

TECHNISCHE UNIVERSITÄT MÜNCHEN

**Institut für Verkehrswesen
Lehrstuhl für Verkehrstechnik**

Netzweite Lichtsignalsteuerung auf Basis Rekurrenter Neuronaler Netze

Dipl.-Ing. Univ. Yining Li

Vollständiger Abdruck der von der Fakultät für Bauingenieur- und Vermessungswesen der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Stephan Freudenstein

Prüfer der Dissertation: 1. Univ.-Prof. Dr.-Ing. Fritz Busch
2. Prof. dr. ir. Hans Hellendoorn,
Delft University of Technology / Niederlande
3. Priv.-Doz. Dr. rer. nat. Andreas Johann

Die Dissertation wurde am 22.11.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Bauingenieur- und Vermessungswesen am 29.03.2012 angenommen.

Danksagung

Für meine Doktorarbeit schulde ich sehr vielen Menschen einen herzlichen Dank. Ganz besonders möchte ich mich bei meinem Doktorvater Herrn Professor Fritz Busch für die Betreuung der Arbeit bedanken. Er öffnete mir die Tür zum Bereich Verkehrstechnik und gab mir immer wertvolle Anregungen.

Für die freundliche Übernahme des Koreferats danke ich Herrn Professor Hans Helendoorn und Herrn Dr. Andreas Johann. Ihre Anmerkungen in der Schlussphase der Arbeit waren sehr hilfreich.

Herrn Dr. Paul Mathias danke ich für die fachliche Betreuung und Unterstützung im Laufe der Entstehung dieser Arbeit. Dies war entscheidend für das Gelingen der Arbeit.

Bei der Siemens AG und dem Universität Bayern e.V. bedanke ich mich für die Förderung des Forschungsvorhabens.

Ich danke allen Kolleginnen und Kollegen für den umfangreichen fachlichen Austausch. Mein besonderer Dank gilt den Kolleginnen und Kollegen, die mir bei dem Korrekturlesen der Arbeit geholfen haben.

Schließlich möchte ich mich recht herzlich bei meinen Eltern, meiner Frau und meiner Schwester für die langjährige Unterstützung bedanken.

Yining Li

Kurzfassung

Um den Verkehr auf der Straße effizient und störungsfrei zu bewältigen, steht neben einer Anzahl wichtiger Maßnahmen die gerechte und optimale Verteilung der Kapazitäten im Vordergrund. In den Städten und Ballungsräumen ist insbesondere die Verkehrssteuerung der Knotenpunkte durch Lichtsignalanlagen (LSA) mit dieser Aufgabe beschäftigt.

Für die Steuerung von LSA werden zur Zeit überwiegend messwert- bzw. logikbasierte Verfahren angewendet. Als Alternative zu den regelbasierten Verfahren kommen modellbasierte bzw. netzweite verkehradaptive Verfahren zum Einsatz. Ein Verkehrsmodell berechnet dabei ein räumlich und zeitlich geordnetes Abbild des aktuellen Verkehrsaufkommens, mit dem es möglich ist, verschiedene Steuerungsalternativen hinsichtlich ihrer Wirkung auf den gesamten Verkehrsablauf zu überprüfen. Dies erfolgt im Allgemeinen als iterative Optimierung von relevanten Kenngrößen, z. B. Verlustzeiten.

Die mit einem dynamischen und korrekten Verkehrsmodell verbundenen Schwierigkeiten deuten auf das wesentliche Problem der netzweiten Steuerungsverfahren hin. Eine integrierte Betrachtung von dynamischer Verkehrsumlegung und netzweiter Verkehrssteuerung ist noch nicht in bisherigen Entwicklungen berücksichtigt.

In diesem Sinne wird in der vorliegenden Arbeit ein neues modellbasiertes netzweites Verkehrssteuerungsverfahren entwickelt. Um das neue Verkehrssteuerungssystem einzuführen, wird zuerst ein dynamisches Verkehrsmodell implementiert. Das Verkehrsmodell basiert auf einem speziellen Rekursiven Σ - Π -Netz und einer modifizierten rekurrenten Rückwärtspropagierung als Lernverfahren. Die zeitveränderliche Routenwahl im Modell ist gegeben durch Knoten-Abbiegeraten, die die Gewichte eines Simulationsnetzes darstellen und vom Fehlerpropagierungsnetz zyklisch korrigiert werden.

In diesem Modell wird entsprechend der vielfältigen Zielvorstellungen eine Vielzahl möglicher Gütebestimmungen für die Optimierung verwendet. Ausschlaggebend für die Auswahl der Kriterien sind einerseits die Möglichkeiten, die die Modelle im Sinne einer realistischen Abbildung zulassen. Andererseits können stadtplanerische oder verkehrspolitische Überlegungen grundsätzlich zu verschiedenen, aber jeweils für sich optimalen Lösungen führen. Hinsichtlich der Optimierung wird zwischen nutzer- und systemoptimalen Lösungen unterschieden.

Hauptgegenstand der Dissertation ist es, das existierende Modell in Richtung netzweiter Verkehrssteuerung zu erweitern. Der grundsätzliche Ansatz dazu ist, als Steue-

rungsparameter anstelle der Knoten-Abbiegeraten Freigabezeit bzw. Link-Kapazitätswerte zu wählen und durch ein entsprechend konstruiertes Fehlerpropagierungsnetz nachführen zu lassen. Hierbei ist zu berücksichtigen, dass die Kapazitätswerte aller auf einen Knoten zuführenden Links über die LSA-Steuerungsprogramme voneinander abhängen.

Neben der Bestimmung von Kapazitäten leisten Netzsteuerungen auch eine Koordination benachbarter LSA entlang belasteter bzw. priorisierter Routen („Grüne Wellen“). Die Optimierung der Versatzzeit im Netz stellt ein komplexes Problem dar, für das in der Forschungsarbeit ein Lösungsansatz erarbeitet werden soll, der mit der vorhandenen Information aus den oben beschriebenen Verkehrsmodellen realisiert werden kann. Die Versatzzeiten werden auf Basis der makroskopischen Informationen über die Größen der Verkehrsströme bestimmt. Dazu wird die Koordination in Abhängigkeit von der mittleren Progressionsgeschwindigkeit durchgeführt. Die Strecken werden so entsprechend der Verkehrsstärke im Schichten-Modell aufgeteilt, dass die Streckenleistungsfähigkeiten auch in den einzelnen Schichten des Rekurrenten Netzes abgebildet werden. Durch ein Wartemodell wird der Zusammenhang zwischen Wartezeit (die abhängig von der Versatzzeit ist) und der so genannten „Streckenleistungsfähigkeit“ hergestellt. Um die Versatzzeit quantitativ zu optimieren, wird ein drittes Fehlerpropagierungsnetz integriert.

Die softwaretechnische Umsetzung wird durch den Prototyp NEURA-NESC (Neural net wide traffic signal control) realisiert. In einer Simulationsstudie konnte NEURA-NESC an einem virtuellen Netz mit acht Knotenpunkten getestet und bewertet werden. Die Untersuchung wird anhand diverser Verkehrssituationen veranschaulichen, welche positive Ergebnisse NEURA-NESC im Netz gegenüber herkömmlichen Verfahren liefern kann.

