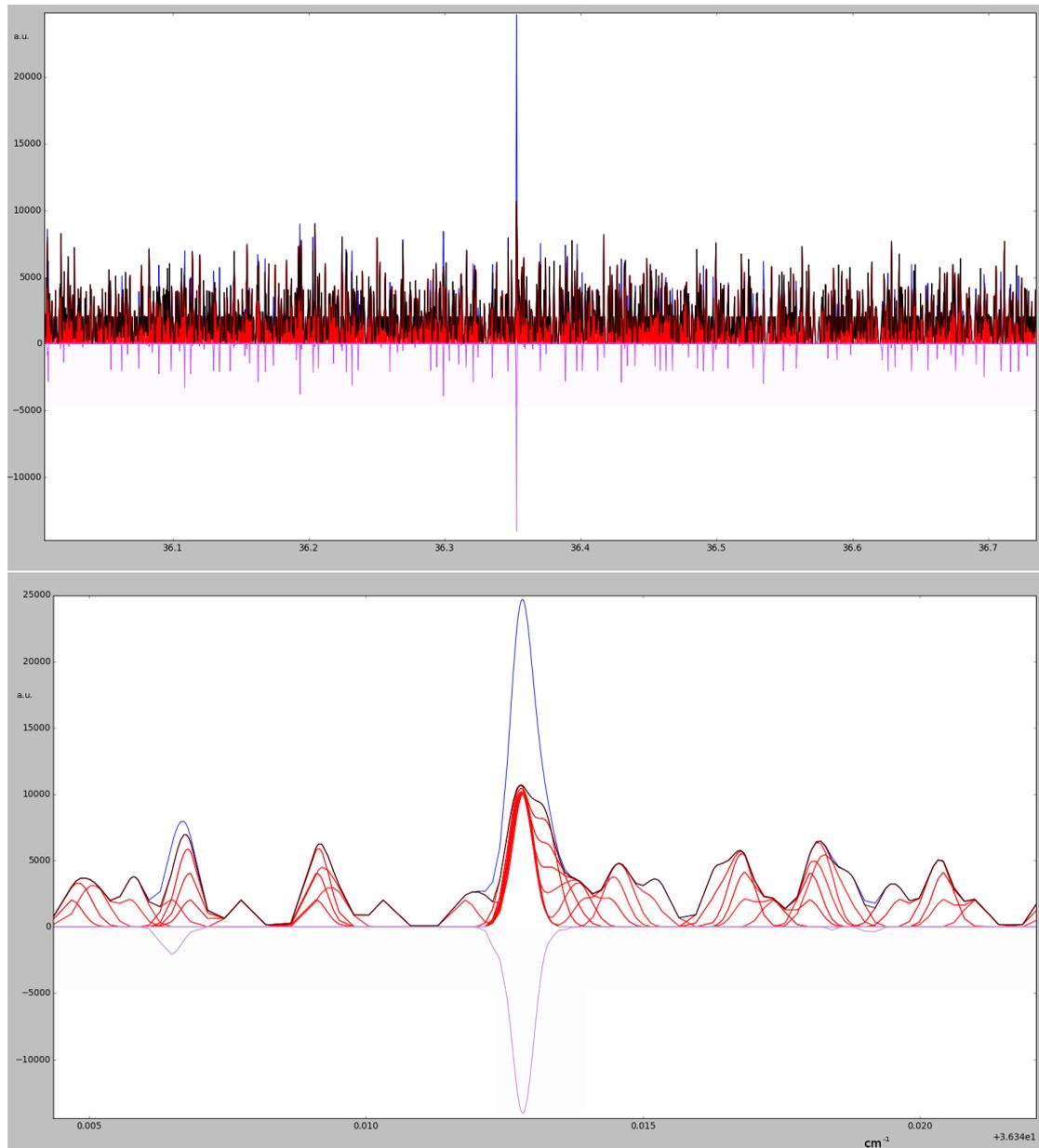


Abbildung 4.2.1: **Rauschunterdrückung mit Enveloppen-Methode**
rot: reduzierte Durchläufe. **blau:** Durchlauf mit vollständigem Datensatz.
schwarz: Enveloppe der Ergebnisse der reduzierten Durchläufe. **violett:** Differenz zwischen Durchlauf mit vollständigem Datensatz und Enveloppe
unten: Ausschnittvergrößerung

5 Von Linien über CDs zu Zuständen

In diesem Abschnitt wird eine während dieser Diplomarbeit erarbeitete Methode vorgestellt, mit deren Hilfe sich aus den Peaks im CD-Spektrum ein Teil der Energieniveaus rekonstruieren lässt. Anders als bei der in [Asvany et al. 2015] verwendeten Methode (vergleiche Abschnitt 1.2.3) sind dabei nicht Messungen bei verschiedenen Temperaturen nötig; vielmehr reicht eine Messung bei einer Temperatur, bei der möglichst viele Vibrationsgrundzustände, aber keine vibrationsangeregten Zustände besetzt sind. Die Messung aus [Asvany et al. 2015] mit einer Nominaltemperatur von 10 K (und einer realen Temperatur von etwa 25 K) erfüllt genau dieses Kriterium und wird im Folgenden ausschließlich verwendet.

5.1 CD-Peaks haben verbundene Beinchen

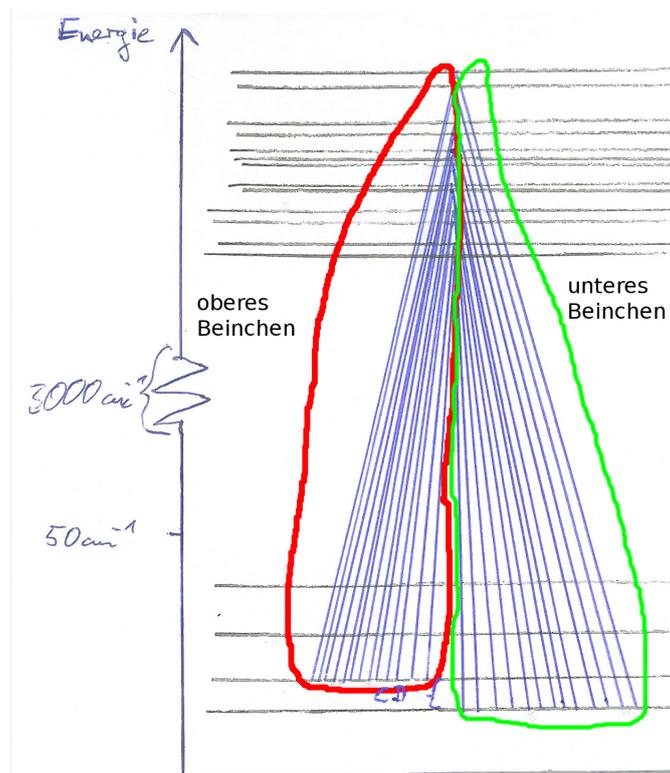


Abbildung 5.1.1: Beinchen eines CD-Peaks

Wie in Abschnitt 1.2.1 erläutert, bestehen die CD-Peaks aus gsCDs. Die Linien, aus denen die CDs des Peaks berechnet wurden, lassen sich in zwei Gruppen, im Folgenden Beinchen des Peaks

5 Von Linien über CDs zu Zuständen

genannt, einteilen (vergleiche Abbildung 5.1.1): Die Energie eines CD-Peaks entspricht ja — wie im Abschnitt 1.2.1.2 erläutert — einem Übergang zwischen zwei Vibrationsgrundzuständen. Die Linien, die den oberen der involvierten Vibrationsgrundzustände mit den vibrationsangeregten Zuständen verbinden, bilden das obere Beinchen des Peaks. Die Linien, die den unteren involvierten Vibrationsgrundzustand mit den vibrationsangeregten Zuständen verbinden, bilden das untere Beinchen.

Zwei verschiedene CD-Peaks können entweder einen oder keinen gemeinsamen Vibrationsgrundzustand haben. (Hätten sie zwei gemeinsame Vibrationsgrundzustände, wären sie ja nicht verschieden.) Geht man zunächst davon aus, dass es keine Auswahlregeln gibt, gilt: Haben zwei CD-Peaks einen gemeinsamen Vibrationsgrundzustand, bestehen die jeweils zugehörigen Beinchen aus denselben Linien. Beinchen, die zu verschiedenen Vibrationsgrundzuständen gehören, haben dagegen keine gemeinsamen Linien (vergleiche Abbildung 5.1.2). Werden zusätzlich Auswahlregeln berücksichtigt, bestehen Beinchen zum selben Vibrationsgrundzustand zwar nicht mehr aus denselben Linien, aber es gibt — zumindest wenn die Auswahlregeln nicht sämtliche zugehörigen Übergänge verbieten — noch immer gemeinsame Linien in den Beinchen.

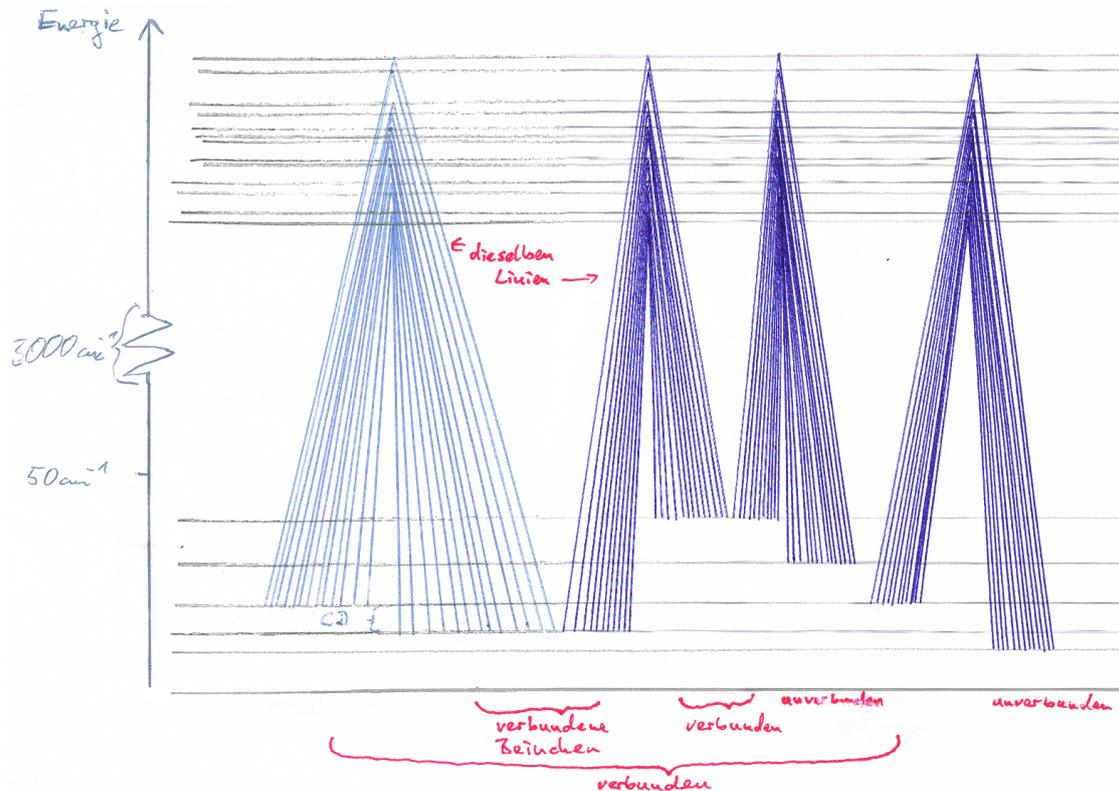


Abbildung 5.1.2: Verbindung der CD-Peaks über gemeinsame Beinchen (ohne Auswahlregeln)

Dies kann man sich zu Nutze machen, indem man zunächst bei allen gefundenen Peaks zusammenstellt, aus welchen Linien ihre Beinchen bestehen. Dazu muss man die CDs, aus denen der Peak besteht, zusammen suchen und die bei der Berechnung dieser CDs verwendeten Linien verwenden. Danach wird innerhalb der verschiedenen Beinchen nach gemeinsamen Linien gesucht. Haben zwei Beinchen gemeinsame Linien, so sind die zugehörigen CD-Peaks durch einen

gemeinsamen Vibrationsgrundzustand des Moleküls verbunden.

5.2 Rekonstruktion der Vibrationsgrundzustände

Wurde bestimmt, welche CD-Peaks durch gemeinsame Beinchen verbunden sind, können die Vibrationsgrundzustände rekonstruiert werden, indem man (vergleiche Abbildung 5.2.1)

1. mit einem beliebigen CD-Peak anfängt, den Vibrationsgrundzustand, an den eines seiner Beinchen ansetzt, als 0 definiert und dem Energie-Niveau-Schema hinzufügt,
2. andere CD-Peaks sucht, die mit dem Start-CD-Peak gemeinsame Beinchen haben, sie anbaut und dem Energie-Niveau-Schema entsprechende Niveaus hinzufügt,
3. Schritt 2 so lange wiederholt, bis alle Beinchen-Verbindungen genutzt sind.

In der Praxis besteht eine Schwierigkeit darin, die zu einem CD-Peak gehörenden CDs eindeutig zu bestimmen: Gemäß den Überlegungen in Abschnitt 1.2.1 liegen die esCDs und aCDs ziemlich zufällig im CD-Spektrum verteilt. Gerade bei Messungen bei verhältnismäßig hohen Temperaturen, bei denen die CDs sehr dicht liegen, ist es deshalb wahrscheinlich, dass auch unerwünschte esCDs und aCDs im Bereich eines CD-Peaks liegen. Sie können den Anschein erwecken, dass Beinchen verbunden wären, obwohl sie eigentlich unverbunden sind. Im entwickelten Programm wird diesem Problem folgendermaßen begegnet:

1. Die Verbindungen werden gewichtet mit der Anzahl der gemeinsamen Linien. Dabei werden gemeinsame Linien, die bei beiden Beinchen zu CDs gehören, die bereits (in vorherigen Programmläufen, vergleiche Abschnitt 5.3.2) als gsCDs identifiziert wurden, wesentlich stärker gewichtet. Die Gewichte werden entlang eines Weges durch das Energieniveau-Schema multipliziert und wenn zwei Wege zu einem widersprüchlichen Ergebnissen führen, wird die Verbindung mit dem geringsten Gewicht im Weg mit dem geringsten Gewicht verworfen. Mit den verbleibenden Verbindungen wird die Rekonstruktion des Energie-Niveau-Schemas komplett von vorne begonnen.
2. Verbindungen werden überhaupt erst berücksichtigt, wenn ihr Gewicht einen gewissen Schwellenwert überschreitet. Dieser Schwellenwert wird einerseits von der Breite der involvierten CD-Peaks und andererseits von der durchschnittlichen Dichte des CD-Spektrums abhängig gemacht. Im Zweifel macht es Sinn, den Schwellenwert lieber etwas zu hoch als zu niedrig anzusetzen: Ein zu niedriger Schwellenwert führt zu falschen Verbindungen, die das gesamte Ergebnis falsch machen können, ein zu hoher Schwellenwert führt zu fehlenden Verbindungen, die aber entweder später ergänzt werden können (siehe Abschnitt 5.2.2), oder zu einem unvollständigen, aber richtigen Energie-Niveau-Schema führen. Es hat sich bewährt, den Schwellenwert zwischen verschiedenen Programmläufen zu variieren: Für eine erste grobe Rekonstruktion ist $\text{CD-Dichte} \cdot \text{Peakbreite} + 1$ ein guter Startwert. Bleiben vereinzelte Peaks übrig, kann der Schwellenwert z. B. im zweiten Lauf reduziert werden, um das Schema zu erweitern. Im dritten Lauf, wenn bereits zahlreiche CDs eindeutig als gsCDs identifiziert wurden (siehe Abschnitt 5.3.2), werden dann nur noch Linien für Verbindungen berücksichtigt, die zu diesen gsCDs gehören, sodass falsche Einordnungen wieder aus dem Schema entfernt werden und insbesondere mehrere Sets, die fälschlicherweise durch aCDs zu einem einzigen Set verbunden wurden, korrekt in mehrere Sets zerfallen. (Siehe Abschnitt 5.2.1.)

