

Diplomarbeit

Simulation einer Multikapillarsäule

Elisabeth Böhmer
15. September 2015

Betreuer:
Prof. Dr. Sven Rahmann
Prof. Dr. Jörg Rahnenführer

Fakultät für Informatik
Algorithm Engineering (Ls11)
Technische Universität Dortmund
<http://ls11-www.cs.tu-dortmund.de>

Inhaltsverzeichnis

1	Einleitung	1
1.1	Ziel der Arbeit	2
1.2	Andere Arbeiten	2
1.3	Aufbau der Arbeit	3
2	Grundlagen	5
2.1	Chromatographie	5
2.1.1	Der chromatographische Prozess	5
2.1.2	Detektion	6
2.1.3	Technische Daten	7
2.2	Peakcharakteristika	8
2.2.1	Lage	8
2.2.2	Breite	9
2.2.3	Form	9
2.3	Probabilistische Arithmetische Automaten	9
2.3.1	Algorithmus zur Berechnung der Wartezeit	11
3	Modell	13
3.1	Vorgaben	13
3.2	2-Zustände Modell	14
3.2.1	Teilchensimulation	14
3.2.2	PAA für das 2-Zustände Modell	15
3.2.3	Grenzen des Modells	16
3.3	3-Zustände Modell	16
3.3.1	PAA	18
3.4	Weitere Modelle	19
4	Methoden	21
4.1	Simulation	21
4.1.1	Step-by-Step	21
4.1.2	By-Event	25
4.2	PAA	28
4.2.1	PAA für das 2-Zustände Modell	28
4.2.2	PAA für das 3-Zustände Modell	30
4.3	Berechnung der Peakdaten	30

5	Implementierung	33
5.1	Teilchensimulation/MCC-Sim	33
5.1.1	Anwendungen	33
5.1.2	Umsetzung der Algorithmen	35
5.1.3	Optimierungen	36
5.1.4	Laufzeitvergleich von Step-by-Step und By-Event	36
5.2	Simulation des PAA	38
5.2.1	Nutzung von MoSDi	38
5.2.2	PAA-Sim	39
5.2.3	Verwertung der Verteilungen	39
5.2.4	Laufzeiten	40
6	Evaluation	43
6.1	Untersuchte Peaks	43
6.2	2-Zustände Modell	45
6.2.1	Einfluss der Parameter auf einzelne Peaks	46
6.2.2	Einfluss abhängig vom anderen Parameter	47
6.2.3	Grenzen	49
6.3	3-Zustände Modell	50
6.3.1	Einfluss auf einzelne Peaks	51
6.3.2	Einfluss abhängig von anderen Parametern	52
6.3.3	Parameter für ähnliche Peaks	59
6.3.4	Grenzen	60
7	Diskussion	63
7.1	Ergebnisse	63
7.2	Ausblick	63
	Abbildungsverzeichnis	68
	Literaturverzeichnis	70
	Erklärung	70

Kapitel 1

Einleitung

Die Chromatographie ist ein seit langem eingesetztes Verfahren in der Chemie zur Auftrennung von Stoffgemischen. Die Auftrennung erfolgt dabei zwischen zwei sogenannten Phasen, der stationären und der mobilen Phase, welche sich in unterschiedlichen Aggregatzuständen befinden und untereinander nicht mischen. Die Teilchen der Stoffgemische interagieren während des chromatographischen Prozesses mit den beiden Phasen, indem sie sehr häufig zwischen den Phasen wechseln. Durch unterschiedliche Eigenschaften interagieren die Moleküle verschiedener Stoffe unterschiedlich mit den Phasen und trennen sich dadurch von den Molekülen anderer Stoffe.

Eine mögliche Anwendung der Chromatographie ist die Präparation, bei der aus dem Stoffgemisch eine bestimmte Substanz isoliert werden soll. Zum Beispiel kann so aus einem verunreinigten Stoff, für die weitere Nutzung gereinigt werden, wenn der Stoff in Reinform zur weiteren Verwendung vorliegen muss.

Eine andere Nutzung liegt in der Analyse von Stoffgemischen. Als Ergebnis einer solchen Analyse entsteht ein Chromatogramm, welches die aufgetrennten Stoffe visualisiert. Dieses enthält einige Peaks, von denen jeder für eine Stoffsorte steht. Wenn bekannt ist, wie sich unterschiedliche Stoffe bei der Chromatographie verhalten und wie sie sich im Chromatogramm zeigen, können daraus Rückschlüsse auf die im Ausgangsgemisch enthaltenen Stoffe gezogen werden.

Praktische Anwendungen dafür finden sich in vielen Bereichen der Chemie und Biochemie. Um nur zwei Beispiele zu nennen, kann in der Analyse von Lebensmitteln mit Hilfe der Chromatographie bestimmt werden, ob Produkte Inhaltstoffe enthalten, die nicht darin enthalten sein sollten. Auch in der Umweltchemie wird das Verfahren genutzt, um beispielsweise festzustellen, ob in einer Wasser- oder Luftprobe bestimmte, möglicherweise gefährliche, Stoffe enthalten sind.

Es gibt verschiedene Arten der Chromatographie, die im Laufe des letzten Jahrhunderts entwickelt wurden. Für diese Arbeit besonders interessant ist die Gaschromatographie in Kapillarsäulen. Die Gaschromatographie gibt es seit den 1950er Jahren, die Kapillartechnik wurde etwa zwanzig Jahre später erfunden. Bei der Gaschromatographie ist die mobile Phase stets ein Gas welches das zu analysierende Stoffgemisch durch eine Trennsäule, beispielsweise eine Kapillarsäule, leitet. In diesem Fall ist die Säule innen mit einer flüssigen stationären Phase beschichtet. Eine Multikapillarsäule, kurz MCC (engl. Multi Capillary Column) besteht aus sehr vielen dieser Kapillaren.

1.1 Ziel der Arbeit

Im Rahmen der Diplomarbeit soll, mit Hilfe eines Modells, der chromatographische Prozess in einer Multikapillarsäule simuliert werden. Dabei soll jeweils nur ein Stoff, charakterisiert durch einige Parameter, simuliert werden. Das Resultat einer Simulation ist jeweils ein einzelner Peak. Sowohl diese simulierten Peaks, als auch die aus realen Messungen können durch bestimmte Eigenschaften charakterisiert werden.

Gesucht ist nun eine Entsprechung dieser Peakcharakteristika zu den Simulationsparametern, die für das Modell verwendet werden. Allerdings ist zunächst nicht klar, ob eine solche Entsprechung zwischen Parametern und Charakteristika existiert. Wenn dies aber der Fall ist, soll es sowohl möglich sein, aus den Simulationsparametern vorherzusagen, welche Eigenschaften ein damit simulierter Peak haben wird. Umgekehrt soll aber auch zu einem vorgegebenen Peak bestimmt werden können, welche Parameter nötig sind, um mit einer Simulation einen möglichst ähnlichen Peak zu erzeugen.

Es muss noch der Begriff der Simulation für diese Arbeit geklärt werden. Meist geht es bei einer Simulation um die Nachahmung realer Phänomene durch Hilfsmittel wie einen Computer. Im Zusammenhang mit der Gaschromatographie könnte man darunter beispielsweise verstehen, den physikalischen Ablauf eines Säulendurchlaufs auf Molekülebene durchzugehen. Dabei müssten Teilchen simuliert werden, die unter Anwendung physikalischer Gesetze miteinander interagieren, beispielsweise aneinander anstoßen und dadurch Richtung und Geschwindigkeit ändern oder sich gegenseitig anziehen oder abstoßen und dabei ebenfalls ihre Bewegung ändern. Auch für Interaktionen mit der chromatographischen Apparatur müssten die Auswirkungen auf die Teilchen berechnet und eingebunden werden.

Doch eine solche physikalische Simulation ist in diesem Fall genausowenig gemeint, wie die beiden Ansätze, die in Abschnitt 1.2 vorgestellt werden.

Statt dessen soll hier mit Hilfe eines Modells der chromatographische Prozess simuliert werden. Es wird dabei von den Details, die bei einer Simulation auf physikalischer Ebene berücksichtigt werden müsste, abstrahiert und nur das grundlegende Prinzip des Wechsels der Teilchen zwischen zwei Phasen berücksichtigt.

Das Problem kann daher als unbekannte Funktion $F : [0,1]^x \rightarrow \mathbb{R}^y$ formuliert werden. Dabei ist x eine zunächst unbekannte Anzahl an Modellparametern p für die Simulation und y die Anzahl an Peakcharakteristika λ mit denen ein Peak beschrieben wird. Schön wäre es, $F(p_1, \dots, p_x) = (\lambda_1, \dots, \lambda_y)$ für alle in echten Daten zu findenden λ_j zu kennen. Zumindest aber soll geklärt werden welcher Parameter p_i welchen Einfluss auf die verschiedenen λ_j hat.

Um dieses Ziel zu erreichen, werden zunächst die Eigenschaften, die ein Peak haben kann, beschrieben. Dann werden mit einem sehr einfachen Modell Simulationen durchgeführt und evaluiert, inwiefern die damit erzeugten Peaks mit realen Peakdaten übereinstimmen und ob es Eigenschaften gibt, die nicht simuliert werden können. Ist dies der Fall, wird das Modell nach Bedarf angepasst oder erweitert. Auch das neue Modell wird getestet und evaluiert.

1.2 Andere Arbeiten

Es existieren bereits Simulatoren für die Chromatographie, welche jedoch Daten auf eine andere Art erzeugen als eine Simulation des chromatographischen Prozesses erzeugen.

Es gibt beispielsweise einen älteren Ansatz (Ghosh et al., 1988), bei dem verschiedene Peaks als Funktionen eingegeben werden können. Die dort zu Grunde liegende Annahme ist, dass sich Peaks, ebenso wie das Hintergrundrauschen einer Messung, durch jeweils eine Funktion darstellen lassen. Diese Eingaben werden dann mit Hilfe einer Tabellenkalkulationssoftware zu einem gemeinsamen Spektrum kombiniert. Ein solcher Ansatz ist aber erst nützlich, wenn die Peaks bereits als Funktionen vorliegen, die erhofften Zusammenhänge zwischen Parametern einzelner Peaks und damit verbundenen Eigenschaften müssen also bereits im Vorhinein vorliegen.

Ein anderer Ansatz wurde von Boswell et al. (2013) beschrieben. Dort geht es zwar um Flüssigchromatographie, die jedoch der Gaschromatographie ausreichend ähnlich ist, um als Vergleich herangezogen zu werden. In der dort beschriebenen Software können verschiedene Parameter der Chromatographie wie Druck oder Temperatur variiert und verschiedene Analyte ausgewählt werden und es werden daraus Chromatogramme erzeugt. Die so gewonnenen Daten werden aber nicht simuliert, sondern aus experimentellen Daten interpoliert. Dazu wurden in vielen Experimenten die Einflüsse der Einstellungen des Chromatographiegerätes untersucht und festgestellt, wie sich beispielsweise die Temperatur auf die resultierenden Peaks auswirkt. Für diese unterschiedlichen Messbedingungen wurden Formeln entwickelt, die in der Simulation auf die Peaks der einzelnen Analyte angewendet werden, um so je nach gewählten Einstellungen modifizierte Chromatogramme zu erhalten.

1.3 Aufbau der Arbeit

Zunächst werden in Kapitel 2 die nötigen Grundlagen für diese Arbeit vorgestellt. Dies ist zum Einen die Chromatographie, insbesondere die Gaschromatographie und der chromatographische Prozess. Zudem werden einige Eigenschaften, welche die resultierenden Peaks haben können, vorgestellt. Zum Anderen wird eine Möglichkeit zur Modellierung des Problems als Probabilistischer Arithmetischer Automat (PAA) in Betracht gezogen.

Kapitel 3 stellt zwei aufeinander aufbauende Modelle für die Simulation vor. Das erste ist ein sehr einfaches Modell, welches mit nur zwei Parametern auskommt, jedoch nur einen Teil der gewünschten Ergebnisse liefert. Das zweite Modell berücksichtigt weitere Details der Chromatographie und verfügt über bis zu sechs Parameter. Für beide Varianten wird jeweils auch eine Modellierung als PAA vorgestellt.

In Kapitel 4 werden Methoden zur Umsetzung der Modelle erklärt. Dabei gibt es zwei Herangehensweisen, die für jeweils das einfache und das komplexere Modell vorgestellt werden, zudem wird auch eine mögliche Umsetzung des PAA beschrieben.

Kapitel 5 gibt einen Überblick über die Implementierung der vorgestellten Methoden.

In Kapitel 6 werden die Ergebnisse der Simulationen ausgewertet. Für beide Modelle wird der Einfluss der einzelnen Parameter auf die Simulationsergebnisse untersucht und mögliche zusammen wirkende Einflüsse der Parameter beleuchtet. Außerdem werden die Grenzen der Modelle aufgezeigt.

Zuletzt werden in Kapitel 7 die Ergebnisse der Arbeit zusammengefasst und ein Ausblick auf mögliche weiterführende Ideen oder Verbesserungen gegeben.

Kapitel 2

Grundlagen

2.1 Chromatographie

Die Chromatographie ist ein Verfahren zur Auftrennung und Analyse von Stoffgemischen. Die Auftrennung eines Gemisches geschieht dabei zwischen zwei sogenannten Phasen, der stationären und der mobilen Phase. Diese befinden sich in unterschiedlichen Aggregatzuständen. Die Analyte werden bewegt durch die chromatographische Apparatur, indem sie von der mobilen Phase mitgenommen werden. Immer wieder geraten sie dabei in Kontakt zur stationären Phase und interagieren mit dieser, indem sie die Phase wechseln. Abhängig von den Eigenschaften der verschiedenen Stoffe, werden sie dadurch kürzer oder länger aufgehalten, sodass sich verschiedene Stoffe voneinander trennen.

Es existieren viele Varianten der Chromatographie. Eine mögliche Unterscheidung findet durch die dabei verwendeten Aggregatzustände der beiden Phasen statt. Zum Einen gibt es die Flüssigchromatographie (LC, engl. Liquid Chromatography), bei der die mobile Phase eine Flüssigkeit ist und die stationäre Phase ein Feststoff. Bei der Gaschromatographie (GC) ist die mobile Phase ein Gas und es wird zusätzlich nach der stationären Phase unterschieden. Ist diese ein Feststoff, so spricht man von gepackten Säulen. Bei der Kapillartechnik hingegen werden die Trennsäulen innen mit einem Flüssigkeitsfilm als stationäre Phase beschichtet.

Beispielsweise kann die GC in einer Multikapillarsäule stattfinden. Sie besteht aus ca. 1000 bis 2000 einzelnen Kapillaren (Łobiński et al., 1999; Baumbach, 2009). Jede davon ist innen mit der stationären Phase beschichtet. Außerdem kommt als mobile Phase ein Trägergas, zum Einsatz, welches die Analyte durch die Säule transportiert. Dieses kann beispielsweise Stickstoff, Helium (Łobiński et al., 1999) oder Luft sein (Baumbach, 2009).

2.1.1 Der chromatographische Prozess

Die nachfolgende Beschreibung des chromatographischen Prozesses gilt für alle Arten der Chromatographie, die Auftrennung der Stoffe durch den häufigen Wechsel zwischen den beiden Phasen ist allen Arten der Chromatographie gemein. Die nachfolgenden Ausführungen beziehen sich jedoch hauptsächlich auf die Gaschromatographie in Kapillarsäulen. Ausführlich beschrieben wurde der chromatographische Prozess beispielsweise von Kolb (2003) oder Kaltenböck (2008). Die für diese Arbeit wesentlichen Punkte sind hier kurz zusammengefasst:

Die Substanzen unterscheiden sich vor allem durch ihre Wechselwirkungen mit der stationären Phase. Während dieser Wechselwirkungen haften die Teilchen an der stationären Phase.

nären Phase, bewegen sich also nicht fort. Finden verhältnismäßig wenig Wechselwirkungen statt, passieren die Teilchen die Säule schneller, als wenn viele Wechselwirkungen stattfinden. Dies beeinflusst die so genannte Retentionszeit, also die Zeit, die zum Durchlaufen der Säule gebraucht wird.

Dabei existieren verschiedene Arten von Wechselwirkungen zwischen den Analytteilchen und der stationären Phase, sowie andere Einfüsse, die sich auf die Retentionszeit auswirken. Dies sind unter anderem:

- Molekülgröße: Besonders bei der Chromatographie in gepackten Säulen spielt die Molekülgröße eine wichtige Rolle. Kleine Moleküle werden weniger stark zurück gehalten, als große.
- Adsorption: Dabei kommen die Teilchen mit der Oberfläche der stationären Phase in Kontakt und reichern sich dort an. Die Adsorption findet dabei auf Grund von physikalischen Kräften wie der Van-Der-Waals-Kräften statt. Dadurch ist die Bindung der Stoffe an die stationäre Phase nicht sehr fest und die Teilchen können sich leicht wieder daraus lösen.
- Löslichkeit: Die Analyte können sich auch in einer flüssigen stationären Phase lösen. Diese Art der Chromatographie wird auch Verteilungschromatographie genannt.
- Ladung: Bei der Ionenaustauschchromatographie ist die stationäre Phase elektrisch geladen, sodass entgegengesetzt geladene Teilchen dort gebunden werden und dadurch länger für den Durchlauf der Trennstrecke brauchen.

Bei der Gaschromatographie in Multikapillarsäulen haben sowohl die Adsorption- als auch die Verteilungschromatographie einen Einfluss auf den Ablauf des Prozesses. Dabei trennen sich gelöste Teilchen schwerer als adsorbierte Teilchen von der stationären Phase. Die Stoffe verweilen dadurch durchschnittlich länger in der stationären Phase und ihre Durchschnittsgeschwindigkeit sinkt stärker.

Auf die eine oder andere Weise diffundieren die Teilchen nun zwischen den beiden Phasen hin und her, wodurch sich ein Gleichgewicht zwischen den Phasen einstellt. Durch dieses Gleichgewicht wandern die Teilchen eines Stoffes als Pulk durch die Säule, bis sie diese verlassen. Für Teilchen vorne im Pulk besteht nur die Möglichkeit, in Wechselwirkung mit der stationären Phase zu treten, bis sich eine Sättigung einstellt. Später ankommende Teilchen können nur dann in Wechselwirkung treten, wenn bereits wieder Teilchen mobil geworden sind, wodurch sich das Gleichgewicht aufrecht erhält. Je nach Verweildauer der einzelnen Teilchen in der stationären Phase, ergibt sich eine für diesen Stoff charakteristische Durchschnittsgeschwindigkeit.

2.1.2 Detektion

Nach Durchlaufen der Säule wird detektiert, welche Menge an Substanzen austritt. Dabei wird nicht die Art des Analyts festgestellt, sondern nur die Menge der zum jeweiligen Zeitpunkt austretenden Stoffe. Während sich die meisten Stoffe ausreichend in ihrem Verhalten zur stationären Phase unterscheiden und dadurch aufgetrennt die Säule verlassen, kann es aber durchaus vorkommen, dass sich zwei Stoffe hinreichend ähnlich sind und zumindest teilweise überlappend austreten.

Zur Detektion existieren mehrere Möglichkeiten. Zum Einen kann ein Detektor wie beispielsweise ein Wärmeleitfähigkeitsdetektor oder Flammenionisationsdetektor ein Chromatogramm aufzeichnen. Dabei wird in bestimmten Zeitintervallen die austretende Menge

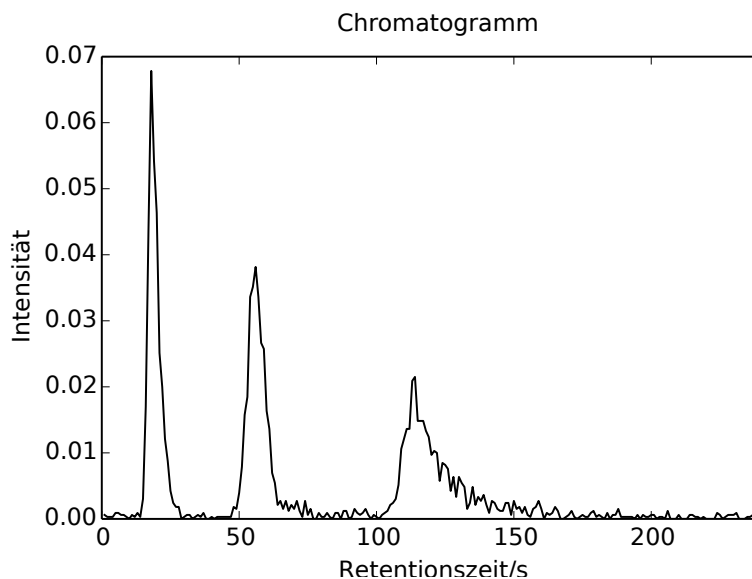


Abbildung 2.1: Beispiel für ein Chromatogramm

Analyte festgestellt. In Abbildung 2.1 ist ein Beispiel für ein Chromatogramm zu sehen, welches aus mehreren simulierten Daten und Hintergrundrauschen erstellt wurde. Typischerweise sind zu Beginn einer Messung die Peaks sehr schmal und werden mit zunehmender Retentionszeit eher breiter.

Alternativ kann die Gaschromatographie auch als Vorverarbeitung für Verfahren wie Massenspektrometrie (MS) oder Ionen-Mobilitäts-Spektrometrie (IMS) dienen. Auch bei diesen Technologien können nicht in jedem Fall alle Stoffe voneinander unterschieden werden. Zum Beispiel existieren Isomere, also Moleküle gleicher Masse aber unterschiedlicher Struktur, die von der MS nicht unterschieden werden können. Durch die zuvor statt findende GC wird jedoch nach einer weiteren Stoffeigenschaft unterschieden, sodass diese ähnlichen Stoffe durch die Vortrennung zu verschiedenen Zeitpunkten mit dem zweiten Analyseverfahren beginnen und so doch getrennt im resultierenden Spektrum erscheinen. Wenn eine solche Weiterverarbeitung statt finden soll, werden die aus der MCC austretenden Moleküle direkt ionisiert und in den entsprechenden Geräten weiter analysiert. Die Kopplung von GC und MS ist beispielsweise von Hübschmann (2009) ausführlich beschrieben worden, näheres zur Kopplung von GC und IMS findet sich zum Beispiel bei Baumbach (2009).

2.1.3 Technische Daten

Es gibt verschiedene Arten von Multikapillarsäulen, die sich in der Anzahl und Größe ihrer Kapillaren, ihrer Länge, ihrer stationären Phase oder der maximalen Temperatur und Druck mit denen sie betrieben werden können, unterscheiden können. Dadurch können sie auch unterschiedliche Gasvolumina aufnehmen und trennen und die Durchlaufzeiten für eine Analyse variieren entsprechend. Außerdem können mit ein und derselben Säule verschiedene Betriebsmodi verwendet werden, die sich beispielsweise durch verwendeten Druck und damit in der Geschwindigkeit des Trägergases oder der Temperatur unterscheiden. Damit können bei gleichen Analyten sehr unterschiedlich aussehende Chromatogramme erzeugt werden, da die Stoffe auch unterschiedlich auf solche Veränderungen reagieren.

Um eine Einschätzung für eine MCC zu geben, sind hier einige mögliche Daten angegeben. Weitere Spezifikationen finden sich beispielsweise auf der Webseite des Herstellers Multichrom Ltd.^{1 2}.

Eine MCC besteht aus ca. 1000–2000 Kapillaren mit je 20–80 μm Durchmesser. Damit ist sie etwa 2–6 mm dick. Die Länge beträgt etwa 20–40 cm. Die stationäre Phase ist ein Flüssigkeitsfilm, der als ca. 0,1–0,8 μm dicke Beschichtung auf die Innenseite der Kapillaren aufgebracht ist.

Der Querschnitt einer MCC ist in Abbildung 2.2 zu sehen.

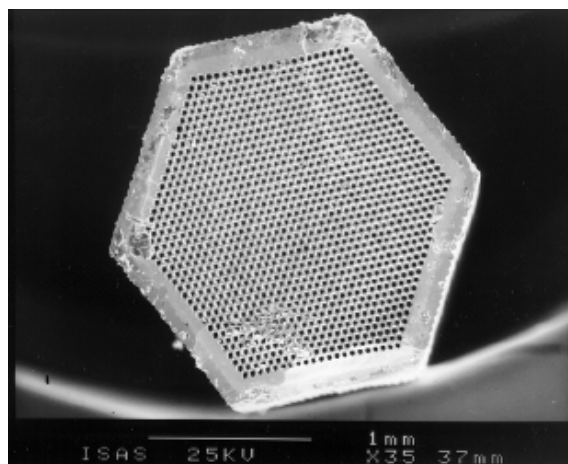


Abbildung 2.2: Querschnitt einer MCC, Quelle: <http://yas.yanaco.co.jp/products/import-gc-ims.html>

2.2 Peakcharakteristika

Die im gaschromatographischen Experiment gewonnenen, sowie später auch die simulierten Peaks, müssen durch einige Eigenschaften charakterisiert werden, um sie vergleichbar zu machen. Dafür seien hier drei solcher Eigenschaften beschrieben: die Lage, Breite und Form eines Peaks.

2.2.1 Lage

Die offensichtlichste Eigenschaft eines Peaks ist seine Lage, genauer gesagt, die Lage seines Maximums. Je nach verwendeter Technologie unterscheiden sich die Messzeiträume für einen chromatographischen Durchlauf stark voneinander. Allerdings gibt es immer einen minimalen Zeitraum der nach Start der Chromatographie vergangen sein muss, bis ein Peak überhaupt auftreten kann. Dieser hängt von der Geschwindigkeit der mobilen Phase ab, da sich die Analyte nicht schneller als die mobile Phase fortbewegen können und daher mindestens dieselbe Zeit benötigen. Außerdem kann für jede einzelne Messung ein Maximalzeitpunkt gewählt werden, nach dem das Experiment beendet wird. Für die Gaschromatographie in Multikapillarsäulen liegen die Retentionszeiten typischerweise bei wenigen Sekunden bis Minuten.

¹<http://www.mcc-chrom.com/catalogue>

²<http://www.mcc-chrom.com/species>

2.2.2 Breite

Peaks können unterschiedlich breit sein. Dabei ist zu beobachten, dass schnelle Teilchen Peaks zu frühen Zeitpunkten erzeugen, die eine relativ geringe Varianz aufweisen, hingegen spätere Peaks tendenziell breiter werden. Da jedoch ein Peak mit größerer Intensität, also möglicherweise größerer Stoffmenge, automatisch auch einen höheren und damit breiteren Peak erzeugt, kann die Halbwertsbreite als Maß für die Peakbreite herangezogen werden. Die Halbwertsbreite wird auf halber Maximalhöhe des Peaks gemessen. Im Fall von auftretendem Tailing ist dieser Wert jedoch möglicherweise nicht mehr aussagekräftig. Außerdem ist die Halbwertsbreite kein robustes Maß für die Breite, sodass statt dessen der Interquartilsabstand (IQR, engl. interquartile range) als Maß für die Breite verwendet werden kann. Auch dieses Maß ist unabhängig von der Stoffmenge, die den Peak verursacht hat.

Der IQR entspricht der Differenz zwischen den Quartilen Q_{25} und Q_{75} :

$$IQR = Q_{75} - Q_{25} \quad (2.1)$$

2.2.3 Form

Als dritte Eigenschaft können sich Peaks in ihrer Form unterscheiden. Im theoretischen Idealfall haben sie die Form einer Gaußkurve. Abweichend davon treten in echten Messungen häufig asymmetrische Peaks auf, die dann ein Tailing aufweisen, diese Eigenschaft wird auch Rechtsschiefe genannt. Dabei steigt die Kurve zunächst stark an, sinkt jedoch nach Erreichen des Maximalwerts deutlich langsamer ab, es entsteht ein Schwanz (engl. Tail). Ursachen für solche schiefen Peaks gibt es viele (Kolb, 2003; Moretti et al., 2004; Giddings, 1963). Einige Beispiele für Ursachen sind zusätzliche Adsorptionseffekte, die beim Altern einer Säule auftreten, technische Ursachen wie kleinste Hohlräume zwischen der Säule und dem Gaseinlass bzw. -austritt, sowie einige Stoffe, die generell zu Tailing neigen.