Inhaltsverzeichnis

1	Einführung	1
1.1	Funktion der Verkehrssteuerung mit Lichtsignalanlagen	1
1.2	Problemstellung	2
1.3	Vorgehensweise	4
2	Grundlagen und Übersicht der Steuerungsverfahren.....	5
2.1	Grundgrößen und Variablen der Lichtsignalsteuerung	5
2.1.1	Signalgruppe, Zwischenzeiten	5
2.1.2	Phaseneinteilung und Phasenübergänge.....	6
2.1.3	Umlaufzeit	7
2.1.4	Freigabezeit	8
2.1.5	Versatzzeit	9
2.2	Steuerungsverfahren.....	9
2.2.1	Festzeitsteuerung	10
2.2.2	Regelbasierte Steuerung	11
2.2.3	Modellbasierte Steuerung.....	12
2.3	Bekannte adaptive Steuerungsverfahren.....	13
2.3.1	Offline-Modelle.....	13
2.3.2	Online-Modelle.....	14
2.4	Zwischenfazit	19
3	Rekurrente Neuronale Netze.....	21
3.1	Das menschliche Gehirn und die Simulation	21
3.2	Grundlage des Rekurrenten Neuronalen Netzes.....	22
3.3	Rekurrente Backpropagation-Lernverfahren	23
3.3.1	Backpropagation für Feedforward-Netze.....	24
3.3.2	Backpropagation für RNN.....	26
3.3.3	Backpropagation für zeitabhängiges RNN	28
4	Das RNN-Verkehrsmodell	31
4.1	Das Simulations-Netzwerk	31
4.1.1	Topologie des Simulations-Netzes	32
4.1.2	Dynamik des Simulations-Netzes	33
4.1.3	Das Schichten-Modell des Simulations-Netzes.....	36
4.2	Das Fehlerpropagierungs-Netz zur statischen Umlegung	38
4.2.1	Optimierungsfunktion.....	38
4.2.2	Berechnung der Verbindungsgewichte.....	40
4.3	Das dynamische RNN-Umlegungsmodell.....	45

4.3.1	Erweitertes dynamisches Simulations-Modell	46
4.3.2	Optimierungsfunktion des dynamischen Modells	46
4.3.3	Das Fehlerpropagierungs-Netz und das Betriebsmodell	47
4.4	Zwischenfazit	48
5	RNN-Verkehrssteuerungsmodell für Freigabezeitanteilloptimierung	51
5.1	Freigabezeitverteilung als Steuerungsparameter	51
5.1.1	Kapazitätswerte des Geradeausfahrers	52
5.1.2	Kapazitätswerte des Abbiegers	53
5.1.3	Festlegung der Umlaufzeit	55
5.2	Optimierungsfunktion für Freigabezeitoptimum	58
5.2.1	Formulierung der Optimierungsfunktion	58
5.2.2	Berechnung der Wartezeit	59
5.2.3	Dynamische Optimierungsfunktion	60
5.3	Erweiterungen des RNN-Modells zur Optimierung der Freigabezeitverteilung	62
5.3.1	Berechnung der Freigabezeitkorrekturen	62
5.3.2	Update der Freigabezeitverteilung	69
5.4	Das Betriebsmodell	72
5.5	Zwischenfazit	74
6	RNN-Verkehrssteuerungsmodell für Versatzzzeitoptimierung	75
6.1	Versatzzzeit als Steuerungsparameter	75
6.1.1	Grunddefinitionen und Konzept	75
6.1.2	Versatzzzeit und Wartezeit	78
6.1.3	Streckenleistungsfähigkeit und Wartezeit	80
6.2	Optimierungsfunktion für Versatzzzeitoptimum	82
6.3	Erweiterungen des RNN-Modells zur Optimierung der Versatzzzeit	83
6.3.1	Dynamische Gleichung der Versatzzzeitoptimierung	83
6.3.2	Berechnung der Versatzzzeitkorrekturen	84
6.3.3	Update der Versatzzzeit	87
6.4	Das gesamte Betriebsmodell der LSA-Steuerung	89
6.5	Zwischenfazit	91
7	Erprobung und Bewertung des Steuerungsverfahrens	93
7.1	Programm NEURA-NESC	93
7.2	Das Testnetz	95
7.3	Simulationsstudie	97
7.3.1	Bewertungskriterien	97
7.3.2	Aufbau der Simulationsstudie	98
7.3.3	Durchführung der Simulationsstudie	99
7.3.4	Ergebnisse der Simulationsstudie	101
7.4	Zwischenfazit	108
8	Fazit und Ausblick	111
8.1	Zusammenfassung	111
8.2	Weiterer Forschungsbedarf	112
	Literaturverzeichnis	115
	Abbildungsverzeichnis	125

Tabellenverzeichnis	127
Symbolverzeichnis.....	129
Anhang A – Berechnungen der Zusammenhänge von Versatzzeit und mittlerer Wartezeit	131
Anhang B – Quell-Ziel-Matrix des virtuellen Straßennetzes.....	141
Anhang C – Anzahl der Simulationsläufe	143

1 Einführung

1.1 Funktion der Verkehrssteuerung mit Lichtsignalanlagen

Der Begriff Verkehrssteuerung steht für eine Reihe von Maßnahmen, die das Ziel anstreben, den Verkehr in einem Netz – also auf Kanten und Knoten - zu lenken und zu verflüssigen. Dabei kommt der Verkehrssteuerung an Knotenpunkten besondere Bedeutung zu, da diese den Verkehrsablauf sich kreuzender bzw. tangierender Streckenabschnitte (Kanten) regelt. Die Lichtsignalsteuerung ist in Ballungsräumen aufgrund ihrer Verbreitung und Auswirkung die allgegenwärtigste Maßnahme zur Steuerung des Verkehrsflusses.

Die Lichtsignalsteuerung ist historisch bedingt in erster Linie mit der Verkehrssicherheit verbunden. Durch die Lichtsignalsteuerung können die Verkehrsströme an Knotenpunkten zeitlich entflochten und insbesondere die nicht-motorisierten Verkehrsteilnehmer besser geschützt werden. Sie dient dazu, gefährliche oder gefahrenträchtige Verkehrssituationen zu entschärfen und Harmonie im Straßenverkehrssystem zu gewährleisten.

Neben der Verkehrssicherheit wird im Zusammenhang mit Lichtsignalanlagen (LSA) deren Einfluss auf die Qualität des Verkehrsablaufs im Straßennetz wahrgenommen. Die Qualität des Verkehrsablaufs in städtischen Netzen wird vor allem von der Kapazität der Knotenpunkte und den dort auftretenden Verlustzeiten bestimmt. So wird es als originäre Aufgabe der Lichtsignalsteuerung betrachtet, eine ausreichende Kapazität des Straßennetzes zu gewährleisten um so die Verlustzeiten zu minimieren.

Die Lichtsignalsteuerung unterstützt in Deutschland den sogenannten 'Stadtfahrplan' [BOHLINGER, 2006] als Instrumentarium zur Bündelung von Hauptverkehrsströmen auf dem Hauptstraßennetz. Dieser Effekt passt auch zu dem Ziel der Verkehrsplanung. Zudem kann die Lichtsignalsteuerung mögliche Defizite des Straßenentwurfs lösen, da die Steuerung im Gegensatz zur Bauausführung auf Probleme kurzfristig und kostengünstig angepasst werden kann.

Ab den 1980er Jahren wurden die Belange des öffentlichen Verkehrs (ÖV) bzw. des öffentlichen Personennahverkehrs (ÖPNV) an LSA mit berücksichtigt. Mit Hilfe von Vorrangschaltungen, ÖV-Priorisierung genannt, konnte die Reisegeschwindigkeit sowie die Pünktlichkeit der ÖV-Fahrzeuge signifikant verbessert werden.

In den letzten zehn Jahren wurde der Umweltaspekt der Verkehrssteuerung immer häufiger betrachtet und berücksichtigt. Neue Steuerungsverfahren können die Möglichkeit schaffen, durch Minimierung der Reisezeiten und Halte Kraftstoffverbrauch und Schadstoffemissionen zu reduzieren.

1.2 Problemstellung

Aufgrund des weiterhin zunehmenden motorisierten Individualverkehrs und der damit verbundenen negativen Auswirkungen auf die Umwelt wird die Optimierung des Verkehrsablaufs gefordert. Im Mittelpunkt dieser Herausforderung steht insbesondere die Optimierung der Lichtsignalsteuerung. Deshalb wurden die Verfahren zur Steuerung von Lichtsignalanlagen in den vergangenen Jahrzehnten fortlaufend weiterentwickelt und verfeinert, so dass es im Wesentlichen gelang, eine Steigerung der Verkehrsleistung ohne eine Verschlechterung der Reisequalität zu erreichen.