5.2.1 Trennung der Sets

Hat das Molekül wie CH_5^+ mehrere unverbundene Sets von Zuständen, wird durch die im Abschnitt 5.2 beschriebene Methode natürlich nur das Energie-Niveau-Schema desjenigen Sets zusammengebaut, aus dem der Start-CD-Peak stammt. Die weiteren Sets können Schritt für Schritt rekonstruiert werden, indem der Rekonstruktionsprozess mit einem bisher nicht involvierten CD-Peak neu gestartet wird.

Die bereits in Abschnitt 5.1 diskutierten falschen Beinchen-Verbindungen, die von esCDs oder aCDs herrühren, können dazu führen, dass mehrere getrennte Sets zu einem falschen zusammengebaut werden. Auch deshalb ist es relevant, den dort diskutierten Schwellenwert eher zu hoch zu wählen als zu niedrig. Ein zu hoch gewählter Schwellenwert hat im schlimmsten Fall die Folge, dass ein Set als mehrere unverbundene Teilsets erscheint. In der Regel führt ein sicherheitshalber zu hoch gewählter Schwellenwert allerdings eher dazu, dass die Sets unvollständig sind und das Ergebnis dafür zahlreiche Sets enthält, die aus nur jeweils zwei Niveaus, die durch einen CD-Peak verbunden sind, bestehen. Diese Sets, die jeweils nur einen einzigen CD-Peak umfassen, können einfach ignoriert werden. Zusätzlich zur bereits beschriebenen Variation des Schwellenwertes bietet sich ein weiterer Programmdurchlauf mit auf Grund von bereits im ersten Durchlauf gewonnener Erkenntnisse selektierten CDs und damit geringerer CD-Dichte zur Vervollständigung an. (Siehe Abschnitt 5.2.2.)

5.2.2 Fehlende CD-Peaks und Verbindungen

Wurde das Energie-Niveau-Schema gemäß Abschnitt 5.2.1 rekonstruiert, kann es sein, dass manche Niveaus nicht durch CD-Peaks miteinander verbunden sind. Dies kann verschiedene Gründe haben:

1. Auf Grund des in Abschnitt 5.1.2 erwähnten Schwellenwertes, der falsche Verbindungen verhindern soll, wurden auch richtige Verbindungen verhindert.
2. CD-Peaks wurden nicht erkannt, sondern sind im Rauschen der esCDs und aCDs untergegangen.
3. Sämtliche Übergänge, die einen vibrationsangeregten Zustand mit den beiden Vibrationsgrundzuständen verbinden könnten, sind durch Auswahlregeln verboten oder haben zu geringe Intensität.

5.2.2.1 Auswahlregeln und Co-CDs

Da im dritten Fall offenbar eine indirekte Verbindung (also über mehrere Schritte im Energie-Niveau-Schema) besteht — denn sonst würden die entsprechenden Niveaus im rekonstruierten Schema ja gar nicht vorkommen —, müssten die entsprechenden Übergänge zu rekonstruieren sein, wenn nicht nur CDs, sondern auch Co-CDs, also Differenzen von CDs, betrachtet werden. Dies ist in der Praxis allerdings unrealistisch, weil es davon zu viele gibt: Zu n Linien gehören wie oben erwähnt bereits $\mathcal{O}(n^2)$ CDs, dazu gäbe es $\mathcal{O}(n^4)$ Co-CDs erster Stufe. Dies würde nicht nur zu unerträglichen Programm-Laufzeiten führen, sondern vor allem wäre die Dichte im Co-CD-Spektrum noch einmal immens höher als im CD-Spektrum und die zur Peak-Erkennung verwendeten Verfahren wären nur noch anwendbar, wenn die Messgenauigkeit der Daten extrem gesteigert werden könnte (siehe Abschnitt 2.3.1). Linien oder CDs zur Senkung von Rechenzeit und Dichte wegzulassen, wäre kontraproduktiv, weil es die gleiche, in Abschnitt 2.3.1.2 diskutierte Auswirkung hätte wie zusätzliche Auswahlregeln. Auf die weitere Auseinandersetzung mit Co-CDs wird deshalb verzichtet, zumal die meisten der darüber potenziell rekonstruierbaren

Übergänge einfach im gemäß Abschnitt 5.2 rekonstruierten Energie-Niveau-Schema abgelesen werden können (vergleiche Abbildung 5.2.1).

5.2.2.2 Rekonstruktion fehlender CD-Peaks und Verbindungen

Um dem Problem im Rauschen untergegangener CD-Peaks zu begegnen, können auf Grund der Zuordnung der CDs die Linien ihrer Beinchen den verschiedenen identifizierten Sets zugeordnet werden. Wird anschließend für jedes Set das gesamte Verfahren neu durchgeführt und werden dabei von Anfang an die bereits jeweils einem anderen Set zugeordneten Linien weggelassen, reduziert sich die Dichte des CD-Spektrums erheblich, sodass das gesamte Verfahren in der zweiten Iteration erheblich besser funktioniert. Dies gilt selbst dann, wenn nur ein geringer Teil der Linien zugeordnet wurde, weil die Anzahl der CDs quadratisch von der Anzahl der Linien abhängt. Indem man alle irgendwie zugeordneten Linien im zweiten Durchgang weg lässt, kann man auch gezielt nach einem zusätzlichen Set suchen, was bei den 10 K-Daten aus [Asvany et al. 2015] zur „Entdeckung“ von zwei neuen Sets führte.

Dieses Verfahren kann nochmals verbessert werden, indem man am Ende eines Durchlaufs die Werte fehlender Verbindungen im Energie-Niveau-Schema abliest und im entsprechenden Bereich des CD-Spektrums die Peak-Erkennung mit höherer Empfindlichkeit durchführt. Findet man auf diese Weise einen Teil der fehlenden CD-Peaks doch, können über die Beinchen der zugehörigen CDs noch mehr Linien einem Set zugeordnet werden, sodass sich der nächste Durchgang umso mehr lohnt.

5.3 Vibrationsangeregte Zustände

Wurde als Ergebnis der Rekonstruktion des Energie-Niveau-Schemas eines Sets einem CD-Peak ein Übergang zwischen zwei Vibrationsgrundzuständen zugeordnet, lassen sich unmittelbar auch vibrationsangeregten Zustände, die in dem CD-Peak involviert sind, rekonstruieren: Indem zum unteren (oberen) Vibrationsgrundzustand des CD-Peaks die Linien seines unteren (oberen) Beinchens addiert werden, erhält man die vibrationsangeregten Zustände (vergleiche Abbildung 5.1.1). Es spielt dabei keinerlei Rolle, ob man den unteren Vibrationsgrundzustand des CD-Peaks und sein unteres Beinchen oder seinen oberen Vibrationsgrundzustand und sein oberes Beinchen verwendet.

5.3.1 Check der Set-Aufteilung und der Verbindungen

Mit dieser Methode lassen sich die vibrationsangeregten Zustände aus jedem CD-Peak unabhängig rekonstruieren, wobei das Ergebnis bei CD-Peaks aus demselben Set gleich sein sollte. Durch Vergleich der unabhängigen Ergebnisse für die einzelnen CD-Peaks lässt sich ihre Zuordnung prüfen und ggf. korrigieren. Sind die rekonstruierten vibrationsangeregten Niveaus von mehreren CD-Peaks eines Sets zwar gleich, aber um einen Offset verschoben, bedeutet das, dass der Peak zwar dem richtigen Set zugeordnet ist, aber innerhalb des Sets falsch mit den anderen CD-Peaks verbunden wurde (vergleiche Abschnitt 5.1). Passen die an Hand eines CD-Peaks rekonstruierten vibrationsangeregten Zustände gar nicht zu den anderen Rekonstruktionen des Sets, gehört der Peak zu einem anderen Set.

Insgesamt ist dabei zu berücksichtigen, dass es ziemlich wahrscheinlich ist, dass mehrere unabhängig voneinander rekonstruierte Vibrationsbanden eines Sets zwar weitgehend zueinander passen, einzelne Niveaus aber dennoch nur in einzelnen Rekonstruktionen vorkommen. Dies liegt einerseits an Auswahlregeln und Übergängen, die wegen ihrer zu geringen Intensität nicht in den Messungen vorkommen. Andererseits liegt das daran, dass mit großer Wahrscheinlichkeit in den

Bereich eines CD-Peaks auch esCDs und aCDs fallen, die gar nicht zum CD-Peak gehören (vergleiche die Diskussion in Abschnitt 5.1) und falsche vibrationsangeregte Zustände produzieren.

5.3.2 Check der CDs eines CD-Peaks

Umfasst die Rekonstruktion der vibrationsangeregten Zustände an Hand eines CD-Peaks ein Niveau, das in den anderen Rekonstruktionen nicht vorkommt, kann daraus nicht direkt der Schluss gezogen werden, dass dieses Niveau auf eine fälschlicherweise zum CD-Peak gezählte CD zurück geht, weil wie beschrieben auch Linien fehlen können. Umgekehrt kann aber bei vibrationsangeregten Zuständen, die in mehreren unabhängigen Rekonstruktionen vorkommen, mit großer Sicherheit davon ausgegangen werden, dass sie stimmen und die zugehörigen CDs auch tatsächlich zu den jeweiligen CD-Peaks (und damit Sets) gehören. Es ist nämlich sehr unwahrscheinlich, dass fälschlicherweise zu einem Peak gezählte CDs einen bestimmten, in der Rekonstruktion mit Hilfe eines anderen CD-Peaks bereits vorhandenen vibrationsangeregten Zustand liefern. Zur Abschätzung dieser Wahrscheinlichkeit sind zwei Fälle zu unterscheiden:

- **Die falsch zugeordnete CD ist unverbunden.** Fälschlicherweise einem CD-Peak zugeordnete aCDs liefern ein zufälliges Niveau im Bereich der Vibrationsbanden. Die Wahrscheinlichkeit, dass dieses Niveau genau ein bestimmtes anderes Niveau trifft, beträgt $\frac{\text{Messfehler}}{\text{Breite der Vibrationsbanden}}$. Da es verschiedene obere Niveaus (sowohl vibrationsangeregte Zustände als auch durch andere falsch zugeordnete CDs produzierte Niveaus) gibt, die getroffen werden könnten, ergibt sich dafür insgesamt die Wahrscheinlichkeit $\# \text{obere Niveaus} * \frac{\text{Messfehler}}{\text{Breite der Vibrationsbanden}}$. Mit den Parametern, die in der Testumgebung den gemessenen Daten ähnliche Testdaten liefern, ergibt sich dafür bei konservativer Abschätzung $900 * \frac{0,0001\text{cm}^{-1}}{450\text{cm}^{-1}} = 0,0002$, also eine vernachlässigbar kleine Wahrscheinlichkeit.
- **Die falsch zugeordnete CD ist unten verbunden.** Der Wert einer esCD entspricht dem Abstand zwischen zwei vibrationsangeregten Zuständen. Damit eine esCD fälschlicherweise einem CD-Peak zugeordnet werden kann, muss es in den Vibrationsbanden zwei Zustände geben, deren Abstand zum CD-Peak, also zum Abstand von zwei Vibrationsgrundzuständen passt. Dann liegen sämtliche zu diesem oberen Abstand gehörenden, esCDs im CD-Peak. Sie liefern zusammen die Vibrationsgrundzustände horizontal gespiegelt als vibrationsangeregte Zustände. Normalerweise sind die vibrationsangeregten Zustände aber keine Spiegelbilder, sondern verschobene Vibrationsgrundzustände mit zudem noch leicht veränderten Rotationskonstanten. Dann entspricht die Wahrscheinlichkeit, dass ein gespiegeltes Vibrationsgrundzustandsniveau ein vibrationsangeregtes Niveau trifft, der Wahrscheinlichkeit, dass eine aCD ein oberes Niveau trifft (vorheriger Punkt). Ist eine Vibrationsbande dagegen wider Erwarten tatsächlich eine Spiegelung der Vibrationsgrundzustände, ist die Wahrscheinlichkeit, dass eine esCD mit gleicher Energie wie eine gsCD ein vorhandenes oberes Niveau trifft, sehr viel größer, nämlich $\frac{1}{\# \text{Anzahl untere Niveaus}}$. Dann richtet die falsche Zuordnung der CD aber auch keinen Schaden an. Insgesamt ist die Wahrscheinlichkeit, dass eine falsch zugeordnete esCD Schaden anrichtet, also noch wesentlich geringer als die Wahrscheinlichkeit, dass eine falsch zugeordnete aCD beim Check mittels vibrationsangeregter Zustände fälschlicherweise als gesichert angesehen wird.

Um ganz sicher zu gehen, wird im Programm folgendermaßen vorgegangen: Wird ein vibrationsangeregtes Niveau von mindestens drei verschiedenen CD-Peaks produziert, wird es als gesichert angenommen. Produziert ein CD-Peak mindestens drei gesicherte angeregte Niveaus, wird er als richtig zugeordnet angesehen. Produziert eine CD eines richtig zugeordneten CD-Peaks ein gesichertes Niveau, wird sie als wahre gsCD und gesichert zum Peak gehörig angesehen.

Außer der Verifikation, dass CDs tatsächlich zu dem Peak passen, dem sie zugeordnet wurden, ist noch eine weitere Plausibilitätsprüfung möglich: Liefert ein CD-Peak als Rekonstruktion der vibrationsangeregten Zustände das Spiegelbild der bereits rekonstruierten Vibrationsgrundzustände, ist dies ein Indiz dafür, dass der Peak in Wirklichkeit aus esCDs besteht und komplett verworfen werden kann.

Der Check der CDs kann genutzt werden, um in einem weiteren Programmdurchlauf (vergleiche Abschnitt 5.2.2.2) Verbindungen (siehe Abschnitt 5.1) sicherer zu identifizieren: Einer Verbindung zwischen zwei Beinchen, die durch sehr sicher zu den CD-Peaks gehörende Linien hergestellt wird, kann ein sehr hohes Gewicht zugeordnet werden.

5.3.3 Rekonstruktion der vibrationsangeregten Zustände

Die Rekonstruktion der vibrationsangeregten Zustände eines Sets dient natürlich nicht nur der Überprüfung der rekonstruierten Vibrationsgrundzustände. Vielmehr sind die Ergebnisse an sich ein großer Fortschritt. Gemäß der Überlegungen aus Abschnitt 5.3.2 können sämtliche Niveaus, die in mindestens drei unabhängigen Rekonstruktionen vorkommen, als gesichert angesehen werden. Alle anderen werden im weiteren Verlauf dieser Arbeit — wenn dies auch übervorsichtig sein mag — zur Sicherheit verworfen.

Dabei ist Folgendes zu beachten: Die gemessenen IR-Übergänge verbinden immer zwei Zustände mit verschiedener Parität. Dennoch haben die durch CDs verbundenen Vibrationsgrundzustände gleiche Parität, weil in ihre Berechnung immer zwei solche Übergänge eingehen. Im Gegensatz dazu haben die rekonstruierten vibrationsangeregten Zustände, jeweils andere Parität als die zugehörigen Vibrationsgrundzustände.