Als Maß für die Schiefe bietet sich der Quantilkoeffizient (Johnson et al., 1994) an. Dieser berechnet sich, aus den Quantilen einer Verteilung und ist definiert als:

$$\frac{Q_{\alpha} + Q_{1-\alpha} - 2 \cdot Q_{50}}{Q_{1-\alpha} - Q_{\alpha}} \quad (2.2)$$

Für $\alpha = 25$ spricht man auch vom Quartilkoeffizienten nach Yule-Bowley und dieser soll später als Maß für die Schiefe der Peaks verwendet werden.

2.3 Probabilistische Arithmetische Automaten

Um die Multikapillarsäule zu simulieren, werden verschiedene Vorgehensweisen angewendet. Am naheliegendsten erscheint es, den chromatographischen Prozess für eine gewisse Anzahl an Teilchen zu simulieren und dabei für jedes einzelne dieser Teilchen mindestens seinen Ort und seinen Zustand festzuhalten. Diese zufallsbasierte Methode hat jedoch den Nachteil, dass verschiedene Durchläufe zu leicht unterschiedlichen Ergebnissen führen können, insbesondere, wenn zu wenige Teilchen simuliert wurden. Außerdem sind sehr viele Teilchen nötig, um glatte Peaks zu bekommen.

Eine andere Möglichkeit besteht darin, die Teilchen als Wahrscheinlichkeitsmasse zu modellieren und für jeden Ort und Zustand festzuhalten, welcher Anteil an Molekülen sich dort befindet. Eine Möglichkeit die Analyte als eine solche Wahrscheinlichkeitsmasse zu betrachten und dadurch eine exakte Verteilung der Teilchen auf Orte und Zustände zu erhalten, bieten Probabilistische Arithmetische Automaten (PAA). Ein PAA nach Marschall

et al. (2012) ist ein Modell, mit dem eine Folge zufälliger Operationen beschrieben werden kann. Für PAA existieren Algorithmen, welche eine gemeinsame Verteilung von Zuständen und Werten oder auch die Verteilung der Wartezeit für einen Wert berechnen. Wie in Kapitel 3 beschrieben wird, können die Modelle zur Simulation einer Multikapillarsäule auch als PAA formuliert werden. In dieser Formulierung entspricht der Wert auf den gewartet wird, der Länge der Säule und die Zeit, die zum Durchlaufen einer Säule gebraucht wird, ist die Wartezeit für diesen Wert.

Zunächst sei hier eine Definition für den PAA gegeben, anschließend wird der Algorithmus zur Berechnung der Wartezeit beschrieben.

2.3.1 Definition (PAA). Ein Probabilistischer Arithmetischer Automat (PAA) ist ein Tupel $\mathcal{P} = (\mathcal{Q}, q_0, T, \mathcal{V}, v_0, \mathcal{E}, (e_q)_{q \in \mathcal{Q}}, (\theta_q)_{q \in \mathcal{Q}})$, dabei ist:

- \mathcal{Q} eine endliche Menge von Zuständen
- $q_0 \in \mathcal{Q}$ der Startzustand
- $T : \mathcal{Q} \times \mathcal{Q} \rightarrow [0,1]$ eine Übergangsfunktion mit $\sum_{q' \in \mathcal{Q}} T(q, q') = 1$ das heißt $(T(q, q'))_{q, q' \in \mathcal{Q}}$ ist eine stochastische Matrix
- \mathcal{E} eine endliche Menge von Emissionen
- $e_q : \mathcal{E} \rightarrow [0,1]$ eine Wahrscheinlichkeitsverteilung der Emissionen für jeden Zustand
- \mathcal{V} eine Menge von Werten
- v_0 der Startwert
- $\theta_q : \mathcal{V} \times \mathcal{E} \rightarrow \mathcal{V}$ eine Operation für jeden Zustand

Dabei entspricht (\mathcal{Q}, q_0, T) einer Markovkette und $(\mathcal{Q}, q_0, T, \mathcal{E}, (e_q)_{q \in \mathcal{Q}})$ einem Hidden Markov Model. Zusätzlich gibt es für jeden Zustand eine Operation θ , wie beispielsweise Addition oder Subtraktion. Mit Hilfe dieser Operation berechnet sich in jedem Schritt t der neue Wert v_t aus dem Wert des vorherigen Schrittes v_{t-1} und der Emission des aktuellen Zustandes.

Aus der Definition für einen PAA ergibt sich der Zustandsprozess $(Q_t)_{t \in \mathbb{N}}$, wobei $Q_0 = q_0$ ist und die Wahrscheinlichkeiten für weitere Zustände nur vom letzten Zustand abhängt:

$$P(Q_{t+1} = q_{t+1} \mid Q_t = q_t, \dots, Q_0 = q_0) = T(q_t, q_{t+1}) \quad (2.3)$$

Die Emissionen hängen nur vom aktuellen Zustand ab, sodass der Emissionsprozess $(E_t)_{t \in \mathbb{N}}$ definiert ist als

$$P(E_{t+1} = \epsilon \mid Q_0 = q_0, \dots, Q_t = q_t, E_0 = \epsilon_0, \dots, E_{t-1} = \epsilon_{t-1}) = e_q(\epsilon) \quad (2.4)$$

Mit Hilfe dieser beiden Prozesse und den Operationen kann dann auch für die Entwicklung der Werte ein Prozess $(V_t)_{t \in \mathbb{N}}$ definiert werden:

$$V_t = \theta_{Q_t}(V_{t-1}, E_t) \quad (2.5)$$

Dabei ist $V_0 = v_0$.

Ein PAA ist zu jedem Zeitpunkt mit einer bestimmten Wahrscheinlichkeit in jedem Zustand und kann dabei mit einer bestimmten Wahrscheinlichkeit jeden Wert annehmen. Es existiert also zu jedem Zeitpunkt eine Zustands-Werte-Verteilung. Ausgehend von einer gegebenen Startverteilung, bei dem sich die gesamte Wahrscheinlichkeitsmasse in Zustand q_0 befindet und den Wert v_0 hat, wird diese Verteilung schrittweise aktualisiert. Dabei findet jeder Übergang gemäß der Wahrscheinlichkeit aus T statt.

2.3.1 Algorithmus zur Berechnung der Wartezeit

Neben der eben vorgestellten schrittweisen Berechnung der Zustand-Werte Verteilung kann es von Interesse sein, die Verteilung der Wartezeit auf einen bestimmten Wert v^* zu kennen. Für die Teilchen in der MCC entspricht dieser Zielwert der Länge der Säule und man ist interessiert in der Verteilung der Anzahl der Schritte die zum Durchlaufen dieser Strecke benötigt werden.

Da es sein kann, dass in einigen Abfolgen von Zuständen, der Wert v^* erst sehr spät oder gar nicht erreicht werden kann, wird eine maximale Wartezeit t_{\max} gesetzt. Zwar können dann Wartezeiten für v^* größer als t_{\max} auftreten, deren genaue Verteilung unbekannt ist, Gesamtwahrscheinlichkeit dieser Zeitpunkte kann jedoch berechnet werden.

Außerdem können innerhalb eines Werteprozesses $(V_t)_{t \in \mathbb{N}}$ mehrere Zeitpunkte existieren, an denen der Wert v^* angenommen werden kann, gesucht ist der kleinste dieser Zeitpunkte. Daher kann die Wartezeit auch als Zufallsvariable aufgefasst werden, die definiert als: $W_{v^*} = \min\{t \in \mathbb{N} \mid V_t = v^*\}$.

Wie oben beschrieben, existiert die Möglichkeit, die Zustands-Werte Verteilung schrittweise für alle Zeitpunkte bis zu einem gewünschten Maximalzeitpunkt t_{\max} zu berechnen. Gesucht ist nun die Verteilung über die Zeitpunkte $t < t_{\max}$ für einen Zielwert v^* . $P(V_t = v^*)$ ist die Wahrscheinlichkeit für jeden solchen Zeitpunkt t , v^* zu erreichen. Dies entspricht der Summe über alle Zustände, die zu Zeitpunkt t den Wert v^* erreichen können. Also ist $P(V_t = v^*) = \sum_{q \in \mathcal{Q}} P(q_t = q, V_t = v^*)$

Zur Berechnung der Verteilung der Wartezeit auf v^* wird ein modifizierter PAA konstruiert. Hierbei werden zwei neue Zustände eingeführt, \bullet und \circ , sodass der neue PAA die Zustände $\mathcal{V}' = \mathcal{V} \cup \{\bullet, \circ\}$ hat. Außerdem werden die Operationen angepasst auf

$$\theta'_q(v, e) = \begin{cases} \theta_q(v, e), & \text{wenn } v \notin \{\bullet, \circ\} \text{ und } \theta_q(v, e) \neq v^*, \\ \bullet, & \text{wenn } v \notin \{\bullet, \circ\} \text{ und } \theta_q(v, e) = v^*, \\ \circ, & \text{wenn } v \in \{\bullet, \circ\} \end{cases}$$

Der Wert \bullet wird also genau dann erreicht, wenn im aktuellen Berechnungsschritt die ursprüngliche Operation den Zielwert v^* erreicht. Also gilt für alle Zeiten $t < t_{\max}$: $P(W_{v^*} = t) = P(V'_t = \bullet)$. Dabei ist V'_t der Werteprozess für den modifizierten PAA.

Die Verteilung für $P(V'_t = \bullet)$ kann wie oben angegeben schrittweise für alle Zeitpunkte t ausgehend von der Startverteilung bis zu t_{\max} berechnet werden. Dadurch, dass im jeweils nächsten Berechnungsschritt der PAA in Zustand \circ gelangt, wird jeweils nur das erste Erreichen von v^* berücksichtigt.

Durch Berechnung der Wahrscheinlichkeiten $P(V'_t = \bullet)$ für alle $t < t_{\max}$ erhält man die Verteilung der Wartezeiten auf v^* .

Kapitel 3

Modell

Im Folgenden werden die Simulationsmodelle für die Chromatographie vorgestellt. Zunächst müssen einige Vorgaben festgelegt werden, wie reale Bedingungen in der Simulation abgebildet werden sollen.

Basis für die anschließend vorgestellten Modelle ist der im letzten Kapitel beschriebene Phasenwechsel der Teilchen im Verlauf der Chromatographie. Nach einem sehr einfachen Modell, das mit lediglich zwei Parametern auskommt, wird ein weiteres, komplexeres Modell beschrieben.

Zusätzlich zur Simulation vieler einzelner Teilchen, welche die Modelle durchlaufen, wird auch jeweils eine Modellierung als PAA vorgestellt.

3.1 Vorgaben

Bei jeder Simulation müssen Einschränkungen gegenüber den realen Bedingungen gemacht werden. In diesem Fall betrifft das vor allem drei Punkte: den Raum, die Zeit und die Anzahl der Analytteilchen.

Für die Simulation wird eine Multikapillarsäule von 20 cm Länge angenommen. Diese wird in 1000 Einheiten von je 0,2 mm unterteilt. Damit nutzt die Simulation ein diskretes Raummodell. Außerdem wird nicht berücksichtigt, dass die Säule eine Röhre und damit dreidimensional ist, sondern wird als gerade Strecke angenommen. An jeder der 1000 Positionen der Strecke können sich jedoch beliebig viele Teilchen aufhalten, ohne aufeinander Einfluss zu nehmen.

Die Zeit wird ebenfalls als diskret und endlich angenommen. Es wird von einem Trägergas ausgegangen, welches nach etwa 0,1 Sekunden die Säule durchlaufen hat. Außerdem soll es sich pro Zeitschritt und damit pro Simulationsschritt je einen Raumschritt fortbewegen, was einer Geschwindigkeit von 2 m/s entspricht. Damit entspricht eine Zeiteinheit 0,1/1000 Sekunden also 0,1 ms. Um die Simulationsdauer zu beschränken und da auch reale Experimente nach einer bestimmten Zeit beendet sind, wird eine maximale Zeit von 240 Sekunden, also 2 400 000 Simulationsschritten gesetzt. Insbesondere diese Grenze lässt sich aber nach Belieben verändern.

Die gemachten Vorgaben sind in Tabelle 3.1 zusammengefasst.

Tabelle 3.1: Festsetzung der Entsprechung der Einheiten zwischen realem Experiment und Simulation

	MCC	Simulation
Länge der Säule	20 cm	1000 Raumschritte
	1 Raumschritt \equiv 0,2 mm	
Durchlaufzeit Trägergas	0,1 s \equiv 2 m/s	1000 Zeitschritte
	1 Zeitschritt \equiv 0,1 ms	
Geschwindigkeit Trägergas	2 m/s	1 Raumschritt / Zeitschritt

Die Anzahl der Analytteilchen pro Stoff im Laborexperiment ist viel zu groß, als dass tatsächlich jedes einzelne simuliert werden könnte und dabei in akzeptabler Zeit Ergebnisse erzielt würden. Daher wurde die Anzahl der Teilchen auf 1000 bzw. 10 000 beschränkt. Wie sich in den simulierten Daten zeigt, ergeben sich bereits bei 1000 Teilchen erkennbare Peaks, die jedoch noch von Simulationsdurchlauf zu Simulationsdurchlauf Unterschiede aufweisen können. Bei 10 000 simulierten Teilchen ergeben sich demgegenüber geglättete Peaks, die sich zwischen den Simulationen nur noch minimal unterscheiden.

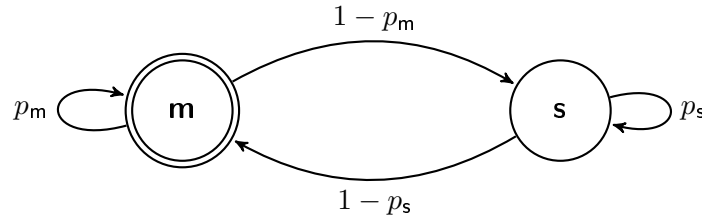
3.2 2-Zustände Modell

Als Grundlage für die Entwicklung eines Simulationsmodells dient die Beobachtung des chromatographischen Prozesses. Dieser ist, wie in Kapitel 2 beschrieben, gekennzeichnet durch viele Teilchen, welche häufig zwischen der stationären und mobilen Phase wechseln. Es liegt also nahe, dass im Modell ebenfalls Teilchen simuliert werden, die zwischen zwei Zuständen, welche die beiden Phasen repräsentieren, wechseln. Im Folgenden bezeichnet m die mobile und s die stationäre Phase. Der Phasenwechsel geschieht jeweils mit einer bestimmten Wahrscheinlichkeit. Dabei kann es möglich sein, dass die Wahrscheinlichkeiten für den Wechsel in die eine oder andere Richtung voneinander abhängig sind. Beispielsweise könnte eine hohe Wahrscheinlichkeit in den stationären Zustand zu wechseln, eine ebenfalls hohe Wahrscheinlichkeit, wieder mobil zu werden, bedingen. Da ein solcher Zusammenhang jedoch nicht gegeben sein muss, wird für das Modell zunächst der Fall von unabhängigen Wechselwahrscheinlichkeiten angenommen. Es sei also p_m die Wahrscheinlichkeit, dass ein Teilchen, welches sich bereits in der mobilen Phase befindet, auch mobil bleibt und $1 - p_m$ die Wahrscheinlichkeit, dass es in die stationäre Phase übergeht. Analog seien p_s und $1 - p_s$ die Wahrscheinlichkeiten, dass ein stationäres Teilchen in der stationären Phase bleibt bzw. zur mobilen Phase wechselt.

3.2.1 Teilchensimulation

Durch die eben gemachten Annahmen ergibt sich als erstes Modell ein einfacher Automat mit zwei Zuständen $\mathcal{Q} = \{m, s\}$. m ist Startzustand, da die Teilchen stets nur in der mobilen Phase in die Säule eintreten können. Dazu kommen die oben beschriebenen Transitionen $T = \begin{pmatrix} p_m & 1 - p_m \\ 1 - p_s & p_s \end{pmatrix}$. Eine graphische Darstellung des Modells zeigt Abbildung 3.1

Für die Simulation müssen viele Teilchen, ausgehend vom mobilen Zustand, den Automaten durchlaufen. Dabei wird zusätzlich zum Zustand der Teilchen auch der Ort, an dem sie sich befinden, verwaltet. Wenn sich ein Teilchen im mobilen Zustand befindet, wird dieser Ortszähler erhöht. Die Simulation eines Teilchens ist beendet, wenn der Ortszähler

**Abbildung 3.1:** Graphische Darstellung des 2-Zustände Modells

den gewünschten Wert, der der Länge der Trennsäule entspricht, erreicht hat. Außerdem werden die dafür benötigten Schritte gezählt, woraus sich dann die Ankunftszeit des Teilchens am Ende der Säule ableitet. Diese Simulation muss für sehr viele Teilchen wiederholt werden, sodass alle Ankunftszeiten zusammen als ein Peak dargestellt werden können.

In Kapitel 4 werden verschiedene Arten der Simulation und jeweils der genaue Ablauf beschrieben.

3.2.2 PAA für das 2-Zustände Modell

Wie in Kapitel 2 erwähnt, kann der chromatographische Prozess auch als PAA modelliert werden. Die nötigen Zustände sind offensichtlich $\mathcal{Q} = \{s, m\}$ und auch hier ist der Startzustand der mobile Zustand. Auch die Transitionen entsprechen denen im oben beschriebenen Modell. Um den Aufenthaltsort der Teilchen zu modellieren, werden die Werte genutzt. Dementsprechend liegen mögliche Werte im Bereich $\mathcal{V} = \{0, \dots, \ell\}$ mit dem Startwert $v_0 = 0$. Mit Hilfe der Emissionen wird die Fortbewegung der Teilchen modelliert: Im mobilen Zustand sollen die Analyte um einen Schritt fortbewegt werden, im stationären Zustand nicht. Damit ergeben sich zwei mögliche Emissionen: $\mathcal{E} = \{0, 1\}$. Emission 0 wird stets und ausschließlich in Zustand **s** emittiert, gleiches gilt für Emission 1 und Zustand **m**. Offensichtlich muss die Emission jeweils auf den Wert addiert werden, daher entspricht θ der Addition.

Dadurch ergibt sich folgende Definition für den PAA für das 2-Zustände Modell:

- $\mathcal{Q} = \{s, m\}$
- $q_0 = m$
- $T = \begin{pmatrix} p_m & 1 - p_m \\ 1 - p_s & p_s \end{pmatrix}$
- $\mathcal{E} = \{0, 1\}$
- $e_s(z) = \llbracket z = 0 \rrbracket$
- $e_m(z) = \llbracket z = 1 \rrbracket$
- $\mathcal{V} = \{0, \dots, \ell\}$
- $v_0 = 0$
- $\theta_s = \theta_m = +$

Dabei ist $\llbracket b \rrbracket$ definiert als $\begin{cases} 1, & \text{wenn } b \text{ wahr ist,} \\ 0, & \text{sonst.} \end{cases}$

Im Abbildung 3.2 ist der PAA für das 2-Zustände Modell graphisch dargestellt.

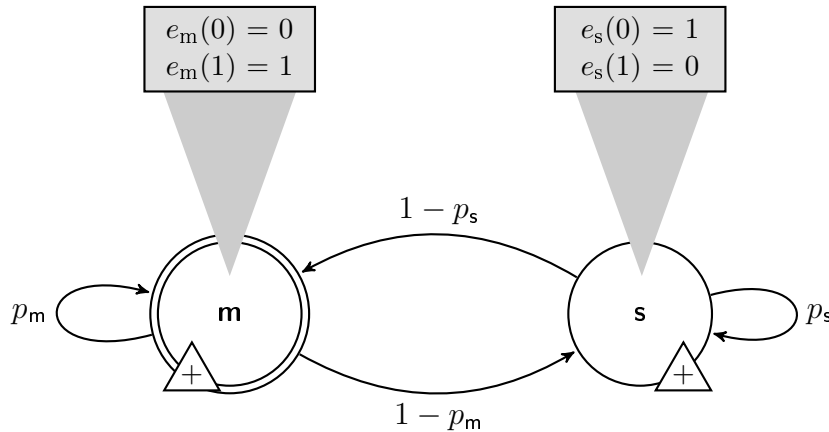


Abbildung 3.2: PAA für das 2-Zustände Modell

3.2.3 Grenzen des Modells

Eine genaue Analyse der Peaks, die mit dem 2-Zustände Modell erzeugt werden können, findet sich in Kapitel 6. An dieser Stelle sei nur vorweg genommen, dass es zwei Hauptprobleme mit dem Modell zu geben scheint.

Das erste mögliche Problem ist, dass die Peaks eine Minimalbreite an einem gegebenen Zeitpunkt t haben. Das heißt, dass mit dem Modell keine Peaks simuliert werden können, die ihrem Maximalzeitpunkt an t haben, jedoch schmaler sind als Breite b . Ob dieser Umstand ein Problem darstellt, muss anhand realer Messdaten herausgefunden werden.

Das andere Problem ist, dass fast keine der simulierten Peaks ein Tailing aufweisen. Lediglich ein sehr stark eingeschränkter Parameterbereich erzeugt ein Tailing. Leider sind die Maximalzeitpunkte der so erzeugten Peaks alle sehr klein, sodass nicht über die gesamte simulierte Zeit Peaks mit Tailing erzeugbar sind. Darüber hinaus ist Tailing in echten Messungen eher bei späten Peaks zu beobachten.

Um dieses Problem zu lösen, wird im Folgenden ein erweitertes Modell mit drei Zuständen eingeführt, welches es ermöglicht, tailende Peaks zu erzeugen.

3.3 3-Zustände Modell

Bisher fand keine Unterscheidung zwischen Adsorption und Lösung der Teilchen an bzw. in der stationären Phase statt, sodass zwei Zustände und Transitionen zwischen diesen beiden Zuständen als Modell ausreichten. In der Realität können beide Wechselwirkungen parallel zueinander statt finden: Wenn ein Teilchen in den stationären Zustand wechselt, wird jeweils entschieden, ob es adsorbiert wird oder sich löst.

Außerdem ist es realistisch anzunehmen, dass sich die Wahrscheinlichkeiten, in einen der beiden Zustände überzugehen oder in die mobile Phase zurückzukehren, unterscheiden. Die Wahrscheinlichkeit, dass ein Teilchen adsorbiert wird, kann viel höher oder niedriger sein als die, dass es zur Lösung übergeht. Umgekehrt kann ein Teilchen viel leichter oder schwerer aus der Adsorption in die mobile Phase zurückkehren, als dies beim gelösten Zustand der Fall ist. Daher liegt es nahe, einen neuen Zustand einzuführen, sodass die Adsorption und die Lösung voneinander getrennt behandelt werden. Die Tatsache, dass Tailing, wie bereits anfangs erwähnt, unter anderem durch zusätzliche Adsorptionseffekte

verursacht sein kann, lässt vermuten, dass durch diesen dritten Zustand ein Tailing in der Simulation verursacht werden kann.

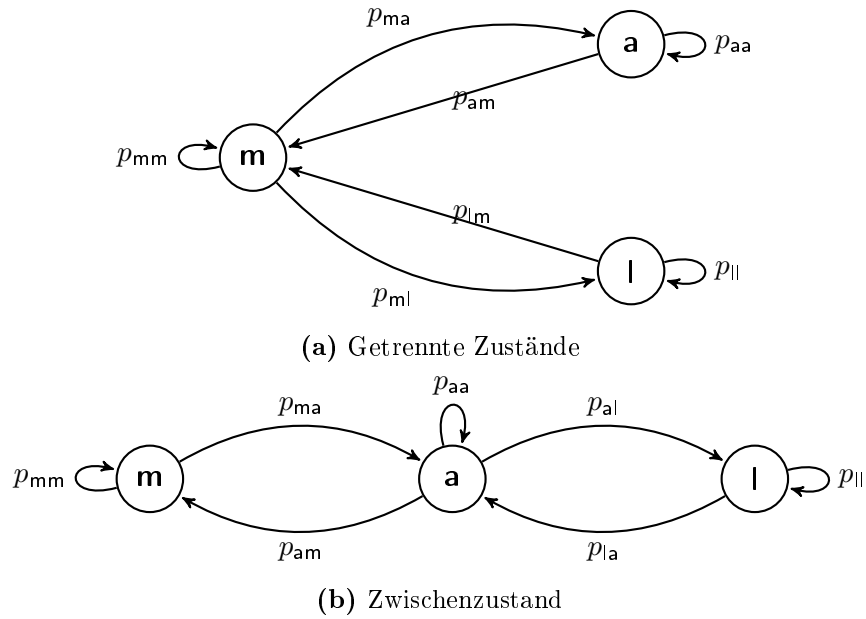


Abbildung 3.3: Mögliche Modelle mit drei Zuständen

In Abbildung 3.3 sind zwei Möglichkeiten gezeigt, dem bisherigen Modell einen neuen Zustand hinzuzufügen. Neben dem mobilen Zustand (m) gibt es in beiden Fällen je zwei stationäre Zustände, einen für die Adsorption (a) und einen für die Lösung (l). Die Übergangswahrscheinlichkeiten zwischen dem alten Zustand i und dem neuen Zustand j sind jeweils p_{ij} .

In Abbildung 3.3a sind die beiden stationären Zustände getrennt voneinander, dazwischen finden keine direkten Übergänge statt. Anschaulich kann man sich vorstellen, dass bei Übertritt in die stationäre Phase schon festgelegt wird, welcher Art der Übergang sein wird. Es scheint intuitiv sinnvoll zu sein, die Übergangswahrscheinlichkeiten zu den beiden Zuständen sowie die beiden Wahrscheinlichkeiten, wieder in die mobile Phase einzutreten, sehr unterschiedlich zu wählen. Damit wird bezweckt, dass sie sich nicht einfach wieder zu einer Gesamtwahrscheinlichkeit aufaddieren, die man auch mit dem 2-Zustände Modell hätte erreichen können. In einen der beiden Zustände sollen die Teilchen also nur seltener kommen, dafür aber dort sehr lange verweilen, in den anderen Zustand wechseln die Teilchen häufiger, bleiben aber auch nicht so lange. Außerdem spiegelt dies die in Kapitel 2 erwähnte Tatsache wieder, dass adsorbierte Teilchen sich leichter wieder in die mobile Phase begeben können, als dies bei gelösten Teilchen der Fall ist.

In Abbildung 3.3b dient der erste stationäre Zustand als Übergangszustand zum zweiten stationären Zustand. Dabei ist der Übergangszustand die Adsorption als Hinweis darauf, dass die Teilchen, die sich in der stationären Phase lösen, zunächst mit deren Oberfläche in Kontakt treten und zumindest kurzzeitig adsorbiert sind, gleiches gilt für die andere Richtung. Nachdem ein Teilchen adsorbiert wurde, besteht die Möglichkeit, dass es sich noch in der stationären Phase löst oder auch adsorbiert bleibt. Auch hier liegt es nahe, die Wahrscheinlichkeiten für die verschiedenen Übergänge so zu gestalten, dass die Teilchen in einem der stationären Zustände seltener aber länger bleiben.

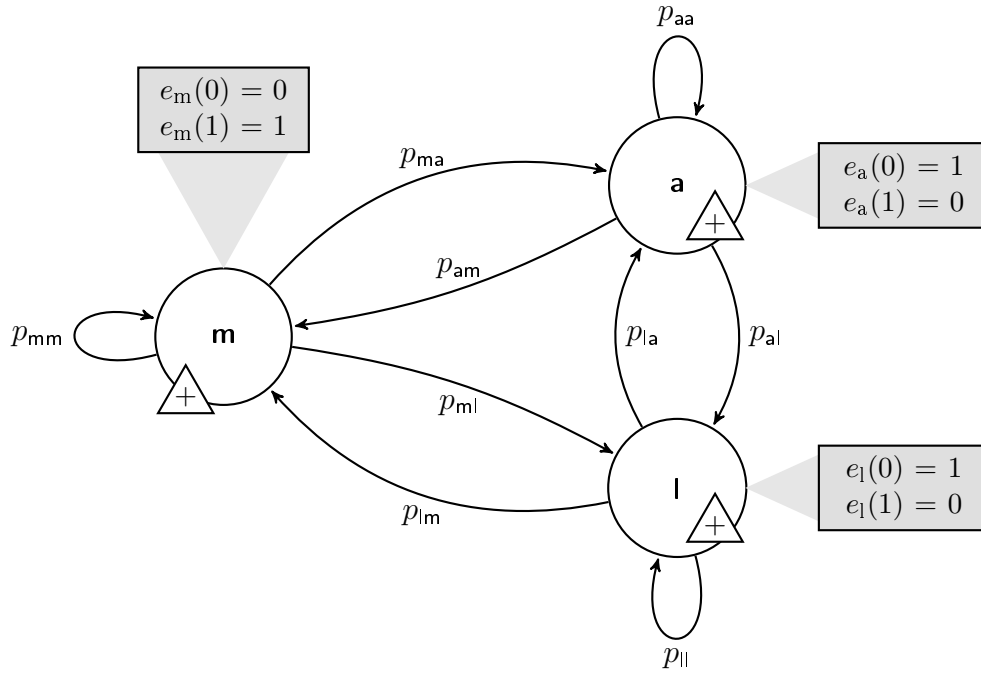


Abbildung 3.4: PAA für das allgemeine 3-Zustände Modell

Beide Fälle können zu einem Gesamtmodell mit drei Zuständen zusammengefasst werden. Das Modell hat die Zustände $\mathcal{Q} = \{m, a, l\}$. m ist wie auch im 2-Zustände Modell der Startzustand. Aus den oben angegebenen Übergangswahrscheinlichkeiten ergibt sich dann die Transitionsmatrix

$$T = \begin{pmatrix} p_{mm} & p_{ma} & p_{ml} \\ p_{am} & p_{aa} & p_{al} \\ p_{lm} & p_{la} & p_{ll} \end{pmatrix} \quad (3.1)$$

In Modell 3a sind die Wahrscheinlichkeiten für p_{al} und p_{la} immer 0 und in 3b gilt dies entsprechend für p_{ml} und p_{lm} .