Verkehrsabhängige Steuerungen in Deutschland sind seit dem Jahr 1934 in Berlin im Einsatz, ihre Wirkungen wurden in zahlreichen Untersuchungen analysiert und mit denen von Festzeitsteuerungen verglichen [HOLROYD, 1971]. Die Ergebnisse zeigen, dass durch verkehrsabhängige Steuerungen günstige Wirkungen, insbesondere bei niedriger und mittlerer Auslastung, resultieren. Im Vergleich zu dieser Aussage wurde durch BRETHERTON ET AL. [1982] gezeigt, dass gewöhnliche verkehrsabhängige Verfahren bei hohen Auslastungen gleiche Ergebnisse wie Festzeitsteuerungen liefern. In diesen Fällen bieten verkehrsabhängige Verfahren keinen Vorteil gegenüber der Festzeitsteuerung.

Um diesen Nachteilen zu begegnen, wurde seit Mitte der 1960er Jahre die neue Technik der adaptiven Steuerung entwickelt, die Ende der 60er Jahre mit dem Verfahren TRANSYT [ROBERTSON, 1969] als Offline-Modell etabliert und eingesetzt wurde. Der Entwicklung von TRANSYT folgte SIGMA [GARBEN ET AL., 1988] in Deutschland. Am Anfang der 80er Jahre wurde auf das offline Modell TRANSYT ein online Modell SCOOT [HUNT ET AL., 1981] aufgebaut. Danach waren im europäischen Ausland die Verfahren SPOT/UTOPIA [MAURO ET AL., 1990], MOTION [BUSCH UND KRUSE, 1993] und BALANCE [FRIEDRICH, 1999] und in den USA das Verfahren OPAC [GARTNER, 1982] im Einsatz.

Die adaptive Verkehrssteuerung ermöglicht eine Reaktion auf kurzfristige Schwankungen des Verkehrsaufkommens im gesamten Netz. Zahlreiche Simulationen und Feldversuche haben gezeigt, dass ein Einsatz solcher Verfahren zu einer Reduzierung von Wartezeiten, Reisezeiten und Anzahl der Halte führen kann.

Netzweite adaptive Verkehrssteuerungen bauen auf ein Verkehrsmodell auf, das ein räumlich und zeitlich geordnetes Abbild des aktuellen Verkehrs berechnet. Mit dem Modell ist es möglich, verschiedene Steuerungsalternativen hinsichtlich ihrer Wirkung auf den gesamten Verkehrsablauf zu überprüfen. Die mit einem Verkehrsmodell ver-

bundenen Schwierigkeiten, auf Basis der begrenzten erfassbaren Informationen ein dynamisches und realitätsnahes Abbild des Verkehrsablaufs zu ermitteln und prognostizieren, deuten auf das wesentliche Problem der netzweiten Steuerungsverfahren hin. Eine integrierte Betrachtung dynamischer Verkehrsumlegungen und netzweiter Verkehrssteuerungsmaßnahmen ist noch nicht in die bisherige Entwicklung der Optimierungsalgorithmen eingeflossen.

Die Lösung des Optimierungsproblems ist eine Kernaufgabe netzweiter Verkehrssteuerungsverfahren. Der Ansatz der Optimierung kann in exakte mathematische Verfahren, Heuristiken und Konstruktionsverfahren eingeteilt werden [MERTZ, 2001]. Die exakten mathematischen Verfahren, z.B. Hill Climbing oder dynamische Programmierung, wurden häufig in bisherigen netzweiten Verkehrssteuerungsverfahren eingesetzt. In den letzten Jahrzehnten wurden Heuristiken wie z.B. Genetische Algorithmen oder Künstliche Neuronale Netze zur Lösung dieses Optimierungsproblems angewendet.

Heuristiken (aus dem Griechischen: heuriskein - finden, entdecken) sind Methoden zur Suche von guten (nahezu optimalen) Lösungen für ein Optimierungsproblem. Bei heuristischen Verfahren werden spezialisierte Methoden oder zusätzliches Wissen zur Lösung des Problems verwendet [PEARL, 1984]. Obwohl keine Aussage getroffen wird, wie nah die gefundene Lösung am Optimum liegt, ist eine gute heuristische Lösung für praktische Probleme ausreichend genau [HÜFTLE, 2006]. Es ist auch besonders vorteilhaft für den komplizierten Optimierungsablauf der netzweiten Verkehrssteuerung, dass die Suche in möglichst kurzer Zeit erfolgt und damit effizienter berechnet werden kann. Bekannte Anwendungen der Heuristiken in der Optimierung der Verkehrssteuerung sind GALOP [BRAUN, 2008], ALMASRI [2006] und CEYLAN [2004,2005], die einen genetischen Algorithmus zur Optimierung verwenden. Neben dem Verfahren 'Genetischer Algorithmus' sind Künstliche Neuronale Netze (KNN) eine der häufigsten benutzten heuristischen Methoden für die Verkehrsoptimierung. KNNs waren auch in junger Vergangenheit in verschiedenen LSA-Steuerungen [SARAF, 1994, FAGHRI ET AL., 1997] in Verwendung. Beiden Verfahren benutzen Feedforward-Netze, um die Signalsteuerung eines einzelnen Knotenpunkts zu optimieren. Ein Verkehrsmodell steht in beiden Verfahren nicht zur Verfügung. Im deutschsprachigen Raum ist ein netzweites adaptives Verkehrssteuerungsverfahren auf Basis neuronaler Netze bislang noch nicht entwickelt worden.

Die vorliegende Arbeit ist vor diesem Hintergrund entstanden und hat zum Ziel, ein ganz neues verkehrsadaptives auf KNNs basiertes Steuerungsverfahren zu entwickeln. Dieses Verfahren bietet die Möglichkeit, eine dynamische Verkehrsumlegung und die Verkehrssteuerung gleichzeitig zu betrachten; mehrere Parameter der Signalprogramme hintereinander zu optimieren und zudem eine optimale Lösung zu finden.

1.3 Vorgehensweise

Ausgehend von den oben dargestellten Schwerpunkten soll in der Arbeit ein neues Verfahren als Forschungsprototyp für die netzweite Verkehrssteuerung entwickelt werden. Im ersten Teil der Arbeit soll ein Überblick über einige verkehrs- und informationstechnische Erkenntnisse und Theorien gegeben werden, so wie sie uns heute zur Verfügung stehen.

Dazu werden zunächst im Kapitel 2 die Grundlagen der Lichtsignalsteuerung und die Begriffe der adaptiven Verkehrssteuerung behandelt. Danach werden heute verfolgte Ansätze zur Steuerung der Lichtsignalanlagen am Beispiel bekannter und etablierter adaptiver Steuerungsverfahren vorgestellt. Weiter wird in diesem Kapitel auf die bekannten Vor- und Nachteile der verschiedenen Steuerungsverfahren eingegangen.

Im Kapitel 3 wird dann die Theorie der Rekurrenten Neuronalen Netze (RANN) mit dem Lernverfahren Rekurrente-Backpropagation (RBP) vorgestellt. Danach wird im Kapitel 4 das neue RANN-basierte Gleichgewichtsverfahren zur Verkehrsumlegung eingeführt und im Detail beschrieben.

Der zweite Teil behandelt den Hauptgegenstand der Arbeit, das existierende Modell in Richtung netzweite Verkehrssteuerung zu erweitern. Im Kapitel 5 wird die Verkehrssteuerung eines einzelnen Knotenpunkts durch RANN eingeführt. Freigabezeitanteile bzw. Link-Kapazitätswerte werden als neue Steuerungsparameter in dem Modell gewählt und durch ein zweites Fehlerpropagierungsnetz nachgeführt. Im nachfolgenden Kapitel 6 wird gezeigt, wie eine Koordinierung benachbarter LSA entlang belasteter Routen optimiert wird. Ein neues ins Modell integriertes drittes Fehlerpropagierungsnetz bestimmt die Versatzzeiten.