5 Von Linien über CDs zu Zuständen

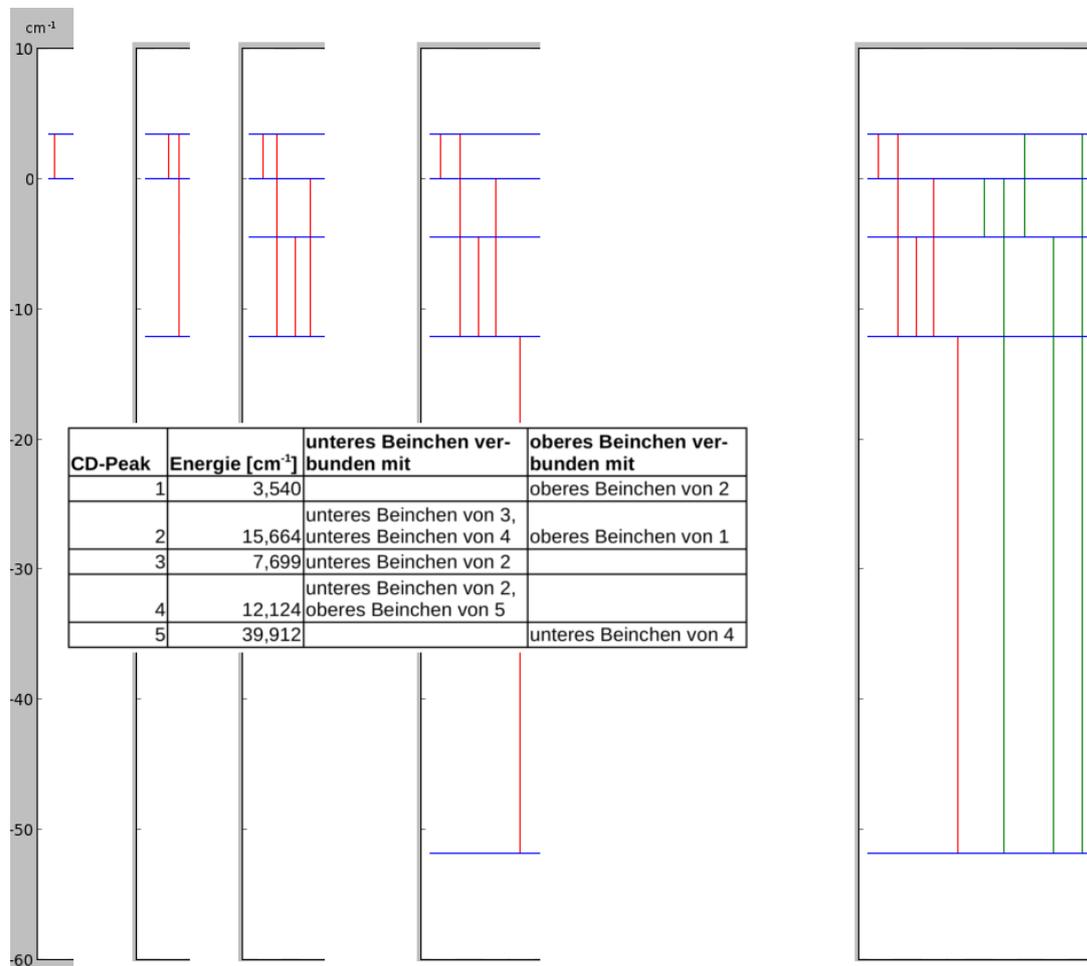


Abbildung 5.2.1: **links:** Rekonstruktion des Energie-Niveau-Schemas aus Beinchen-Verbindungen. Von links nach rechts: Schema zu Beginn, Schema nach Auswertung der Verbindungen in der ersten Tabellenzeile, Schema nach Auswertung der ersten beiden Tabellenzeilen, Endergebnis
rechts: Energie-Niveau-Schema nach Ergänzung fehlender Übergänge (grün) gemäß Abschnitt 5.2.2.1

6 Das Programm

Grundsätzlich ist das entwickelte Python-Script darauf ausgelegt, dass bestimmte Schritte mehrfach ausgeführt werden, wobei bei jedem Durchlauf die Erkenntnisse der vorherigen Durchläufe genutzt werden, um zu jeweils besseren Ergebnissen zu kommen. Zudem werden in den einzelnen Durchläufen nach vielen Schritten Plausibilitäts-Checks durchgeführt, die es ermöglichen, unsichere oder gar widersprüchliche Ergebnisse entweder direkt zu korrigieren oder zunächst zu verwerfen, um sie im nächsten Durchlauf unter durch neue Vorkenntnisse verbesserten Bedingungen sicher zu ermitteln. Am Ende jedes Durchlaufs kann darauf Einfluss genommen werden, welche Ergebnisse (zunächst) verworfen und welche als verifiziert angesehen werden können. Dafür gibt das Script an den entsprechenden Stellen eine Menge an meistens hilfreichen Informationen aus. Dies führt nicht nur dazu, dass insgesamt mehr Informationen aus den Daten gezogen werden können als bisher und z. B. alle gemäß [Schmiedt et al. 2016] zu erwartenden vier Sets gefunden wurden und nicht nur zwei wie bisher von [Asvany et al. 2015]. Vielmehr ist das iterative Vorgehen auch relativ robust. Kleinere Fehler werden im Laufe der Durchläufe korrigiert; ungeschickt gewählte Parameter in den einzelnen Auswertungsschritten führen nicht zu anderen Endergebnissen, sondern vor allem dazu, dass mehr Durchläufe notwendig sind, um zum gleichen Ergebnis zu kommen. Fast immer stellt sich nach einer Reihe von Durchläufen ein Ergebnis ein, das konsistent ist und sich durch weitere Programmdurchläufe auch nicht mehr verändert. Um Zeit zu sparen, werden zahlreiche Zwischenergebnisse automatisch abgespeichert und bei den verschiedenen Durchläufen nur dann neu berechnet, wenn sich an den Eingangsparametern des entsprechenden Programmschrittes etwas verändert hat, was das Script automatisch entscheidet.

Über alle Programmdurchläufe hinweg werden zwei Listen geführt, die festhalten, welche CDs gesichert als gsCDs identifiziert wurden und welche Linien gesichert einem bestimmten Set zugeordnet werden konnten. Mit diesen Listen arbeiten drei Scripte (diese drei Programme sowie weitere Dateien, die eine Sammlung von allen drei Programmen genutzter Routinen enthalten, finden sich im Anhang A.3): `generate_line_list.py` erzeugt Listen gemessener Linien, aus denen Linien, die eindeutig bestimmten, wählbaren Species zugeordnet werden konnten, ausgefiltert sind. Um ein bestimmtes Set genauer zu untersuchen, kann man so alle Linien, die auf jeden Fall zu einem anderen Set gehören, vorher aussortieren. Oder man sortiert, um ein neues Set zu finden, alle überhaupt einem Set zugeordneten Linien aus. `generate_CDs.py`, das im Wesentlichen eine Weiterentwicklung eines von [Asvany et al. 2015] genutzten Scripts ist, erzeugt aus einer gefilterten Linienliste CDs, die dann mit `analyze.py` ausgewertet werden. Sämtliche Scripte verarbeiten, wenn sie auf Daten aus der Testumgebung angewendet werden, die jeweils zu erwartenden Ergebnisse parallel mit, sodass z. B. immer eine Liste mit den bei der angewandten Linienfilterung zu erwartenden Peaks im CD-Spektrum angelegt wird, die `analyze.py` in den CD-Spektren wie in Abbildung 2.0.3 mit plotten kann.

Zusätzlich zu dieser Gesamtiteration kann innerhalb des Scripts `analyze.py` die Analyse, Zuordnung und Bewertung der einzelnen Peaks des CD-Spektrums wiederholt durchlaufen werden. In diesem Script kommen die oben erläuterten, im Rahmen dieser Diplomarbeit (weiter-) entwickelten Auswertungsmethoden zur Anwendung.

6.1 Erstellung und Auswertung des CD-Spektrums

Zur Auswertung des CD-Spektrums führt das Script `analyze.py` folgende Schritte durch:

1. Analyse der CD-Liste zur Bestimmung einer sinnvollen Bandbreite für die Kerndichteschätzung wie in Abschnitt 3.2 beschrieben
2. Berechnung eines CD-Spektrums aus vollem Datensatz durch Kerndichteschätzung
3. Rauschunterdrückung gemäß Abschnitt 4.2 zur Vorbereitung der Peaksuche:
 - a) Berechnung mehrerer Kerndichteschätzung mit zufällig reduzierten Datensätzen
 - b) Berechnung der Enveloppe dieser Kerndichteschätzungen mit den reduzierten Datensätzen
 - c) Berechnung einer rauschreduzierten Variante des CD-Spektrums mit vollem Datensatz, indem vom Ergebnis des 2. Schrittes die Enveloppe abgezogen wird.
4. Peaksuche. Im CD-Spektrum mit reduziertem Rauschen, das in Schritt 3 erstellt wurde, werden dazu Stellen gesucht, die eine bestimmte Höhe überschreiten.
5. Analyse, Zuordnung und Bewertung der Peaks durch wiederholtes Durchlaufen der folgenden Schritte:
 - a) Bestimmung der genauen Lage der Peaks: An der Stelle der einzelnen Peaks wird im CD-Spektrum ohne Rauschunterdrückung, das in Schritt 2 erstellt wurde, zunächst das größte lokale Maximum gesucht und als tatsächliche Position des Peaks festgehalten. Um diese Position herum wird ein Bereich bestimmt, innerhalb dessen die einzelnen CDs zum Peak gezählt werden. Dafür wird von der Peakposition soweit in die beiden verschiedenen Richtungen gegangen, bis entweder die Höhe der Kerndichteschätzung auf 10 % abgefallen ist oder auf Grund eines benachbarten Peaks wieder zu steigen beginnt.
 - b) Ermittlung, welche Linien zu den einzelnen Beinchen der Peaks gehören (vergleiche Abschnitt 5.1)
 - c) Ermittlung der Verbindungen zwischen den Beinchen der CD-Peaks (Abschnitt 5.1) und ihrer Gewichte. Bei der Ermittlung der Gewichte werden gemeinsame Linien, die bei beiden Beinchen von CDs stammen, die in vorherigen Läufen als gsCDs verifiziert wurden, erheblich stärker gewichtet. Verbindungen, die ein bestimmtes Gewicht unterschreiten, werden ignoriert. Siehe Abschnitt 5.2.
 - d) Anzeige der ermittelten Verbindungen und Möglichkeit für die Nutzerin / den Nutzer, zusätzliche Verbindungen mit hohem Gewicht einzugeben.
 - e) Rekonstruktion der Energie-Niveau-Schemata gemäß Abschnitt 5.2 und 5.2.1. Wenn dabei Widersprüche auftreten, werden die unsichersten in den Widerspruch verwickelten Verbindungen gelöscht und von Neuem begonnen.
 - f) Verwerfen von Sets, die aus nur einem Peak bestehen.
 - g) Ergänzen fehlender Übergänge zwischen den rekonstruierten Vibrationsgrundzuständen gemäß Abschnitt 5.2.2.2.
 - h) Testen, ob die Peaks, die zu den in Schritt 5f verworfenen Sets gehören, zu den in Schritt 5g ergänzten Übergängen passen und ggf. Einordnung an entsprechender Stelle in die rekonstruierten Energie-Niveau-Schemata

6 Das Programm

- i) Rekonstruktion der vibrationsangeregten Zustände der einzelnen Sets gemäß Abschnitt 5.3 unabhängig für jeden einem Set zugeordneten CD-Peak.
 - j) Bestimmung der gesicherten vibrationsangeregten Zustände, der sicher zugeordneten Peaks und Erweiterung der Liste der gesicherten gsCDs gemäß Abschnitt 5.3.2
 - k) Präsentation der Zwischenergebnisse in Form von Tabellen und als Plot des Energieniveau-Schemas. Abfrage des Namens des Sets
 - l) Abfrage, ob weiterer Durchlauf von Schritt 5 erwünscht ist. Falls nicht, weiter mit Schritt 6, sonst weiter mit Schritt 5m.
 - m) Möglichkeit des Nutzers, die Peakliste für den nächsten Durchlauf zu bearbeiten. Dabei stehen folgende Optionen zur Verfügung: Löschen aller Peaks bestimmter Sets, Löschen aller nicht gemäß Abschnitt 5.3.2 verifizierter Peaks, Ergänzung aller fehlenden Übergänge als mögliche Peakpositionen, Löschen einzelner Peaks, Ergänzen einzelner Peakpositionen
 - n) Möglichkeit des Nutzers, den Schwellenwert für das Gewicht, ab dem eine Verbindung in Schritt 5c gezählt wird, zu ändern.
Weiter mit Schritt 5a
6. Aktualisieren der abgespeicherten Liste der gesicherten gsCDs¹
 7. Aktualisieren der abgespeicherten Liste mit Zuordnungen der Linien zu Sets an Hand der in Schritt 5k abgefragten Namen der Sets

Insgesamt arbeiten die Scripte nicht vollautomatisch; ohne sinnvolle Nutzereingaben zwischen den einzelnen Durchläufen bekommt man keine sinnvollen Ergebnisse. Im Laufe der Entwicklung dieser Arbeit hat sich der Arbeitsschwerpunkt immer mehr zu den kombinatorischen Teilen im Schritt 5 verlagert. Abgesehen von der Wahl der richtigen Bandbreite der Kernschätzung sind nämlich für das Endergebnis alle Details der vorherigen Schritte belanglos. Sie liefern für die Ermittlung der Ergebnisse in Schritt 5 vor allem mehr oder weniger gute Startbedingungen, deren Bedeutung allerdings auch nicht zu unterschätzen ist.

6.2 Verbesserungsmöglichkeiten

Im Wesentlichen gibt es vier Möglichkeiten, den Programmablauf weiter zu verbessern:

- Verschieben der Schritte 5m und 5n an den Anfang von Schritt 5, sodass bereits vor dem ersten Durchlauf eingegriffen werden kann.
- Erweiterung von Schritt 5d, sodass nicht nur zusätzliche Verbindungen eingegeben, sondern auch bestehende verworfen und ihr Gewicht verändert werden kann.
- Einbau der Möglichkeit, die automatische Peaksuche bei Schritt 5m zu wiederholen.
- Die in Schritt 5d von Hand ergänzten Verbindungen müssen in jedem Durchlauf von Schritt 5 ggf. neu eingegeben werden. Da sich die programminterne Nummerierung der Peaks zwischen den einzelnen Durchläufen ändern können, können die Eingaben nicht von

¹Die abgespeicherte Liste der gsCDs könnte auch direkt in Schritt 5j aktualisiert werden. Das Abspeichern der Ergebnisse ganz am Schluss ermöglicht aber, das Programm abubrechen, bevor die abgespeicherten Listen verändert wurden, sodass ein Programmmlauf, der keinen Fortschritt gebracht hat oder bei dem etwas falsch durch den Nutzer eingegeben wurde, die vorher ermittelten Ergebnisse nicht beeinträchtigt.

6 Das Programm

einem Durchlauf zum nächsten übernommen werden. In der Regel ist es aber nur selten notwendig, die gleichen Verbindungen mehrmals von Hand vorzugeben: Existiert ein Peak im CD-Spektrum, ist es eigentlich nur dann notwendig, eine Verbindung per Hand vorzugeben, wenn entweder die Verbindung zwar existiert, aber nur aus so wenig Linien besteht, dass ihr Gewicht unter dem Schwellenwert liegt, den man aber wiederum aus anderen Gründen nicht absenken möchte. Oder eine Verbindung existiert, wird aber bei der Rekonstruktion des Energie-Niveau-Schemas in einen Widerspruch verwickelt und in Folge dessen in Schritt 5e verworfen, obwohl besser eine andere Verbindung verworfen werden sollte. In beiden Fällen reicht es in der Regel, die Verbindung einmalig von Hand zu ergänzen, denn sobald ein Peak einmal korrekt eingeordnet wurde, ist die Wahrscheinlichkeit sehr hoch, dass die richtigen Verbindungen bei weiteren Programmdurchläufen auf Grund gesicherter gsCDs automatisch so sehr gestärkt sind, dass er immer wieder korrekt eingeordnet wird.

Während die Implementierung der ersten beiden Verbesserungen einfach möglich sein sollte, erfordert die letzte größere Umbauten am Programm. Bisher wurden alle drei Verbesserungsmöglichkeiten aus Zeitgründen nicht umgesetzt, zumal sie nicht zu anderen / besseren Ergebnissen führen, sondern lediglich die Bedienung des Programms effizienter (und weniger nervig) machen würden.

7 Ergebnisse

Mit dem entwickelten Programm wurden vor allem die Messwerte von Asvany et al. bei nominal 10 K für CH_5^+ untersucht. Die Ergebnisse finden sich in Abschnitt 7.2. Kurz vor Ende der Bearbeitung dieser Diplomarbeit lagen zusätzlich neue Messwerte für CD_2H^+ vor, siehe [Jusko et al. 2016]. Es handelt sich um den gleichen Typ von Messungen (ebenfalls mit Hilfe eines Frequenzkamms hochaufgelöst gemessene Übergänge zwischen Vibrationsgrundzuständen und vibrationsangeregten Zuständen), das Spektrum ist aber weit weniger dicht (lediglich 117 Linien). Für CD_2H^+ gibt es gute Modelle, sodass die Auswertung mit herkömmlichen Methoden keine Schwierigkeiten bereitet. Das in dieser Diplomarbeit entwickelte modelllose Verfahren wurde deshalb versuchs halber auf die Daten angewendet und mit den Ergebnissen des herkömmlichen Verfahrens verglichen. Die Ergebnisse finden sich in Abschnitt 7.1.