In den späteren Abschnitten der Arbeit ist mit dem 3-Zustände Modell oft das Modell 3a gemeint, da sich in ersten Auswertungen der beiden Modelle ergeben hat, dass sich damit tailende Peaks erzeugen lassen, für das Modell 3b wurden bei diesen oberflächlichen Untersuchungen keine solche Peaks gefunden. Die später vorgestellten Methoden und deren Implementierungen können aber alle Varianten des 3-Zustände Modells simulieren.

Die Parameter, die für das Modell 3a betrachtet werden, sind p_{mm} , p_{ml} , p_{aa} und p_{ll} , die übrigen drei Parameter lassen sich aus diesen berechnen.

3.3.1 PAA

Analog zum PAA für nur zwei Zustände kann für drei Zustände ein PAA definiert werden. Aus Gründen der Übersicht wird hier nur das Gesamtmodell mit drei Zuständen und allen dazwischen möglichen Übergängen berücksichtigt. Für die Modelle 3a und 3b müssen auch für den PAA nur die entsprechenden Wahrscheinlichkeiten angepasst werden.

Eine graphische Darstellung des PAA für drei Zustände ist in Abbildung 3.4 gegeben. Der stationäre Zustand wurde in zwei Zustände aufgeteilt, sodass $\mathcal{Q} = \{m, a, l\}$. m bleibt Startzustand. Die Transitionen wurden entsprechend der Transitionsmatrix 3.1 angepasst. An den Werten ändert sich im Vergleich zum PAA für 2-Zustände nichts, da weiterhin die

gleiche Säulenlänge ℓ zu erreichen ist. Ebenso bleibt die Addition für alle Zustände z als Operation θ_z bestehen. Bezüglich der Emissionen ändert sich am mobilen Zustand nichts und die beiden stationären Zustände übernehmen das gleiche Verhalten des bisherigen stationären Zustandes, da sich in beiden die Analyte nicht fortbewegen. Der PAA für das 3-Zustände Modell ist also definiert als:

- $\mathcal{Q} = \{m, a, l\}$
- $q_0 = m$
- $T = \begin{pmatrix} p_{mm} & p_{ma} & p_{ml} \\ p_{am} & p_{aa} & p_{al} \\ p_{lm} & p_{la} & p_{ll} \end{pmatrix}$
- $\mathcal{E} = \{0, 1\}$
- $e_m(z) = \llbracket z = 1 \rrbracket$
- $e_a(z) = e_l(z) = \llbracket z = 0 \rrbracket$
- $\mathcal{V} = \{0, \dots, \ell\}$
- $v_0 = 0$
- $\theta_m = \theta_a = \theta_l = +$

3.4 Weitere Modelle

Eine weitere Möglichkeit, das Modell zu verändern, besteht darin, das Gleichgewicht, welches sich zwischen den beiden Phasen aufbaut, zu berücksichtigen. Damit müssten die Wahrscheinlichkeiten, den Zustand zu wechseln, nicht mehr fest vorgegeben sein. Statt dessen müssten sie sich für jeden Ort aus der aktuellen Verteilung der Teilchen auf die Zustände berechnen. Diese Berechnung müsste dynamisch während der Simulation für jeden Schritt erfolgen. Das würde dazu führen, dass Teilchen, die sich zu Beginn des Pulks befinden, mit einer höheren Wahrscheinlichkeit in die stationäre Phase oder eine der stationären Phasen übergehen, Teilchen in der Mitte haben demgegenüber eine deutlich geringere Wahrscheinlichkeit, stationär zu werden, da nun das Gleichgewicht gehalten werden muss. Die Veränderung der Übergangswahrscheinlichkeiten hinge demnach auch von der Quote der Teilchen ab, die dann schon wieder mobil geworden sind. Teilchen am Ende des Pulks sollten demnach nur sehr selten noch in eine stationäre Phase wechseln können. Genau diese Teilchen würden jedoch einen Tail verursachen, da sie sich, nach dem Übergang in die mobile Phase, hinter dem Pulk befinden würden.

Eine sehr ähnliche Idee ist es, noch eine Sättigung der freien Plätze zur Adsorption einzubauen. Unabhängig vom Gleichgewicht zwischen den Phasen, kann es möglich sein, dass sehr viele Teilchen bereits mit der Oberfläche interagieren. Weitere Teilchen, die sich an den Rand der Säule begeben haben dann möglicherweise kaum noch Kontaktmöglichkeiten mit der stationären Phase, sodass sich ihre Wahrscheinlichkeit, mobil zu bleiben, erhöht.

Kapitel 4

Methoden

In diesem Kapitel werden für die in Kapitel 3 vorgestellten Modelle Möglichkeiten beschrieben, diese zu simulieren. Zunächst werden zwei Herangehensweisen erläutert, das 2-Zustände und das 3-Zustände Modell als Simulation vieler Einzelteilchen umzusetzen.

Im zweiten Abschnitt des Kapitels wird beschrieben, wie der PAA umgesetzt werden kann. Dabei liegt der Fokus auf einer eingeschränkten Variante des PAA, bei dem einige nicht benötigte Funktionalitäten, wie beispielsweise je Zustand verschiedene Operationen oder mehrere mögliche Emissionen, nicht berücksichtigt werden.

Die Ankunftszeiten der simulierten Teilchen bzw. die Verteilung der Wartezeiten auf das Erreichen der Säulenlänge, können dann als Peak aufgefasst und dargestellt werden. Wie aus diesen Daten außerdem werden die resultierenden Peakdaten berechnet werden können, steht im letzten Abschnitt beschrieben.

4.1 Simulation

Offensichtlich ist es möglich, tatsächlich jeden einzelnen Zeitschritt für alle Teilchen zu betrachten. Jedes Teilchen durchläuft also den Automaten aus Abbildung 3.1 und es wird in jedem Schritt entschieden, ob es sich fortbewegt. Diese schrittweise Art der Simulation wird im Folgenden als “Step-by-Step” bezeichnet.

Wenn Zustandswechsel jedoch nicht sehr häufig vorkommen, kann es viel Zeit einsparen, nur zu relevanten Zeitpunkten, also für jeden Zustandswechsel, zu simulieren und dabei den Zeitpunkt des nächsten Wechsels zu notieren. Diese Zeitpunkte werden als Ereignisse verwaltet, daher wird diese Art der Simulation im Folgenden auch “By-Event” genannt.

4.1.1 Step-by-Step

Der Ablauf einer schrittweisen Simulation ist in Algorithmus 4.1 gegeben und unterscheidet sich zunächst nicht für das 2-Zustände und 3-Zustände Modell.

Simuliert werden n Teilchen. Für jedes Teilchen p muss jeweils der aktuelle Zustand z_p und Ort o_p festgehalten werden, außerdem muss in jedem Schritt der aktuelle Zeitpunkt t vermerkt werden. Die Zustände z_p und Orte o_p der Teilchen werden jeweils in einem Array festgehalten. Die Identifikation der Teilchen geschieht dabei über ihre Position, die in beiden Arrays identisch ist.

Simuliert wird, bis alle Teilchen das Säulenende erreicht haben, also die Vektoren eine Länge von 0 haben. Als alternative Abbruchbedingung kann eine maximale Anzahl von

Eingabe: Anzahl der Teilchen n , Länge der Säule ℓ , Zustandsliste $z = [z_1, \dots, z_n]$, Ortsliste $o = [o_1, \dots, o_n]$, Zeit $t = 0$, Maximalzeit t_{\max} , Transitionen T

Ausgabe: Liste von Ankunftszeiten r

```

1:  $r = \text{list}()$ 
2: while  $t < t_{\max}$  and  $\text{len}(z) > 0$  do
3:   for all Teilchen  $p$  do
4:     Aktualisiere  $z_p$  und  $o_p$  anhand von  $T$ 
5:     if  $o_p > \ell$  then
6:        $r.\text{append}(t)$ 
7:       Lösche  $z_p$  und  $o_p$ 
8:    $t \leftarrow t + 1$ 
9: return  $r$ 

```

Algorithmus 4.1: Ablauf der Step-by-Step Simulation

Simulationsschritten gewählt werden, nach deren Ablauf die Simulation beendet ist, wenn nur Peaks von Interesse sind, die vor diesem Zeitpunkt im Chromatogramm erscheinen.

Es wird nun jeweils ein Zeitschritt für alle Teilchen simuliert. Nach jedem solchen Simulationsschritt wird getestet, ob Teilchen bereits den Wert von ℓ erreicht und damit die ganze Säule durchlaufen haben. Die entsprechende Anzahl dafür benötigter Schritte entspricht der Ankunftszeit der Teilchen und wird festgehalten, sodass als Ergebnis am Ende der Simulation eine Liste von Ankunftszeiten aller Teilchen entsteht. Teilchen, die angekommen sind, werden aus beiden Vektoren gelöscht.

Nach Beendigung der Simulation, ist die Rückgabe die Liste r der Ankunftszeiten aller Teilchen, die das Säulenende erreicht haben. Anhand der Länge von r kann leicht festgestellt werden, ob die Parameter p_s und p_m einen vollständigen Peak innerhalb des Maximalzeitraumes erzeugt haben.

Simulation des 2-Zustände Modells

Da nur zwei Zustände existieren und für alle p gilt $z_p \in \{0,1\}$, bietet es sich an, als Zustandsvektor ein Array mit booleschen Werten zu verwenden. 0 steht dabei für stationär und 1 für mobil.

Die Berechnung eines Simulationsschrittes läuft wie folgt: Die Teilchen wechseln in jedem Schritt mit einer bestimmten Wahrscheinlichkeit den Zustand. Diese sei wie in Kapitel 3 erwähnt, p_m für das Verweilen in der mobilen Phase, p_s für das Verweilen in der stationären Phase und $1 - p_m$ bzw. $1 - p_s$ für die beiden Möglichkeiten des Phasenwechsels. Daher muss pro Schritt und Teilchen eine Zufallszahl gezogen werden, anhand derer sich entscheidet, wie es sich verhält. Außerdem muss in jedem Schritt der Ort für jedes mobile Teilchen um eins erhöht werden.

Zum Ablauf der Simulation wurde entschieden, dass zunächst die Entscheidung über die Phase gefällt wird und sich die Teilchen anschließend gemäß ihres neuen Zustandes bewegen. Alle Teilchen starten ganz zu Beginn in der mobilen Phase, da sie nur so in das System eintreten können. Außerdem beginnen sie an Ort 1, da sie bereits als in der Säule befindlich angenommen werden. Das heißt, die Eingaben für Algorithmus 4.1 sind $z = [z_1 = 1, \dots, z_n = 1]$ für die Zustände und $o = [o_1 = 1, \dots, o_n = 1]$ für die Orte.

In Abbildung 4.1 ist eine schematische Darstellung eines Simulationsschritts zu finden. Die angenommenen Wahrscheinlichkeiten für das Verweilen in den Zuständen sind hier $p_s = 0,8$ und $p_m = 0,5$.

Phasen z	0	0	1	1	1	1	0
Orte o	0	4	4	1	6	6	3
ZV	0,9	0,6	0,2	0,9	0,7	0,2	0,7
bleibe stat.: s	0	1	1	0	1	1	1
bleibe mobil: m	0	0	1	0	0	1	0
Phasen z'	1	0	1	0	0	1	0
Orte o'	1	4	5	1	6	7	3

Abbildung 4.1: Prinzip der schrittweisen Simulation mit den Parametern $p_s = 0,8$, $p_m = 0,5$. Beispiel anhand eines Schrittes: Ausgehend von dem Zustandsvektor z und Ortsvektor o mit den gezogenen Zufallszahlen ZV, werden zunächst Hilfsvektoren s und m berechnet und anhand des Ausgangszustandes die neuen Zustands- und Ortsvektoren z' und o' berechnet. Farblich markiert sind alle vier Möglichkeiten, die es zum Phasenübergang gibt.

Um eine parallele Bearbeitung aller Teilchen zu ermöglichen, soll nicht jeweils eine explizite Zustandsabfrage gemacht werden. Statt dessen werden für alle n gezogenen Zufallszahlen die daraus resultierenden Übergänge für beide Zustände berechnet und in zwei Hilfsvektoren gespeichert, in s wird jeweils vermerkt, ob ein Teilchen stationär bleibt und in m , ob es mobil bleibt, so es denn im jeweiligen Zustand war. Beispielsweise wurde für die violett und orange markierten Teilchen eine 0,7 gezogen, was für mobile Teilchen wie das Violette bedeutet, dass sie den Zustand wechseln müssen, für stationäre Teilchen wie das Orangene jedoch, dass sie stationär bleiben. Dadurch, dass alle Übergänge vorberechnet wurden, werden die neuen Zustände für alle Teilchen p über die Formel: $(z_p \wedge m_p) \vee (\neg z_p \wedge \neg s_p)$ berechnet.

Für die neuen Orte kann einfach der neue Zustandsvektor auf den alten Ortsvektor addiert werden.

Simulation des 3-Zustände Modells

Durch die Hinzunahme eines weiteren Zustandes können nun nicht mehr boolesche Werte zur Zustandsbeschreibung herangezogen werden. Nun gilt Zustand $z \in \{0, 1, 2\}$ wobei 0 für den mobilen Zustand steht, 1 für die Adsorption und 2 für die Lösung. Außerdem werden die Wahrscheinlichkeiten entsprechend auf eine Transitionsmatrix wie in Definition 3.1 angepasst.

Damit weiterhin eine logische Formel für die Zustandsberechnung genutzt werden kann und so die neuen Zustände parallel berechnen können, sind einige Hilfsarrays nötig. Um aus den gezogenen Zufallszahlen die passenden Arrays zu erstellen, aus denen die neuen Zustände berechnet werden können, wird zunächst eine Hilfstabelle K mit den kumulierten Wahrscheinlichkeiten erstellt.

Startzustände	0	1	0	2	2	1	0	2	1
M_0	1	0	1	0	0	0	1	0	0
M_1	0	1	0	0	0	1	0	0	1
M_2	0	0	0	1	1	0	0	1	0
ZV	0,3	0,9	0,7	0,8	0,2	0,1	0,9	0,1	0,3
U_{00}	0	1	1	1	0	0	1	0	0
U_{01}	0	1	0	0	0	0	1	0	0
$M_0 \cdot (U_{00} + U_{01})$	0	0	1	0	0	0	2	0	0
U_{10}	1	1	1	1	0	0	1	0	1
U_{11}	0	1	1	1	0	0	1	0	0
$M_1 \cdot (U_{10} + U_{11})$	0	2	0	0	0	0	0	0	1
U_{20}	1	1	1	1	1	0	1	0	1
U_{21}	0	1	1	1	0	0	1	0	0
$M_2 \cdot (U_{20} + U_{21})$	0	0	0	2	1	0	0	0	0
Endzustände	0	2	1	2	1	0	2	0	1

Abbildung 4.2: Berechnung eines Zustandswechsels bei der schrittweisen Simulation im 3-Zustände Modell. Farblich gleich markierte Felder zeigen einen gemeinsamen Ausgangszustand mit den drei Möglichkeiten zum Phasenübergang.

$$K = \begin{pmatrix} p_{mm} & (p_{ma} + p_{mm}) & (p_{ml} + p_{ma} + p_{mm}) \\ p_{am} & (p_{aa} + p_{am}) & (p_{al} + p_{aa} + p_{am}) \\ p_{lm} & (p_{la} + p_{lm}) & (p_{ll} + p_{la} + p_{lm}) \end{pmatrix} \quad (4.1)$$

Beispielsweise wird aus den Transitionen

$$T = \begin{pmatrix} 0,5 & 0,3 & 0,2 \\ 0,2 & 0,2 & 0,6 \\ 0,15 & 0,15 & 0,7 \end{pmatrix} \text{ die folgende Hilfstabelle: } K = \begin{pmatrix} 0,5 & 0,8 & 1 \\ 0,2 & 0,4 & 1 \\ 0,15 & 0,3 & 1 \end{pmatrix}$$

In Abbildung 4.2 ist der Ablauf der Simulation eines Zustandswechsels für mehrere Teilchen exemplarisch zu sehen. Farblich markiert sind jeweils für einen gleichen Startzustand die relevanten Berechnungsschritte. Die Übergangswahrscheinlichkeiten entsprechen denen im obigen Beispiel. Zunächst wird pro Zustand z eine Maske M_z erstellt, die eine 1 an den Positionen enthält, an denen sich ein Teilchen in z befindet.

Außerdem werden analog zum 2-Zustände Modell jeweils aus den Zufallszahlen alle möglichen Übergänge berechnet. U_{xy} steht dabei für den Übergang von Zustand x nach Zustand y . Die Übergänge für U_{x2} würden für alle x auf Grund der kumulierten Wahrscheinlichkeiten nur Nulleinträge enthalten, daher können sie ausgelassen werden.

Der neue Zustand eines Teilchens, das sich in Startzustand x befindet, ergibt sich dann aus der Summe der beiden Arrays U_{x0} und U_{x1} . Über diese Summe wird nun noch die passende Maske M_x gelegt, damit jeweils nur der richtige Zielzustand für den Startzustand ausgewählt wird. Dadurch entsteht pro Zustand ein Array, welches für jedes in diesem Zustand befindliche Teilchen den neuen Zustand nach dem Schritt beinhaltet. Diese Arrays brauchen nun nur noch kombiniert zu werden, um die Endzustände für alle Teilchen zu erhalten.

Die Berechnung der neuen Orte der Teilchen ist dann wiederum einfach. Da die Maske für den mobilen Zustand bereits alle mobilen Teilchen markiert, kann diese auf den Ortsvektor addiert werden. Um M_0 nicht nach den Zustandswechseln neu berechnen zu

müssen, findet das Fortbewegen im Simulationsschritt vor dem Zustandswechsel statt und die Teilchen starten dafür an Position 0 statt an 1 wie es bei der 2-Zustände Simulation der Fall war. Die Eingabe der Orte für Algorithmus 4.1 ist also $o = [o_1 = 0, \dots, o_n = 0]$.

4.1.2 By-Event

Der Ablauf der ereignisbasierten Simulation ist in 4.2 angegeben. Auch bei dieser Simulationsart unterscheidet sich das 2-Zustände Modell vom 3-Zustände Modell nur in der Realisierung eines Simulationsschrittes, in diesem Fall der Bearbeitung eines Events.

Eingabe: Anzahl der Teilchen n , Länge der Säule ℓ , Zeit $t = 0$, Maximalzeit t_{\max} , Eventliste $e = [e_0 = [(z_0, o_0), \dots, (z_n, o_n)]]$ mit $z_p = 0, o_p = 0 \forall p$, Transitionen T

Ausgabe: Ankunftszeiten r

```

1:  $r = \text{list}()$ 
2: while  $t < t_{\max}$  and Länge ( $e$ )  $> 0$  do
3:   for all  $(z_p, o_p)$  in  $e[t]$  do
4:     if  $o_p > \ell$  then
5:       Berechne Ankunftszeit  $t_a(p)$ 
6:        $r.\text{append}(t_a(p))$ 
7:       Lösche  $(z_p, o_p)$ 
8:     else
9:       Simuliere  $(z_p, o_p)$ 
10:      Aktualisiere  $e$ 
11:     $t \leftarrow t + 1$ 
12: if Länge ( $e$ )  $> 0$  then
13:   for all Event  $e[t]$  in  $e$  do
14:     for all  $(z_p, o_p)$  in  $ev$  do
15:       if  $o_p > \ell$  then
16:         Berechne Ankunftszeit  $t_a(p)$ 
17:          $r.\text{append}(t_a(p))$ 
18: return  $r$ 

```

Algorithmus 4.2: Ablauf der by-Event Simulation

Wie bei der schrittweisen Simulation werden hier n Teilchen simuliert. Jedes davon besteht dabei aus seinem Ort und seinem Zustand. Zunächst starten alle Teilchen bei Zeitpunkt 0, an Ort 0 und stationär. Dadurch werden sie im ersten Simulationsschritt mobil und bewegen sich direkt fort. Im Gegensatz zur schrittweisen Simulation wird nun eine Liste mit Zeitpunkten verwaltet. Jeder Zeitpunkt beinhaltet eine Liste von Teilchen, die zu diesem Zeitpunkt simuliert werden müssen. Nach erfolgter Simulation aller Teilchen der Liste, wird der Zeitpunkt gelöscht und der nächste Zeitpunkt, an dem etwas passieren soll, bearbeitet.

Simuliert wird auch hier bis entweder eine maximale Zeit erreicht ist oder alle Teilchen die gesamte Säule durchlaufen haben. Auch das Ergebnis ist wie bei der schrittweisen Simulation eine Liste mit Ankunftszeiten aller Teilchen, die das Säulenende erreicht haben.

Ein Event beinhaltet alle Teilchen eines Zeitpunktes. Für jedes wird zunächst getestet, wo sich das Teilchen befindet. Falls es noch nicht das Säulenende erreicht hat, wird es simuliert. Diese Simulation umfasst den Wechsel des Zustands, die Bestimmung des Zeitpunkts des nächsten Zustandswechsels sowie bei mobilen Teilchen eine Änderung des Ortes

um eine entsprechende Anzahl Schritte. Anschließend wird das Teilchen der Liste seines nächsten Aktionszeitpunktes hinzugefügt.

2-Zustände Modell

Wann ein Teilchen das nächste Mal seinen Zustand wechselt, wird mit Hilfe einer geometrisch verteilten Zufallszahl mit Parameter $p = 1 - p_s$ bzw. $p = 1 - p_m$ bestimmt: Der nächste Zufallswechsel von Zustand z nach $\neg z$ findet statt an Zeitpunkt $t = \text{geom}(1 - p_z)$. Daher lohnt sich diese Art der Simulation insbesondere bei Parametern nahe 1. Wenn ein Teilchen im aktuellen Schritt in den mobilen Zustand gewechselt hat, muss auch sein Ort angepasst werden. Da sich Teilchen eine Raumeinheit pro Zeitschritt bewegen, wird der Ort entsprechend der gezogenen Zufallszahl erhöht.

Allerdings gestaltet sich das Testen auf Erreichen der Säulenlänge bei dieser Methode etwas schwieriger, da ℓ nur sehr selten genau getroffen wird. In vielen Fällen wird ein Teilchen von einem Ort kleiner ℓ direkt auf einen viel größeren Ort befördert. Daher muss zwischen allen Simulationsschritten getestet werden, ob ein Teilchen bereits ℓ überschritten hat. Wenn das der Fall ist, muss mit Hilfe der Differenz zum aktuellen Simulationszeitpunkt die korrekte Ankunftszeit des Teilchens berechnet werden. Ob diese Berechnung nach jedem Simulationsschritt, also vor Neueintragung in die Liste geschieht, oder vor jedem Schritt, ist egal. Allerdings muss im zweiten Fall beachtet werden, dass, falls als Abbruchbedingung eine Maximalzeit gewählt wurde, sichergestellt sein muss, dass alle übrigen Teilchen noch einmal auf Erreichen von ℓ getestet werden.

In Abbildung 4.3 ist beispielhaft die Simulation eines Events gezeigt. Wie in 4.3a zu sehen, ist das nächste Event e_3 . Es beinhaltet die Teilchen $p_1 = (z_1, o_1)$, $p_4 = (z_4, o_4)$ und $p_7 = (z_7, o_7)$. Keines der Teilchen hat bereits die Säulenlänge erreicht, sodass für alle ein Schritt simuliert wird. Gezogen werden die Zufallszahlen $x = 2$, $y = 13$ und $z = 6$. Damit werden die Teilchen den Events e_5 , e_{16} und e_9 hinzugefügt. p_7 war mobil und geht nun in den stationären Zustand über. Deshalb ändert sich auch nichts an seinem Ort. p_1 und p_4 waren stationär und werden beide mobil. Darum muss ihr Ort noch um 2 bzw. 13 angepasst werden. Der Zustand der Zeitpunkteliste nach Abarbeitung der drei Teilchen ist in 4.3b abgebildet. Im nächsten Durchlauf wird das Event e_4 bearbeitet.

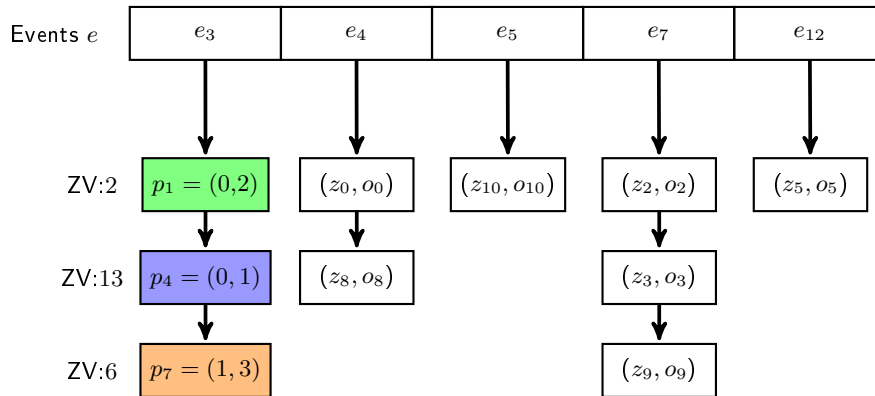
3-Zustände Modell

Bei dieser Simulationsart sind nur minimale Anpassungen notwendig. Wie in der Step-by-Step Simulation wird ein weiterer Zustand eingeführt, sodass $z_p \in \{0,1,2\}$ und die Wahrscheinlichkeiten werden angepasst.

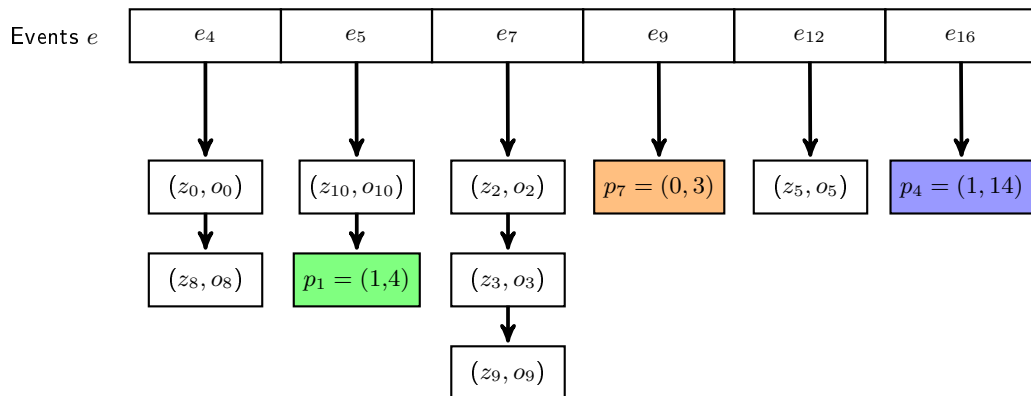
Die Berechnung, wann ein Teilchen von Zustand z in einen neuen Zustand z' mit $z' \neq z$ wechselt, bleibt gleich, da sie nur vom aktuellen Zustand bzw. der Wahrscheinlichkeit in diesem Zustand zu bleiben abhängt. Unverändert bleibt auch die Bestimmung des neuen Ortes sowie der Test, ob ein Teilchen die Säule vollständig durchquert hat.