Das Verfahren wird mit Hilfe von einem virtuellen Netz in einer Simulationsstudie gegenüber anderen herkömmlichen Steuerungsverfahren verglichen und bewertet. Diese Ergebnisse werden im Kapitel 7 in graphischer Form dargestellt. Der dritte und letzte Teil der Arbeit schließt im Kapitel 8 mit der Zusammenfassung und dem zukünftigen Forschungsbedarf ab.

2 Grundlagen und Übersicht der Steuerungsverfahren

2.1 Grundgrößen und Variablen der Lichtsignalsteuerung

2.1.1 Signalgruppe, Zwischenzeiten

Der Steuerungsablauf der Lichtsignalanlagen kann mit einem Signalprogramm definiert werden. In einem Signalprogramm werden Dauer und Positionen der Freigabezeit und Sperrzeit für alle Signalgruppe einer LSA festgelegt. In einer Signalgruppe werden die Signalgeber zusammenfasst. Abbildung 2-1 veranschaulicht diese Definition. Die Signalgruppen für Fußgänger werden bei diesem einfachen Beispiel nicht berücksichtigt.

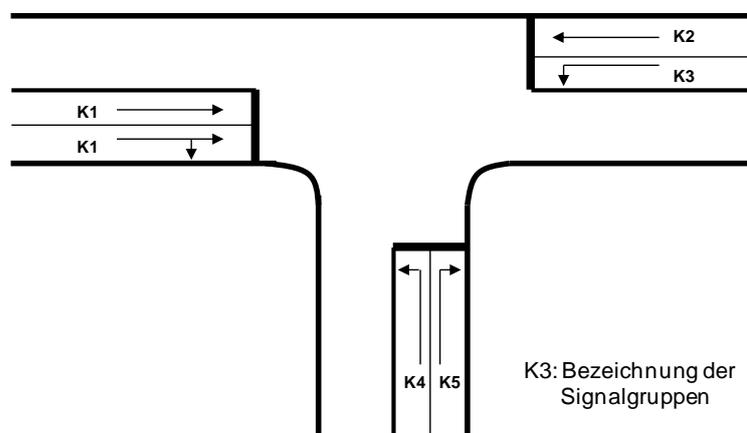


Abbildung 2-1: Modellknotenpunkt und die zugehörigen Signalgruppen

Die Verträglichkeiten und Zwischenzeiten zwischen den Signalgruppen sind für die Berechnung des Signalprogramms maßgebend. In Abbildung 2-2 wird die Verriegelungsmatrix, die die Verträglichkeiten aller Signalgruppen beschreibt, dargestellt. Die Verriegelungsmatrix weist im Allgemeinen einen symmetrischen Aufbau auf.

		beginnende Signalgruppen				
		K1	K2	K3	K4	K5
endende Signalgruppen	K1		+	-	x	x
	K2	+		+	x	+
	K3	-	+		x	+
	K4	x	x	x		+
	K5	x	+	+	+	

+ : verträglich
- : bedingt verträglich
x : nicht verträglich

Abbildung 2-2: Verriegelungsmatrix des Modellknotenpunkts

Das Ende der Freigabezeit einer Signalgruppe und der Beginn der Freigabezeit einer dazu nicht verträglichen Signalgruppe wird durch eine Zwischenzeit getrennt. Während der Zwischenzeit kann der endende Strom die gesamte Konflikfläche räumen und der beginnende Strom einfahren. Die Zwischenzeiten, die abhängig von den Fahrvorgängen und den Knotenpunktabmessungen sind, müssen für alle Kombinationen der nicht verträglichen Ströme berechnet werden. Für den Modellknotenpunkt kann sich die in Abbildung 2-3 dargestellte Matrix der Zwischenzeiten ergeben.

		beginnende Signalgruppen				
		K1	K2	K3	K4	K5
endende Signalgruppen	K1			5	4	5
	K2				5	
	K3	4			5	
	K4	4	5	4		
	K5	5				

Abbildung 2-3: Zwischenzeitenmatrix des Modellknotenpunkts

Wenn der Knotenpunktsentwurf erfolgt ist, werden die Signalgruppen sowie Zwischenzeiten der Signalgruppen nicht mehr als Variablen des Steuerungsprogramms betrachtet. Auf Basis dieser Grundgrößen werden die verschiedenen Variablen festgelegt und lässt sich die Signalsteuerung entwerfen.

2.1.2 Phaseneinteilung und Phasenübergänge

Unter Phaseneinteilung versteht man die Gruppierung verträglicher Verkehrsströme in Phasen und die Festlegung ihrer Reihenfolge. Die Ermittlung aller möglichen Phaseneinteilungen stellt ein kombinatorisches Problem dar [TULLY, 1977 UND MÖLLER, 1987]. Die Grundregel für die Auswahl der Phasen ist, dass verträgliche und bedingt verträgliche Signalgruppen in einer Phase liegen können und nicht verträgliche in getrennten Phasen liegen müssen. Außerdem soll jede Signalgruppe mindestens in einer Phase Freigabe erhalten. Um eine sinnvolle Phaseneinteilung festzulegen, muss ein weiteres Kriterium beachtet werden. Die Anzahl der Phasen und die damit verbundenen nicht

effektiv nutzbaren Zwischenzeiten sollen minimiert werden. Abbildung 2-4 zeigt eine mögliche Phaseneinteilung für den oben dargestellten Modellknotenpunkt.

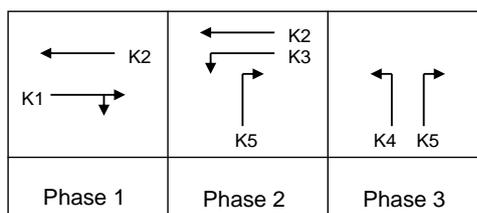


Abbildung 2-4: Mögliche Phaseneinteilung des Modellknotenpunkts

Wegen des Ab- und Einschaltens der Signalgruppen und den dadurch einzuhaltenden Zwischenzeiten existiert zwischen zwei Phasen ein Phasenübergang. Der Phasenübergang beginnt mit der zuerst endenden Signalgruppe der endenden Phase und endet mit der zuletzt beginnenden Signalgruppe der darauffolgenden Phase. Daher ist der Phasenübergang mindestens gleich der für den Phasenwechsel erforderlichen Zwischenzeit. Es kann auch zweckmäßig sein, in den Phasenübergang Randbedingungen für Freigabe- und Sperrzeit zu integrieren. In Abbildung 2-5 sind für den oben dargestellten Knotenpunkt die Phasenübergänge veranschaulicht.

2.1.3 Umlaufzeit

Während einer Umlaufzeit des Signalprogramms laufen die Phasen und die dazwischen liegenden Phasenübergänge einmal vollständig ab. Die für ein Signalprogramm mindestens erforderliche Umlaufzeit ergibt sich im Allgemeinen aus der Summe von maßgebenden Freigabezeiten und den dazwischen liegenden erforderlichen Zwischenzeiten.

$$tu_{erf} = \sum_{i=1}^p tg_i + \sum_{i=1}^p t_{z,i} \tag{2-1}$$

- mit
- tg_i = maßgebende Freigabezeit
 - p = Anzahl der Phasen
 - $t_{z,i}$ = maßgebende Zwischenzeit für einen Phasenwechsel

Die mindestens erforderliche Umlaufzeit berechnet sich gemäß RiLSA aufgrund eines deterministischen Modellansatzes. Die Anzahl der während einer Umlaufzeit einfahrenden Fahrzeuge soll gleich der Anzahl der während der Freigabezeit abgefertigten Fahrzeuge sein.

$$\frac{q_{maßg,i}}{3600} tu_{erf} = \frac{S}{3600} tg_i \tag{2-2}$$

- mit
- S = Sättigungsverkehrsstärke
 - $q_{maßg,i}$ = maßgebende Fahrstreifenverkehrsstärke der Phase i

Aus dem Zusammenhang von (Gl. 2-1) und (Gl. 2-2) ergibt sich die erforderliche Umlaufzeit.

$$tu_{erf} = \frac{\sum_{i=1}^p t_{z,i}}{\left(1 - \sum_{i=1}^p \frac{q_{maßg,i}}{S}\right)} \quad (2-3)$$

Aus der Koordinierung der Signalprogramme mehrerer Knotenpunkte können weitere Randbedingungen für die Umlaufzeit berücksichtigt werden. In den RiLSA werden auch minimale und maximale Umlaufzeiten als Richtwert definiert.