Zwar ist das entwickelte Auswertungsverfahren stark auf die hohe Genauigkeit der Messungen angewiesen, liefert aber selbst keine besonders genauen Ergebnisse: Die Energie der CD-Peaks lässt sich zwar durch die verbesserte Kerndichteschätzung noch etwas genauer bestimmen als in [Asvany et al. 2015], die Energien des rekonstruierten Energie-Niveau-Schemas sind aber mit einem Fehler von etwa 0.00035 cm^{-1} behaftet und damit knapp eine Größenordnung größer. Dies liegt daran, dass der Algorithmus, mit dessen Hilfe das Energie-Niveau-Schema rekonstruiert wird (Abschnitt 5.2), sich an Hand der CD-Peaks Stück für Stück durch das Schema arbeitet und sich dabei die Fehler der einzelnen CD-Peaks aufsummieren. Um diese Fehler zu reduzieren wäre es notwendig, nach erfolgter Zuordnung die einzelnen Energien nochmal neu so zu bestimmen, dass die Fehler minimiert werden. Darauf wurde im Rahmen dieser Diplomarbeit allerdings verzichtet. Einerseits weil andere Verbesserungen spannender zu sein schienen, andererseits weil dafür bei Bedarf ohne Weiteres bereits bestehende Programme wie das von [Furtenbacher et al. 2015] verwendet werden können.

7.1 CD_2H^+

Die Auswertung des CD_2H^+ -Spektrums aus [Jusko et al. 2016] mit Hilfe des in dieser Arbeit entwickelten Verfahrens war problemlos möglich und die Ergebnisse sind vollständig kompatibel mit den Ergebnissen der von Jusko et al. mit Hilfe eines Modells und der Software [pgopher] vorgenommenen Auswertung. Insbesondere konnten auch die vier verschiedenen Sets reproduziert werden, die dadurch zu Stande kommen, dass CD_2H^+ wie Wasserstoff in einer para- und einer ortho-Species vorkommt und die Zustände jeweils zwei verschiedene Paritäten haben können.

Allerdings fehlen eine Reihe von Übergängen. Dies hat im Wesentlichen zwei Gründe:

- Das modellbasierte Auswertungsverfahren bestimmt an Hand der Messwerte die Parameter des Modells und berechnet daraus die Energien der Zustände. Manche der Übergänge, die zu den so erhaltenen Zuständen gehören, liegen außerhalb des von Jusko et al. vermessenen Wellenzahlenbereichs und können deshalb mit der modellfreien Auswertung nicht gefunden werden.
- Manche auf Grund der Auswertung von Jusko et al. zu erwartenden CD-Peaks tauchen im CD-Spektrum nicht auf, obwohl sie im vermessenen Bereich liegen. Dies liegt vermutlich

7 Ergebnisse

daran, dass die Übergänge, aus denen die zugehörigen CDs zu berechnen wären, durch Auswahlregeln verboten oder zu schwach sind. Diese Hypothese muss aber noch einmal an Hand des von Jusko et al. verwendeten Modells verifiziert werden, wozu die Zeit im Rahmen dieser Diplomarbeit nicht mehr ausreichte. Da CD_2H^+ weniger vibrationsangeregte Zustände hat als CH_5^+ und CD-Peaks deshalb aus nur wenigen CDs bestehen, ist es aber durchaus plausibel, dass Auswahlregeln und geringe Übergangswahrscheinlichkeiten dazu führen, dass die CD-Peaks nicht auftauchen, was bei CH_5^+ eher unwahrscheinlich ist. Dennoch lassen sich viele der zugehörigen Differenzen zwischen Vibrationsgrundzuständen gemäß der Abschnitte 5.2.2 und 5.2.2.2 rekonstruieren. Sie sind in den Plots grün und in den Tabellen kursiv gekennzeichnet. Bezieht man diese Rekonstruktionen mit ein, konnten mit folgenden Ausnahmen alle theoretisch erwarteten CD-Peaks rekonstruiert werden: Beim A1-Set fehlen in der Rekonstruktion die CD-Peaks, deren Bottom-Level 0 cm^{-1} ist. Dies ist plausibel, weil es nur sehr wenige vom 0 cm^{-1} -Level aus erlaubte Übergänge gibt. Außerdem fehlen in der Rekonstruktion CD-Peaks, bei denen mindestens eines der involvierten Level so hoch liegt ($\geq 95\text{ cm}^{-1}$), dass es bei der Messtemperatur so gut wie unbesetzt ist.

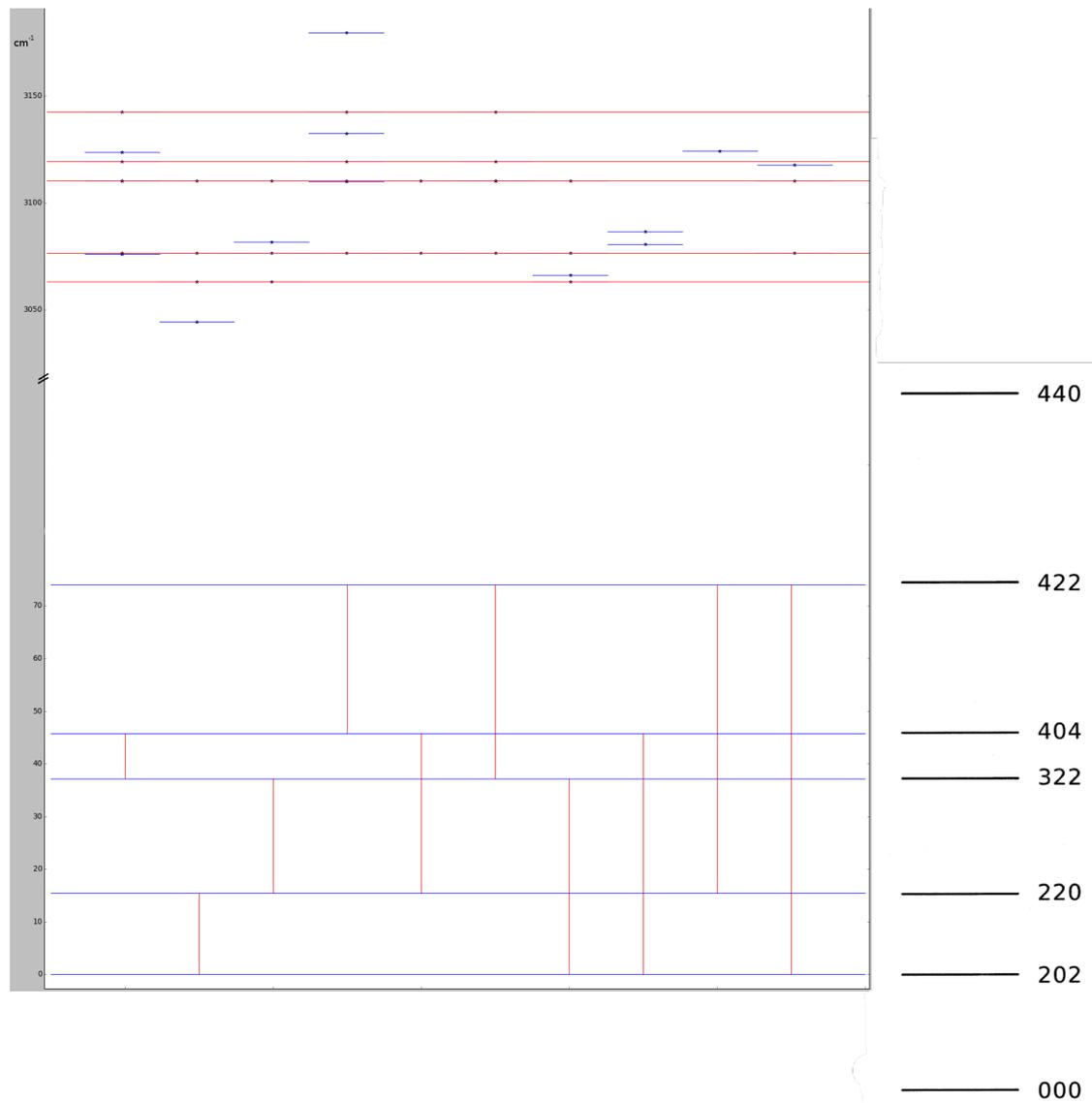
Es ist relativ wahrscheinlich, dass sich die modellos ermittelten Energie-Niveau-Schemata durch eine verbesserte Parameterwahl, weitere Programmdurchläufe etc. noch weiter vervollständigen lassen. So brachte etwa eine gegenüber der vom Script automatisch gewählten auf $\frac{1}{3}$ reduzierte Bandbreite deutlich bessere Ergebnisse. (Dies ist ein Hinweis darauf, dass die automatische Bandbreitenwahl noch weiter optimiert werden könnte, um bei weniger dichten Spektren bessere Ergebnisse zu liefern. Aber auch jetzt schon kann man auch bei weniger dichten Spektren mit der automatisch gewählten Bandbreite arbeiten.) Die für die Auswertung des CD_2H^+ -Spektrums verwendeten Parameter finden sich in Tabelle 7.1.1. Sie sind allerdings aus Zeitgründen nicht so sorgfältig gewählt und erprobt wie die für die anderen Spektren in Tabelle 4.2.1.

Parameter	Wert
Bandbreite	$0,3 * \text{automatische Bandbreite}$
# Durchläufe	15
Erwartungswert Sprungweite	3
Erwartungswert Clustergröße	1,5
für Enveloppe verwendete Percentile	80
Höhen-Schwellenwert	$0,5/\text{Bandbreite}$

Tabelle 7.1.1: Für die Auswertung des CD_2H^+ -Spektrums verwendete Parameter

7.1.1 A1

Ergebnisse des Sets, das bei der modellgebundenen Auswertung A1-Symmetrie hat.



Rekonstruiertes Energie-Niveau-Schema des A1-Sets

Die CD-Peaks sind als Übergänge zwischen den Vibrationsgrundzuständen in rot geplottet. Jeweils oberhalb in blau mit Stern finden sich die aus dem jeweiligen Peak rekonstruierten vibrationsangeregten Zustände. Gesicherte vibrationsangeregte Zustände sind durch rote horizontale Linien markiert. Rechts daneben die Ergebnisse von Jusko et al.

7 Ergebnisse

Grundzustände Jusko et al.			Grundzustände modellunabhängige Auswertung			
gsCD-Energie	Bottom-Level	Top-Level	CD-Peak-Energie $\pm 0,00005 \text{ cm}^{-1}$	Bottom-Level $\pm 0,00035 \text{ cm}^{-1}$	Top-Level $\pm 0,00035 \text{ cm}^{-1}$	# gesicherte es
7,1265	124,3689	131,4954				
8,5356	59,0153	67,5509	8,53619	59,01473 (37,21583)	67,55092 (45,75202)	5
15,4220	21,7989	37,2209	15,42205	21,79890 (0,00000)	37,21971 (15,42081)	3
21,7944	37,2209	59,0153	21,79441	37,21971 (15,42081)	59,01413 (37,21523)	3
21,7989	0,0000	21,7989				
28,3211	67,5509	95,8720	28,32108	67,55092 (45,75202)	95,87200 (74,07310)	4
28,4969	95,8720	124,3689				
30,3300	37,2209	67,5509	30,32997	37,22095 (15,42205)	67,55092 (45,75202)	2
35,6234	95,8720	131,4954				
36,8567	59,0153	95,8720	36,85666	59,01413 (37,21523)	95,87078 (74,07188)	5
37,2164	21,7989	59,0153	37,21523	21,79890 (0,00000)	59,01413 (37,21523)	3
37,2209	0,0000	37,2209				
39,0168	131,4954	170,5122				
45,7520	21,7989	67,5509	45,75335	21,79890 (0,00000)	67,55092 (45,75202)	0
46,1433	124,3689	170,5122				
56,8180	67,5509	124,3689				
58,6511	37,2209	95,8720	58,65203	37,21971 (15,42081)	95,87200 (74,07310)	0
59,0153	0,0000	59,0153				
63,9445	67,5509	131,4954				
65,3536	59,0153	124,3689				
67,5509	0,0000	67,5509				
72,4801	59,0153	131,4954				
74,0731	21,7989	95,8720	74,07310	21,79890 (0,00000)	95,87200 (74,07310)	2
74,6402	95,8720	170,5122				
87,1480	37,2209	124,3689				
94,2745	37,2209	131,4954				
95,8720	0,0000	95,8720				
102,5700	21,7989	124,3689				
102,9613	67,5509	170,5122				
109,6965	21,7989	131,4954				
111,4969	59,0153	170,5122				
124,3689	0,0000	124,3689				
131,4954	0,0000	131,4954				
133,2913	37,2209	170,5122				
148,7133	21,7989	170,5122				
170,5122	0,0000	170,5122				

A1

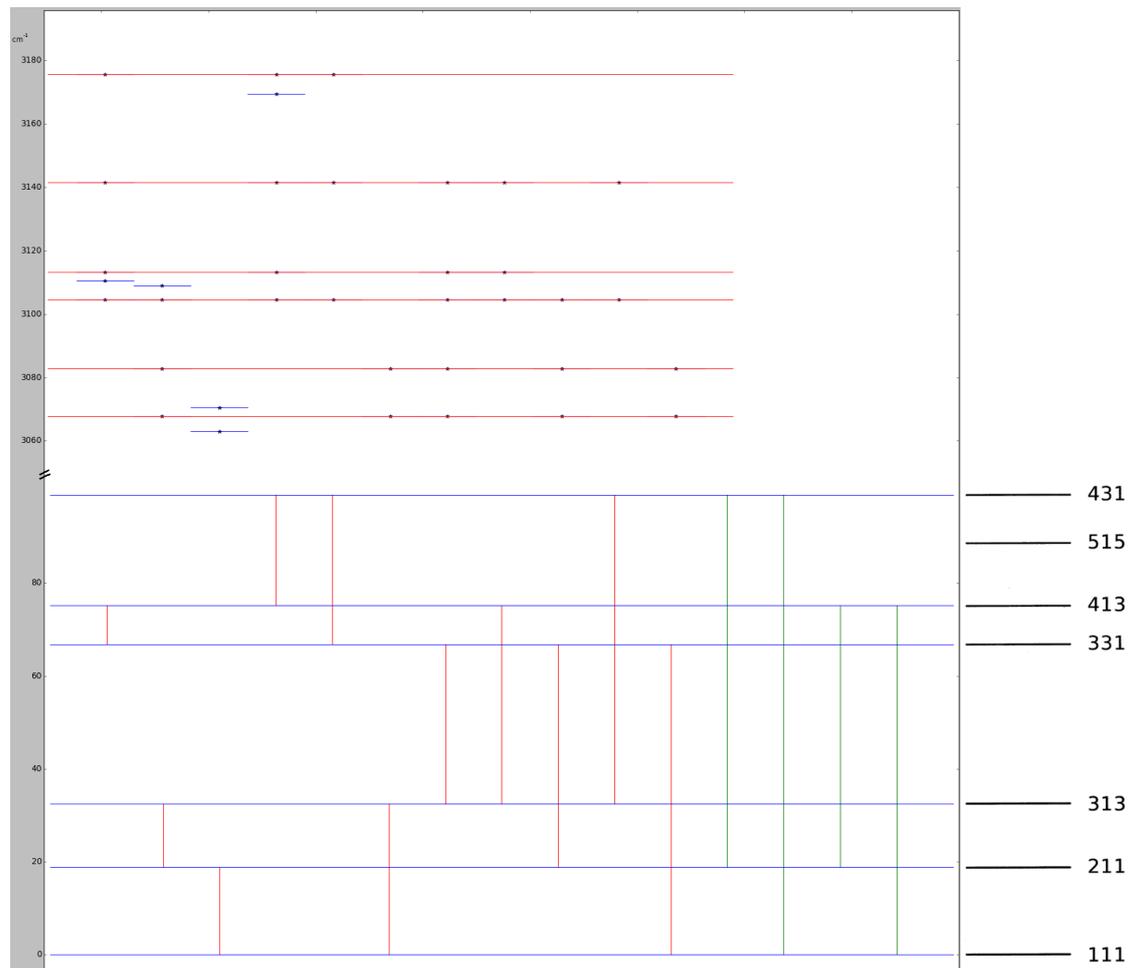
alle Energien in cm^{-1}

Modellunabhängig rekonstruierte Zustandsenergien sind zum besseren Vergleich um $21,7989 \text{ cm}^{-1}$ verschoben.
In Klammern Originalwerte, die zum Plot passen.