Es muss lediglich eine Entscheidung, in welchen der beiden anderen Zustände das Teilchen wechselt, eingeführt werden. Ein Ereignis beinhaltet stets einen Zustandswechsel, also hat das Teilchen jeweils zwei mögliche Folgezustände. Die Übergangswahrscheinlichkeit von z nach z ist also immer 0. Die Wahrscheinlichkeiten für die Wechsel in einen anderen der beiden Zustände müssen entsprechend ihres Anteils an der Restwahrscheinlichkeit angepasst werden.

Dafür kann vorab eine Wechselmatrix berechnet werden, in der sich die Wahrscheinlichkeiten für Folgezustände ablesen lassen.



(a) Eventliste vor der Simulation des Events e_3 , für welches 3 Teilchen simuliert werden müssen



(b) Eventliste nach der Simulation, Event e_3 wurde abgearbeitet und gelöscht, die Teilchen entsprechend der gezogenen Zufallszahlen simuliert und den passenden Events neu zugeordnet, bzw neue Events angelegt

Abbildung 4.3: Prinzip der ereignisbasierten Simulation bei zwei Zuständen, am Beispiel eines Events

$$\text{Wechselmatrix } W = \begin{pmatrix} 0 & p_{ma}/(1-p_{mm}) & p_{ml}/(1-p_{mm}) \\ p_{am}/(1-p_{aa}) & 0 & p_{al}/(1-p_{aa}) \\ p_{lm}/(1-p_{ll}) & p_{la}/(1-p_{ll}) & 0 \end{pmatrix} \quad (4.2)$$

Aus der Transitionsmatrix

$$T = \begin{pmatrix} 0,5 & 0,3 & 0,2 \\ 0,2 & 0,2 & 0,6 \\ 0,15 & 0,15 & 0,7 \end{pmatrix} \text{ wird beispielsweise } W = \begin{pmatrix} 0 & 0,6 & 0,4 \\ 0,25 & 0 & 0,75 \\ 0,5 & 0,5 & 0 \end{pmatrix}$$

Nun muss nur noch eine weitere Zufallszahl gezogen werden, anhand derer sich mit Hilfe der Wechselmatrix der Folgezustand bestimmt.

4.2 PAA

Die im Folgenden vorgestellte Umsetzung des PAA beinhaltet einige Einschränkungen gegenüber dem allgemeinen Modell, da nicht alle Funktionalitäten benötigt werden.

Dies betrifft die Möglichkeit je Zustand mehrere verschiedene Emissionen mit einer jeweiligen Wahrscheinlichkeit zuzulassen. Es wird, wie in den Abbildungen 3.2 und 3.4 zu sehen ist, nur jeweils ein Wert mit der Wahrscheinlichkeit 1 emittiert. Außerdem wird keine Möglichkeit weitere Operationen zuzulassen benötigt, Addition genügt. Da darüber hinaus nur im mobilen Zustand überhaupt eine Veränderung des Wertes erfolgt, wird der PAA auf diese Möglichkeit beschränkt.

Außerdem wird nicht jeweils die vollständige Zustands-Werte-Verteilung betrachtet, sondern nur der tatsächlich relevante Bereich. Dieser Bereich umfasst zu Beginn einer Simulation lediglich die bereits erreichbaren Werte. Für alle $n < \ell$ ist n der maximal erreichbare Wert in Schritt n .

Zusätzlich werden Werte sehr nah bei 0 ignoriert. Solche Werte werden aus der Verteilung gelöscht und die dadurch benötigten Arrays verkürzt, wodurch wiederum Simulationszeit eingespart wird.

4.2.1 PAA für das 2-Zustände Modell

Eingabe: Startverteilungen s und m , Maximalzeit t_{\max} , Zielwert ℓ , Parameter T , Startzeit $t = 0$

Ausgabe: Verteilung der Wartezeit

```

1: while  $t < t_{\max}$  do
2:   Aktualisiere  $s$  und  $m$ 
3:   if  $\text{length}(m) > \ell$  then
4:      $r \leftarrow (s[\ell] + m[\ell])$ 
5:     Lösche  $s[\ell]$  und  $m[\ell]$ 
6:   else
7:      $r \leftarrow 0$ 
8:    $t+ = 1$ 
9: return  $r$ 
```

Algorithmus 4.3: Berechnung der Wartezeit für den PAA im 2-Zustände Modell

In Algorithmus 4.3 ist der Ablauf der Simulation mit einem PAA beschrieben. Die Verteilungen werden in jedem Schritt in s und m gespeichert. Die Position im Array entspricht

dabei dem Wert und damit der Position in der Säule. Beispielsweise bezeichnet $s_t[0]$ die Wahrscheinlichkeit, mit der sich zum Zeitpunkt t Teilchen im stationären Zustand und an Position 0 in der Säule befinden.

Die Simulation beginnt mit $s_0 = [0]$ und $m_0 = [1]$. Alle Teilchen befinden sich zunächst im mobilen Zustand und am Beginn der Säule. Weitere Einträge würden nur aus Wahrscheinlichkeiten von 0 bestehen und werden daher ausgelassen.

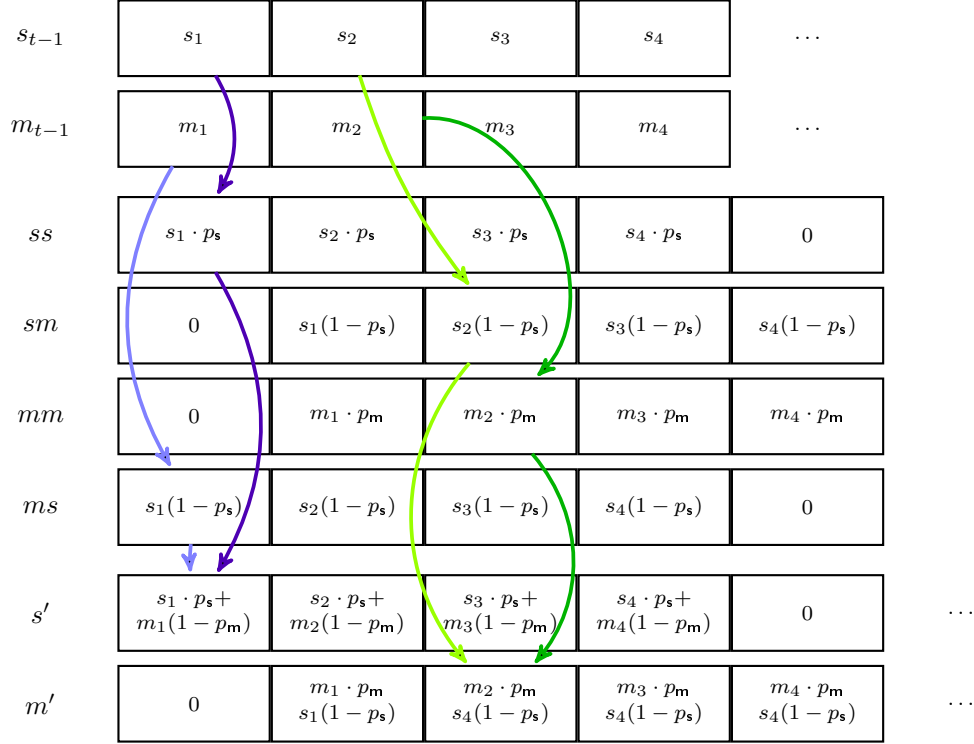


Abbildung 4.4: Berechnung der neuen Zustände des PAA im 2-Zustände Modell, Entstehung der neuen Wahrscheinlichkeit für den stationären Zustand durch blaue Pfeile gekennzeichnet. Diese setzt sich zusammen aus stationär bleibenden (dunkelblau) und vorher mobilen (hellblau) Anteilen. Analog die Berechnung für mobile Wahrscheinlichkeiten, gekennzeichnet durch grüne Pfeile, unter Berücksichtigung der Ortsänderung

In jedem Schritt werden die beiden Verteilungen aktualisiert. Diese Berechnung ist in Abbildung 4.4 exemplarisch dargestellt.

Gegeben sind die alten Wahrscheinlichkeitsverteilungen s_{t-1} und m_{t-1} . Um auf die neue Verteilung für den stationären Zustand zu kommen, muss für jede Position und jeden möglichen Ausgangs- und Folgezustand berechnet werden, welchen Anteil an Teilchen er enthält. In ss sind diejenigen Anteile enthalten, die stationär waren und es bleiben, in ms diejenigen, die aus dem mobilen Zustand in den stationären Zustand wechseln. In Summe ergeben ss und sm die neuen Verteilungen für den stationären Zustand. In sm und mm sind dann die Anteile enthalten, die vom stationären Zustand mobil werden, beziehungsweise mobil bleiben. Da in diesen Fällen die Teilchen anschließend mobil sind, müssen die Felder um eine Position verschoben werden. Zum Ausgleich werden den neuen stationären Verteilungen jeweils ein Feld mit der Wahrscheinlichkeit von 0 angehängt. Daher vergrößern sich die beiden Verteilungsvektoren um jeweils ein Feld pro Schritt.

Eine Berechnung der Zustände-Werte Verteilung würde nur aus dieser schrittweisen Aktualisierung entstehen. Um die Verteilung der Wartezeiten zu berechnen, wird nach je-

dem solchen Berechnungsschritt getestet, ob Anteile der Teilchenmasse bereits den Zielwert ℓ erreicht haben. Dies ist der Fall, wenn sich ein Eintrag an Position ℓ befindet, also kann einfach auf die Länge der Verteilungsvektoren getestet werden. Im positiven Fall würde der theoretische PAA in Zustand \bullet übergehen. Hier wird die Wahrscheinlichkeit an Stelle ℓ ins Ergebnis geschrieben. Außerdem wird der nächste Übergang in den theoretischen Zustand \circ , aus dem kein Übergang wieder heraus führt, dadurch realisiert, dass die Werte an Position ℓ gelöscht werden. Ist die Länge der Verteilungsvektoren kürzer als ℓ , sind keine Anteile der Teilchenmasse angekommen und es wird entsprechend eine Wahrscheinlichkeit von 0 ins Ergebnis geschrieben. Dadurch entsteht letztendlich eine Liste von Ankunfts-wahrscheinlichkeiten für jeden Zeitpunkt.

Bei Erreichen der maximalen Wartezeit wird die Berechnung beendet. Ob die gesamte Wahrscheinlichkeitsmasse erreicht wurde kann anschließend berechnet werden, um herauszufinden, ob mit den getesteten Parametern ein Ergebnis innerhalb der vorgegebenen Wartezeit erzielt werden konnte.

Wie bereits erwähnt, sollen zu kleine Wahrscheinlichkeiten ignoriert werden. Daher wird nach jedem Schritt getestet, ob an der ersten oder letzten Position von s und m Werte unterhalb eines Schwellwerts liegen. Ist das der Fall, werden beide Werte entfernt, wodurch sich die Verteilungsvektoren um ein Feld verkürzen. Wenn die letzte Position gelöscht wird, hat dies auf den sonstigen Ablauf keine weiteren Auswirkungen. Wenn die erste Position gelöscht wird, muss jedoch ein zusätzliche eingeführter Index erhöht werden, der beim Testen auf Erreichen des Zielwertes auf die Länge der Arrays addiert werden muss, um korrekte Ergebnisse zu erhalten.

4.2.2 PAA für das 3-Zustände Modell

Der PAA wird auf drei Zustände erweitert, indem einfach ein weiterer Verteilungsvektor eingeführt wird. Es gibt dann also für jeden der drei Zustände m , a und l ein Array m , a und l , in welchem die Zustands-Werte-Verteilung gespeichert wird. Wie auch beim PAA für das 2-Zustände Modell ist dann zum Beispiel in $a_t[o]$ die Wahrscheinlichkeit enthalten, dass sich zum Zeitpunkt t Teilchen an Position o im adsorbierten Zustand befinden. Algorithmus 4.3 muss dann dahingehend angepasst werden, dass alle drei dieser Verteilungen als Eingabe akzeptiert und in jedem Schritt aktualisiert werden. Die Startverteilungen sind offensichtlich $m_0 = [1]$, $a_0 = [0]$ und $l_0 = [0]$.

Die Aktualisierung der Verteilungen wird ähnlich wie im Fall von nur zwei Zuständen behandelt, nur dass nun neun statt vier Hilfsvektoren nötig sind, um alle möglichen Zustandsübergänge zu berechnen.

Alternativ kann jedes Feld der drei Verteilungen einzeln berechnet werden. Ein neues Feld $z_{t+1}[o]$, welches die Wahrscheinlichkeit an Position o in Zustand z an Zeitpunkt $t+1$ beschreibt, berechnet sich dann aus den drei alten Verteilungen m , a und l nach folgender Formel: $z_{t+1}[o] = m_t[o-1] \cdot p_{mz} + a_t[o] \cdot p_{az} + l_t[o] \cdot p_{lz}$

Nach jedem Schritt werden auch hier zu kleine Werte entfernt, wobei natürlich jeweils überprüft werden muss, ob der Eintrag an der ersten oder letzten Position für jeweils alle drei Vektoren zu klein ist.

4.3 Berechnung der Peakdaten

Sowohl für die Simulation vieler Teilchen als auch bei der Verwendung eines PAA müssen nach der Simulation noch die Eigenschaften der Peaks berechnet werden.

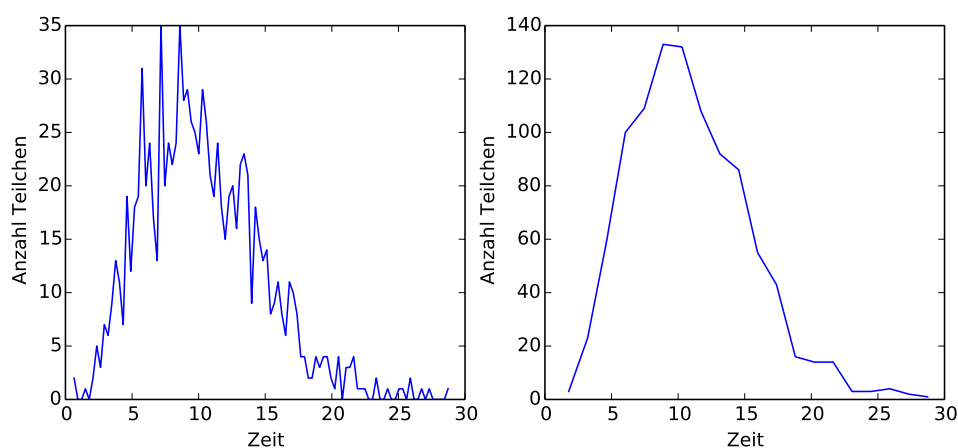


Abbildung 4.5: Beispielpeak vor (links) und nach (rechts) der Zusammenfassung der Datenpunkte

Bei den Teilchensimulationen liegen die Peaks als Liste von Ankunftszeiten vor, beim PAA hingegen als Anteil der Gesamtwahrscheinlichkeit zu jedem Zeitpunkt.

Besonders bei den Teilchensimulationen ist es vorteilhaft, jeweils mehrere Zeitpunkte zusammen zu fassen, um die Peaks zu glätten. Durch die zufälligen Bewegungen der Teilchen kann es sonst vorkommen, dass die Peaks sehr verrauscht sind. Außerdem erfassen Detektoren im Experiment auch nur in gewissen Zeitintervallen die ankommenden Stoffe, sodass dadurch kein Nachteil der Simulation entsteht. Werden die Datenpunkte nicht zusammengefasst, kann es insbesondere bei der Bestimmung des Maximalwertes eines Peaks, zu deutlichen Verschiebungen kommen, wenn zu einem beliebigen Zeitpunkt zufällig mehr Teilchen den Säulendurchlauf beenden, als am Maximalpunkt des zusammengefassten Peaks.

Bei der Simulation, welche die Peaks in Abbildung 4.5 erzeugt hat, wurden 1000 Teilchen simuliert. Es ist auf den ersten Blick nicht klar, wo sich die Maximalstelle des Peaks befindet, sowohl bei $t_1 = 7$ als auch bei $t_2 = 8,5$ erreicht die Kurve ihren Maximalwert. Nach der Zusammenfassung von Ankunftszeiten wird klar, dass die Maximalstelle eher bei $t_2 = 8,5$ liegen muss. Die Berechnung der Lage der Peaks bezieht sich daher auf die zusammengefassten Daten.

Die Lage des Peaks und damit seine Retentionszeit bestimmt sich als am häufigsten vorkommende Ankunftszeit bei der Teilchensimulation bzw. als Lage des Maximalwertes beim PAA.

Für die Bestimmung von Interquartilsabstand und Quartilkoeffizient müssen zunächst die Quartile der Verteilung berechnet werden. Daraus lassen sich die beiden Werte direkt mit den in Kapitel 2 angegebenen Formeln berechnen.

Kapitel 5

Implementierung

Zur Simulation der Modelle wurde die Software MCC-Sim entwickelt, die auf der beigelegten CD zu finden ist. Sie ist unterteilt in zwei Bereiche, die Teilchensimulation und die Simulation des PAA und setzt jeweils die im Kapitel 4 beschriebenen Ansätze zur Simulation um.

5.1 Teilchensimulation/MCC-Sim

5.1.1 Anwendungen

Die für die Teilchensimulation zur Verfügung stehenden Anwendungen sind `simulation.py`, `2s_simulate_and_plot.py` sowie `3s_simulate_and_plot.py`. Diese wurden in Python 3.4 implementiert ¹ und werden im Folgenden vorgestellt.

simulation.py Durch den Aufruf von `simulation.py` können einzelne Simulationen gestartet und deren Ergebnisse angezeigt werden. Die Parameter für die Simulation werden dabei über Kommandozeilenparameter eingegeben. Erster Parameter ist dabei das gewünschte Modell, es besteht die Auswahl zwischen **2s**, **3s** und **3a**, danach müssen Simulationsparameter eingegeben werden. Für das 2-Zustände Modell sind dies p_s und p_m . Für das allgemeine 3-Zustände Modell wird die Übergangsmatrix 3.1 als Liste der neun Parameter **pmm**, **pma**, **pml**, **pam**, **paa**, **pal**, **plm**, **pla**, **pll** übergeben. Wird das Modell 3a ausgewählt, können auch nur die vier nötigen Parameter **pmm**, **pml**, **paa**, **pll** übergeben werden. Die Simulationsart wird über den Kommandozeilenparameter **-a** ausgewählt, dabei stehen **E** für die By-Event Simulation und **S** für die Step-by-Step Simulation zur Auswahl. Ein Aufruf

```
python3 simulation.py 3a 0.5 0.0001 0.9993 0.99999 -a S
```

startet beispielsweise eine schrittweise Simulation des 3-Zustände Modells 3a mit den Parametern $p_{mm} = 0,5$, $p_{ml} = 0,0001$, $p_{aa} = 0,9993$ und $p_{ll} = 0,99999$. Als Ergebnis der Simulation wird anschließend eine Liste von Ankunftszeiten der simulierten Teilchen, sowie die daraus berechneten Peakdaten angezeigt. Eine weitere Ausgabeoptionen ist **-p**, womit zusätzlich eine graphische Darstellung der Ankunftszeiten in Form eines Histogramms angezeigt wird, **-q** unterdrückt die textuelle Ausgabe.

Als Standard für die Simulationen werden eine Säulenlänge von 1000 Einheiten, ein Zeitraum von 240s und 1000 zu simulierende Teilchen angenommen. Diese Einstellungen

¹<http://python.org>

können über die Parameter `-n` (Anzahl der Teilchen), `-l` (Länge der Säule) und `-m` (Maximalzeit) verändert werden. Der Aufruf

```
python3 simulation.py 2s 0.9999 0.01 -a E -n 10000 -m 400
```

startet daher eine ereignisbasierte Simulation im 2-Zustände Modell, bei der 10 000 Teilchen über einen Maximalzeitraum von 400 Sekunden simuliert werden.

2s_simulate_and_plot.py Diese Anwendung erlaubt es für das Modell mit zwei Zuständen, viele Simulationen zu starten und deren Ergebnisse durch verschiedene Plots darzustellen. Die Ergebnisse der einzelnen Simulationen werden dafür abgespeichert und stehen damit auch späteren Auswertungen zur Verfügung. Bereits vorhandene Simulationen werden daher auch nicht neu simuliert.

Mit dem Parameter `-p` wird eine Menge von Parameterkombinationen ausgewählt. Zur Verfügung stehen dabei **small_set**, **medium_set** und **large_set**, außerdem können mittels **random** zufällige Kombinationen erstellt werden, die Anzahl dieser zufälligen Kombinationen wird über den Parameter `-cn` festgelegt. Darüber hinaus können die Parameter `-n`, `-l`, `-m` und `-a` mit der gleichen Funktion wie bei der Anwendung **simulation.py** benutzt werden, um Einstellungen an den Simulationen vorzunehmen. Dabei ist für den Ansatz `-a` die schrittweise Simulation als Standard gesetzt, da diese sich für das 2-Zustände Modell als weniger laufzeitintensiv erwiesen hat, siehe auch (Abschnitt 5.1.4)

Geplottet werden kann auch hier ein einzelner Peak zu gegebenen Parametern, der Aufruf dafür ist

```
python3 2s_simulate_and_plot.py -pp 0.999 0.5
```

Mit der zusätzlichen Option `-q` werden außerdem die Quartile der Verteilung angezeigt.

Zusätzlich möglich sind ein Chromatogramm für das gewählte Set, allerdings nur bis maximal 30 Peaks. Der Befehl dafür ist `--pc` und mit `--an` wird diesem Chromatogramm zusätzlich Rauschen hinzugefügt. Es können auch alle mit dem gewählten Set erreichten Werte für eine Peakeigenschaft dargestellt werden (Abbildung 6.5). Ein Beispiel für einen solchen Aufruf, um die Lage der resultierenden Peaks anzuzeigen, ist

```
python3 2s_simulate_and_plot.py -p medium_set --pt loc
```

Alternativ kann statt `loc` auch `iqr` für die Breite oder `qk` für die Schiefe ausgewählt werden. Des Weiteren können auch mögliche Parameterkombinationen für einen Peak mit gegebener Retentionszeit `-r` und davon erlaubter Abweichung `-e` ausgegeben werden. Der Aufruf

```
python3 2s_simulate_and_plot.py --ppt -r 50 -e 2
```

erzeugt einen Plot, der alle Parameterkombinationen für Peaks mit Maximalzeitpunkten im Intervall [48, 52] enthält. Zuletzt ist mit der Option `--pr` eine Übersicht über alle erreichbaren Zeit-Breiten-Kombinationen verfügbar. Für die beiden letzten Plots werden jeweils nicht nur eine bestimmte Auswahl an Parameterkombinationen, sondern alle vorhandenen Simulationen verwendet.

`3s_simulate_and_plot.py` Analog zur `2s_simulate_and_plot.py` können mit dieser Anwendungen Simulationen des 3-Zustände Modells gestartet und die Ergebnisse davon geplottet werden. Es gibt ebenfalls die vier beschriebenen Möglichkeiten, Parameterkombinationen auszuwählen und die Einstellung an den Simulationen vorzunehmen. Allerdings ist hier der ereignisbasierte Ansatz für `-a` voreingestellt. Es können auch für das 3-Zustände Modell Einzelpeaks mit oder ohne Quartile, sowie Spektren mit oder ohne Rauschen erstellt werden. Die anderen Plots der 2-Zustände Simulation stehen auf Grund der höheren Parameteranzahl nicht zur Verfügung.

5.1.2 Umsetzung der Algorithmen

Die Algorithmen 4.1 und 4.2 wurden jeweils für zwei und drei Zustände implementiert.

Um insbesondere bei der Step-by-Step Implementierung eine schnelle Verarbeitung der Zustands- und Ortsarrays zu gewährleisten, wurden dafür **NumPy**-Arrays verwendet. Außerdem stellt **NumPy**² Methoden zur Berechnung von Percentilen, geometrisch verteilte Zufallszahlen und Bestimmung der Position des größten Wertes einer Liste zur Verfügung, welche für die Implementierung genutzt wurden. Für die graphische Ausgabe der Peaks und weitere geplottete Darstellungen im Bereich der Auswertung der Simulationen wurde **matplotlib**³ genutzt.

Die Datei `simulation.py` enthält alle Funktionen zur Simulation und Berechnung der Peakdaten. Wichtigster Bestandteil ist die abstrakte Klasse **Simulation** und die zwei Unterklassen **Simulation_2s** und **Simulation_3s**. Wie im vorherigen Kapitel gesehen, unterscheiden sich sowohl bei der Step-by-Step, als auch bei der By-Event Variante die Algorithmen nur durch die Umsetzung eines Berechnungsschrittes, sodass die Methoden zur Simulation eines Einzelschrittes bzw. Events in der Superklasse abstrakt sind und jeweils für das Modell passend in den Unterklassen implementiert wurden. In der Klasse **Simulation_3s** sind zusätzlich auch die Vorberechnungen für die kumulierte Parametermatrix aus Definition 4.1 der Step-by-Step Simulation und die Wechselmatrix (4.2) der By-Event Simulation zu finden.

Um eine Simulation zu starten, wird zunächst eine Instanz der Simulationsklasse erstellt:

```
new_Sim = (params, model, approach, length, number, maxtime, times)
```

Für das 2-Zustände Modell sind die Parameter **params** ein Tupel aus (p_s, p_m) . Für das allgemeine 3-Zustände Modell wird die Übergangsmatrix 3.1 als Liste von Listen

`[[pmm, pma, pml], [pam, paa, pal], [plm, pla, pll]]` übergeben. Wird als **model** das Modell 3a ausgewählt, können die vier nötigen Parameter auch als Liste `[pmm, pml, paa, pll]` übergeben werden, die dann passend umgewandelt werden. Zulässige Eingaben für **model** sind `2s`, `3s` und `3a`, womit das entsprechende Modell ausgewählt wird. Über **approach** kann zwischen `S` und `E` gewählt werden, wodurch der Simulationsansatz gewählt wird. Mit **length**, **number** und **maxtime** können die Rahmenparameter für die Simulation gesetzt werden, diese entsprechen ℓ , n und t_{\max} in den Algorithmen 4.1 und 4.2. Als Defaultwerte sind auch hier eine Länge von 1000, sowie 1000 Teilchen und eine maximale Zeit von 240 Sekunden gesetzt.

Zusätzlich ist es möglich, eine bereits vorhandene Liste von Ankunftszeiten mit zu übergeben, aus der beispielsweise die Maße des Peaks berechnet werden sollen.

Mit

²<http://www.numpy.org/>

³<http://matplotlib.org/>

```
new_Sim = ([0.5 0.0001 0.9993 0.99999], "3a", "S", number = 10000)
```

wird beispielsweise eine Simulationsinstanz mit den Parametern $p_{mm} = 0,5$, $p_{ml} = 0,0001$, $p_{aa} = 0,9993$ und $p_{ll} = 0,99999$ für das Modell 3a erzeugt. Diese soll für 10 000 Teilchen mit der Step-by-Step Simulation simuliert werden.

Anschließend wird mit `newSim.simulate()` die Simulation gestartet. Anhand des gesetzten `model` und `approach` entscheidet sich, welche Funktionen zur Simulation aufgerufen werden. Die Peakdaten werden anschließend mittels `newSim.calculate()` berechnet.

5.1.3 Optimierungen

Sowohl bei der Umsetzung der Step-by-Step als auch der By-Event Methode kann mit der Implementierung des 3-Zustände Modells auch das 2-Zustände Modell simuliert werden, indem nicht benötigte Übergangswahrscheinlichkeiten auf einen Wert von 0 gesetzt werden. Die 2-Zustände Simulation ist aber vor allem bei der Step-by-Step Implementierung deutlich schneller, da nur auf boolschen Werten gearbeitet werden muss und deutlich weniger ungenutzte Information in den Zwischenschritten berechnet wird.

Im Laborexperiment wird nur in gewissen Intervallen die Ankunft von Teilchen detektiert, bei der MCC-IMS-Kopplung wird beispielsweise nur zwei Mal pro Sekunde eine Messung durchgeführt. Daher genügt es auch in der schrittweisen Simulation, wenn nicht nach jedem Simulationsschritt die Orte aller Teilchen getestet werden, um ein ausreichend exaktes Ergebnis zu erhalten. Es können also zunächst viele Simulationsschritte durchgeführt werden, bevor getestet wird, ob Teilchen die Säulenlänge ℓ erreicht haben. Nach den oben angegebenen Einheiten können so zwischen jedem Test 50 Simulationsschritte erfolgen, was 0,5 s entspricht. Dadurch entfallen viele Abfragen und es kann Simulationszeit eingespart werden.