2.1.4 Freigabezeit

Wenn die Umlaufzeit festgelegt wurde, können die Freigabezeiten für die maßgebende Signalgruppe ermittelt werden. Die Berechnung der von der Umlaufzeit abhängigen erforderlichen Freigabezeiten erfolgt nach Gleichung (Gl. 2-4):

$$tg_i = \frac{q_{maßg,i}}{S} tu_{erf} \quad (2-4)$$

Falls eine gewählte Umlaufzeit länger als die erforderliche Umlaufzeit dauert, errechnet sich die Freigabezeit nach folgender Gleichung:

$$tg_i = \frac{tu_{gew} - \sum_{i=1}^p t_{z,i}}{\sum_{i=1}^p \frac{q_{maßg,i}}{S}} \quad (2-5)$$

In Abbildung 2-5 werden die Freigabezeiten aller Signalgruppen in einem Signalzeitenplan während eines Umlaufs für das obige Beispiel dargestellt.

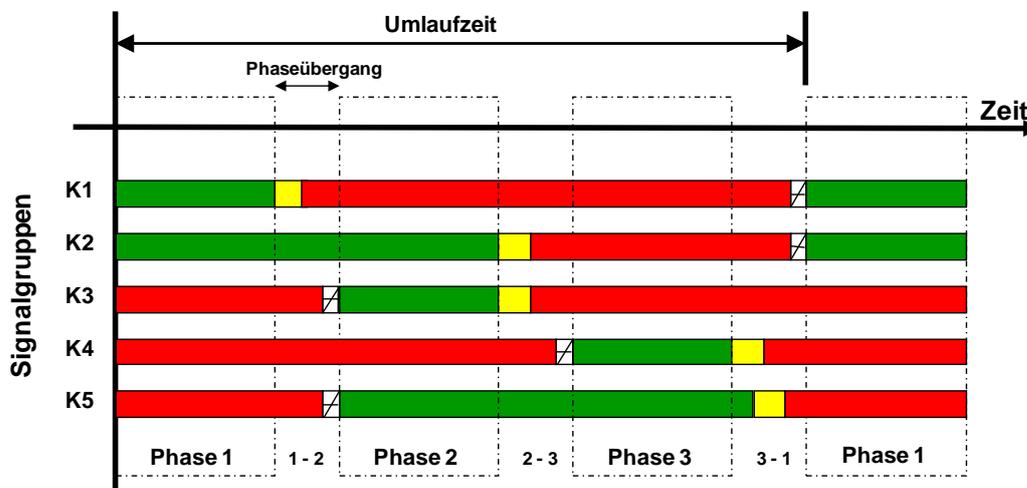


Abbildung 2-5: Signalprogramm des Modellknotenpunkts

2.1.5 Versatzzeit

Das Signalprogramm für einen einzelnen Knotenpunkt kann schon durch die oben beschriebenen Grundgrößen und Variablen der Signalsteuerung festgelegt werden. Für die netzweite Verkehrssteuerung müssen noch die Freigaben hintereinanderliegender Knotenpunkte aufeinander abgestimmt werden.

Für diese Aufgabe ist die Versatzzeit zuständig, die zeitliche Differenz des Beginns der Freigabezeiten zweier aufeinanderfolgender Signalgruppen auf einer Strecke oder Fahrbeziehung. Über die Versatzzeit können die Freigabezeiten benachbarter Signalanlagen koordiniert werden. Die Koordinierung der Signalanlagen beschäftigt sich mit der Erzeugung der „Grünen Welle“ für maßgebende Verkehrsströme, damit die Mehrzahl der Fahrzeuge unter Einhaltung einer bestimmten Geschwindigkeit mehrere Knotenpunkte ohne Halt passieren kann. Die Koordinierung im Netz dient hauptsächlich dazu, die Summe aller fahrzeugbezogenen Reisezeiten im System zu verringern und die Anzahl der Halte zu reduzieren, in gewissem Umfang die Kapazität zu erhöhen, den Fahrkomfort zu verbessern und die Umwelt durch Lärm und Schadstoffe weniger zu beeinträchtigen.

Die Koordinierung und damit die Optimierung der Versatzzeiten basiert auf der Progressionsgeschwindigkeit, der Geschwindigkeit eines Verkehrsstroms zwischen zu koordinierenden Signalanlagen und der Distanz zwischen zwei Haltelinien. Wenn der Stau vernachlässigt wird, ist die optimale Versatzzeit der Freigabezeiten zwischen beiden Knotenpunkten gleich der Reisezeit des Verkehrsstroms.

Die Koordinierung im Streckenzug, bzw. die Grüne Welle, kann in einem Zeit-Weg-Diagramm anschaulich dargestellt werden. Dieses Diagramm zeigt die Bewegung der Kraftfahrzeugströme in Form von sogenannten Grünbändern. Bei der Planung müssen die Belange des individuellen Kraftfahrzeugverkehrs (IV), der öffentlichen Verkehrsmittel (ÖV) sowie des Radverkehrs beachtet werden. Der Konflikt bei der Koordinierung von Hin- und Rückrichtung sowie der sich zwischen den sich kreuzenden Koordinierungsbändern ergebenden zusätzlichen Zwänge sind wesentliche Probleme bei Erstellung Grüner Wellen. Aufgrund dieser Berücksichtigung kann nur eine Kompromisslösung gefunden werden. Nicht alle Verkehrsströme können von der gewählten Koordinierung profitieren. Deshalb ist eine Realisierung eines Systemoptimums im Sinne einer Zielfunktion für die netzweite Signalsteuerung notwendig.

2.2 Steuerungsverfahren

Unter Steuerungsverfahren versteht man die Weise, in der die Elemente der Lichtsignalsteuerung eingesetzt werden, um die Steuerung des Verkehrsablaufs zu realisieren. Die bisher verwendeten Steuerungsverfahren können anhand ihrer Eigenschaften in insgesamt drei Kategorien eingeteilt werden. In Abbildung 2-6 ist die Kategorisierung

der drei Steuerungsverfahren über die spezifischen Eigenschaften mittels eines Einteilungsbaums grafisch veranschaulicht.

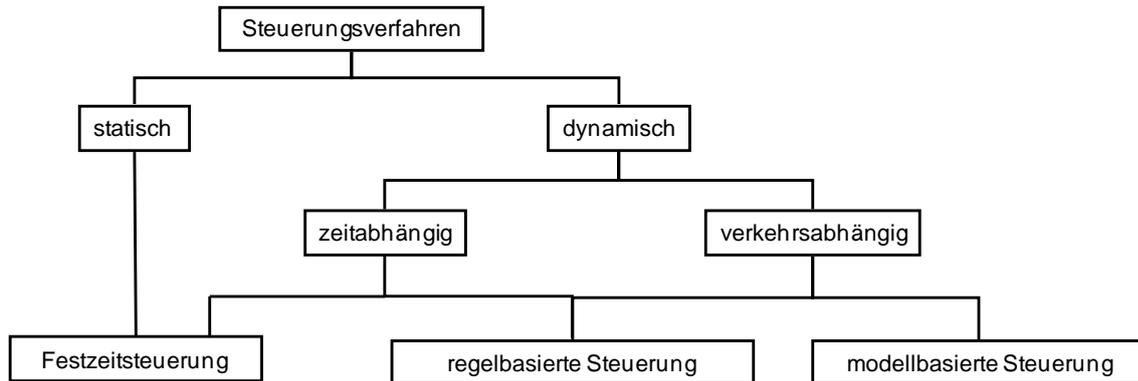


Abbildung 2-6: Einteilung der Steuerungsverfahren

Eine weitere gebräuchliche Einteilung kann anhand des Aggregationsgrades der gemessenen Verkehrskenngrößen erfolgen. Sogenannte makroskopische Verfahren beschäftigen sich auf Gebiets- bzw. Netzebene mit der Auswahl der Signalprogramme, der Umlaufzeit und der Versatzzzeit. Mikroskopische Verfahren hingegen beziehen sich in der Regel auf Einzelknoten. Es handelt sich um sekundweise Anpassung und Auswahl von einzelnen Phasen oder Signalgruppen. Anhand der Art der Steuerungsdefinition unterscheidet man das phasenorientierte und signalgruppenorientierte Verfahren. Für das obige einfache Beispiel wurde das phasenorientierte Verfahren verwendet, das auch in der Praxis größtenteils eingesetzt wird.