Gesicherte angeregte Zustände	
Energie $\pm 0,00035 \text{ cm}^{-1}$	
3084,83591 (3063,03701)	
3098,36277 (3076,56387)	
3131,983921 (3110,185021)	
3140,9963962 (3119,1974962)	
3164,2599512 (3142,4610512)	

7.1.2 A2

Ergebnisse des Sets, das bei der modellgebundenen Auswertung A2-Symmetrie hat.



Rekonstruiertes Energie-Niveau-Schema des A2-Sets

Die CD-Peaks sind als Übergänge zwischen den Vibrationsgrundzuständen in rot geplottet. In grün Differenzen, die zwar rekonstruiert wurden, an deren Stelle im CD-Spektrum aber kein Peak zu finden ist. Jeweils oberhalb in blau mit Stern finden sich die aus dem jeweiligen Peak rekonstruierten vibrationsangeregten Zustände. Gesicherte vibrationsangeregte Zustände sind durch rote horizontale Linien markiert. Rechts daneben die Ergebnisse von Jusko et al.

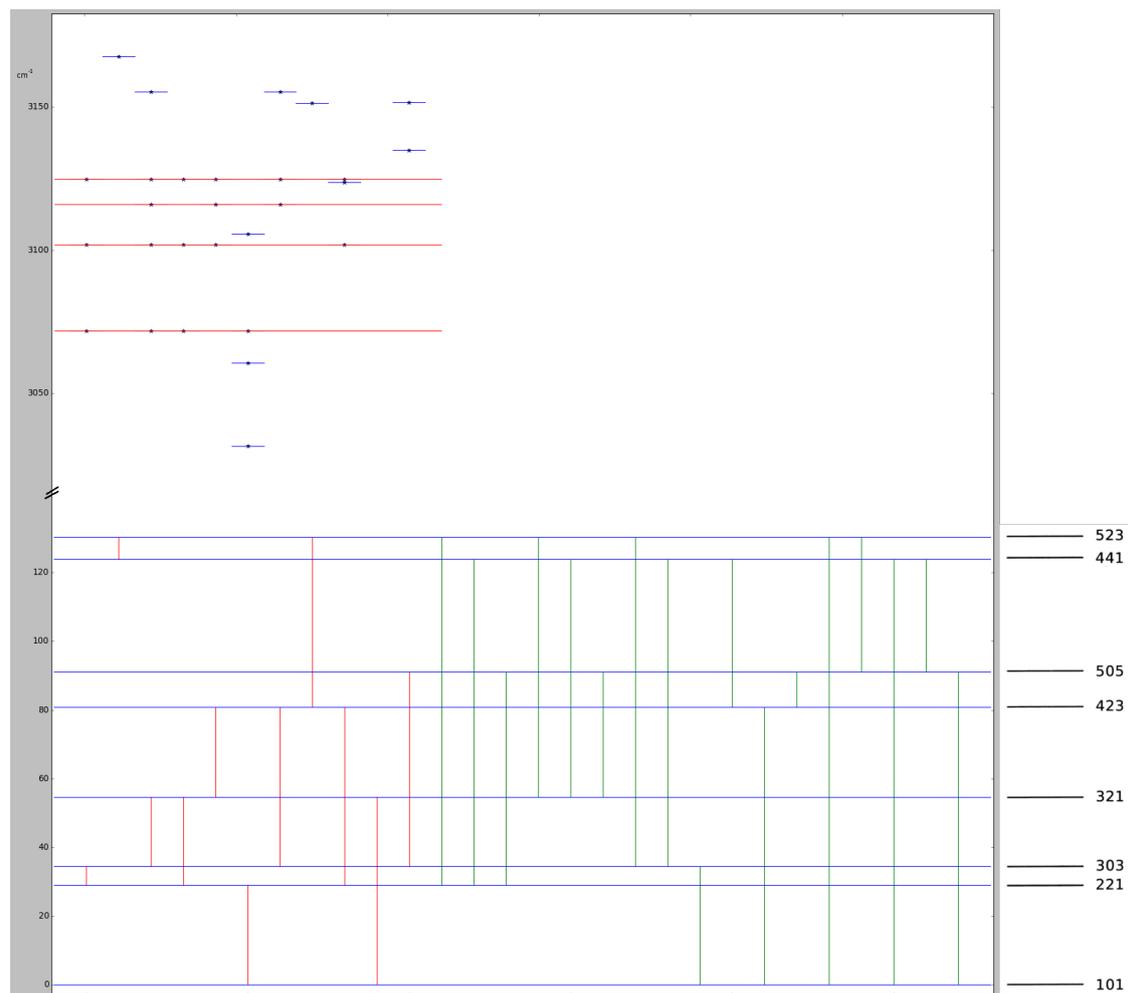
7 Ergebnisse

Grundzustände Jusko et al.			Grundzustände modellunabhängige Auswertung			
gsCD-Energie	Bottom-Level	Top-Level	CD-Peak-Energie $\pm 0,00005 \text{ cm}^{-1}$	Bottom-Level $\pm 0,00035 \text{ cm}^{-1}$	Top-Level $\pm 0,00035 \text{ cm}^{-1}$	# gesicherte es
8,4185	76,8636	85,2821	8,41850	76,86354 (66,79574)	85,28204 (75,21424)	4
10,3611	98,7175	109,0786				
13,4354	85,2821	98,7175				
13,6941	28,8243	42,5184	13,69321	28,82516 (18,75736)	42,51837 (32,45057)	3
18,7565	10,0678	28,8243	18,75769	10,06780 (0,00000)	28,82516 (18,75736)	0
21,8539	76,8636	98,7175				
23,7965	85,2821	109,0786	23,79651	85,28204 (75,21424)	109,07856 (99,01076)	4
32,2150	76,8636	109,0786	32,21503	76,86353 (66,79573)	109,07856 (99,01076)	3
32,4506	10,0678	42,5184	32,45057	10,06780 (0,00000)	42,51837 (32,45057)	2
34,3452	42,5184	76,8636	34,34517	42,51837 (32,45057)	76,86353 (66,79573)	5
37,4497	109,0786	146,5283				
42,7637	42,5184	85,2821	42,76368	42,51837 (32,45057)	85,28204 (75,21424)	3
47,8108	98,7175	146,5283				
48,0393	28,8243	76,8636	48,03926	28,82516 (18,75736)	76,86442 (66,79662)	3
53,8433	146,5283	200,3716				
56,1991	42,5184	98,7175				
56,4578	28,8243	85,2821	<i>56,45688</i>	<i>28,82516 (18,75736)</i>	<i>85,28204 (75,21424)</i>	
61,2462	85,2821	146,5283				
66,5602	42,5184	109,0786	66,56018	42,51837 (32,45057)	109,07856 (99,01076)	2
66,7958	10,0678	76,8636	66,79574	10,06780 (0,00000)	76,86353 (66,79573)	2
69,6647	76,8636	146,5283				
69,8932	28,8243	98,7175				
75,2143	10,0678	85,2821	<i>75,21424</i>	<i>10,06780 (0,00000)</i>	<i>85,28204 (75,21424)</i>	
80,2543	28,8243	109,0786	<i>80,25340</i>	<i>28,82516 (18,75736)</i>	<i>109,07856 (99,01076)</i>	
88,6497	10,0678	98,7175				
91,2930	109,0786	200,3716				
99,0108	10,0678	109,0786	<i>99,01076</i>	<i>10,06780 (0,00000)</i>	<i>109,07856 (99,01076)</i>	
101,6541	98,7175	200,3716				
104,0099	42,5184	146,5283				
109,0786	0,0000	109,0786				
115,0895	85,2821	200,3716				
117,7040	28,8243	146,5283				
123,5080	76,8636	200,3716				
136,4605	10,0678	146,5283				
157,8532	42,5184	200,3716				
171,5473	28,8243	200,3716				
190,3038	10,0678	200,3716				

<p>alle Energien in cm^{-1}</p>	<p>A2</p>							
<p>Modellunabhängig rekonstruierte Zustandsenergien sind zum besseren Vergleich um $10,0678 \text{ cm}^{-1}$ verschoben.</p> <p>In Klammern Originalwerte, die zum Plot passen.</p> <p><i>Kursive</i> Werte sind rekonstruiert ohne zugehörigen CD-Peak.</p>	<p>Gesicherte angeregte Zustände</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Energie $\pm 0,00035 \text{ cm}^{-1}$</th> </tr> </thead> <tbody> <tr><td>3077,85740 (3067,78960)</td></tr> <tr><td>3092,94496 (3082,87716)</td></tr> <tr><td>3114,62452 (3104,55672)</td></tr> <tr><td>3123,23900 (3113,17120)</td></tr> <tr><td>3151,52754 (3141,45974)</td></tr> <tr><td>3185,76610 (3175,69830)</td></tr> </tbody> </table>	Energie $\pm 0,00035 \text{ cm}^{-1}$	3077,85740 (3067,78960)	3092,94496 (3082,87716)	3114,62452 (3104,55672)	3123,23900 (3113,17120)	3151,52754 (3141,45974)	3185,76610 (3175,69830)
Energie $\pm 0,00035 \text{ cm}^{-1}$								
3077,85740 (3067,78960)								
3092,94496 (3082,87716)								
3114,62452 (3104,55672)								
3123,23900 (3113,17120)								
3151,52754 (3141,45974)								
3185,76610 (3175,69830)								

7.1.3 B1

Ergebnisse des Sets, das bei der modellgebundenen Auswertung B1-Symmetrie hat.



Rekonstruiertes Energie-Niveau-Schema des B1-Sets

Die CD-Peaks sind als Übergänge zwischen den Vibrationsgrundzuständen in rot geplottet. In grün Differenzen, die zwar rekonstruiert wurden, an deren Stelle im CD-Spektrum aber kein Peak zu finden ist. Jeweils oberhalb in blau mit Stern finden sich die aus dem jeweiligen Peak rekonstruierten vibrationsangeregten Zustände. Gesicherte vibrationsangeregte Zustände sind durch rote horizontale Linien markiert. Rechts daneben die Ergebnisse von Jusko et al.

7 Ergebnisse

Grundzustände Jusko et al.			Grundzustände modellunabhängige Auswertung			
gsCD-Energie	Bottom-Level	Top-Level	CD-Peak-Energie $\pm 0,00005 \text{ cm}^{-1}$	Bottom-Level $\pm 0,00035 \text{ cm}^{-1}$	Top-Level $\pm 0,00035 \text{ cm}^{-1}$	# gesicherte es
5,4375	36,5054	41,9429	5,43631	36,50940 (29,00380)	41,94571 (34,44011)	3
6,4207	131,4556	137,8763	6,42291	131,45802 (123,95242)	137,88093 (130,37533)	0
10,2155	88,4429	98,6584	<i>10,21859</i>	<i>88,44568 (80,94008)</i>	<i>98,66428 (91,15868)</i>	
20,1510	41,9429	62,0939	20,15103	41,94571 (34,44011)	62,09675 (54,59115)	4
25,5885	36,5054	62,0939	25,58672	36,50940 (29,00380)	62,09675 (54,59115)	3
26,3490	62,0939	88,4429	26,34894	62,09675 (54,59115)	88,44568 (80,94008)	3
28,9998	7,5056	36,5054	28,99924	7,50560 (0,00000)	36,50484 (28,99924)	1
32,7972	98,6584	131,4556	<i>32,79374</i>	<i>98,66428 (91,15868)</i>	<i>131,45802 (123,95242)</i>	
32,9694	137,8763	170,8457				
34,4373	7,5056	41,9429	<i>34,44011</i>	<i>7,50560 (0,00000)</i>	<i>41,94571 (34,44011)</i>	
36,5645	62,0939	98,6584	<i>36,56753</i>	<i>62,09675 (54,59115)</i>	<i>98,66428 (91,15868)</i>	
39,2179	98,6584	137,8763	<i>39,21665</i>	<i>98,66428 (91,15868)</i>	<i>137,88093 (130,37533)</i>	
39,3901	131,4556	170,8457				
43,0127	88,4429	131,4556	<i>43,01233</i>	<i>88,44568 (80,94008)</i>	<i>131,45802 (123,95242)</i>	
46,5000	41,9429	88,4429	46,49996	41,94571 (34,44011)	88,44568 (80,94008)	2
49,4334	88,4429	137,8763	49,43524	88,44568 (80,94008)	137,88093 (130,37533)	0
51,9375	36,5054	88,4429	51,93787	36,50940 (29,00380)	88,44568 (80,94008)	2
54,5883	7,5056	62,0939	54,59115	7,50560 (0,00000)	62,09675 (54,59115)	0
56,7155	41,9429	98,6584	56,71857	41,94571 (34,44011)	98,66428 (91,15868)	0
62,1530	36,5054	98,6584	<i>62,15426</i>	<i>36,50940 (29,00380)</i>	<i>98,66428 (91,15868)</i>	
69,3617	62,0939	131,4556	<i>69,36127</i>	<i>62,09675 (54,59115)</i>	<i>131,45802 (123,95242)</i>	
72,1873	98,6584	170,8457				
75,7824	62,0939	137,8763	<i>75,78418</i>	<i>62,09675 (54,59115)</i>	<i>137,88093 (130,37533)</i>	
80,9373	7,5056	88,4429	<i>80,94008</i>	<i>7,50560 (0,00000)</i>	<i>88,44568 (80,94008)</i>	
82,4028	88,4429	170,8457				
89,5127	41,9429	131,4556	<i>89,51231</i>	<i>41,94571 (34,44011)</i>	<i>131,45802 (123,95242)</i>	
91,1528	7,5056	98,6584	<i>91,15868</i>	<i>7,50560 (0,00000)</i>	<i>98,66428 (91,15868)</i>	
94,9502	36,5054	131,4556	<i>94,94800</i>	<i>36,50940 (29,00380)</i>	<i>131,45802 (123,95242)</i>	
95,9334	41,9429	137,8763	<i>95,93522</i>	<i>41,94571 (34,44011)</i>	<i>137,88093 (130,37533)</i>	
101,3709	36,5054	137,8763	<i>101,37091</i>	<i>36,50940 (29,00380)</i>	<i>137,88093 (130,37533)</i>	
108,7518	62,0939	170,8457				
123,9500	7,5056	131,4556	<i>123,95242</i>	<i>7,50560 (0,00000)</i>	<i>131,45802 (123,95242)</i>	
128,9028	41,9429	170,8457				
130,3707	7,5056	137,8763	<i>130,37533</i>	<i>7,50560 (0,00000)</i>	<i>137,88093 (130,37533)</i>	
134,3403	36,5054	170,8457				
163,3401	7,5056	170,8457				

alle Energien in cm^{-1}

B1

Modellunabhängig rekonstruierte Zustandsenergien sind zum besseren Vergleich um $7,5056 \text{ cm}^{-1}$ verschoben.

In Klammern Originalwerte,
die zum Plot passen.