Da außerdem klar ist, dass Teilchen eine Säule der Länge ℓ frühestens nach ℓ Schritten überhaupt durchlaufen haben können, muss vor diesem Zeitpunkt auch kein Test auf Erreichen der Länge durchgeführt werden, wodurch weitere Zeit eingespart werden kann.

Diese Optimierungen wurden im Laufzeitvergleich im folgenden Abschnitt bereits berücksichtigt.

5.1.4 Laufzeitvergleich von Step-by-Step und By-Event

Die Laufzeit einer Simulation hängt stark von den gewählten Parametern ab, außerdem gibt es gravierende Unterschiede zwischen dem 2- und 3-Zustände Modell.

In Tabelle 5.1 ist eine Übersicht über die Laufzeiten der verschiedenen Simulationsarten gegeben. Jede Zeile enthält die Zeiten für die gegebenen Parameter, jede Spalte entspricht einer Kombination aus Simulationsart und Teilchenanzahl n .

Die Zeiten wurden auf einem Rechner mit einem Intel® Core™ i7-4500U Prozessor und 8 GB RAM unter jeweils gleichen Bedingungen gemessen und sind in Sekunden angegeben. Grau hinterlegte Parameter deuten an, dass diese Simulationen keinen vollständigen Peak innerhalb des Zeitraumes von 240 s liefern und daher nach Ablauf der maximalen Anzahl an Simulationsschritten beendet werden.

Wie zu erwarten war, steigt die Laufzeit an, je größer p_s und je kleiner p_m ist. Das liegt offensichtlich an der Abbruchbedingung des Algorithmus, der terminiert, wenn alle Teilchen über die gesamte Strecke von ℓ Schritten simuliert wurden. Simulationsparameter mit kleinem p_s oder großem p_m sorgen für weniger Verweilen der Teilchen im stationären Zustand und damit eine schnelle Passage der Strecke, wodurch die Simulationsdauer kürzer wird.

Tabelle 5.1: Laufzeitvergleich für die 2-Zustände Simulation mit den Parametern p_s und p_m für n Teilchen, Zeitangaben in Sekunden. Grau hinterlegte Parameter zeigen einen Peak innerhalb von 240 Sekunden an.

p_s	p_m	Step-by-Step $n = 1000$	By-Event $n = 1000$	Step-by-Step $n = 10\,000$	By-Event $n = 10\,000$
0,997	0,001	8,4	20,9	47,4	98,2
0,997	0,3	6,1	15,4	33,7	69,3
0,997	0,6	3,6	8,9	19,4	40,1
0,997	0,95	0,6	1,3	2,6	5,4
0,999	0,001	25,5	40,2	142,7	124,5
0,999	0,3	18,6	27,8	100,4	88
0,999	0,6	11,0	16,1	57,8	51
0,999	0,95	1,7	2,2	7,6	7,1
0,9993	0,001	37,6	47,6	204,1	140,4
0,9993	0,3	26,6	32,7	143,4	99,3
0,9993	0,6	15,3	19,6	82,6	57,6
0,9993	0,95	2,3	2,5	10,8	8,1
0,9996	0,001	59,9	54,3	341,2	167,8
0,9996	0,3	46,2	39,5	252,6	125,4
0,9996	0,6	26,7	22,4	150,2	72,8
0,9996	0,95	4,1	3,0	18,8	10,2

Bei der Simulation des 2-Zustände Modells ist die Step-by-Step Variante für die meisten Parameterkombinationen deutlich schneller, als die By-Event Implementierung. Der große Vorteil von Step-by-Step liegt hier in der Verwendung von booleschen Werten und Nutzung von NumPy-Arrays, auf welchen die Ausführung logischer Operationen sehr schnell erfolgen kann. Diese Zeiten der Step-by-Step Implementierung würden sich allerdings fast verdoppeln, wenn tatsächlich nach jedem Simulationsschritt eine Abfrage auf Erreichen der Säulenlänge und entsprechende Bearbeitung der fertigen Teilchen erfolgen würde und nicht erst nach den oben erwähnten 50 Schritten.

Mit größer werdendem p_m verringert sich der Unterschied zwischen den beiden Simulationsarten, ab einem sehr großen Wert für p_m , ab etwa 0,99 ist die By-Event Variante geringfügig schneller. Das liegt daran, dass in diesen Fällen besonders selten Ereignisse erzeugt werden müssen und sich die Teilchen in großen Sprüngen forbewegen können.

Anders sieht die Situation bei der Simulation des 3-Zustände Modells 3a aus. Tabelle 5.2 bietet eine Übersicht über die Laufzeiten mit verschiedenen Parametern. Grau hinterlegte Parameter zeigen auch hier an, dass kein vollständiger Peak innerhalb der Retentionszeit von 240 Sekunden erzeugt werden kann, in der Spalte % ist in diesem Fall ist der Prozentsatz der angekommenen Teilchen angegeben.

Auch hier ist für beide Simulationsarten zu erkennen, dass bei größerem p_{mm} oder kleineren Werten für die anderen Parametern deutlich geringere Laufzeiten erreicht werden.

Jedoch müssen bei der Step-by-Step Implementierung sehr viele Einträge für die Hilfsarrays erzeugt werden, die für die Berechnung der neuen Arrays nicht benötigt werden. Durch diese überflüssigen Berechnungen verliert die Step-by-Step Implementierung ihren Vorteil gegen über der By-Event Variante, sodass diese schneller wird. Besonders bei denjenigen Simulationen, die einen sehr späten Peak erzeugen oder gar keinen vollständigen Peak innerhalb der 240 Sekunden erzeugen, wird der Vorteil der By-Event Simulation deutlicher.

Tabelle 5.2: Laufzeitvergleich für die 3-Zustände Simulation mit den Parametern p_{mm} , p_{ml} , p_{aa} und p_{ll} für n Teilchen, Zeitangaben in Sekunden. Grau hinterlegte Kombinationen zeigen einen unvollständigen Peak innerhalb von 240 Sekunden an, in der Spalte % ist für diese der Prozentsatz angekommener Teilchen vermerkt.

p_{mm}	p_{ml}	p_{aa}	p_{ll}	%	S, 1000	E, 1000	S, 10 000	E, 10 000
0,01	0,00005	0,997	0,99995		26,2	23,9	123,7	196,1
0,01	0,00005	0,997	0,999995		47,4	24,4	150	198,9
0,01	0,00005	0,9996	0,99995	18	170,8	35,9	870	219,4
0,01	0,00005	0,9996	0,999995	18	169,7	35,4	869	220,5
0,01	0,003	0,997	0,99995		36,2	25,2	150,6	199,4
0,01	0,003	0,997	0,999995	99	118,8	31,5	390,7	218,3
0,01	0,003	0,9996	0,99995	8	170	34,8	870	217,5
0,01	0,003	0,9996	0,999995	2	170	29,9	871	188,5
0,9	0,00005	0,997	0,99995		5	2,6	15,5	20,7
0,9	0,00005	0,997	0,999995		28,4	2,7	44,3	21,4
0,9	0,00005	0,9996	0,99995		22,5	3,9	98	24,3
0,9	0,00005	0,9996	0,999995		49,9	4	126,1	24,9
0,9	0,003	0,997	0,99995		16,9	3,3	43	22,6
0,9	0,003	0,997	0,999995	99	108,3	4,3	294,2	30,6
0,9	0,003	0,9996	0,99995		30,1	4,0	120	25,3
0,9	0,003	0,9996	0,999995	99	115,6	4,7	370,6	32,4

In diesen Fällen finden meist seltener Wechsel von der stationären in die mobile Phase statt. Bei der Step-by-Step Variante wird dennoch jeder Schritt simuliert, ohne dass sich die Teilchen bewegen. By-Event lässt diese Schritte einfach aus, wodurch deutlich Simulationszeit gespart wird.

5.2 Simulation des PAA

5.2.1 Nutzung von MoSDi

Es existiert bereits eine Implementierung für PAA. Diese wurde innerhalb der Software MoSDi ⁴ realisiert, eine Beschreibung findet sich in Marschall (2011). Um eine Simulation eines PAA mit MoSDi zu starten, muss für die darin enthaltene abstrakte Klasse PAA eine Unterklasse geschrieben werden, welche die Zustände, Werte, Operationen und Transitionswahrscheinlichkeiten für den konkreten PAA enthält. In der beigelegten CD finden sich für beide Modelle jeweils eine solche Unterklasse `myPAA_2s` und `myPAA_3s`. Zum Starten des PAA für das 2-Zustände Modell müssen in der Kommandozeile nacheinander die Befehle

```
javac -classpath mosdi-1.3.jar myPAA_2s.java
java -classpath .:mosdi-1.3.jar myPAA_2s
```

eingegeben werden. Dabei ist darauf zu achten, dass sich `mosdi.jar` im gleichen Verzeichnis wie `myPAA_2s.java` befindet. Analog kann der PAA für das 3-Zustände Modell aufgerufen werden. Damit werden jeweils eine Menge von Parameterkombinationen simuliert, die dem `small_set` der Teilchensimulation entsprechen. Weitere Kombinationen können in der

⁴<https://bitbucket.org/tobiasmarschall/mosdi>

Methode `combineParams()` erzeugt werden. Anschließend werden die Wahrscheinlichkeitsverteilungen zur weiteren Verwendung abgespeichert.

5.2.2 PAA-Sim

Wie im Kapitel 4 erwähnt, werden für die Modelle der Multikapillarsäule jedoch nicht alle Funktionalitäten eines PAA benötigt, sodass zusätzlich noch eine eigene Implementierung erstellt wurde. Diese wurde in Julia⁵ implementiert und ist in den Programmen `2s_PAA.jl` und `3s_PAA.jl` zu finden. Darin sind jeweils die beschriebenen Methoden für die beiden Modelle umgesetzt. Ein Aufruf von

```
julia 2s_PAA.jl bzw.
```

```
julia 3s_PAA.jl
```

startet jeweils die Simulationen, deren Ergebnisse ebenfalls abgespeichert werden. In der Methode `combineParams(setsize)` sind wieder `small_set`, `medium_set` und `large_set` als `setSize` verfügbar.

5.2.3 Verwertung der Verteilungen

Sowohl bei der Simulation des PAA mittels MoSDi als auch mittels PAA-Sim wird das Ergebnis, die Ankunftswahrscheinlichkeiten für jeden Zeitpunkt, als einfache Liste der Wahrscheinlichkeiten in einer `.csv`-Datei gespeichert. Zur Weiterverarbeitung und zum Auswerten der Ergebnisse gibt es die Module `process_simulations.py` und `plottings_PAA.py`, welche wiederum in Python 3.4 implementiert wurden.

process_simulations.py Das Programm `process_simulations` bereitet die von MoSDi oder PAA-Sim erzeugten Ergebnisse zur weiteren Verwendung auf. Dabei werden die Ergebnisse der PAA Simulationen zunächst komprimiert, da sich bei 240 Sekunden und 10 000 Simulationsschritten pro Sekunde immerhin Einzelwahrscheinlichkeiten für 2 400 000 Zeitpunkte ergeben. Wie auch bei der Teilchensimulation ist diese Genauigkeit jedoch nicht nötig. Außerdem werden die Peakdaten berechnet und zusammen mit den komprimierten Daten und den Simulationsparametern abgespeichert. Durch den Aufruf von

```
python3 process_simulations.py 2s
```

werden die Simulationsergebnisse des 2-Zustände Modells verarbeitet, wenn das Programm statt mit `2s` mit `3a` gestartet wird, entsprechend die Ergebnisse des Modells `3a`. Als weitere Option steht hier `-s` zur Verfügung, mit der die Quelle der Simulationsdaten (MoSDi oder PAA-Sim) ausgewählt werden kann.

plottings_PAA.py Für das Plotten der Ergebnisse der PAA Simulationen existieren ähnliche Möglichkeiten wie bei der Teilchensimulation. Im 2-Zustände Modell stehen neben den bekannten Optionen `--pp`, `--pr` und `--ppt` auch die Option `--pfp` zur Verfügung. Damit wird ein Plot erstellt, bei dem ein Parameter auf einen festen Wert gesetzt, der andere jedoch variiert wird. Durch den Aufruf

```
python3 plottings_PAA.py 2s --pfp ps 0.999
```

⁵julialang.org

wird der Parameter p_s auf den Wert 0,999 gesetzt und für p_m werden eine Reihe vorgegebener Werte eingesetzt. Die daraus resultierenden Peakdaten lassen die Darstellung links in Abbildung 6.8 entstehen.

Im 3-Zustände Modell ist ein ähnlicher Plot mit jeweils drei festen Parametern vorgesehen. Dazu muss neben der Option `--p3` die Werte für die festen Parameter mittels `--pmm`, `--pml`, `--paa` bzw. `--pll` eingegeben werden.

Des Weiteren kann mit der Option `--pd` ein Plot erstellt werden, der zu gegebenen Peakdaten mögliche Parameter anzeigt. Dazu müssen jeweils Intervalle für die Lage, Breite und Schiefe des gewünschten Plots angegeben werden. Ein Beispiel für einen solchen Aufruf ist:

```
python3 plottings_PAA.py 3a --pd -r 50 60 --iqr 1 5 --qk 0.2 0.4
```

Damit werden in allen vorhandenen Simulationen Peak gesucht, die im Intervall $[50; 60]$ ihren Maximalzeitpunkt haben, sowie eine Breite von $[1; 5]$ und Schiefe von $[0,2; 0,4]$ haben. Ein solcher Plot kann genutzt werden, um verschiedene mögliche Parameterkombinationen für einen gewünschten Peak anzuzeigen. Mit weiteren Simulationen können diese Parameter verändert werden, um die Zielpeakdaten weiter anzunähern.

5.2.4 Laufzeiten

Die Laufzeiten von PAA-Sim hängen, wie bei der Teilchensimulation auch, unter anderem von den Simulationsparametern ab. Erheblich zur Verkürzung der Laufzeit trägt hier das Entfernen von Wahrscheinlichkeiten, die unter einem bestimmten Schwellwert liegen, bei. Dadurch werden die verwendeten Arrays kürzer und es müssen deutlich weniger Berechnungen angestellt werden. Außerdem kann mit dem zusätzlich eingeführten Index ein vorzeitiger Abbruch der Simulation erfolgen, wenn die gesamte Wahrscheinlichkeitsmasse die Strecke durchlaufen hat. Anderenfalls werden auf jeden Fall alle Schritte bis zum Erreichen der Maximalzeit simuliert, auch wenn der Peak bereits vollständig wäre.

In Tabelle 5.3 ist eine Übersicht über die Laufzeiten wie in Abschnitt 5.1.4 gegeben. Darin werden die Laufzeiten von PAA-Sim einmal mit dem Abschneiden kleiner Wahrscheinlichkeiten und einmal ohne dieses Vorgehen, sowie MoSDi verglichen. Bei Zeiten, die mit einem * gekennzeichnet sind, konnte die Simulation vorzeitig beendet werden, da bereits die gesamte Wahrscheinlichkeitsmasse den Zielwert erreicht hat.

Auch im 3-Zustände Modell hängen die Laufzeiten von den verwendeten Parametern ab und auch hier kann durch das Entfernen sehr kleiner Werte Zeit eingespart werden. Wie in Tabelle 5.4 zu sehen ist, ist dennoch die Simulation mit MoSDi für dieses Modell schneller.

Tabelle 5.3: Laufzeitvergleich für die Simulation des PAA mit zwei Zuständen, mit den Parametern p_s und p_m , Zeitangaben in Sekunden. Grau hinterlegte Parameter zeigen unvollständigen Peak innerhalb von 240 Sekunden an. Verglichen werden MoSDi und zwei Varianten der eigenen Implementierung des PAA, in der ersten Spalte wurden Wahrscheinlichkeiten unter dem Schwellwert von 10^{-17} abgeschnitten, in der zweiten nicht. Mit * gekennzeichnete Zeiten deuten auf vorzeitiges Abbrechen auf Grund eines vollständigen Peaks hin.

p_s	p_m	PAA	PAA ohne cut	MoSDi
0,997	0,001	2,6*	96	195
0,997	0,3	2,8*	97	197
0,997	0,6	2,1*	98	205
0,997	0,95	0,7*	99	220
0,999	0,001	8,0*	86	79
0,999	0,3	7,2*	89	93
0,999	0,6	6,1*	92	115
0,999	0,95	2,3*	97	161
0,9993	0,001	11,3*	80	43
0,9993	0,3	10,5*	84	47
0,9993	0,6	8,4*	88	60
0,9993	0,95	3,2*	93	126
0,9996	0,001	16,8	76	20
0,9996	0,3	18,2	70	20
0,9996	0,6	15,0*	76	20
0,9996	0,95	5,6*	89	32

Tabelle 5.4: Laufzeitvergleich für die Simulation des PAA mit drei Zuständen, mit den Parametern p_{mm} , p_{ml} , p_{aa} und p_{ll} , Zeitangaben in Sekunden. Grau hinterlegte Parameter zeigen unvollständigen Peak innerhalb von 240 Sekunden an. Verglichen werden MoSDi und zwei Varianten der eigenen Implementierung des PAA, in der ersten Spalte wurden Wahrscheinlichkeiten unter dem Schwellwert von 10^{-17} abgeschnitten, in der zweiten nicht. Mit * gekennzeichnete Zeiten deuten auf vorzeitiges Abbrechen auf Grund eines vollständigen Peaks hin.

p_{mm}	p_{ml}	p_{aa}	p_{ll}	%	PAA	PAA	MoSDi
0,01	0,00005	0,997	0,99995		40*	180	26
0,01	0,00005	0,997	0,999995		132	133	27
0,01	0,00005	0,9996	0,99995	18,2	65	154	30
0,01	0,00005	0,9996	0,999995	17,7	96	142	32
0,01	0,003	0,997	0,99995		58*	157	27
0,01	0,003	0,997	0,999995	98,8	131	134	28
0,01	0,003	0,9996	0,99995	8,2	73	148	30
0,01	0,003	0,9996	0,999995	1,6	94	142	32
0,9	0,00005	0,997	0,99995		35*	184	26
0,9	0,00005	0,997	0,999995		138	129	27
0,9	0,00005	0,9996	0,99995		44*	177	27
0,9	0,00005	0,9996	0,999995		143	127	26
0,9	0,003	0,997	0,99995		55*	162	27
0,9	0,003	0,997	0,999995	99,5	140	131	27
0,9	0,003	0,9996	0,99995		58	157	28
0,9	0,003	0,9996	0,999995	99	129	130	26

Kapitel 6

Evaluation

Im folgenden Kapitel wird betrachtet, was mit den vorgestellten Modellen simuliert werden kann. Es werden die Einflüsse der Simulationsparameter auf die Peakcharakteristika beschrieben, Zusammenhänge zwischen den verschiedenen Parametern untersucht und Grenzen der einzelnen Modelle aufgezeigt.

6.1 Untersuchte Peaks

Mit den beiden vorgestellten Modellen ist es möglich, eine Vielzahl verschiedener Verteilungen der Ankunftszeiten von Teilchen zu simulieren. Dazu wurden zunächst im gesamten Parameterraum Simulationen durchgeführt und die aus den verschiedenen Kombinationen resultierenden Verteilungen betrachtet.

Einige Kombinationen würden zwar Peaks erzeugen, diese lägen jedoch teilweise oder vollständig jenseits der 240 Sekunden, die als Maximalzeit gesetzt wurden. Diese Verteilungen werden in den Evaluationen nicht weiter berücksichtigt, da die Simulationen vorher abgebrochen wurden und daher auch die Eigenschaften der Peaks nicht oder nur teilweise berechnet wurden. Insbesondere betrifft dies Simulationen, in denen die Wahrscheinlichkeit, mobil zu bleiben (p_m bzw. p_{mm}), sehr gering ist und die Wahrscheinlichkeit(en), stationär zu bleiben, sehr hoch ist. Ein Beispiel für eine Parameterkombination im 2-Zustände Modell, die einen Peak erzeugt, der teilweise außerhalb des Chromatogramms liegt, ist links in Abbildung 6.1 gezeigt.

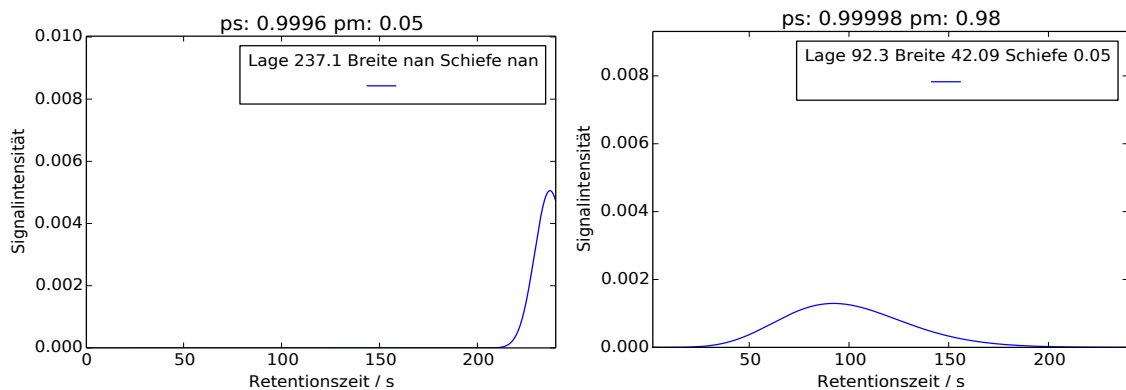


Abbildung 6.1: Nicht berücksichtigte Verteilungen, links teilweise außerhalb des Chromatogramms, daher Breite und Schiefe nicht berechenbar, rechts zu breit

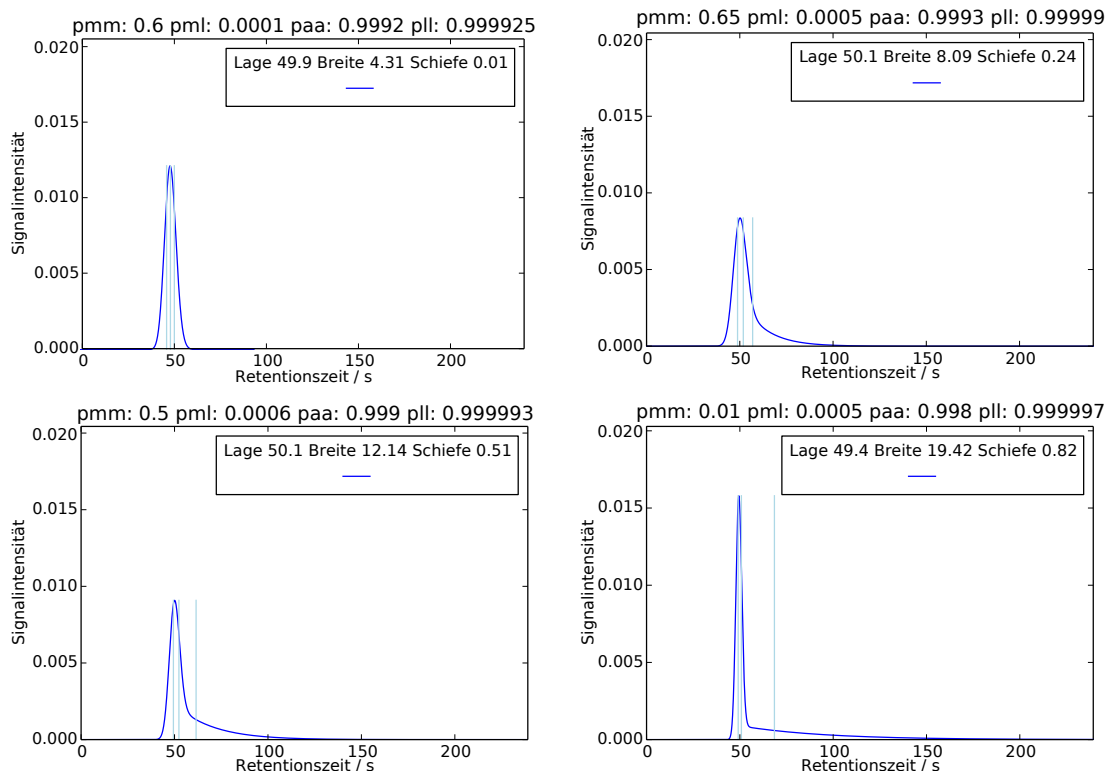


Abbildung 6.2: Peaks mit verschiedenen Schiefen, zusätzlich Quartile in hellblau eingezeichnet

Auch von den übrigen, innerhalb des Zeitraumes von 240 Sekunden liegenden Verteilungen, sind nicht alle als Peak zu bezeichnen. Beispielsweise kann eine Verteilung extrem breit sein und über das gesamte Spektrum reichen. Damit würde sie in einer echten Messung wohl eher als Hintergrundrauschen erkannt werden. Ein Beispiel dafür ist in rechts in Abbildung 6.1 zu sehen. Derartige Peaks werden nur teilweise in den folgenden Auswertungen erfasst. Ähnliches gilt für extrem frühe Peaks, die ihren Maximalzeitpunkt bei einer Retentionszeit von $t \approx 0,1$ s haben. Diese dienen zwar als Anhaltspunkt für die Entstehung von Schiefe, da hierbei allerdings nur wenige Teilchen überhaupt in Kontakt mit der stationären Phase treten, können die Ergebnisse insbesondere bei der Teilchensimulationen auch leicht verfälscht werden.

Für die Schiefe eines Peaks sollen keine Einschränkungen gemacht werden, es werden alle Werte im möglichen Intervall $[0,1]$ für den Quartilskoeffizienten untersucht werden. Um die Schiefe verschiedener Peaks besser einschätzen zu können, sind in Abbildung 6.2 Peaks mit verschie- dener Schiefe gezeigt.

Beim Peak unten rechts in Abbildung 6.2 mit der Schiefe von 0,82 fällt auf, dass die Breite ebenfalls sehr groß ist, obwohl ein menschlicher Betrachter ihn wohl eher als schmaler als die anderen abgebildeten Peaks bezeichnen würde. Dies liegt an den verwendeten Maßen für Schiefe und Breite. Der Tail des Peaks hat zwar eine sehr niedrige Intensität im Vergleich zum Maximalwert, reicht jedoch über einen Großteil des restlichen Spektrums. Es kommt noch ein großer Teil der Teilchen deutlich nach dem Hauptpulk an, wodurch sich das Q_{75} Quantil deutlich nach hinten verschiebt. Zur Verdeutlichung wurden die Quartile als senkrechte Linien mit in die Plots aufgenommen.

Wenn ein solcher Peak in einer tatsächlichen Messung vorkäme, könnte es gut sein, dass der Tail auf Grund seiner niedrigen Intensität jedoch gar nicht mehr als zum Peak gehörig erkannt würde, sondern eher dem Hintergrundrauschen zugerechnet würde.

6.2 2-Zustände Modell

Zunächst wird der Einfluss der beiden Parameter auf die Peaks beschrieben. Da sich die Intensität dieses Einflusses je nach Wert des anderen Parameters verändert, wird anschließend der gemeinsame Einfluss der beiden Parameter betrachtet. Außerdem wird beschrieben, was nicht mit dem Modell simuliert werden kann.