2.2.1 Festzeitsteuerung

Als erstes wird hier das älteste und einfachste LSA-Steuerungsverfahren aufgeführt. Die Festzeitsteuerung besteht darin, ein Signalprogramm mit festen Freigabezeiten zu verwenden. Innerhalb einer festgelegten zyklischen Umlaufzeit werden alle Verkehrsströme mindestens einmal mit vorgegebener Dauer freigegeben. Alle Signalgruppen des Signalprogramms wiederholen sich periodisch in gleicher Weise in jedem Umlauf und bleiben von äußeren Situationen - wie z.B. von der aktuellen Verkehrsnachfrage - unbeeinflusst.

Eine Festzeitsteuerung kann auch als dynamisches Steuerungsverfahren eingesetzt werden. Für verschiedene Tageszeiten werden oft eigene Signalprogramme entwickelt und geschaltet, um die unterschiedlichen Verkehrsbelastungen bestmöglich zu berücksichtigen. Spezifische Signalprogramme werden meist für den morgendlichen und abendlichen Spitzenverkehr, für den Tagesverkehr und ggf. für den Nachtverkehr entworfen. Mittels einer sogenannten Wochenautomatik findet zeitplanabhängig eine Umschaltung der einzelnen Signalprogramme statt. Außerdem kann die Umschaltung der Signalprogramme auch auf Basis verkehrsabhängiger Erhebungen durchgeführt werden, indem der Schaltzeitpunkt der Signalprogramme mit Hilfe von Messungen des Verkehrsablaufs variabel festgelegt wird.

Aufgrund der Einfachheit, Robustheit und geringen Bau- und Instandhaltungskosten sind Festzeitsteuerungen bis heute noch ein häufig verbreitetes Steuerungsverfahren. Nachteilig wirken sich Festzeitsteuerungen bei veränderlicher Belastung aus, da die zur Verfügung stehenden Kapazitäten dann nicht optimal genutzt werden können. Als Folge können Stauungen in den einzelnen Zufahrten entstehen. Außerdem ist eine besondere Berücksichtigung der öffentlichen Verkehrsmittel auf diese Weise nicht möglich. Wegen dieser Nachteile der Festzeitsteuerung werden seit mehreren Jahrzehnten verstärkt verkehrsabhängige Steuerungsverfahren eingesetzt.

2.2.2 Regelbasierte Steuerung

In Deutschland werden verkehrsabhängige Steuerungen hauptsächlich in Form von regelbasierten Steuerungen auf Knotenpunktebene eingesetzt. Das durch Fahrzeuge gesteuerte Verfahren benutzt meistens mikroskopische Verkehrskenngrößen, die direkt und kontinuierlich aus dem Verkehrsablauf gesammelt und in der Steuerungslogik verarbeitet werden. Deswegen werden diese Verfahren häufig auch mikroskopische Steuerung oder regelbasierte Steuerung genannt. In der Neufassung der RiLSA [FGSV, 2010] wurde der in der RiLSA Teilfortbeschreibung [FGSV, 2003] verwendete Begriff der messerwertbasierten Steuerung in regelbasierte Steuerung abgeändert [BOLTZE ET AL., 2007].

Anforderung und Bemessung sind zwei wichtige Elemente für die Logik-Entwicklung einer regelbasierten Steuerung. Die Anforderung erfasst die Kenngrößen der Anwesenheit und Belegung. Für die Bemessung stehen zwei Verfahren zur Verfügung: Zum einen die Zeitlücken-Bemessung und zum anderen die Belegungs-Bemessung. Bei der Zeitlücken-Bemessung werden die Zeitlücken zwischen zwei hintereinander fahrenden Fahrzeugen innerhalb einer Zufahrt gemessen. Wenn der Wert der Zeitlücke über einen Schwellenwert liegt und die Anforderung der folgenden Phase erfüllt ist, wird die Freigabezeit nach einer Minimaldauer abgebrochen und der entsprechende Phasenübergang geschaltet. Andernfalls verlängert sie sich bis zu einer festgelegten Maximaldauer. Die Alternative zur Zeitlücken-Bemessung ist die Belegungs-Bemessung, die bei höherem Schwerverkehrsanteil oder wechselnder Geschwindigkeit verwendet wird.

Eine regelbasierte Steuerung bietet die Möglichkeit, auf kurzfristige stochastische Schwankungen schnell zu reagieren und sich direkt anzupassen. Bei schwacher bis mittlerer Belastung erzielt dieses Verfahren ein sehr zufriedenstellendes Ergebnis [KURZAK, 1974].

Für eine regelbasierte Steuerung wird kein Verkehrsmodell verwendet, eine Modellierung der verkehrlichen Wirkungen und ihre Bewertung mit Hilfe eines Performance Index (PI) erfolgt ebenfalls nicht. Aus diesem Grund beschränkt sich die Wirkungsweise der Verfahren auf die Reaktion aktueller Anforderungen [FRIEDRICH, 1999]. Außerdem besitzt die regelbasierte Steuerung kein Vorteil gegenüber einer Festzeitsteuerung bei hoher Auslastung [BRETHERTON ET AL., 1982; BRILON, 1994]. Bei hoher Dichte von beschleunigten ÖV-Fahrzeugen sind die regelbasierten Steuerungen nicht so

leistungsfähig, so dass sich für den Individualverkehr ggf. vermeidbare Verlustzeiten ergeben [FRIEDRICH, 1999].

Aufgrund dieser Nachteile der regelbasierten Steuerungen und insbesondere wegen des zuvor beschriebenen Festzeitsteuerungseffekts wurde ab Mitte der 60er Jahre eine neue Generation von Steuerungsverfahren entwickelt, die im deutschsprachigen Raum als 'modellbasierte Steuerung' bezeichnet werden.

2.2.3 Modellbasierte Steuerung

Im Vergleich zu einer regelbasierten Steuerung arbeiten modellbasierte Steuerungen mit einem Verkehrsflussmodell, das den Verkehrsablauf simuliert. Der Verkehrsablauf ergibt sich aus zunächst vorgegebenen Steuerungsgrößen (Freigabezeiten und Umlaufzeit). Der simulierte Verkehrsablauf wird durch Kenngrößen wie Zufluss- und Abflussprofile beschrieben. Dies sind die Eingangsgrößen für das Verkehrswirkungsmodell, mit dem es möglich ist, verschiedene Steuerungsalternativen hinsichtlich ihrer Wirkungsgröße, z.B. Anzahl der Halte und Wartezeiten auf dem gesamten Verkehrsnetz, zu überprüfen. Eine iterative Optimierung einer festgelegten Zielfunktion wird mit Hilfe von verschiedenen Optimierungsmethoden durchgeführt. Abbildung 2-7 zeigt den prinzipiellen Aufbau von modellbasierten Steuerungsverfahren.