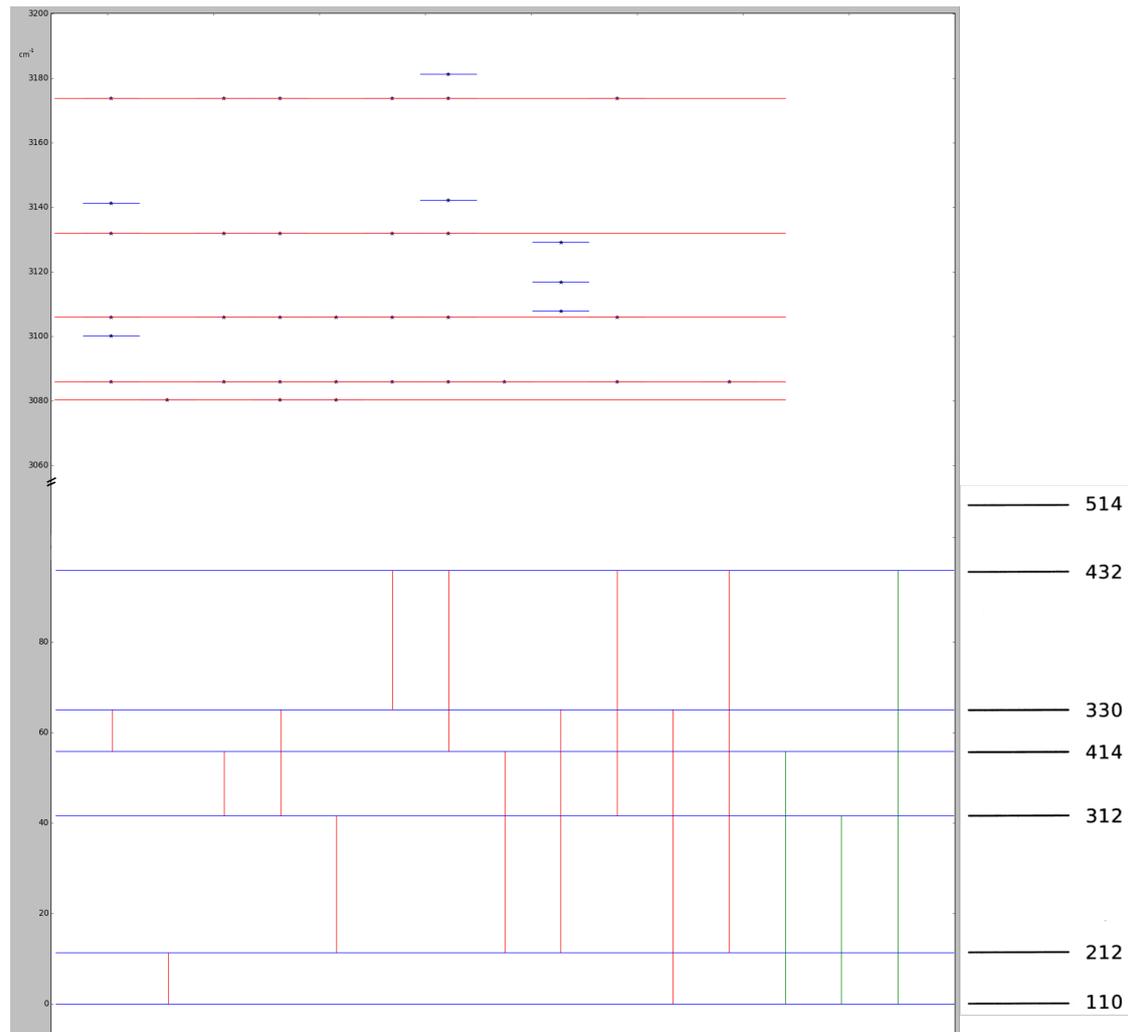
Kursiv: rekonstruiert ohne CD-Peak

Gesicherte angeregte Zustände

Energie $\pm 0,00035 \text{ cm}^{-1}$
3079,1794976 (3071,6738976)
3109,4829036 (3101,9773036)
3123,4132202 (3115,9076202)
3132,2099696 3124,7043696)

7.1.4 B2

Ergebnisse des Sets, das bei der modellgebundenen Auswertung B2-Symmetrie hat.



Rekonstruiertes Energie-Niveau-Schema des B2-Sets

Die CD-Peaks sind als Übergänge zwischen den Vibrationsgrundzuständen in rot geplottet. In grün Differenzen, die zwar rekonstruiert wurden, an deren Stelle im CD-Spektrum aber kein Peak zu finden ist. Jeweils oberhalb in blau mit Stern finden sich die aus dem jeweiligen Peak rekonstruierten vibrationsangeregten Zustände. Gesicherte vibrationsangeregte Zustände sind durch rote horizontale Linien markiert. Rechts daneben die Ergebnisse von Jusko et al.

7 Ergebnisse

Grundzustände Jusko et al.			Grundzustände modellunabhängige Auswertung			
gsCD-Energie	Bottom-Level	Top-Level	CD-Peak-Energie $\pm 0,00005 \text{ cm}^{-1}$	Bottom-Level $\pm 0,00035 \text{ cm}^{-1}$	Top-Level $\pm 0,00035 \text{ cm}^{-1}$	# gesicherte es
9,3044	67,7455	77,0499	9,30486	67,74505 (55,80305)	77,04992 (65,10792)	4
11,2620	11,9420	23,2040	11,26205	11,94200 (0,00000)	23,20405 (11,26205)	1
14,1787	53,5668	67,7455	14,17822	53,56687 (41,62487)	67,74509 (55,80309)	4
14,9981	107,9040	122,9021				
23,4831	53,5668	77,0499	23,48305	53,56687 (41,62487)	77,04992 (65,10792)	5
27,4758	122,9021	150,3779				
30,3628	23,2040	53,5668	30,36282	23,20405 (11,26205)	53,56687 (41,62487)	3
30,8541	77,0499	107,9040	30,85424	77,04992 (65,10792)	107,90416 (95,96216)	4
40,1585	67,7455	107,9040	40,15855	67,74505 (55,80305)	107,90416 (95,96216)	4
41,6248	11,9420	53,5668	<i>41,62487</i>	<i>11,94200 (0,00000)</i>	<i>53,56687 (41,62487)</i>	
42,4739	107,9040	150,3779				
44,5415	23,2040	67,7455	44,54149	23,20405 (11,26205)	67,74509 (55,80309)	1
45,8522	77,0499	122,9021				
50,0013	150,3779	200,3792				
53,8459	23,2040	77,0499	53,84729	23,20405 (11,26205)	77,04992 (65,10792)	0
54,3372	53,5668	107,9040	54,33725	53,56687 (41,62487)	107,90416 (95,96216)	3
55,1566	67,7455	122,9021				
55,8035	11,9420	67,7455	<i>55,80309</i>	<i>11,94200 (0,00000)</i>	<i>67,74509 (55,80309)</i>	
65,1079	11,9420	77,0499	65,11079	11,94200 (0,00000)	77,04992 (65,10792)	0
69,3353	53,5668	122,9021				
73,3280	77,0499	150,3779				
77,4771	122,9021	200,3792				
82,6324	67,7455	150,3779				
84,7000	23,2040	107,9040	84,70006	23,20405 (11,26205)	107,90416 (95,96216)	1
92,4752	107,9040	200,3792				
95,9620	11,9420	107,9040	<i>95,96216</i>	<i>11,94200 (0,00000)</i>	<i>107,90416 (95,96216)</i>	
96,8111	53,5668	150,3779				
99,6981	23,2040	122,9021				
110,9601	11,9420	122,9021				
123,3293	77,0499	200,3792				
127,1739	23,2040	150,3779				
132,6337	67,7455	200,3792				
138,4359	11,9420	150,3779				
146,8124	53,5668	200,3792				
177,1752	23,2040	200,3792				
188,4372	11,9420	200,3792				

alle Energien in cm^{-1}

Modellunabhängig rekonstruierte Zustandsenergien sind zum besseren Vergleich um $11,9420 \text{ cm}^{-1}$ verschoben.

In Klammern Originalwerte, die zum Plot passen.

Kursive Werte sind rekonstruiert ohne zugehörigen CD-Peak.

B2

Gesicherte angeregte Zustände

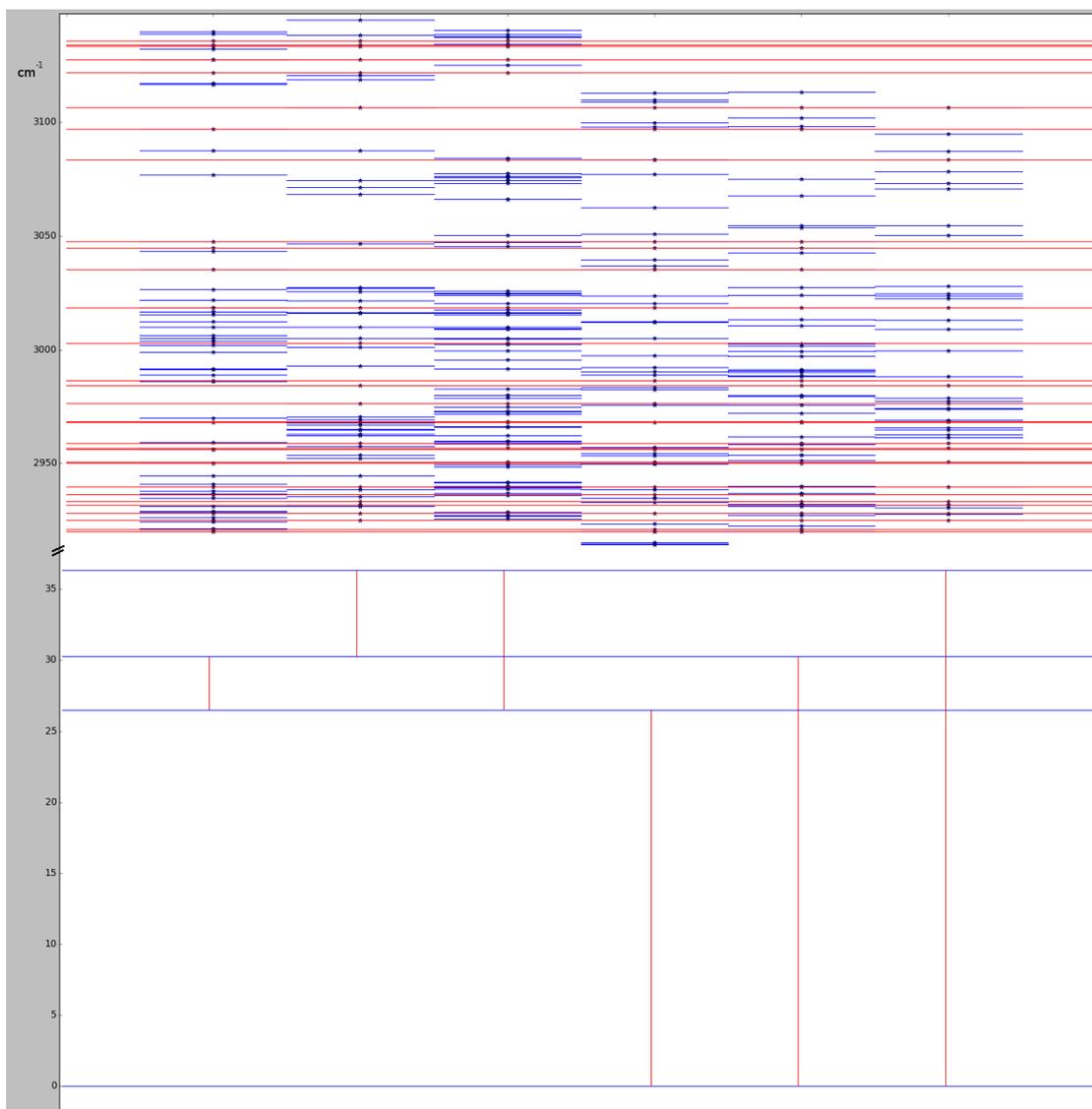
Energie $\pm 0,00035 \text{ cm}^{-1}$
3092,20433 (3080,26233)
3097,84091 (3085,89891)
3117,89369 (3105,95169)
3143,90630 (3131,96430)
3185,71674 (3173,77474)

7.2 CH₅⁺

Im Rahmen dieser Diplomarbeit konnten sämtliche CD-Peaks, die in [Asvany et al. 2015] einem Set zugeordnet wurden, reproduziert und auf gleiche Weise zugeordnet werden. Darüber hinaus wurden die Sets erweitert und die beiden anderen nach [Schmiedt et al. 2016] zu erwartenden Sets gefunden. Dabei konnten die bereits von Asvany et al. gefundenen, aber nicht zugeordneten Peaks bei 5.46036 cm^{-1} , 8.25267 cm^{-1} und 29.76214 cm^{-1} und zahlreiche neue CD-Peaks zugeordnet werden. Schwierigkeiten bereiten dagegen die von Asvany et al. bei 7.04935 cm^{-1} , 36.91995 cm^{-1} und 46.46478 cm^{-1} gefundenen (und nicht zugeordneten) Peaks: Sie stehen nur bei Verwendung eines Kastenkerne hervor und auch dann nur sehr unzuverlässig und stark abhängig von der Wahl der Parameter. Daher liegt die Vermutung nahe, dass ihre Auffälligkeit im CD-Spektrum vor allem ein Artefakt ist, das wie in Abschnitt 3.1 beschrieben typisch für Kastenkerne ist. Alle drei problematischen CD-Peaks sind zwar auch in CD-Spektren, die mit einem Gaußkern oder einem Kastenkern mit anderen Parametern erstellt wurden, prinzipiell auffindbar, unterscheiden sich dort aber nicht von anderen Stacheln des Rauschens. Verwendet man die in Abschnitt 4.2 beschriebene Rauschunterdrückung, verschwinden die CD-Peaks bei 7.04935 cm^{-1} und 36.91995 cm^{-1} komplett, wohingegen der Peak bei 46.46478 cm^{-1} auch mit Rauschunterdrückung erhalten bleibt. Dies legt nahe, dass es sich bei letzterem tatsächlich um die Differenz zweier Vibrationsgrundzustände handelt, die aber auf Grund ihrer hohen Energie bei den Messtemperaturen nur schwach sichtbar ist und noch nicht zugeordnet werden konnte.

Im Folgenden werden die vier rekonstruierten Sets dargestellt. Darüber hinaus findet sich in Anhang A.1 eine Tabelle, in der vermerkt ist, welche von Asvany et al. gemessenen Linien welchem Set zugeordnet werden konnten. In Anhang A.2 findet sich zudem eine Liste der als gesichert identifizierten gsCDs und ihrer Zuordnungen zu den Sets. Dabei ist jeweils gekennzeichnet, welche Ergebnisse bereits von Asvany et al. ermittelt wurden.

7.2.1 Set 1

**Rekonstruiertes Energie-Niveau-Schema von Set 1**

Die CD-Peaks sind als Übergänge zwischen den Vibrationsgrundzuständen in rot geplottet. Jeweils oberhalb in blau mit Stern finden sich die aus dem jeweiligen Peak rekonstruierten vibrationsangeregten Zustände. Gesicherte vibrationsangeregte Zustände sind durch rote horizontale Linien markiert.