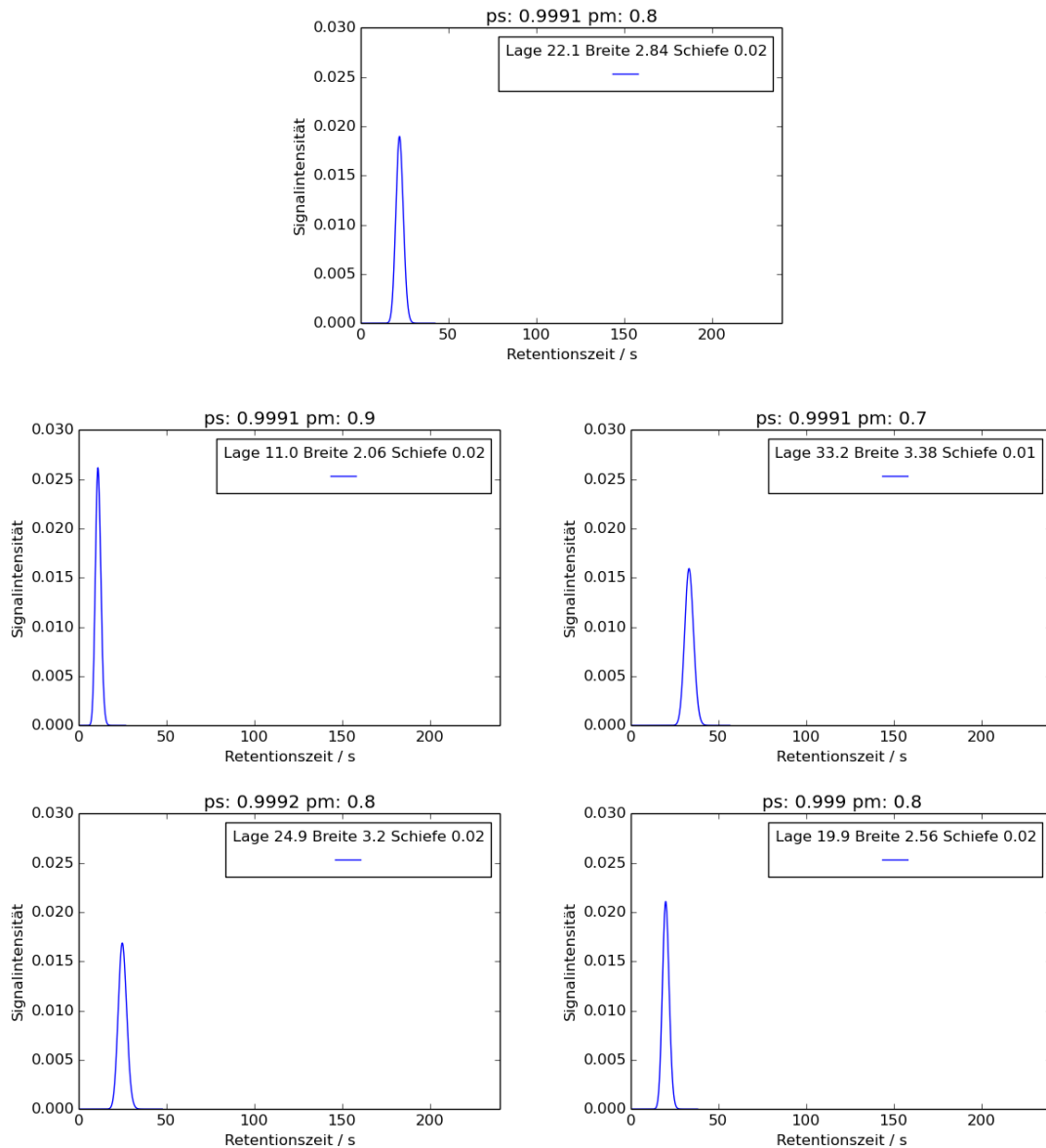


Abbildung 6.3: Einfluss von p_m und p_s auf einen Peak, oben der Ausgangspeak mit $p_m = 0,8$ und $p_s = 0,9991$, veränderte Parameter: in der Mitte links $p_m = 0,9$, rechts $p_m = 0,7$, unten links $p_s = 0,9992$, rechts $p_s = 0,999$

6.2.1 Einfluss der Parameter auf einzelne Peaks

Es wird an zwei einzelnen Peaks gezeigt, wie sich die Veränderung von jeweils einem Parameter auf die Peakcharakteristika auswirkt. Beide dieser Peaks könnten so in einem echten Chromatogramm beobachtet worden sein.

In Abbildung 6.3 ist ein Peak zu sehen, der mit den Parametern $p_m = 0,8$ und $p_s = 0,9991$ simuliert wurde. Sein Maximalwert liegt bei etwa $t = 22$ s und er hat eine Breite von etwa 2,8. Abbildung 6.4 zeigt einen Peak, dessen Maximalwert bei $t = 100$ s liegt und der eine Breite von 7,4 hat. Er wurde mit den Parametern $p_m = 0,5$ und $p_s = 0,9995$ simuliert.

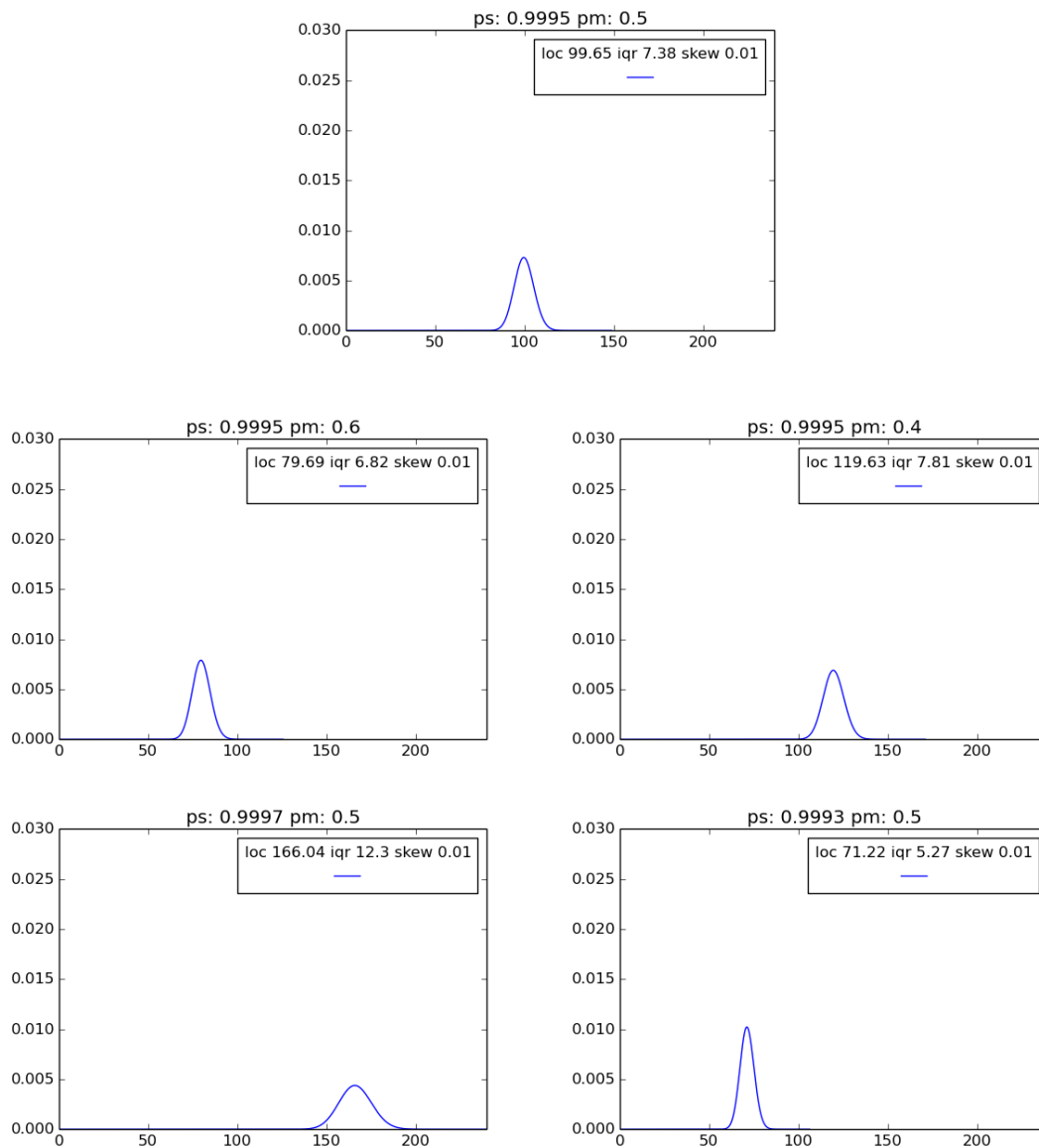


Abbildung 6.4: Einfluss von p_m und p_s auf einen Peak, oben der Ausgangspeak mit $p_m = 0,5$ und $p_s = 0,9995$, veränderte Parameter: in der Mitte links $p_m = 0,6$, rechts $p_m = 0,4$, unten links $p_s = 0,9997$, rechts $p_s = 0,9993$

In beiden Fällen zeigen sich bei Veränderungen an den Parametern die gleichen Effekte am Peak. Wenn p_m größer oder p_s kleiner wird, resultiert ein schmalerer und früherer Peak. Umgekehrt bewirkt eine Verkleinerung von p_m oder eine Vergrößerung von p_s , einen späteren und breiteren Peak. Die Schiefe verändert sich, wenn überhaupt, nur minimal.

Unter anderen ergibt sich der Zusammenhang von Lage und Breite daraus, dass sich in diesem Fall auch der Maximalzeitpunkt nach hinten verschiebt und spätere Peaks generell breiter sind, als frühe Peaks. Durch mehr Interaktionen, deren Länge jeweils zufällig ist, wird der Pulk breiter.

Abbildung 6.5 zeigt eine Übersicht darüber, welche Peakdaten sich bei verschiedenen Parameterkombinationen ergeben. Auf den x-Achsen sind jeweils verschiedene Werte für p_m und auf der y-Achse für p_s aufgetragen. An der Farbe des entsprechenden Feldes ist zu erkennen, welchen Wert die Eigenschaft für die jeweilige Parameterkombination annimmt. Das weiße Feld oben links steht für eine Simulation, die keinen Peak innerhalb der Maximalzeit ergeben hat.

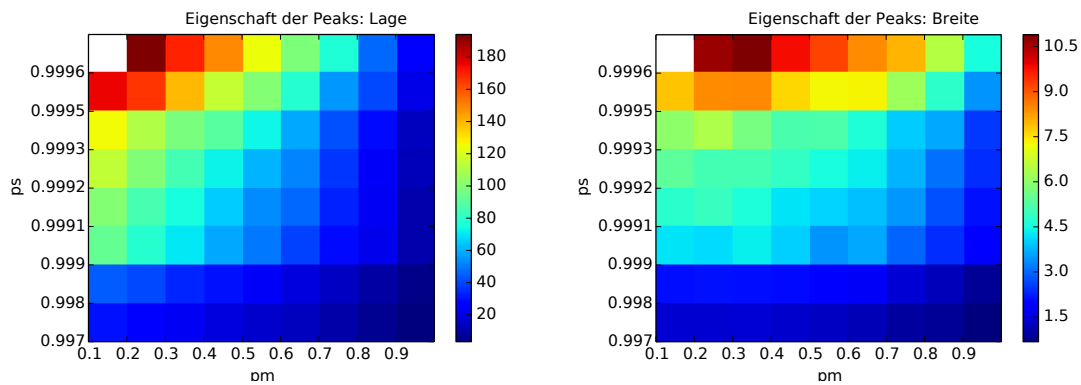


Abbildung 6.5: Die resultierenden Zeiten und Breiten einiger Parameterkombinationen

Offensichtlich kann eine bestimmte Retentionszeit durch verschiedene Kombinationen erreicht werden, die sich dann jeweils in der Breite des Peaks unterscheiden. Dies wird in Abbildung 6.6 noch einmal deutlich. Abgebildet sind verschiedene Kombinationen aus p_s und p_m , mit denen ein Peak etwa zum Zeitpunkt $t = 75$ erreicht werden kann. In diesem Fall liegen alle Zeitpunkte der Peaks im Intervall $[72, 78]$. Für den Parameter $p_s = 0,9996$ gibt es offenbar zwei Möglichkeiten, p_m zu wählen, um einen Peak in der Nähe von $t = 75$ zu erzeugen. Eine weitere Simulation mit dem Parameter $p_s = 0,9996$ und einem Wert für p_m , der zwischen diesen Möglichkeiten liegt, beispielsweise $p_m = 0,73$ würde einen Peak näher an der gewünschten Retentionszeit liefern. Die Breite des neuen Peaks liegt dann auch zwischen den Breiten der beiden vorhandenen Peaks.

Auf diese Weise ist es möglich, Peaks mit gewünschter Retentionszeit und Breite schrittweise anzunähern.

Die Stärke der Veränderung an den Peaks unterscheidet sich für Änderungen an p_s oder p_m . Daher wird im nächsten Abschnitt untersucht, inwiefern der Einfluss eines Parameters größer oder kleiner wird, abhängig davon, welchen Wert der andere Parameter annimmt.

6.2.2 Einfluss abhängig vom anderen Parameter

Um den Einfluss der Parameter in Abhängigkeit vom anderen Parameter zu untersuchen, werden im Folgenden Plots verwendet, bei denen ein Parameter fest ist, der andere jedoch verschiedene Werte annimmt. Alle Plots wurden mit der gleichen Achseneinteilung

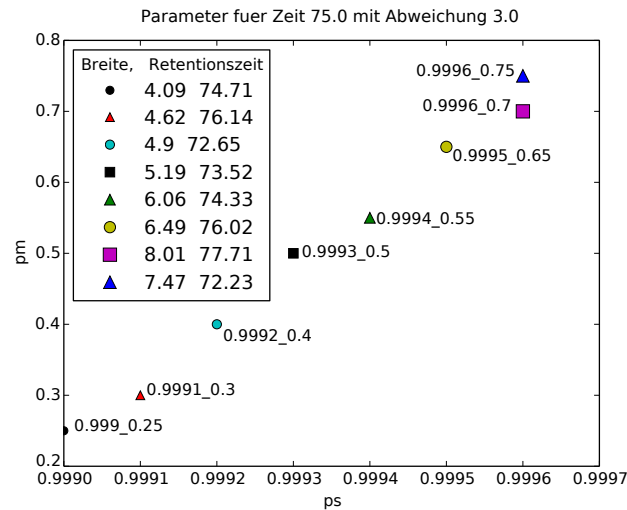


Abbildung 6.6: Mögliche Parameterkombination für Retentionszeit $t \approx 75$

gemacht, um eine bessere Vergleichbarkeit zu ermöglichen. Jeder Peak wird dargestellt durch einen Punkt an den Koordinaten seines Zeitpunktes und seiner Schiefe. Die Größe des Punktes wird festgelegt durch die Breite des entsprechenden Peaks. Zusätzlich ist die Breite in der Legende noch einmal aufgeführt. Beschriftet ist jeder Punkt mit dem variablen Parameter des Vergleichsplots, in der Überschrift ist jeweils der Wert des festen Parameters zu finden.

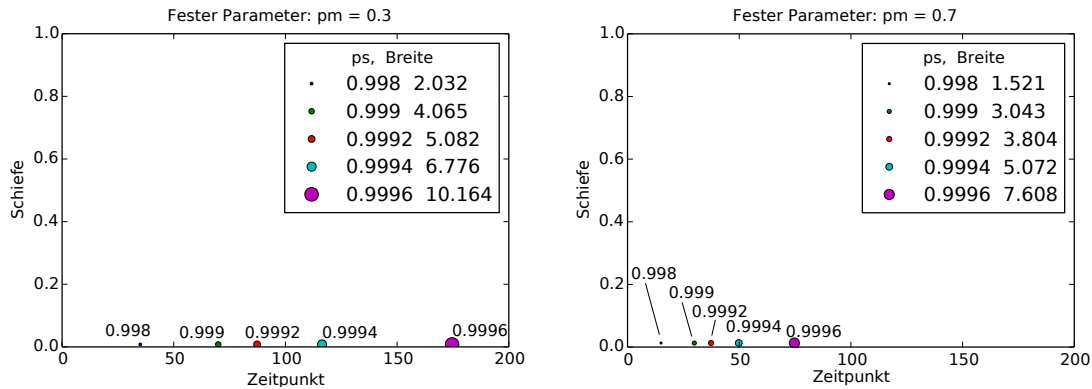
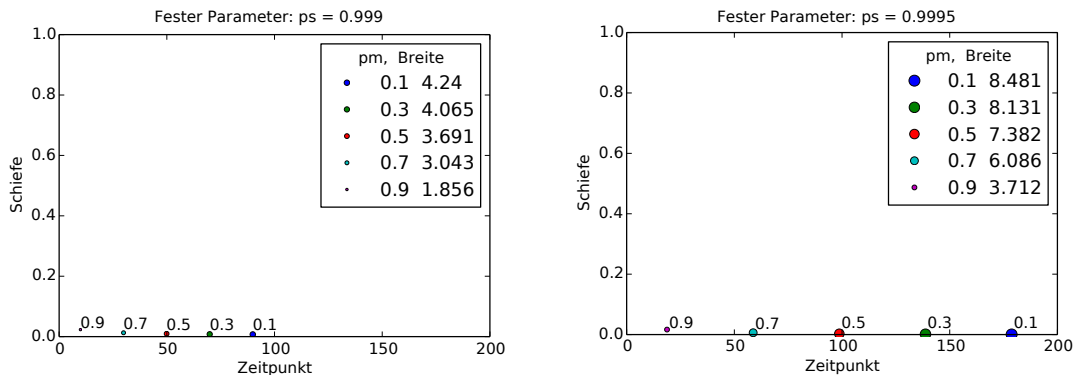


Abbildung 6.7: Der Einfluss von p_s auf die Peaks abhängig von p_m

In Abbildung 6.7 ist der Einfluss des Parameters p_s in Abhängigkeit von p_m gezeigt. Neben der bereits im letzten Abschnitt gewonnenen Erkenntnis, dass die Peaks mit zunehmendem p_s breiter werden und später erscheinen, ist zu sehen, dass der Einfluss von p_s größer wird, wenn p_m kleiner ist. Dies gilt sowohl für den Zeitpunkt, als auch die Breite des Peaks. Die Schiefe wird in beiden Fällen nicht wahrnehmbar beeinflusst.

Der Einfluss des Parameters p_m in Abhängigkeit von p_s ist in Abbildung 6.8 dargestellt. In beiden Fällen werden die Peaks mit zunehmendem p_m schmaler und erscheinen zu einem früheren Zeitpunkt. Dies verstärkt sich bei größerem p_s . Zu erkennen ist hier auch eine geringe Zunahme der Schiefe mit steigendem p_m . Diese Veränderung ist jedoch nicht abhängig von p_s .

Abbildung 6.8: Der Einfluss von p_m auf die Peaks abhängig von p_s

6.2.3 Grenzen

Mit diesem Modell sind kaum tailende Peaks erzeugbar. Schiefe von mehr als 0,05 tritt nur selten auf und immer bei Peaks, die ihren Maximalzeitpunkt ganz am Anfang des Spektrums, knapp über der Trägergaszeit von 0,1 Sekunden haben. Der Effekt wird verursacht durch einen sehr großen Wert für p_m , der bei ungefähr 0,999 liegt. Dadurch gibt es viele Teilchen die (bei einer Säulenlänge von 1000 Schritten) gar nicht in Wechselwirkung mit der stationären Phase geraten und dadurch mit Trägergasgeschwindigkeit durch die Säule wandern. Die wenigen Teilchen, die doch in die stationäre Phase eintreten, verursachen den Tail, da sie deutlich später als der Pulk die Säule verlassen.

Eine weitere Grenze ist jedoch die Minimalbreite der erzeugten Peaks. Damit ist gemeint, dass mit zunehmender Zeit die Minimalbreite der Peaks, die an diesem Zeitpunkt ihr Maximum haben, ansteigt. Zur Verdeutlichung ist in Abbildung 6.9 eine Übersicht über die erreichbaren Zeit-Breiten Kombinationen gegeben. Jeder Punkt steht für einen Peak, der mit der entsprechenden Retentionszeit und Breite erzeugt wurde. Zu sehen ist, dass es zu jedem Zeitpunkt möglich ist, Peaks mit dem Modell zu erzeugen. Auch zwischen den einzelnen Punkten sind Peaks erzeugbar, wenn die Parameter entsprechend angepasst werden.

Dass die untere Kante der Punktwolke jedoch eine Grenze für die Breite darstellt, ist in Abbildung 6.10 zu sehen, die einen Ausschnitt aus Abbildung 6.9 darstellt. Zu sehen sind drei Reihen von simulierten Daten, die mit jeweils dem gleichen Parameter p_s erzeugt wurden. Die Breite nimmt mit fallendem p_m zu. Diese Beobachtung wurde bereits im letzten Abschnitt gemacht. Jedoch ist hier zu sehen, dass sich bei einer Annäherung von p_m an 0 keine deutlich späteren Peaks ergeben, sondern der Zeitpunkt auf einem Wert stagniert. Das liegt daran, dass bei so kleinen Werten für p_m in fast jedem Schritt, der in der mobilen Phase startet, ein Wechsel zur stationären Phase statt findet. Der Zeitpunkt wird dann also maßgeblich durch p_s beeinflusst. Da höhere Werte für p_s nicht nur für einen späteren Zeitpunkt sorgen, sondern auch für eine größere Breite, ergibt sich dadurch eine Minimalbreite für alle Zeitpunkte.

Auch die obere Kante der Punktwolke aus Abbildung 6.9 stellt gewissermaßen eine Grenze für die mit diesem Modell erreichbaren Zeit-Breiten Kombinationen dar. Einen passenden Ausschnitt zeigt Abbildung 6.11. Zwar können in diesem Bereich ganz zu Beginn des Spektrums – der Bereich, der Tailing aufweist – fast beliebig breite Peaks erzeugt werden. Dazu muss p_m sehr groß gewählt werden. Wird jedoch bei gleichem p_s der Wert

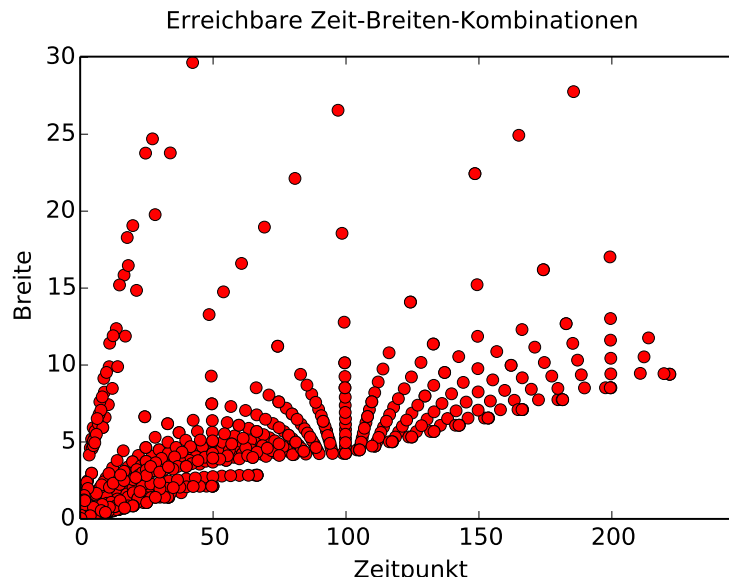


Abbildung 6.9: Mit dem 2-Zustände Modell erreichbare Zeitpunkte und Breiten

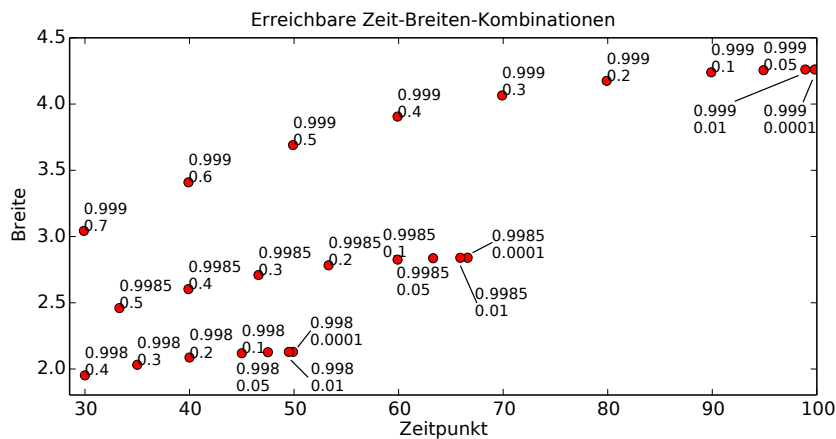


Abbildung 6.10: Ausschnitt der erreichbaren Zeiten und Breiten

für p_m noch ein wenig erhöht, ist der Maximalzeitpunkt des Peaks bei $t = 0,1$, da nicht mehr ausreichend Teilchen überhaupt in Kontakt mit der stationären Phase kommen.

6.3 3-Zustände Modell

Die wenigen Peaks, die beim 2-Zustände Modell Schiefe aufweisen dienen als Hinweis darauf, wie das Tailing entsteht: Einige wenige Teilchen kommen deutlich nach dem Hauptpulk an, da sie in der stationären Phase verweilen, während die Teilchen des Pulks ohne Interaktion die Säule durchqueren. Wenn nun bei einer beliebigen, nicht-tailenden Verteilung, die dadurch entsteht, dass immer wieder Teilchen mit der stationären Phase interagieren, ein zusätzlicher Zustand hinzugefügt wird, in den die Teilchen seltener, aber dafür viel länger eintreten, müsste genau dieser Effekt auch bei anderen als den ganz frühen Verteilungen entstehen. Genau dieser Zustand steht im 3-Zustände Modell zur Verfügung.

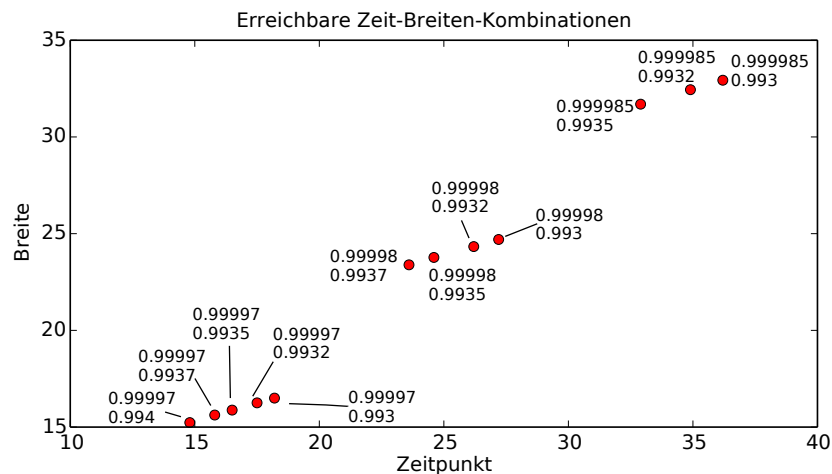


Abbildung 6.11: Ausschnitt der erreichbaren Zeiten und Breiten

Untersucht wurde vor allem das Modell aus Abbildung 3.3a, bei dem die beiden stationären Zustände getrennt sind.

Im Folgenden werden die Einflüsse der Parameter dieses Modells untersucht, dabei beschränkt sich diese Untersuchung auf p_{mm} , p_{ml} , p_{aa} und p_{ll} . Die anderen Parameter p_{ma} , p_{am} und p_{lm} können daraus berechnet werden.

6.3.1 Einfluss auf einzelne Peaks

Um den Einfluss der Parameter auf Peaks zu untersuchen, wurden zunächst einige realistische Ausgangspeak gewählt. Der erste hat eine Retentionszeit von 50 s, eine Breite von 12 und Schiefe von 0,46. Der zweite hat eine Retentionszeit von 100 s, eine Breite von 8 und Schiefe von 0,23. Die für die beiden Peaks verwendeten Simulationsparameter wurden dann nur wenig verändert und beobachtet, wie die Maße der Peaks darauf reagierten. In 6.1 und 6.2 sind diese Veränderungen jeweils gezeigt. Gelb markiert ist immer der Parameter, der in der jeweiligen Simulation verändert wurde.