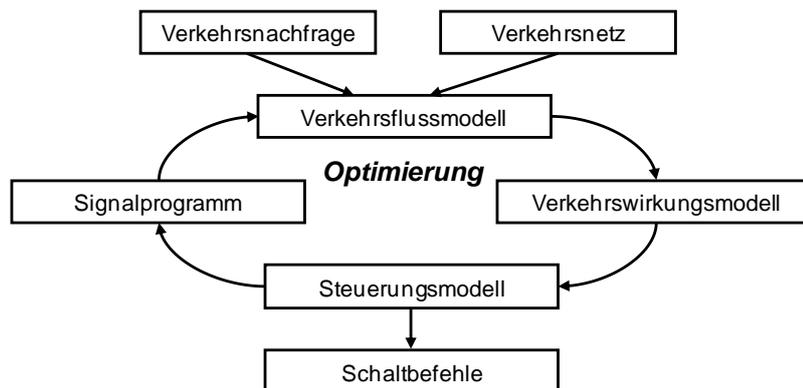


Abbildung 2-7: Aufbau der modellbasierte Steuerungen [RiLSA, 2010]

Dieser Ansatz besitzt gegenüber den regelbasierten Verfahren den Vorteil, dass alle Anforderungen der verschiedenen Verkehrsströme bei der Optimierung berücksichtigt werden können, wodurch eine ausgeglichene Bedienung erfolgen kann. Deswegen kann die modellbasierte Steuerung eine optimale Anpassung an die aktuelle Verkehrssituation im gesamten Netz erreichen. Weiterhin können einzelne wichtige Verkehrsströme durch eine Gewichtung in der Zielfunktion bevorrechtigt werden, damit die Belange der Stadtplanung Berücksichtigung finden.

Während in Deutschland bis 2002 fast ausschließlich regelbasierte Steuerungsverfahren zum Einsatz kamen [KELLER ET AL., 2002], wurden in Großbritannien diese Verfahren schon seit Anfang der 70er Jahre eingesetzt, die auf die wechselnde Verkehrsnachfrage anpassungsfähig (adaptiv) reagieren. In den letzten sechs Jahren sind aber auch in Deutschland modellbasierte Verfahren (BALANCE, MOTION) zum Einsatz ge-

kommen. In der Neufassung der RiLSA [2010] werden bei den verkehrsabhängigen Steuerungsverfahren regelbasierte und modellbasierte Steuerungen unterschieden. In der vorliegenden Arbeit werden in Anlehnung an die Fachliteratur 'modellbasierte Steuerungsverfahren' auch 'adaptive Steuerungsverfahren' genannt.

2.3 Bekannte adaptive Steuerungsverfahren

Die Entwicklung der adaptiven Steuerungsverfahren wurde in den letzten 40 Jahren laufend vorangetrieben. Vom im Jahr 1969 entwickelten TRANSYT [ROBERTSON, 1969] bis zur neuen Version des MOTION 4 [MÜCK, 2007] entstand eine Reihe von modellbasierten Steuerungsverfahren. Sie sind in Abbildung 2-8 chronologisch aufgeführt.

Verfahren (Entwickler)	seit	Systemarchitektur	Installationen
TRANSYT (Robertson)	1969	Offline Planungs-Tool	USA, England
OPAC (Gartner)	1979	Online, zentral / dezentral	Arlington, Tucson (USA)
TASS	1980	Online, zentral	weltweit
SCOOT (Hunt, et al.)	1981	Online, zentral	weltweit
SCATS (Lowrie)	1982	Online, zentral / dezentral	weltweit, Schwerpunkt Asien
PRODYN (Henry et al.)	1983	Online, dezentral	Frankreich
UTOPIA (Donati et al.)	1987	Online, zentral / dezentral	Italien, Europa
SIGMA (Garben et al.)	1988	Offline Planungs-Tool	Deutschland
MONARC	1990	Online, zentral	USA
ADONIS (Busch et al.)	1990	Online, dezentral	Deutschland (F&E-Prototyp)
MOTION (Busch et al.)	1993	Online, zentral / dezentral	Deutschland, Europa
RHODES (Head et al.)	1994	Online, zentral / dezentral	USA
BALANCE (Friedrich et al.)	1994	Online, zentral / dezentral	Deutschland, England
TRACON, TRASSIS	1995	Online	Deutschland
TUC (Diakaki et al.)	1999	Online, zentral / dezentral	Griechenland
ADAPT2 (Zackor et al.)	2001	Online, dezentral	Deutschland (F&E-Prototyp)
LAATCS	2001	Online, zentral / dezent	USA
LISA / ADANTES	2002	Offline	Deutschland (F&E-Prototyp)
GALOP (Braun, Weichenmeier)	2005	Offline (Weiterentwicklung zu online)	Deutschland (F&E-Prototyp)
MOTION 4 (Mück)	2007	Online, zentral / dezentral	Deutschland, Europa

Abbildung 2-8: Überblick der modellbasierten LSA-Steuerungsverfahren [BUSCH, 2011]

Im Folgenden werden die Verfahren TRANSYT als Beispiel für Offline-Modelle und SCOOT, UTOPIA, MOTION als Beispiel für Online-Modelle näher beschrieben.

2.3.1 Offline-Modelle

Ein Offline-Modell besteht aus einem Verkehrsmodell und einem Optimierungsmodell. Das Verkehrsmodell simuliert den Verkehrsablauf an den signalisierten Knotenpunkten und den dazwischenliegenden Verbindungsstrecken und berücksichtigt aufgrund des offline Betriebs keine aktuelle Schwankung der Verkehrssituation.

TRANSYT

TRANSYT (Traffic Network Study Tool) wurde im Jahre 1969 von Robertson entwickelt und in seiner ersten Version im 'Glasgow Experiment' eingesetzt. Es ist das erste und auch das bekannteste adaptive offline Steuerungsverfahren. Danach entstand eine Reihe von Überarbeitungen, die das Verkehrsmodell, die Schnittstelle und Optimierungsmethoden betreffen. Ein Cell Transmission Model [ZHANG, 2008], [ALMASRI ET AL., 2005] wird in der neuesten Version TRANSYT 13.0.3 [BINNING, 2008] angewendet. In der U.S. Version TRANSYT-7F [HALE, 2005] kommt ein genetischer Algorithmus als Optimierungsmethode zum Einsatz.

TRANSYT koordiniert festzeitgesteuerte Lichtsignalanlagen in einem Straßennetz, das durch ein deterministisches makroskopisches Verkehrsmodell simuliert wird. Es ermittelt die Wartezeit und die Anzahl der Halte mit Hilfe von einem Untermodell der Pulkauflösung für jeden Optimierungsschritt. Abbildung 2-9 zeigt den Ablauf dieses Verfahrens.

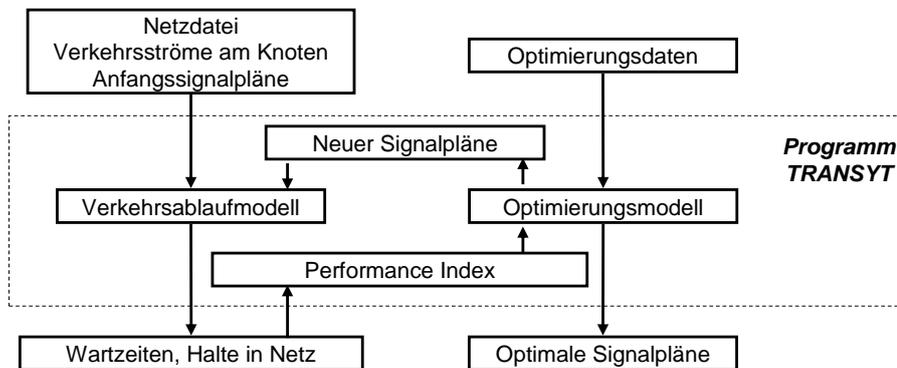


Abbildung 2-9: Ablaufdiagramm des Steuerungsverfahrens TRANSYT

TRANSYT setzt voraus, dass alle LSA mit einheitlicher und fester Umlaufzeit betrieben werden. Die Freigabezeitanteil und die Versatzeiten zwischen den LZA werden entsprechend einem Performance Index (PI, siehe Gl. 2-6) optimiert. Die Minimierung des PI erfolgt mit Hilfe eines sogenannten Hill Climbing Algorithmus.

$$PI = \sum_{i=1}^n (t_w + k \cdot H_i) \quad (2-6)$$

mit t_w : mittlere Wartezeit
 H_i : mittlere Zahl der Halte in Zufahrt i
 k : Gewichtungsfaktor

2.3.2 Online-Modelle

Im Gegensatz zu den Offline-Modellen können Online-Modelle auf die aktuelle und veränderbare Verkehrssituation reagieren. Auf Basis aktueller Messwerte werden online Verkehrsmodellierung und Optimierung durchgeführt. Es werden in den folgenden Abschnitten die Steuerungsverfahren beschrieben, die auf der Welt sehr verbreitet sind

und jeweils über besondere Ausprägungen verfügen. Das Verfahren SCOOT steht für ein zentrales System, die Verfahren SCATS, UTOPIA und MOTION für eine verteilte Systemarchitektur. SCATS verwendet kein Verkehrsmodell. Das Verfahren UTOPIA bietet zwei verschiedene Steuerungsstrategien in Abhängigkeit von lokalen Anforderungen und Datenerfassung. MOTION ist das erste im deutschsprachigen Raum entwickelte adaptive Netzsteuerungsverfahren, das mittlerweile weltweit zum Einsatz kommt.