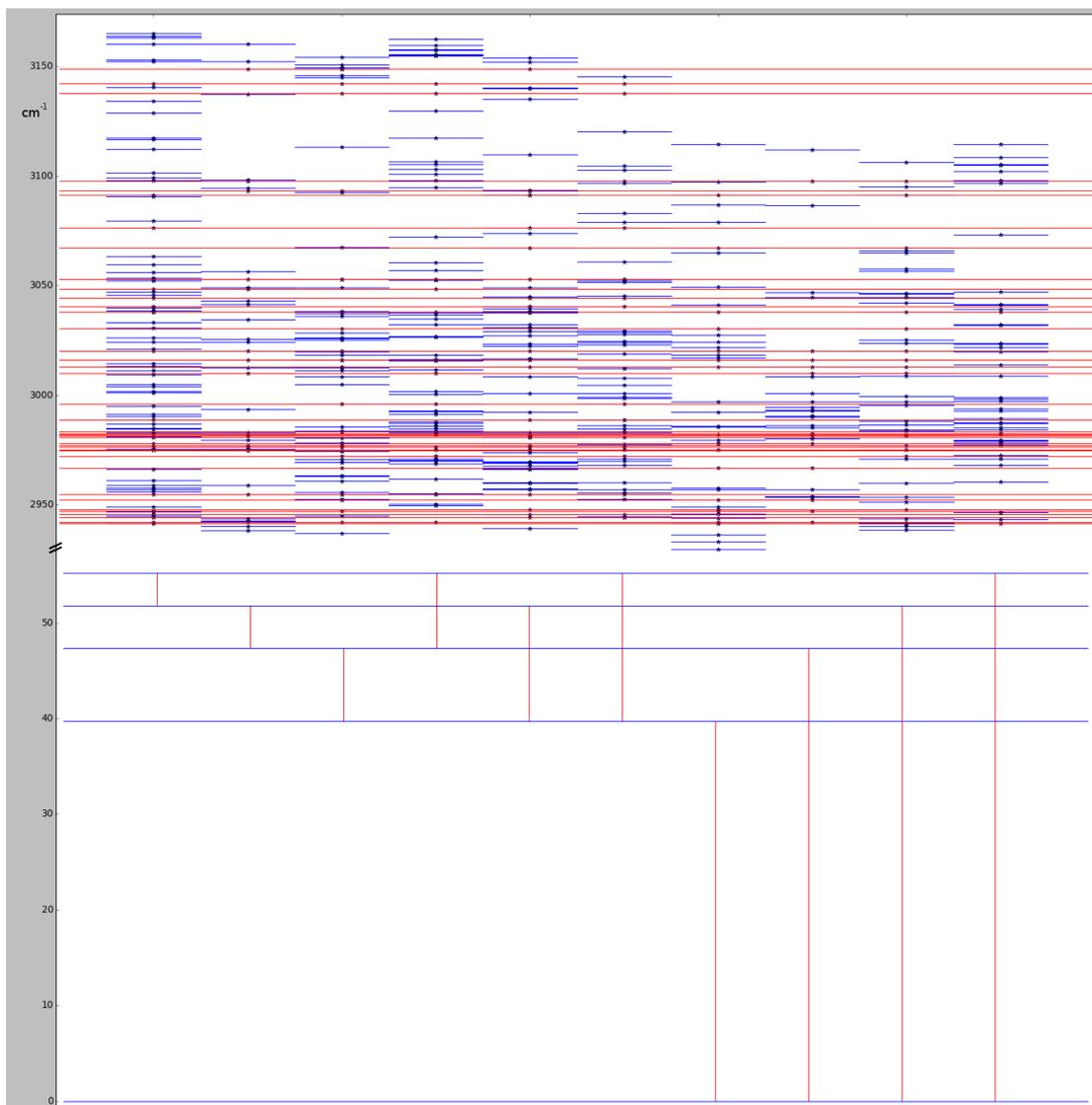
**Gesicherte
vibrationsangeregte Zustände von Set 1**

Energie [cm⁻¹] $\pm 0,00035 \text{ cm}^{-1}$	
2920,10239	2976,32378
2920,34089	2984,27141
2921,17118	2986,45637
2925,15153	3003,06672
2928,12339	3018,60883
2931,72084	3035,45938
2933,21981	3044,79549
2936,29310	3047,53703
2939,83950	3083,70769
2950,15092	3097,18596
2950,73413	3106,66326
2956,20562	3121,65004
2956,83667	3127,72767
2958,86260	3133,36849
2968,23137	3133,88125
2968,36578	3135,84135

Rekonstruierte Vibrationsgrundzustände von Set 1

CD-Peak-Energie [cm ⁻¹] $\pm 0,00005 \text{ cm}^{-1}$	Bottom-Level [cm ⁻¹] $\pm 0,00035 \text{ cm}^{-1}$	Top-Level [cm ⁻¹] $\pm 0,00035 \text{ cm}^{-1}$	# gesicherter vibrationsangeregter Zustände	von Asvany et al. gefunden und zugeordnet
3,79733	26,47250	30,26978	21	x
6,08304	30,26978	36,35282	18	
9,88030	26,47250	36,35282	14	x
26,47250	0,00000	26,47250	19	x
30,26979	0,00000	30,26979	22	x
36,35282	0,00000	36,35282	14	x

7.2.2 Set 2

**Rekonstruiertes Energie-Niveau-Schema von Set 2**

Die CD-Peaks sind als Übergänge zwischen den Vibrationsgrundzuständen in rot geplottet. Jeweils oberhalb in blau mit Stern finden sich die aus dem jeweiligen Peak rekonstruierten vibrationsangeregten Zustände. Gesicherte vibrationsangeregte Zustände sind durch rote horizontale Linien markiert.

7 Ergebnisse

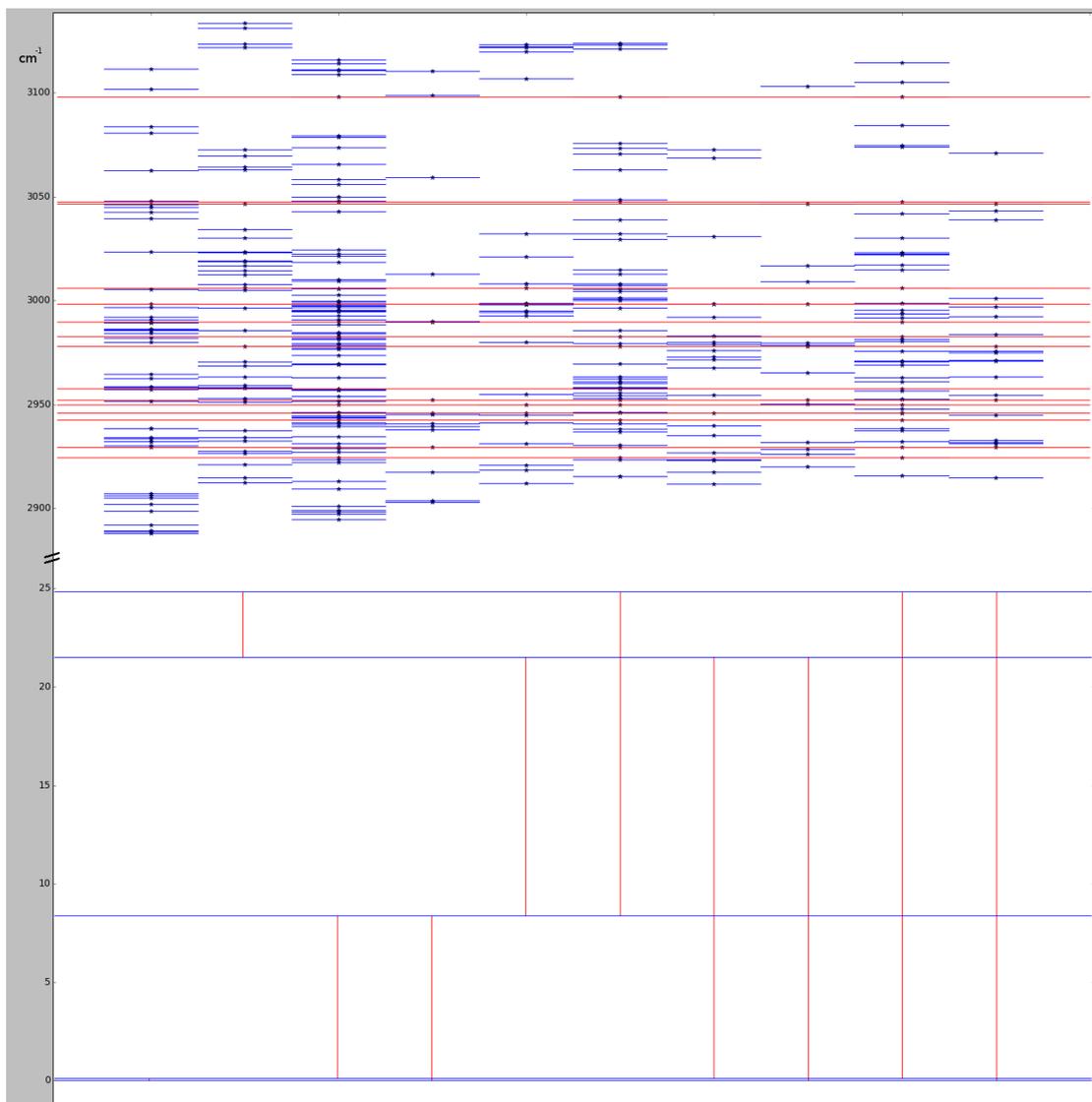
**Gesicherte
vibrationsangeregte Zustände von Set 2**

Energie [cm⁻¹] ±0,00035 cm ⁻¹	
2941,63379	2988,74430
2942,34574	2996,23410
2944,42281	3010,11904
2945,64199	3013,05975
2947,39739	3016,21931
2947,77156	3020,31964
2952,28244	3030,39276
2954,86838	3038,07911
2966,94439	3040,65973
2972,45707	3044,41448
2974,92971	3048,38110
2975,27852	3052,85525
2976,41522	3067,20326
2977,11171	3076,38728
2978,11323	3091,25189
2980,90053	3093,43162
2981,72328	3097,85922
2982,17796	3137,58677
2982,41876	3142,12635
2983,49322	3148,78704

Rekonstruierte Vibrationsgrundzustände von Set 2

CD-Peak- Energie [cm ⁻¹] ±0,00005 cm ⁻¹	Bottom- Level [cm ⁻¹] ±0,00035 cm ⁻¹	Top- Level [cm ⁻¹] ±0,00035 cm ⁻¹	# gesicherter vibrations- angeregter Zustände	von Asvany et al. gefunden und zugeordnet
3,37831	51,80073	55,17903	15	
4,47901	47,32216	51,80118	15	
7,63710	39,68506	47,32216	22	
7,85722	47,32216	55,17938	12	
12,11616	39,68506	51,80122	22	x
15,49431	39,68506	55,17938	17	
39,68506	0,00000	39,68506	12	x
47,32217	0,00000	47,32217	12	x
51,80115	0,00000	51,80115	12	x
55,17903	0,00000	55,17903	4	

7.2.3 Set 3

**Rekonstruiertes Energie-Niveau-Schema von Set 3**

Die CD-Peaks sind als Übergänge zwischen den Vibrationsgrundzuständen in rot geplottet. Jeweils oberhalb in blau mit Stern finden sich die aus dem jeweiligen Peak rekonstruierten vibrationsangeregten Zustände. Gesicherte vibrationsangeregte Zustände sind durch rote horizontale Linien markiert.

7 Ergebnisse

Gesicherte vibrationsangeregte Zustände von Set 3

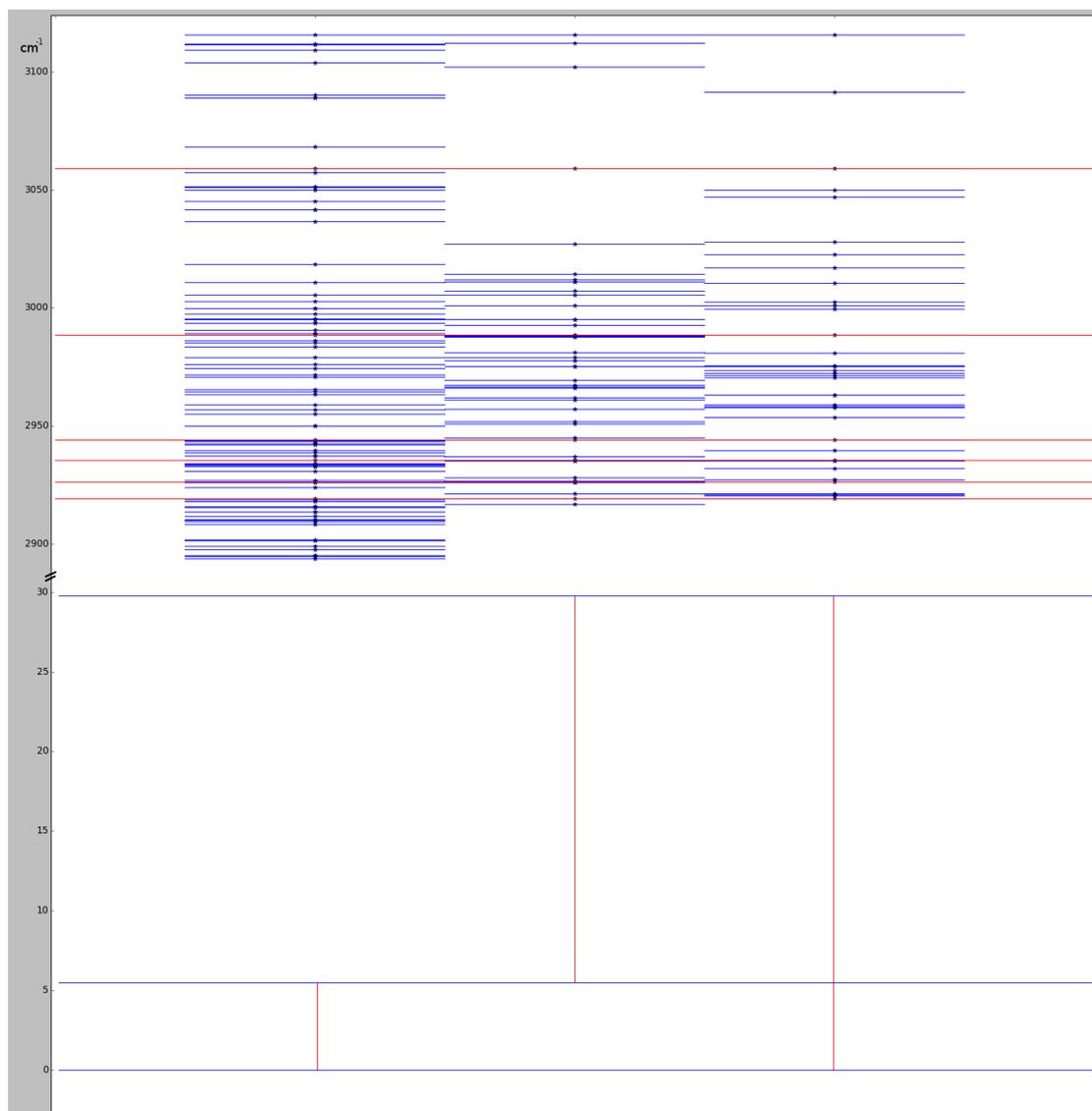
Energie [cm ⁻¹] ±0,00035 cm ⁻¹
2924,38037
2929,61895
2942,61488
2945,77441
2950,08653
2952,25153
2957,53917
2978,00756
2982,78661
2989,73539
2998,52340
3005,97509
3046,44068
3047,46409
3098,00780

Rekonstruierte Vibrationsgrundzustände von Set 3

CD-Peak- Energie [cm ⁻¹] ±0,00005 cm ⁻¹	Bottom- Level [cm ⁻¹] ±0,00035 cm ⁻¹	Top- Level [cm ⁻¹] ±0,00035 cm ⁻¹	# gesicherter vibrations- angeregter Zustände	von Asvany et al. gefunden aber nicht zugeordnet
0,13033	0,00000	0,13033	3	
3,30813	21,52500	24,83313	4	
8,25264	0,13033	8,38297	10	x
8,38295	0,00000	8,38295	3	
13,14129	8,38295	21,52425	3	
16,44998	8,38295	24,83293	9	
21,39406	0,13033	21,52439	2	
21,52500	0,00000	21,52500	4	
24,70234	0,13033	24,83267	9	
24,83271	0,00000	24,83271	4	

Set 3 wurde von Asvany et al. nicht gefunden.

7.2.4 Set 4

**Rekonstruiertes Energie-Niveau-Schema von Set 4**

Die CD-Peaks sind als Übergänge zwischen den Vibrationsgrundzuständen in rot geplottet. Jeweils oberhalb in blau mit Stern finden sich die aus dem jeweiligen Peak rekonstruierten vibrationsangeregten Zustände. Gesicherte vibrationsangeregte Zustände sind durch rote horizontale Linien markiert.

7 Ergebnisse

Gesicherte vibrationsangeregte Zustände von Set 4

Energie [cm ⁻¹] ±0,00035 cm ⁻¹
2919,24222
2926,24919
2935,42112
2944,12722
2988,55330
3059,13646

Rekonstruierte Vibrationsgrundzustände von Set 4

CD-Peak- Energie [cm ⁻¹] ±0,00005 cm ⁻¹	Bottom- Level [cm ⁻¹] ±0,00035 cm ⁻¹	Top- Level [cm ⁻¹] ±0,00035 cm ⁻¹	# gesicherter vibrations- angeregter Zustände	von Asvany et al. gefunden aber nicht zugeordnet
5,46034	0,00000	5,46034	6	x
24,30169	5,46034	29,76204	6	
29,76215	0,00000	29,76204	6	x

Set 4 wurde von Asvany et al. nicht gefunden.

8 Fazit und Ausblick

Im Rahmen dieser Diplomarbeit hat sich heraus gestellt, dass der verfolgte Ansatz erfolgversprechend ist. Einerseits können die ermittelten Resultate genutzt werden, um CH_5^+ weiter zu verstehen. Andererseits lässt sich die entwickelte Auswertungsmethode verbessern und zu einem auch bei anderen Molekülen nützlichen Werkzeug weiterentwickeln.

8.1 Was bedeuten die Ergebnisse für die Arbeit an CH_5^+ ?

Bei der weiteren Untersuchung von CH_5^+ haben die Ergebnisse dieser Arbeit zweierlei Bedeutung:

Wie eingangs erwähnt funktionieren die Standardmodelle für CH_5^+ nicht. [Schmiedt et al. 2016] arbeiten deshalb an einer verallgemeinerten Theorie und können damit bereits Vibrationsgrundzustände von CH_5^+ berechnen. Die Ergebnisse dieser Arbeit ermöglichen es besser als bisher, die Vorhersagen dieser neuen Theorie mit dem Experiment zu vergleichen. Die ersten Versuche in diese Richtung sehen bereits vielversprechend aus.