Tabelle 6.1: Einfluss der Parameter auf einen Peak, ausgehend von den Parametern der ersten Zeile wurde jeweils ein Parameter verändert (gelb markiert) und die daraus resultierenden Peakdaten berechnet.

p_{mm}	p_{ml}	p_{aa}	p_{ll}	Lage	Breite	Schiefe
0,6	0,0006	0,9992	0,999993	50,14	12,34	0,46
0,5	0,0006	0,9992	0,999993	62,69	12,45	0,43
0,7	0,0006	0,9992	0,999993	37,59	12,2	0,5
0,6	0,0005	0,9992	0,999993	50,09	9,7	0,38
0,6	0,0007	0,9992	0,999993	50,19	14,82	0,5
0,6	0,0006	0,9991	0,999993	44,55	12,18	0,5
0,6	0,0006	0,9993	0,999993	57,33	12,57	0,41
0,6	0,0006	0,9992	0,99999	50,25	9,2	0,33
0,6	0,0006	0,9992	0,999995	50,05	16,66	0,58

Zu erkennen ist, dass eine Veränderung von p_{mm} und p_{aa} jeweils die Lage des Peaks deutlich beeinflusst, hingegen eine Änderung von p_{ml} oder p_{ll} kaum einen Einfluss auf die

Tabelle 6.2: Einfluss der Parameter auf einen Peak, ausgehend von den Parametern der ersten Zeile wurde jeweils ein Parameter verändert (gelb markiert) und die daraus resultierenden Peakdaten berechnet.

p_{mm}	p_{ml}	p_{aa}	p_{ll}	Lage	Breite	Schiefe
0,1	0,0005	0 9991	0 99999	100,22	8,1	0,23
0,05	0,0005	0 9991	0 99999	105,76	8,11	0,229
0,2	0,0003	0 9991	0 99999	89,1	8,05	0,235
0,1	0,0007	0 9991	0 99999	100,04	6,22	0,12
0,1	0,0005	0 9991	0 99999	100,4	10,89	0,34
0,1	0,0005	0 999	0 99999	90,17	7,68	0,26
0,1	0,0005	0 9992	0 99999	112,77	8,64	0,2
0,1	0,0005	0 9991	0 999975	100,41	5,91	0,08
0,1	0,0005	0 9991	0 999993	100,12	9,93	0,34

Lage hat. Umgekehrt wird die Breite viel deutlicher durch p_{ml} und p_{ll} verändert, p_{mm} und p_{aa} haben hier auch einen Einfluss, der jedoch viel geringer erscheint.

Dabei ist deutlich, dass p_{mm} die gleichen Auswirkungen auf die Peaks bezüglich Lage und Breite hat, wie p_m im 2-Zustände Modell. Die gleiche Entsprechung findet sich für p_{aa} und p_s . Für den Maximalzeitpunkt des Peaks sowie seine Breite gilt, dass beide größer werden, wenn die Parameter p_{ml} , p_{aa} und p_{ll} größer werden oder p_{mm} kleiner wird.

Die Schiefe wird von allen Parametern zumindest geringfügig beeinflusst. Dabei wird die Schiefe größer, wenn p_{mm} , p_{ml} oder p_{ll} größer oder p_{aa} kleiner werden.

Insgesamt scheint es, dass die Einflüsse auf Zeitpunkt und Breite der Peaks weniger komplex sind, als die Einflüsse auf die Schiefe. Das ist auch damit erklärbar, dass für Zeitpunkt und Breite eigentlich nur die zwei Parameter des einfachen Modells nötig sind und die Aufspaltung des stationären Zustands daran nicht viel ändert. Die Schiefe jedoch hängt immer von mehreren Parametern und deren Kombination ab.

6.3.2 Einfluss abhängig von anderen Parametern

Bei diesem Modell ergeben sich, durch die größere Anzahl an Parametern, komplexere Zusammenhänge zwischen den Parametern und den Peakdaten. Teilweise hängt die Stärke des Einflusses eines Parameters auf ein Peakcharakteristikum von den anderen Parametern ab. Das heißt, der Einfluss eines Parameters auf die resultierenden Peaks kann größer oder kleiner sein, abhängig davon, welchen Wert ein oder mehrere der anderen Parameter annehmen.

Im Folgenden ist für jeden der vier Parameter eine detaillierte Übersicht gegeben welche Eigenschaften der Peaks er in Abhängigkeit anderer Parameter beeinflusst. Die zur Verdeutlichung verwendeten Plots sind denen aus der Auswertung des 2-Zustände Modells sehr ähnlich, hier wurden drei der vier Parameter fest gehalten, der vierte wird variiert. Auch diese Plots wurden zur besseren Vergleichbarkeit weiterhin mit der gleichen Achsen-einteilung erstellt.

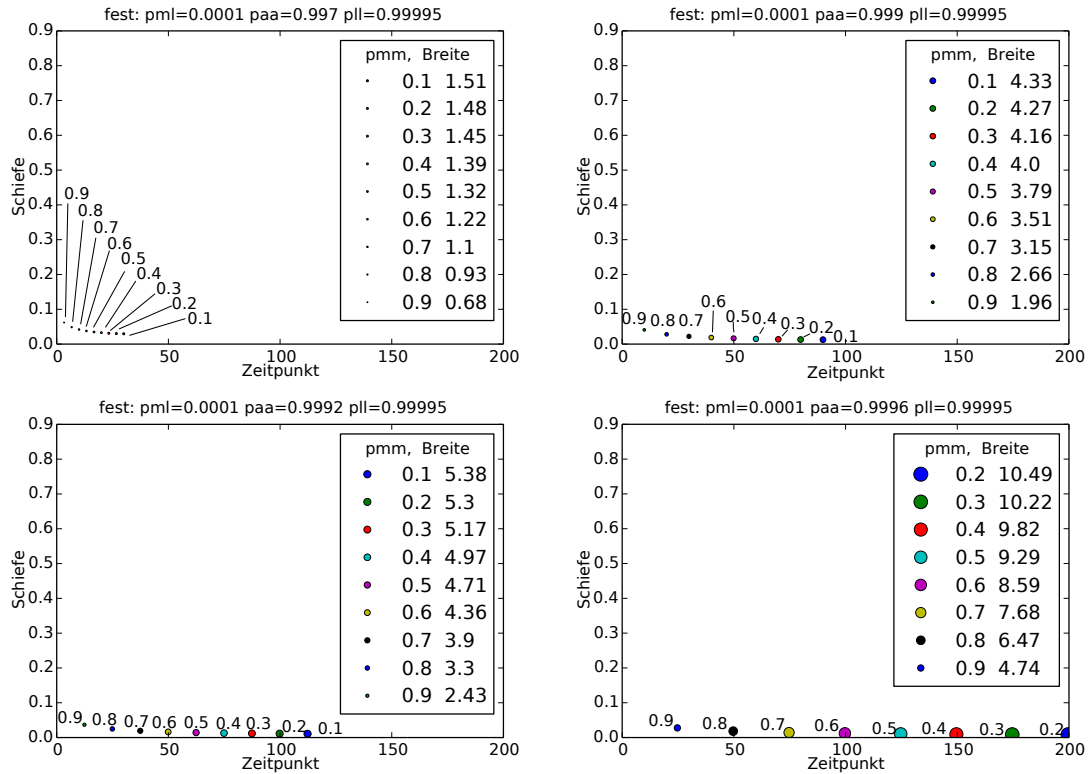


Abbildung 6.12: Der Einfluss von p_{mm} auf die Peaks abhängig von p_{aa} , welches von oben links nach unten rechts größer wird

Einfluss des Parameters p_{mm} Wie in Abbildung 6.12 zu sehen ist, ist der Einfluss von p_{mm} auf den Zeitpunkt von p_{aa} abhängig: Je größer p_{aa} ist, desto stärker wird der Zeitpunkt von p_{mm} beeinflusst. Wenn p_{mm} und p_{aa} klein sind, entstehen nur Peaks zu Beginn des Spektrums.

Der Einfluss von p_{mm} auf die Breite ist ebenfalls von p_{aa} abhängig. Bei größerem p_{aa} steigt dieser Einfluss etwas, wie ebenfalls in Abbildung 6.12 zu erkennen ist.

Die Schiefe steigt mit steigendem p_{mm} . Dieser Einfluss hängt noch von p_{aa} ab, bei kleinem p_{aa} hat p_{mm} einen größeren Einfluss auf die Schiefe.

Außerdem ist der Einfluss von p_{mm} auf die Schiefe am größten, wenn p_{ml} und p_{ll} beide nicht zu klein und nicht zu groß sind. Dieser Zusammenhang ist für p_{ml} in Abbildung 6.13 dargestellt.

Zusätzlich wird der Einfluss von p_{ml} und p_{ll} auf die Abhängigkeit der Schiefe von p_{mm} größer, wenn p_{aa} klein ist.

Dieser Zusammenhang kann wie folgt erklärt werden: Durch den gemeinsamen Einfluss von p_{ml} und p_{ll} wird das Tailing gesteuert. Dieser Effekt kann sich bei großem p_{mm} besser zeigen. Dann sind die Teilchen im Durchschnitt schneller. Ein kleiner Wert für p_{aa} sorgt ebenfalls für eine höhere Durchschnittsgeschwindigkeit. Dadurch geraten im Gesamtverlauf der Säulendurchquerung Teilchen seltener in den gelösten Zustand, wodurch sie das Tailing erzeugen. Bei durchschnittlich langsameren Teilchen lösen sich die Teilchen zu häufig und statt Schiefe entsteht wiederum Breite. Ist p_{aa} jedoch sehr klein und p_{ll} sehr groß entstehen Peaks, die eine extrem großen Schiefe haben.

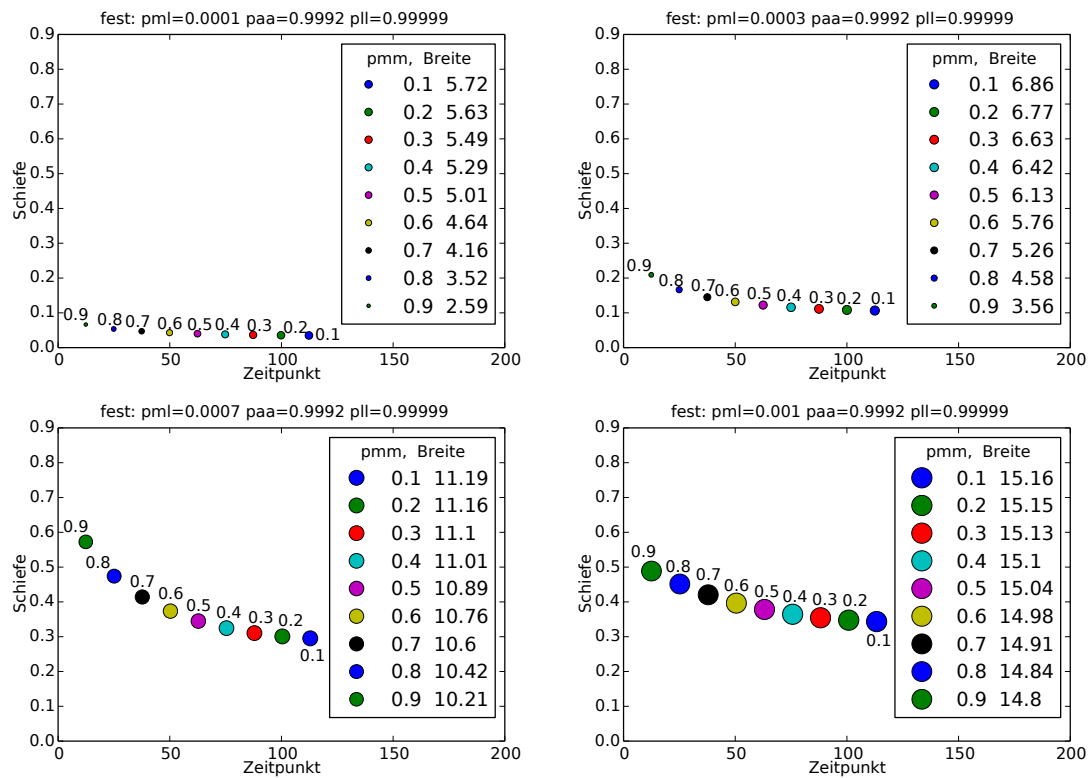


Abbildung 6.13: Der Einfluss von p_{mm} auf die Peaks abhängig von p_{ml} , welches von oben links nach unten rechts größer wird

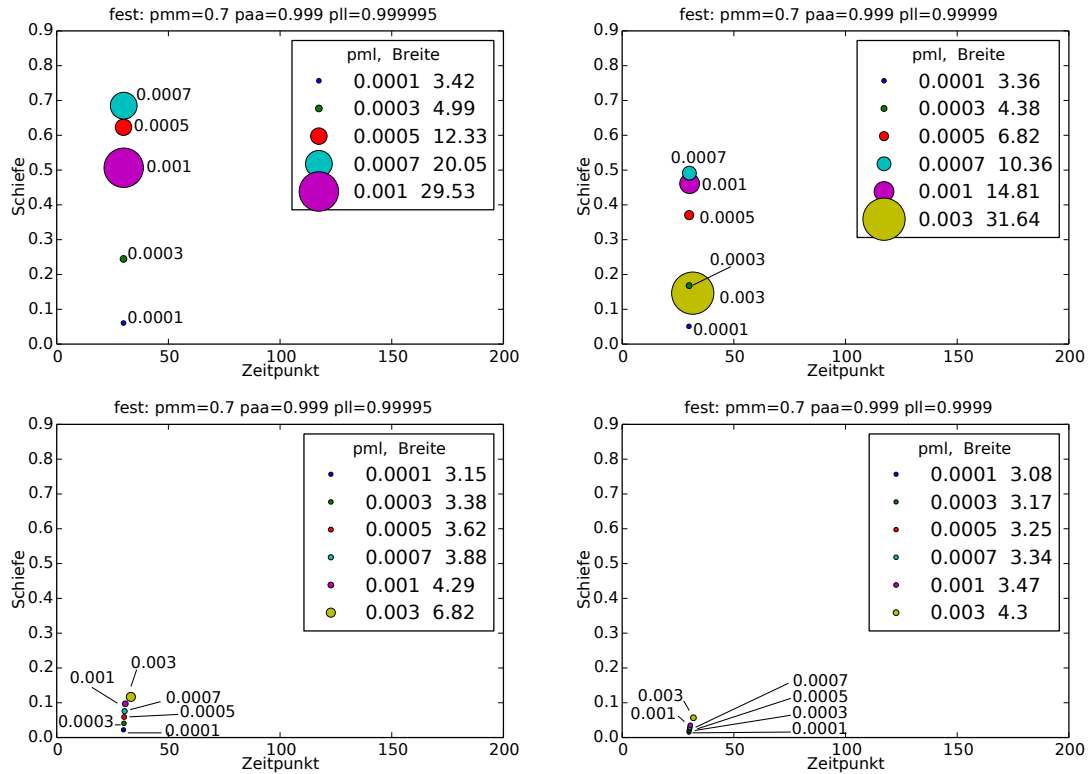


Abbildung 6.14: Der Einfluss von p_{ml} auf die Schiefe und Breite abhängig von p_{ll} , welches von links oben nach rechts unten kleiner wird.

Einfluss des Parameters p_{ml} Der Einfluss von p_{ml} auf die Peakdaten ist in den Abbildungen 6.14 und 6.15 gezeigt. Schiefe und Breite sind häufig deutlich von p_{ml} beeinflusst, der Zeitpunkt hingegen kaum. Allenfalls sehr große Werte für p_{ml} sorgen für einen erkennbar höheren Zeitpunkt, dann aber auch für deutlich mehr Breite und weniger Schiefe.

Auf die Schiefe und Breite ist der Einfluss von p_{ml} sehr groß, wenn auch p_{ll} groß ist, wie in Abbildung 6.14 zu erkennen. Hier sind p_{mm} und p_{aa} auf allen vier Plots gleich gewählt, nur p_{ll} nimmt von oben links über oben rechts und unten links bis zu unten rechts immer weiter ab und damit schwindet auch der Einfluss von p_{ml} auf Schiefe und Breite.

Wie in Abbildung 6.15 gezeigt, spielen auch p_{mm} und p_{aa} für den Einfluss von p_{ml} auf Schiefe und Breite eine Rolle. Hier ist p_{ll} in allen vier Plots konstant, p_{mm} und p_{aa} werden variiert. Zu erkennen ist, dass bei kleinem p_{aa} , der Einfluss von p_{ml} auf Schiefe und Breite größer wird, ist p_{mm} jedoch klein, wird auch der Einfluss von p_{ml} kleiner.

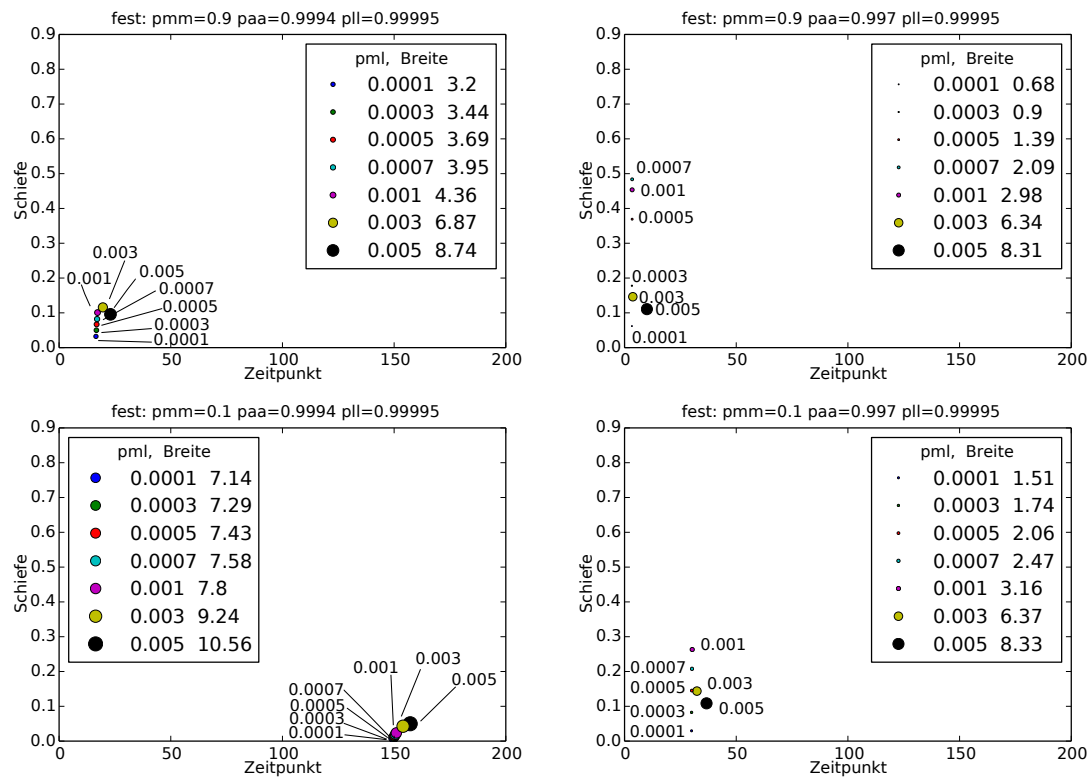


Abbildung 6.15: Der Einfluss von p_{ml} auf die Schiefe und Breite abhängig von p_{mm} und p_{aa} , p_{mm} ist oben klein und unten groß, p_{aa} ist links groß und rechts klein.

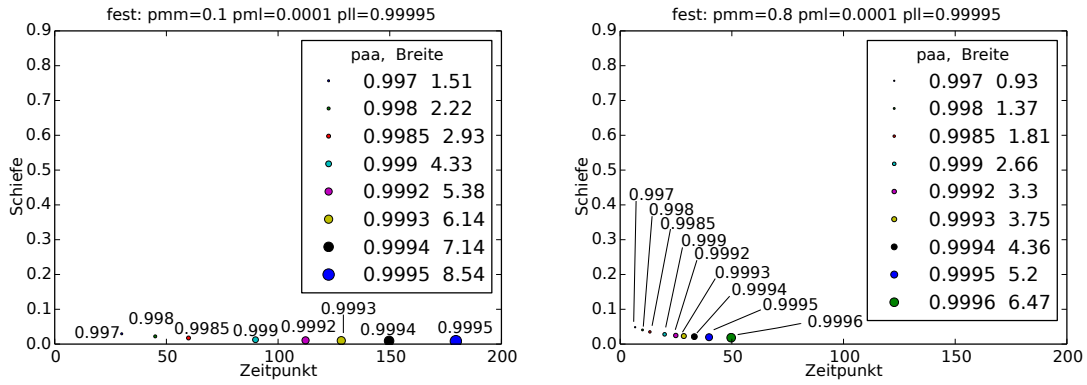


Abbildung 6.16: Der Einfluss von p_{aa} auf die Peaks abhängig von p_{mm} , welches links klein und rechts groß ist

Einfluss des Parameters p_{aa} In Abbildung 6.16 ist beispielhaft der Einfluss von Parameter p_{aa} in Abhängigkeit von p_{mm} auf die Peaks gezeigt. Der Einfluss von p_{aa} auf den Zeitpunkt wird mit steigendem p_{mm} immer kleiner und reicht dadurch von minimalem Einfluss und recht frühen Peaks bei jeder Wahl von p_{aa} bis zu sehr großem Einfluss und Peaks, die je nach p_{aa} über das ganze Spektrum verteilt sind. Damit hat p_{aa} einen ähnlichen Einfluss wie p_{mm} .

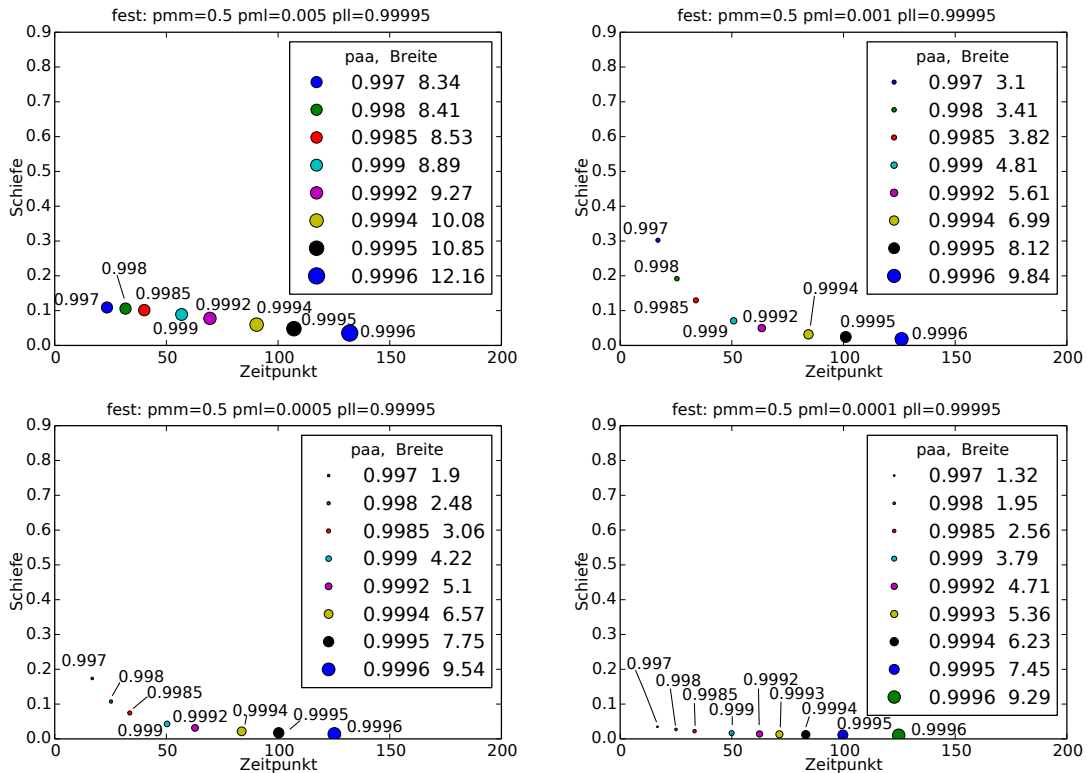


Abbildung 6.17: Der Einfluss von p_{aa} auf die Peaks abhängig von p_{ml} , welches von links oben nach rechts unten kleiner wird

Wie in 6.17 zu sehen ist, hat p_{aa} nur dann einen relevanten Einfluss auf die Schiefe, wenn p_{ml} weder zu klein, noch zu groß ist. Bei zu großem p_{ml} werden Peaks statt dessen

wieder nur breit und bei zu kleinem p_{ml} hat p_{aa} kaum Einfluss auf die Schiefe. Gleiches gilt für p_{ll} , welches auch den größten Einfluss von p_{aa} zulässt, wenn es weder sehr groß noch sehr klein ist.

Der Einfluss auf die Breite ist mäßig und wird größer, wenn p_{ml} und p_{ll} klein sind, für p_{ml} ist dies ebenfalls in Abbildung 6.17 zu erkennen.

Einfluss des Parameters p_{ll} Der Einfluss von p_{ll} auf den Zeitpunkt ist ähnlich wie bei p_{ml} nur minimal.

Auf Breite und Schiefe hat p_{ll} einen von p_{ml} abhängigen Einfluss: Wenn p_{ml} mittelgroß ist, ist die Wirkung von p_{ll} am größten. Das ist in Abbildung 6.18 gezeigt.

Wenn p_{ml} nicht zu klein oder groß ist, hat p_{ll} einen großen Einfluss auf Schiefe und Breite. Dieser Einfluss wächst, je größer p_{mm} und je kleiner p_{aa} ist. Dies ist in 6.19 für p_{mm} und in 6.20 für p_{aa} zu sehen.

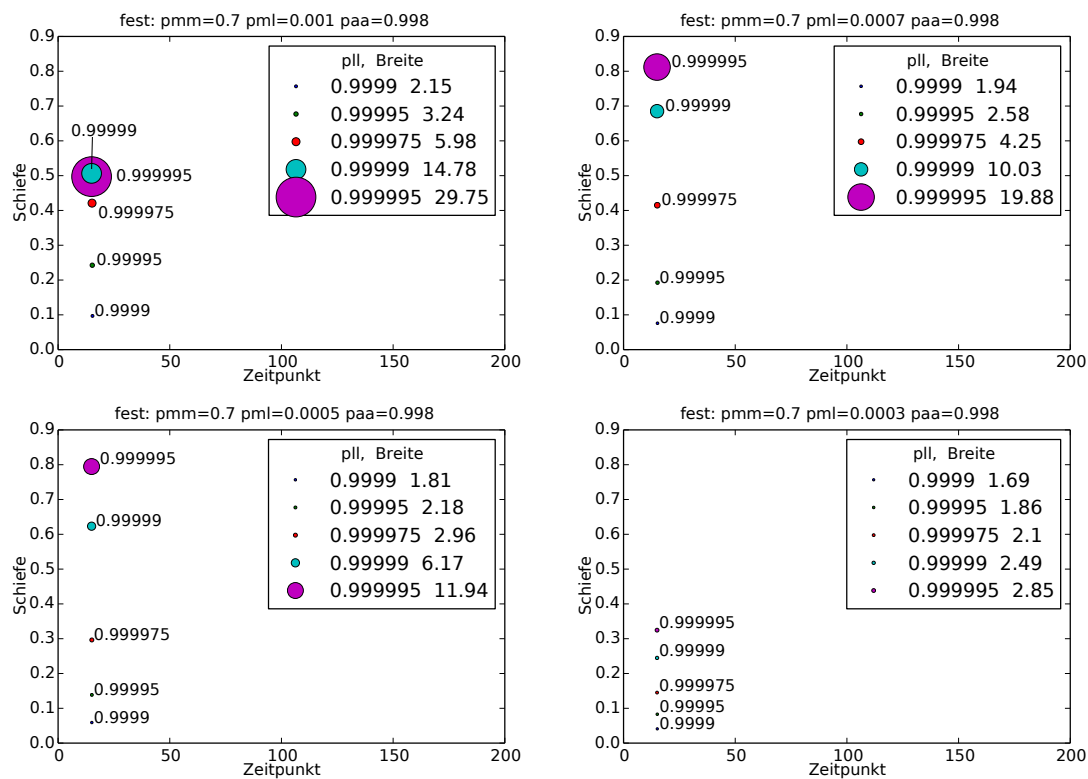


Abbildung 6.18: Der Einfluss von p_{ll} auf die Schiefe und Breite abhängig von p_{ml} , welches von links oben nach rechts unten kleiner wird.

Der Einfluss der Parameter kann wie folgt zusammengefasst werden:

Da bei einem Wert von 0 für p_{ml} effektiv das 2-Zustände Modell zum Einsatz kommt, ist hier fast nie Schiefe zu beobachten, außerdem hat der Parameter p_{ll} keinen Einfluss.

p_{aa} hat einen stärkeren Einfluss auf den Zeitpunkt als p_{ll} und p_{ml} . Dieser wird noch stärker, wenn p_{mm} klein ist. p_{mm} hat einen großen Einfluss auf den Zeitpunkt und nur einen minimalen Einfluss auf die Breite.

p_{ml} und p_{ll} haben zunächst beide einen deutlichen Einfluss auf die Schiefe. Erst, wenn beide sehr groß sind, sinkt die Schiefe wieder (und die Breite steigt extrem an) Das ist

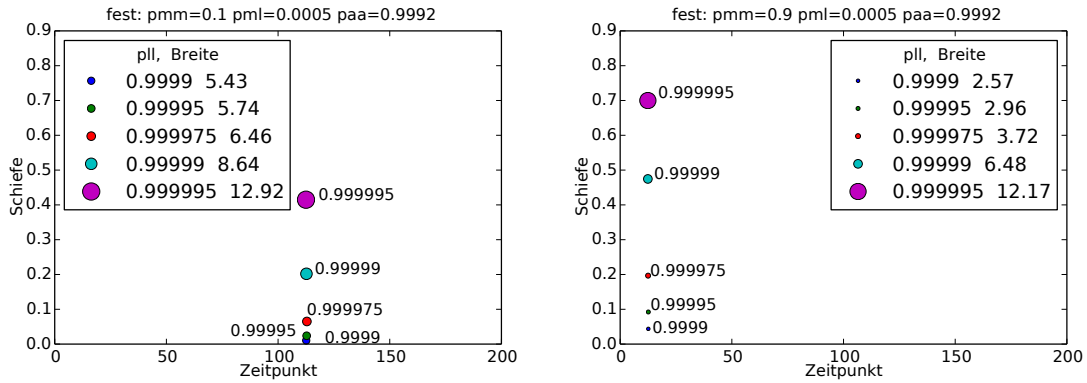


Abbildung 6.19: Der Einfluss von p_{II} auf die Peaks abhängig von p_{mm} welches links klein und rechts groß ist.

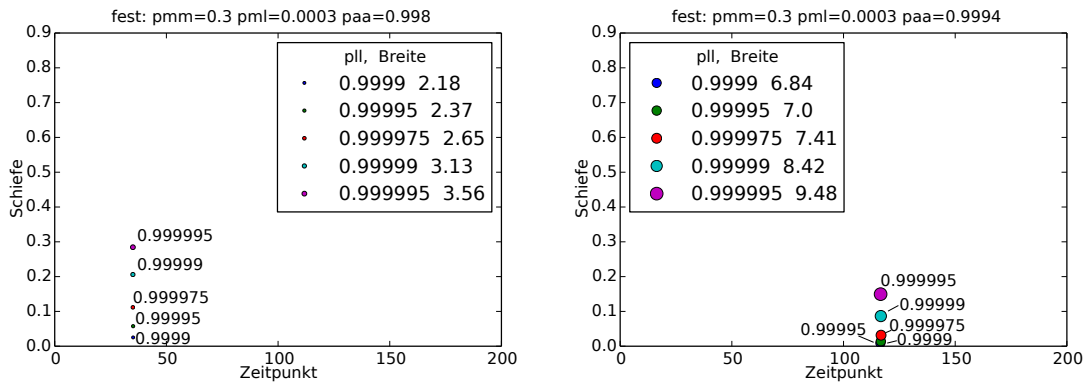


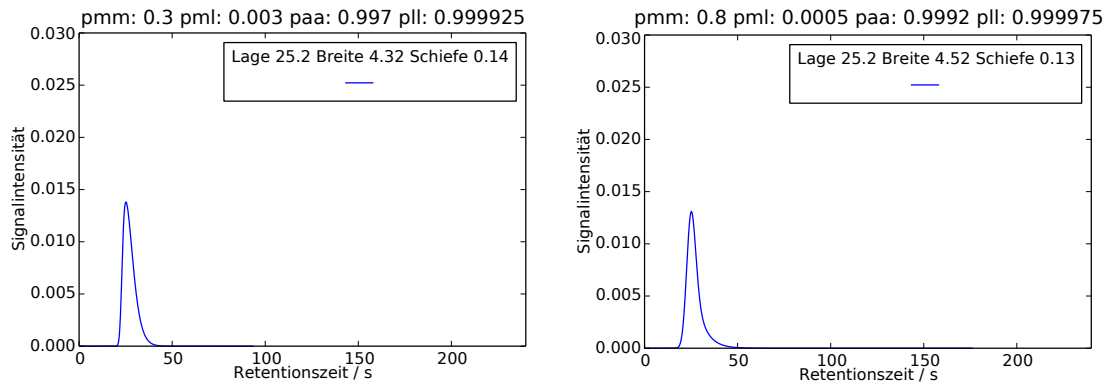
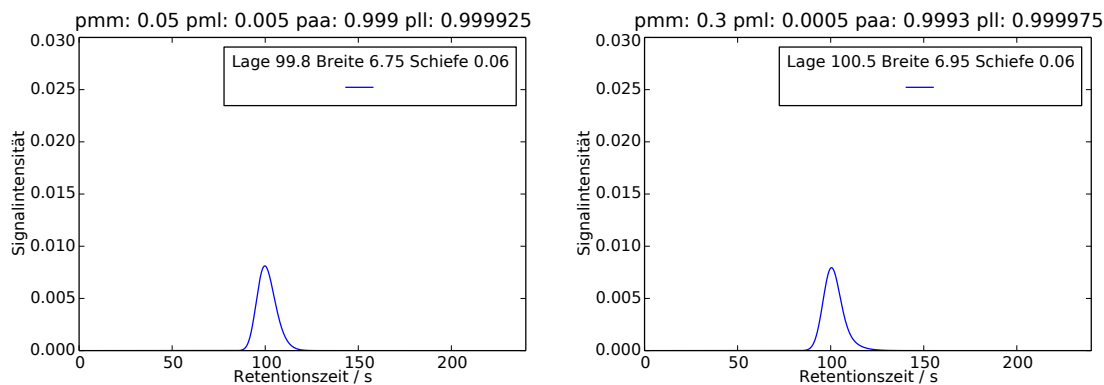
Abbildung 6.20: Der Einfluss von p_{II} auf die Peaks abhängig von p_{aa} , welches links klein und rechts groß ist.

wohl dadurch zu erklären, dass sich nicht mehr genügend Teilchen im mobil-adsorbierten System befinden, wodurch ein normaler Peak entstehen kann. Die gelösten Teilchen verursachen nun keine weitere Schiefe mehr, sondern dominieren so sehr, dass statt dessen die Peaks breiter werden. Der Maximalzeitpunkt ist zum großen Teil durch die Durchschnittsgeschwindigkeit der Teilchen bestimmt, die sich aus dem mobilen und adsorbierten Zustand ergibt. Sind sie hauptsächlich mobil, entstehen frühe Peaks, sind sie oft und lange stationär, wandern die Peaks nach hinten. Tailing und damit Schiefe entsteht immer da, wo nur wenige Teilchen langsamer sind, als die große Masse, was der Fall ist, wenn sich Teilchen nur selten lösen, dafür aber lange gelöst bleiben. Sind viele Teilchen sehr lange stationär, werden die Peaks statt dessen breit.

6.3.3 Parameter für ähnliche Peaks

Durch die vielfältigen und unterschiedlich starken Einflüsse der einzelnen Parameter, ist es möglich, fast gleiche Peaks mit völlig unterschiedlichen Parameterkombinationen zu erzeugen. Zwei Beispiele dafür sind in den Abbildungen 6.21 und 6.22 gezeigt.

In beiden Fällen unterscheiden sich die resultierenden Peaks kaum, sie könnten durch minimale Parameterveränderungen sogar noch stärker angenähert werden. Die Zuordnung von Simulationsparameter zu Peakcharakteristika ist also nicht eindeutig.

Abbildung 6.21: Sehr ähnliche Peaks bei $t \approx 25$ Abbildung 6.22: Sehr ähnliche Peaks bei $t \approx 100$

6.3.4 Grenzen

Mit dem 2-Zustände Modell war es nur eingeschränkt möglich, Tailing zu erzeugen. Insbesondere tritt kein Tailing nicht zu den späteren Zeitpunkten im Chromatogramm auf. Diese Einschränkung konnte, wie gezeigt wurde, mit dem 3-Zustände Modell aufgehoben werden.

Daher wäre es interessant zu wissen, welche Kombinationen aus Schiefe und Breite zu einem gegebenen Zeitpunkt realisierbar sind. Dazu sind in Abbildung 6.23 für die Zeitpunkte $t \approx 20$ und $t \approx 100$ jeweils eine Übersicht über die erreichbaren Breiten und Schiefen gegeben.

Generell können sehr schmale Peaks nicht sehr schief sein, was mit der Definition der beiden Maße zusammen hängt. Außerdem ist hier noch einmal die Mindestbreite für spätere Peaks zu sehen. Für $t \approx 100$ werden keine Peaks mit einer Breite von unter 4 beobachtet.

Bei sehr frühen Peaks ist mit zunehmender Breite auch immer Schiefe zu beobachten. Das liegt daran, dass vor dem Erreichen des Maximalwertes nur eine eher kurze Zeit vergangen sein kann, bei einem breiten Peak muss also einen großen Teil seiner Teilchen erst nach diesem Zeitpunkt das Ende der Säule erreichen. Genau das verursacht aber die beobachtbare Schiefe.

Dass bei eher späten Peaks nicht beliebig große Werte für die Schiefe erreicht werden, hängt mit dem Abbruch der Simulationen nach der maximalen Wartezeit zusammen. Es kann zum Zeitpunkt $t = 100$ durchaus Peaks mit einer Schiefe von 0,9 geben, deren Tail

wäre jedoch so lang, dass er über die 240 Sekunden hinaus ragt und damit hier nicht berücksichtigt wird.

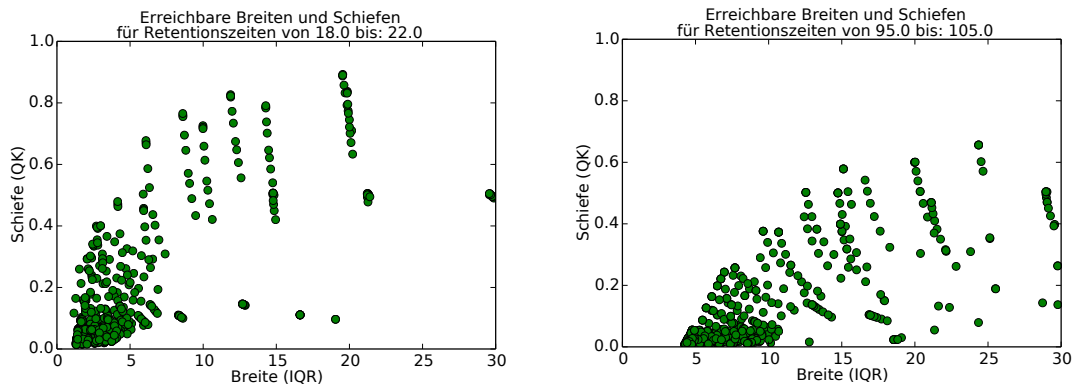


Abbildung 6.23: Mit dem 3-Zustände Modell erreichbare Breiten und Schiefen zu den Zeitpunkten $t \approx 20$ und $t \approx 100$

Die Erzeugung von breiteren Peaks ganz zu Beginn des Spektrums ist jedoch durch die tailenden Peaks möglich geworden. Dazu ist in Abbildung 6.24 eine Übersicht über die erreichbaren Zeiten und Breiten mit dem 3-Zustände Modell gegeben.

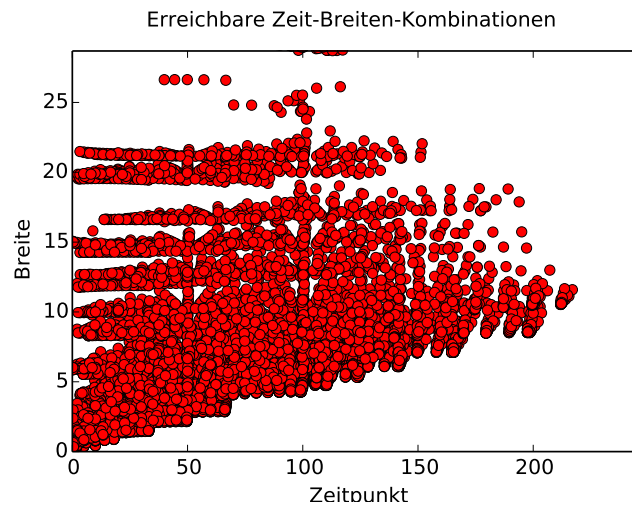


Abbildung 6.24: Mit dem 3-Zustände Modell erreichbare Zeitpunkte und Breiten

Das andere Problem jedoch, dass Peaks zu jedem Zeitpunkt eine Minimalbreite haben, konnte dadurch jedoch nicht gelöst werden, wie ebenfalls in Abbildung 6.24 zu sehen ist. Die Peaks können durch Hinzunahme eines weiteren stationären Zustandes höchstens breiter werden.

Kapitel 7

Diskussion

7.1 Ergebnisse

Es wurden zwei aufeinander aufbauende Modelle vorgestellt, mit denen der chromatographische Prozess simuliert werden kann. Bereits mit dem einfachen Modell können viele verschiedene Kombinationen aus zwei der drei Peakcharakteristika in den resultierenden Peaks gefunden werden. Es existiert auch ein deutlicher Zusammenhang zwischen den verwendeten Parametern und den Peakdaten. Daher kann ein gewünschter Peak, der bestimmte Werten für Retentionszeit und Breite aufweisen soll, schnell angenähert werden, indem die Parameter von Peaks mit ähnlichen Eigenschaften etwas variiert werden. Dabei müssen die Retentionszeit und Breite innerhalb eines gewissen Rahmens bleiben, welche die Grenzen des Modells darstellen.

Um auch Peaks mit Schiefe simulieren zu können, wurde das einfache Modell erweitert und dadurch eine Unterscheidung zwischen Adsorption an und Lösung der Teilchen in der stationären Phase eingeführt. Auch mit diesem Modell wurde ein Zusammenhang zwischen den Simulationsparametern und Peakdaten gefunden. Auch diese gefundenen Zusammenhänge erlauben eine Annäherung von gewünschten Peaks.

Es existiert zwar noch keine Formel, die eine exakte Entsprechung beschreibt, jedoch ist der Einfluss der Parameter auf die untersuchten Peakcharakteristika bekannt.

Aktuell kann für die in der Einleitung erwähnte, unbekannte Funktion $F : [0,1]^x \rightarrow \mathbb{R}^y$ gesagt werden, dass für $x = 4$ und $y = 3$ schon gute Annäherungen an reale Peaks erreicht werden können. Wie in Abschnitt 6.3.3 gezeigt, ist diese Funktion mit dem vorgestellten 3-Zustände Modell nicht bijektiv, da für den gleichen Peak völlig unterschiedliche Parameter im Modell gefunden wurden.

7.2 Ausblick

Ausgehend von den bisherigen Ergebnissen sind viele Ansätze für eine Fortsetzung der Arbeit denkbar.

Andere Maße zur Peakbeschreibung Wie in Abbildung 6.2 zu sehen war und dort bereits erwähnt wurde, widersprechen die Maße zur Beschreibung der Peaks in manchen Fällen der Wahrnehmung eines menschlichen Betrachters. Insbesondere ist dies der Fall, wenn das Tailing eines Peaks, wie unten rechts in Abbildung 6.2, eine sehr niedrige Intensität relativ zur maximalen Peakhöhe aufweist. Dadurch wächst auch der IQR deutlich an, jedoch nicht die wahrgenommene Breite eines Peaks.

Um die Auswirkungen dieser Art von Tailing zu untersuchen, müssten Chromatogramme mit realistischem Rauschen erstellt werden. Anschließend müssten die Peaks wieder extrahiert werden, um festzustellen, ob diese nach wie vor tailen beziehungsweise, welchen Wert die gemessene Schiefe anschließend hat. Die daraus resultierenden Konsequenzen für den IQR müssten ebenfalls untersucht werden.

Um dieses Problem zu umgehen, könnten die Peaks auch weitere Maße für Breite und Schiefe erhalten. Eine Möglichkeit wäre es, doch die Halbwertsbreite als Maß für die Breite heranzuziehen. Zumindest in den Beispielen aus Abbildung 6.2 würde die Halbwertsbreite den optischen Eindruck zur Breite der Peaks bestätigen, nämlich, dass der Peak unten rechts der schmalste der Peaks ist.

Zur Beschreibung der Schiefe der Peaks müsste dann ebenfalls ein neuer Koeffizient gebildet werden. Dazu werden die Zeiten vor und nach dem Maximalzeitpunkt t , zu denen der Peak die halbe Maximalintensität erreicht hat, ermittelt. Diese werden nun mit t_l und t_r für links und rechts des Maximums bezeichnet. Anschließend werden die Abstände der beiden Zeiten zum Maximalzeitpunkt gebildet: $a_l = t - t_l$ und $a_r = t_r - t$. Das Verhältnis dieser Abstände gibt einen Hinweis auf die Schiefe: Wenn $a_l < a_r$, dann liegt ein Tailing vor. Alternativ zur halbem Maximalhöhe können die Werte für t_l und t_r auch auf 30% oder 10% der Maximalhöhe des Peaks ermittelt werden um weitere Kenngrößen zu erhalten.

Nach Einarbeitung dieser Maße und eventuell weiterer Peakcharakteristika, müsste erneut eine Evaluierung der vorhandenen Simulationen statt finden.

Peaks als Funktionen Bisher wurden Peaks durch mehrere Maße charakterisiert. Jedoch ist insbesondere die Form eines Peaks schwer zu beschreiben. Ein Peak, der einen sehr langen, aber wenig intensiven Tail aufweist, wird damit möglicherweise mit den gleichen Werten beschrieben wie ein Peak mit einem kurzen Tail, der aber eine höhere Intensität aufweist. Mit der Halbwertsbreite und dem neuen Schiefe-Koeffizienten könnte dieses Problem auch nur teilweise behoben werden, da die relevanten Maße nur punktuell gemessen werden, nämlich auf 50% oder 10% der Peakhöhe. Statt eine Beschreibung der Form eines Peaks durch diese Maße, könnte statt dessen auch versucht werden, die Peaks als Funktionen aufzufassen und die Funktionswerte, die den Peak darstellen, möglichst genau zu simulieren.

Dies bedeutet aber weitreichende neue Untersuchungen, wie sich die Parameter auf diese Funktionswerte auswerten, um daraus Schlüsse ziehen zu können, welche Parameter für einen gewünschten Peak vonnöten sind.

Verifikation des Modells in größerem Rahmen Da mit dem Simulationsmodell echte Daten simuliert werden sollen, wäre es sinnvoll, einen Vergleich von simulierten und echten Daten in größerem Rahmen durchzuführen. Dazu wäre ein Vergleich der simulierten Peaks mit verschiedenen echten Messungen nützlich. Diese unterschiedlichen Messungen sollten dann Ergebnisse aus Säulen unterschiedlicher Hersteller berücksichtigen, die sich in Eigenschaften wie Länge, Durchmesser und Anzahl der Kapillaren, Art der Beschichtung, sowie Betriebsparameter wie Temperatur und Druck unterscheiden.

Bei einem solchen Vergleich sollte sich dann herausstellen, ob die minimale Breite, die ein mit dem Modell erzeugter Peak hat, ein Problem darstellt, wenn nämlich tatsächlich schmalere Peaks zum gleichen Zeitpunkt existieren.

Außerdem wurde bisher von einer konstanten Länge der Säule ausgegangen, ein Vergleich verschiedener Längen und die Auswirkungen davon auf die Peaks wäre interessant.

Es ist zu erwarten, dass für andere Säulenlängen die Parameter verändert werden müssen, um ähnliche Peaks zu erhalten.

Im Rahmen dieses Vergleiches wird es sicher nötig sein, die bisherige Maximalzeit von 240 Sekunden anzupassen und zu überprüfen, ob die sich die Ergebnisse auch auf spätere Zeitpunkte übertragen lassen überprüfen.

Wenn sich bei einem solchen Vergleich mit vielen Messungen und angepasster Säulenlänge herausstellt, dass mit dem Modell einige der Peaks nicht simuliert werden können, kann zunächst das 3-Zustände Modell voll ausgeschöpft werden, indem alle sechs möglichen Parameter berücksichtigt werden. Eventuell ergeben sich dadurch weitere, bisher nicht realisierbare Peaks.

Wenn auch das nicht ausreicht, muss eine erneute Veränderung des Modells erfolgen. In Kapitel 3 wurden am Ende einige Ideen vorgestellt, wie das Modell verändert werden könnte. Sowohl die Berücksichtigung des Gleichgewichts der Phasen als auch eine maximale Kapazität für die Adsorption könnten für weiteres Tailing sorgen. Dieses neue Tailing könnte sich insbesondere in Bezug auf die im letzten Abschnitt angedachten anderen Beschreibungen der Peaks anders verhalten, als das bisher erzeugte Tailing. Außerdem ist es auch gut vorstellbar, dass so schmalere Peaks zum gleichen Zeitpunkt entstehen, da sich für einige Teilchen die Wahrscheinlichkeit, mobil zu bleiben erhöht und der Teilchenpulk enger beieinander bleibt.

Entsprechung von Simulationsparametern zu Peakcharakteristika Bisher ist der Einfluss der Parameter auf die Peakcharakteristika bekannt, es fehlt aber noch eine genaue Formel, mit der diese aus den Simulationsparametern berechnet werden können.

Vorteilhaft wäre es daher, ein Verfahren zu entwickeln, um einen Peak mit bestimmten, vorgegebenen Eigenschaften zu simulieren. Bisher funktioniert das händisch unter Anwendung der Erkenntnisse, wie sich die Parameter auf die Peaks auswirken. Wenn also ein Peak mit bestimmten Eigenschaften gewünscht ist, wird bisher die grobe Umgebung abgesucht. Die Parameter der dort gefundenen Peaks werden in die passende Richtung abgewandelt (zb. er ist etwas zu früh, dann wird p_{mm} verringert oder zu schmal, dann wird p_{ll} erhöht). Anschließend wird überprüft, wo der neue Peak liegt, um eventuell in einer weiteren Iteration die Parameter anzupassen. Um das zu ersparen, müsste entweder der komplette Parameterraum noch besser abgedeckt werden, was sehr viel Laufzeit und Speicherplatz kosten würde, oder die Vorgehensweise automatisiert werden.

Laufzeit Da sich die Umsetzung der schrittweisen Simulation für mehr als zwei Zustände als eher laufzeitintensiv erwiesen hat, könnte eine neue Methode zur Simulation eines Einzelschrittes entwickelt werden. Bei der bisherigen Vorgehensweise werden sehr viele, danach nicht mehr benötigte Zwischeninformationen berechnet. Möglicherweise wäre daher ein Ansatz, der schleifenbasiert jedes Feld der neuen Zustandsvektoren berechnet, lohnenswert.

Abbildungsverzeichnis

2.1	Beispiel für ein Chromatogramm	7
2.2	Querschnitt einer MCC	8
3.1	Graphische Darstellung des 2-Zustände Modells	15
3.2	PAA für das 2-Zustände Modell	16
3.3	Mögliche Modelle mit drei Zuständen	17
3.4	PAA für das allgemeine 3-Zustände Modell	18
4.1	Prinzip der schrittweisen Simulation	23
4.2	Berechnung eines Zustandswechsels bei der schrittweisen Simulation im 3-Zustände Modell	24
4.3	Prinzip der ergebnisbasierten Simulation bei zwei Zuständen	27
4.4	Berechnung der neuen Zustände des PAA im 2-Zustände Modell	29
4.5	Zeiten zusammenfassen	31
6.1	Nicht berücksichtigte Verteilungen	43
6.2	Peaks mit verschiedenen Schiefen	44
6.3	Einfluss von p_m und p_s auf einen Peak	45
6.4	Einfluss von p_m und p_s auf einen Peak	46
6.5	Die resultierenden Zeiten und Breiten einiger Parameterkombinationen	47
6.6	Mögliche Parameterkombination für Retentionszeit $t \approx 75$	48
6.7	Der Einfluss von p_s auf die Peaks abhängig von p_m	48
6.8	Der Einfluss von p_m auf die Peaks abhängig von p_s	49
6.9	Mit dem 2-Zustände Modell erreichbare Zeitpunkte und Breiten	50
6.10	Ausschnitt der erreichbaren Zeiten und Breiten	50
6.11	Ausschnitt der erreichbaren Zeiten und Breiten	51
6.12	Der Einfluss von p_{mm} auf die Peaks abhängig von p_{aa}	53
6.13	Der Einfluss von p_{mm} auf die Peaks abhängig von p_{ml}	54
6.14	Der Einfluss von p_{ml} auf die Schiefe und Breite abhängig von p_{ll}	55
6.15	Der Einfluss von p_{ml} auf die Schiefe und Breite abhängig von p_{mm} und p_{aa}	56
6.16	Der Einfluss von p_{aa} auf die Peaks abhängig von p_{mm}	57
6.17	Der Einfluss von p_{aa} auf die Peaks abhängig von p_{ml}	57
6.18	Der Einfluss von p_{ll} auf die Schiefe und Breite abhängig von p_{ml}	58
6.19	Der Einfluss von p_{ll} auf die Peaks abhängig von p_{mm}	59
6.20	Der Einfluss von p_{ll} auf die Peaks abhängig von p_{aa}	59
6.21	Sehr ähnliche Peaks bei $t \approx 25$	60
6.22	Sehr ähnliche Peaks bei $t \approx 100$	60
6.23	Mit dem 3-Zustände Modell erreichbare Breiten und Schiefen	61

6.24 Mit dem 3-Zustände Modell erreichbare Zeitpunkte und Breiten	61
---	----

Literaturverzeichnis

- Jörg Ingo Baumbach. Ion mobility spectrometry coupled with multi-capillary columns for metabolic profiling of human breath. *Journal of breath research*, 3(3):034001, September 2009. ISSN 1752-7163. doi: 10.1088/1752-7155/3/3/034001. URL <http://www.ncbi.nlm.nih.gov/pubmed/21383463>.
- Paul G. Boswell, Dwight R. Stoll, Peter W. Carr, Megan L. Nagel, Mark F. Vitha, and Gary A. Mabbott. An advanced, interactive, high-performance liquid chromatography simulator and instructor resources. *Journal of Chemical Education*, 90(2):198–202, 2013. doi: 10.1021/ed300117b. URL <http://dx.doi.org/10.1021/ed300117b>.
- Amit Ghosh, David S. Morison, und Robert J. Anderegg. Simulation of a gas chromatography-mass spectrometry experiment with a commercial spreadsheet program. *Journal of Chemical Education*, 65(6):A154, 1988. doi: 10.1021/ed065pA154. URL <http://dx.doi.org/10.1021/ed065pA154>.
- J. C. Giddings. Kinetic Origin of Tailing in Chromatography. *Analytical Chemistry*, 35(13):1999–2002, Dezember 1963. ISSN 0003-2700. doi: 10.1021/ac60206a008. URL <http://pubs.acs.org/doi/abs/10.1021/ac60206a008>.
- Hans-Joachim Hübschmann. *Handbook of GC/MS: Fundamentals and Applications*. Wiley-VCH, Weinheim, 2. Edition, 2009. ISBN 9783527314270. doi: 10.1002/9783527625215.
- N.L. Johnson, S. Kotz, und N. Balakrishnan. *Continuous univariate distributions*. Number Bd. 1 in Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley & Sons, 1994.
- Karl Kaltenböck. *Gaschromatographie für Einsteiger*. Wiley-VCH, Weinheim, 1. Edition, 2008.
- Bruno Kolb. *Gaschromatographie in Bildern [eine Einführung]*. Wiley-VCH, Weinheim, 2. Edition, 2003.
- Tobias Marschall. Algorithms and statistical methods for exact motif discovery. 2011. URL <http://eldorado.tu-dortmund.de/handle/2003/27760>.
- Tobias Marschall, Inke Herms, Hans-Michael Kaltenbach, und Sven Rahmann. Probabilistic arithmetic automata and their applications. *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, 9(6):1737–1750, November 2012. ISSN 1545-5963. doi: 10.1109/TCBB.2012.109. URL <http://dx.doi.org/10.1109/TCBB.2012.109>.
- P. Moretti, S. Vezzani, E. Garrone, und G. Castello. Evaluation and prediction of the shape of gas chromatographic peaks. *Journal of Chromatography A*, 1038(1-2):171–181, Juni

2004. ISSN 00219673. doi: 10.1016/j.chroma.2004.03.039. URL <http://linkinghub.elsevier.com/retrieve/pii/S002196730400490X>.

R. Łobiński, V. Sidelnikov, Y. Patrushev, I. Rodriguez, und A. Wasik. Multicapillary column gas chromatography with element-selective detection. *TrAC - Trends in Analytical Chemistry*, 18(7):449–460, 1999.

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet sowie Zitate kenntlich gemacht habe.

Dortmund, den 15. September 2015

Elisabeth Böhmer