Wie ebenfalls eingangs erwähnt müssen die von Asvany et al. verwendeten Techniken LIR und LIICG weiter entwickelt werden, um Übergänge zwischen Vibrationsgrundzuständen direkt zu vermessen. Eine Möglichkeit dafür ist die Verwendung Infrarot-mm-Wellen-Doppelresonanzmethode, die auch in [Jusko et al. 2016] verwendet wurde. Für diese Methode setzt man aber zumindest grundlegende Kenntnisse über die Lage der vibrationsangeregten Zustände voraus (siehe Fußnote 3 in der Einführung). Ein Teil dieser vibrationsangeregten Zustände ist in dieser Arbeit erstmals bestimmt worden, sodass nun eine Grundlage zur Vermessung der Übergänge zwischen den Vibrationsgrundzuständen besteht. Dies ist im Hinblick auf die von Schmiedt et al. entwickelte neue Theorie besonders interessant, weil diese für die Übergänge zwischen den Vibrationsgrundzuständen andere Auswahlregeln vorhersagt als die bisherigen (bei CH_5^+ nicht funktionierenden) Theorien.

8.2 Potenzial und Weiterentwicklung der modellosen Auswertungsmethode

Die Bewertung der Relevanz der Rekonstruktion der Energieniveaus eines Moleküls im Einzelnen sprengt den Rahmen dieser Arbeit. Prinzipiell kann sie aber aus folgenden Gründen grundlegend zum Verständnis des Moleküls beitragen:

- Die Energien der Zustände enthalten mit Ausnahme der Auswahlregeln die gleichen Informationen wie das gemessene Spektrum, bestehen aber aus deutlich weniger Daten: Ein Molekül mit n Zuständen hat ein Spektrum mit $\mathcal{O}(n^2)$ Linien.
- Dies vereinfacht die Entwicklung und Anpassung eines Modells, z. B. in Form eines praktisch verwendbaren Hamiltonians, erheblich.
- Sind die Energieniveaus rekonstruiert und dabei gleichzeitig die gemessenen Linien zugeordnet, kennt man auch Auswahlregeln (im Spektrum vorhandene versus fehlende Übergänge)

bzw. Übergangswahrscheinlichkeiten (Linienintensitäten). Dies ist ebenfalls für die Modellentwicklung hilfreich. Beispielsweise kann die Reihenfolge der Energieniveaus bei einem Modell von der Wahl der Modellparameter abhängen, sodass die Zuordnung von Energieniveaus des Modells zu den aus Messungen rekonstruierten Niveaus nicht eindeutig ist; hier können Informationen über die Auswahlregeln falsche Möglichkeiten auszuschließen helfen.

[Furtenbacher et al. 2015] präsentieren eine möglichst genaue Rekonstruktion der Energie-Niveaus zahlreicher Moleküle auf Grundlage von aus verschiedenen Quellen stammenden Messdaten. Sie greifen dabei auf bekannte Zuordnungen der gemessenen Übergänge zurück. Diese Zuordnungen stammen im Wesentlichen aus Plausibilitätsbetrachtungen auf Grund ganz unterschiedlicher Argumente und Modelle und sind über zahlreiche Veröffentlichungen verstreut. Dies zeigt, dass eine systematische Zuordnung von Übergängen allein auf Grund von Messwerten ohne weitere Modellannahmen bisher nicht stattgefunden hat. Mit Hilfe der in dieser Arbeit (weiter-)entwickelten Methode könnte dies für viele gemessene, aber nicht zugeordnete Übergänge nachgeholt werden. Auch wenn diese Methode die Energien der Zustände selbst mit rekonstruiert, kann sie bezüglich der Genauigkeit mit der von Furtenbacher et al. verwendeten Methode, die z. B. die Messfehler der einzelnen, aus verschiedenen Quellen stammenden Daten in einem elaborierten Verfahren mit berücksichtigt, sicherlich nicht mithalten. Deshalb scheint es sinnvoll zu sein, zunächst die Zuordnung und grobe Bestimmung der Energien der Molekülzustände mit dem hier vorgestellten Verfahren vorzunehmen und darauf aufbauend mit dem von Furtenbacher et al. verwendeten Verfahren weiter zu arbeiten, um die Energien noch einmal präziser zu bestimmen.

Grundsätzlich lässt sich die vorgestellte Methode zur Rekonstruktion des Energie-Niveaus-Schemas eines Moleküls immer dann verwenden, wenn Übergänge zwischen einem Bereich des Energie-Niveau-Schemas mit vielen und einem Bereich mit wenigen Zuständen vermessen wurden. Dann nämlich tauchen die Abstände innerhalb des Bereiches mit wenigen Zuständen als Peaks im CD-Spektrum auf (vergleiche Abschnitt 1.2.1). Diese Situation lässt sich fast immer herstellen: Dafür sind Absorptionsmessungen bei einer verhältnismäßig niedrigen Temperatur durchzuführen und sicherzustellen, dass der Messbereich so groß ist, dass die Anzahl der in die Messung involvierten oberen Niveaus wesentlich größer ist als die Anzahl der bei Messtemperatur besetzten Niveaus. Dies kann ggf. auch dadurch erreicht werden, dass Messungen in verschiedenen Spektralbereichen ggf. mit unterschiedlichen Techniken, die aber bei etwa der gleichen Temperatur stattgefunden haben, zusammen verwendet werden.

Weiterentwicklung des Programms Soll das entwickelte Verfahren breitere Anwendung finden, sollte das im Rahmen dieser Diplomarbeit erstellte Programm zuvor noch weiterentwickelt werden:

- Ähnlich wie die Implementierung einer halbwegs treffsicheren automatischen Bandbreitenwahl gelungen ist, könnte viel Zeit und eher aufreibende Arbeit gespart werden, wenn die Parameter der Rauschunterdrückung sinnvoll automatisch bestimmt würden. Wie bei der Bandbreitenwahl ist es sicherlich nicht trivial eine vollkommen treffsichere Automatik zu implementieren; so liefert die automatische Bandbreitenwahl — wie sich an den CD_2H^+ -Daten gezeigt hat — bei Spektren geringer Dichte eine tendenziell etwas zu große Bandbreite. Dennoch dürfte es möglich sein, eine automatische Parameterwahl zu entwickeln, die zumindest sinnvoll verwendbare Parameter, die bei Bedarf per Hand weiter verbessert werden können, liefert.
- Ebenso kämen die bereits in Abschnitt 6.2 dargelegten Verbesserungen des Programmablaufs der unkomplizierten Nutzbarkeit sehr zu Gute.

8 Fazit und Ausblick

- Auch dann noch erfordert die Nutzung des Programms ziemlich viel Mitdenken, Notizen usw., denn viele Informationen werden zwar vom Programm zu im Allgemeinen nützlichen Zeitpunkten präsentiert, aber später nochmal gebraucht etc. Deshalb böte sich die Entwicklung einer interaktiven graphischen Benutzeroberfläche an, was aber faktisch auf eine komplette Reimplementierung hinaus liefe: Der Programmumfang ist mit der Zeit erheblich gewachsen. Deshalb sind sowohl der prozedurale Programmablauf als auch die programminterne Zwischenspeicherung der Daten für so ein Projekt eher schlecht geeignet. Eine interaktive Bedienung führt dazu, dass einzelne Programmschritte unabhängig voneinander neu ausgeführt und die jeweils zugehörigen Daten aktualisiert werden. Dies führte mit der bisherigen Programmkonzeption dazu, dass das Programm extrem komplex, unübersichtlich und fehleranfällig würde. Sinnvoll wäre für solch ein Projekt deshalb eine objektorientierte Programmgestaltung und die Sammlung sämtlicher Daten in einer externen Datenbank. Dies würde es z. B. auch auf einfache Art erlauben, Bearbeitungsschritte rückgängig zu machen, Daten einfach und ohne zusätzlichen Programmcode zu im- und exportieren etc.

A Ergänzendes Material

Auf der eingelebten CD finden sich zusätzliche Materialien.

A.1 Zuordnung der Linien von CH_5^+ zu Sets

In der Datei `CH5+_10K_LIR-Zuordnungen.csv` im Ordner `Ergebnisse_CH5+` ist zu jeder Linie des 10 K-Spektrums von [Asvany et al. 2015] aufgeführt, welchem Set sie im Rahmen dieser Diplomarbeit zugeordnet wurde. In einer weiteren Spalte finden sich die Zuordnungen von Asvany et al., die bisher nur als Plot veröffentlicht wurden. Zusätzlich enthält die Datei `CH5+_10K_LIR.dat_zugeordnet.npy` eine Binärversion der Zuordnungen, die man mit Hilfe der Skripte zur Weiterarbeit verwenden kann.

A.2 Gesicherte gsCDs von CH_5^+

Die Datei `CH5+_10K_LIR-gesicherte_gsCDs.csv` im Ordner `Ergebnisse_CH5+` enthält eine Liste von aus dem 10 K-Spektrum von [Asvany et al. 2015] berechneten CDs, die als gsCDs identifiziert werden konnten, sowie die Linien, aus denen sie berechnet wurden, und deren Zuordnungen zu einem Set. Zusätzlich enthält die Datei `CH5+_10K_LIR.dat_wahreGSCDs.npy` eine Binärversion der Liste, die man mit Hilfe der Skripte zur Weiterarbeit verwenden kann.

A.3 Skripte

Im Ordner `scripts` finden sich sämtliche für diese Arbeit verwendeten Skripte in der jeweils zuletzt verwendeten Version und in einer Ordnerstruktur, in der sie arbeitsfähig sind.

Literaturverzeichnis

- [Asvany et al. 2005] O. Asvany, P. Kumar, B. Redlich, I. Hegemann, S. Schlemmer, D. Marx: *Understanding the Infrared Spectrum of Bare CH₅⁺*. Science Volume 309 Issue 5738, 2005, Seiten 1219–1222
- [Asvany et al. 2014] O. Asvany, S. Brünken, L. Kluge, S. Schlemmer: *COLTRAP: a 22-pole ion trapping machine for spectroscopy at 4 K*. Applied Physics B Volume 114, Issue 1, 2014, Seiten 203–211
- [Asvany et al. 2015] O. Asvany, K. M. T. Yamada, S. Brünken, A. Potapov, S. Schlemmer: *Experimental ground-state combination differences of CH₅⁺*. Science Volume 347 Issue 6228, 2015, Seiten 1346–1349
- [Camps-Valls et al. 2007] G. Camps-Valls, J. L. Rojo-Álvarez, M. Martínez-Ramón (Herausgeber): *Kernel Methods in Bioengineering, Signal and Image Processing*. Idea Group Publishing, London, 2007
- [Furtenbacher et al. 2015] T. Furtenbacher, A. Császár, J. Tennyson: *MARVEL: measured active rotational–vibrational energy levels*. Journal of Molecular Spectroscopy Volume 245 Issue 2, 2007, Seiten 115–125
- [Guidoum 2014] A. C. Guidoum: *kedd: Kernel estimator and bandwidth selection for density and its derivatives*. R package version 1.0.3, 2015, Permanent-URL <http://CRAN.R-project.org/package=kedd>
- [Huang et al. 2006] X. Huang, A. B. McCoy, J. M. Bowman, L. M. Johnson, C. Savage, F. Dong, D. J. Nesbitt: *Quantum Deconstruction of the Infrared Spectrum of CH₅⁺*, Science Volume 311 Issue 5757, 2006, Seiten 60–63
- [Jin et al. 2006] Z. Jin, B. J. Braams, J. M. Bowman: *An ab Initio Based Global Potential Energy Surface Describing CH₅⁺ → CH₃⁺ + H₂*. Journal of Physical Chemistry A Volume 110 No. 4, 2006, Seiten 1569–1574
- [Jusko et al. 2016] P. Jusko, S. Thorwirth, S. Brünken, S. Schlemmer, O. Asvany: *High-resolution vibrational and rotational spectroscopy of CD₂H⁺ in a cryogenic ion trap*. In Vorbereitung, Publikation voraussichtlich 2016
- [Kumar et al. 2006] P. Kumar, D. Marx: *Understanding hydrogen scrambling and infrared spectrum of bare CH₅⁺ based on ab initio simulations*. Physical Chemistry Chemical Physics Volume 8 Issue 5, 2006, Seiten 573–586
- [pgopher] Projektwebseite: <http://pgopher.chm.bris.ac.uk>
- [Schlemmer et al. 2002] S. Schlemmer, E. Lescop, J. v. Richtofen, D. Gerlich: *Laser induced reactions in a 22-pole ion trap: C₂H₂⁺ + hν₃ + C₂H₃⁺ + H*. Journal of Physical Chemistry Volume 110 No. 5, 2002, Seiten 2068–2075

Literaturverzeichnis

- [Schmiedt et al. 2015] H. Schmiedt, S. Schlemmer, P. Jensen: *Symmetry of extremely floppy molecules: Molecular states beyond rotation-vibration separation*. Journal of Chemical Physics Volume 143, Number 15, 2015
- [Schmiedt et al. 2016] H. Schmiedt, P. Jensen, S. Schlemmer: *A unified theory of rotational and vibrational motion in floppy molecules*. In Vorbereitung, Publikation voraussichtlich 2016
- [Silverman 1986] B. W. Silverman: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, London, 1986
- [Simonoff 1996] J. S. Simonoff: *Smoothing methods in statistics*. Springer-Verlag, New York 1996
- [Wand et al. 1995] M. P. Wand, M. C. Jones: *Kernel Smoothing*. Chapman & Hall, London, 1995
- [Wertz 1978] W. Wertz: *Statistical Density Estimation: A Survey*. Vandenhoeck & Ruprecht, Göttingen, 1978
- [White et al. 1999] E. T. White, J. Tang, T. Oka: CH_5^+ : *The Infrared Spectrum Observed*. Science Volume 284 Issue 5411, 1999, Seiten 135–137
- [Wikipedia 2008] Wikipedia, von Briam Amberg (eigenes Werk): *A range of different kernels*. Licensed under the Creative Commons Attribution-ShareAlike 3.0 License Permanent-URL: <https://commons.wikimedia.org/w/index.php?title=File:Kernels.svg&oldid=138917163> Artikel mit Bildbeschreibung: https://en.wikipedia.org/wiki/Kernel_%28statistics%29
- [Wikipedia 2010] Wikipedia, von Driest (eigenes Werk): *Comparison of 1D bandwidth selectors*. Licensed under the Creative Commons Attribution-ShareAlike 3.0 License Permanent-URL: https://en.wikipedia.org/w/index.php?title=File:Comparison_of_1D_bandwidth_selectors.png&oldid=547231554 Artikel mit Bildbeschreibung: https://en.wikipedia.org/wiki/Kernel_density_estimation
- [Wolfgang et al. 2004] W. Härdle, M. Müller, S. Sperlich, A. Werwatz: *Nonparametric and Semiparametric Models*. Springer-Verlag, Berlin Heidelberg, 2004
- [Zambom et al. 2012] A. Z. Zambom, R. Dias: *A Review of Kernel Density Estimation with Applications to Econometrics*. Eingereicht bei arXiv von Ronaldo Dias, 12.2012, <http://arxiv.org/abs/1212.2812v1>

Danksagung

Bei allen, die durch Rat und Tat zum Gelingen dieser Diplomarbeit beigetragen haben, möchte ich mich herzlich bedanken, insbesondere bei Oskar Asvany, Hanno Schmiedt und Stephan Schlemmer. Entscheidend für meine Motivation war stets, dass die Arbeit nicht nur dazu diente, zu beweisen, dass man eine Arbeit schreiben kann. Vielmehr hatte ich die Möglichkeit, an einer relevanten Fragestellung zu arbeiten, bei der die Kollegen echtes Interesse an den Ergebnissen haben. Dies bildete eine Einheit mit der kollegialen Zusammenarbeit in der Arbeitsgruppe, besonders mit den Kollegen des Problemlösebüros.

„Früher war mehr Lametta.“

Opa Hoppenstedt

Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Sämtliche Zitate habe ich kenntlich gemacht.

Köln, den