SCOOT

SCOOT (Split Cycle Offset Optimization Technique) [HUNT ET AL., 1981] wurde vom Transport and Road Research Laboratory (TRL) im Jahr 1973 entwickelt. Mittlerweile beteiligen sich Siemens und Peek Traffic Ltd an der Weiterentwicklung des Systems. In der laufenden Version wurde auch die besondere Berücksichtigung des ÖPNV, Stauererkennung und Datenbank ASTRID/INGRID¹ in dem System integriert [CHILUKURI ET AL. 2004].

Der Baustein von SCOOT stammt aus dem Offline-Modell TRANSYT. Auf Basis einer Real-Time-Erfassung der Verkehrsdaten optimierte SCOOT die Umlaufzeit (Cycle Length), Freigabezeitaufteilung (Split) und Versatzzeit (Offset). SCOOT hat zum Ziel, eine Zielfunktion aus Wartezeit und Anzahl der Halte² im Netz zu minimieren und damit die Ausbreitung von Stauungen über kritische Punkte hinaus zu vermeiden.

Die Verkehrsdaten werden mittels Induktionsschleifen in 4s-Intervallen an den Zufahrten erfasst und zu mesoskopischen Verkehrsflussprofilen oder sogenannten 'Cyclic Flow Profiles' aufbereitet. Die Daten aus vorangegangenen Umläufen werden auch berücksichtigt. Aus diesem Modell (Abbildung 2-10) können Wartezeiten, Anzahl der Halte und Staulängen abgeleitet werden, die als Ausgangsdaten für die Ermittlung der Zielgrößen dienen.

Voraussetzung für die Optimierung ist die Festlegung der Phaseneinteilung für jeden Knotenpunkt. Die Optimierung wird in kleinen, aber häufigen Schritten durchgeführt. Wenige Sekunden vor dem Ende jeder Freigabezeit wird die optimale Freigabezeiteinteilung knotenpunktbezogen in Abhängigkeit von der größten auftretenden Auslastung neu bestimmt. Die Versatzzeitoptimierung erfolgt einmal pro Umlaufzeit unter Verwendung der aus dem Verkehrsmodell abgeleiteten Zielfunktion und wird sofort umgesetzt. Die günstigste Umlaufzeit wird maximal alle 2,5 Minuten entwickelt und die bestehende Umlaufzeit wird in Schritten von 4 bis 16 Sekunden zum berechneten Optimalwert hin verändert.

¹ ASTRID steht für **A**utomatic **S**coot **T**raffic **I**nformation **D**atabase; INGRID steht für **I**Nte**G**rated **I**ncident **D**etection

² Die Zielfunktion ist ähnlich wie (2-6) von TRANSYT

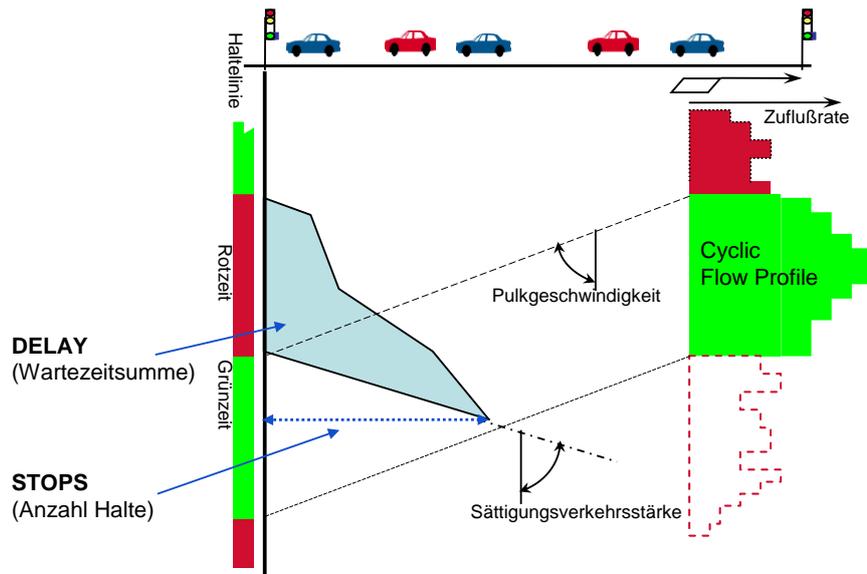


Abbildung 2-10: Prinzipien des Verkehrsmodells SCOOT [SCOOT, 2008]

SCOOT war das erste in der Praxis erfolgreich eingesetzte adaptive Verfahren und schaffte auch den Durchbruch der adaptiven Steuerungsverfahren. In zahlreichen Feldversuchen wurde seine Überlegenheit gegenüber Festzeitsteuerungen und regelbasierten Verfahren bewiesen [STEVANOVIC ET AL., 2007]. Bis heute ist SCOOT in mehr als 100 Städten im Einsatz und ist das am weitesten verbreitete adaptive Steuerungsverfahren.

Als wichtigster Kritikpunkt an diesem Verfahren ist anzumerken, dass die Wirkungsanalyse ausschließlich für die Optimierung der Versatzeiten verwendet wird. Umlaufzeit und Freigabezeiteinteilung werden auf Basis der Auslastung ermittelt. Weiterhin kann das Verfahren aufgrund der Veränderung der Variablen in kleinen Schritten keine schnelle Reaktion auf stochastische Schwankungen leisten.

UTOPIA/SPOT

Das Steuerungssystem UTOPIA (Urban Traffic Optimization by Integrated Automation) /SPOT (System for Priority and Optimisation of Traffic) [MAURO ET AL., 1989,1990] wurde Anfang der 1980er in Turin erstmals erprobt und danach weitentwickelt. Seither verbreitet sich das Verfahren in Italien und wird in mehr als 30 Städte in Europa eingesetzt.

Wie die neueren Steuerungsverfahren baut UTOPIA/SPOT auf einer hierarchisch-dezentralisierten Systemarchitektur auf. Zwei Ebenen kommen zum Einsatz:

1. Auf der Netzebene (UTOPIA) wird ein makroskopisches Verkehrsmodell verwendet, in dem anhand einer geschätzten Herkunft-Zielbeziehung eine netzweite Wirkungsanalyse durchgeführt wird. Die zeitliche Auflösung beträgt 5 Minuten und der Prognosezeitraum liegt bei 15 Minuten.
2. Auf der lokalen Ebene (SPOT) wird ein mikroskopisches Verkehrsmodell mit einer zeitlichen Auflösung von 3 Sekunden und einem Prognose-Horizont von 120 Se-

kunden verwendet. Jede SPOT-Einheit modelliert sämtliche Zu- und Ausfahrten. Der Serviceprozess wird durch eine konstante Sättigungsverkehrsstärke dargestellt. Für die Prognose der Fahrzeugankünfte werden die Verkehrsabflüsse aus den benachbarten Knotenpunkten verwendet. Auf diese Weise findet eine Versatzzeitoptimierung zwischen benachbarten Kreuzungen auf lokaler Ebene statt.

Zum Einsatz kommen für UTOPIA/SPOT zwei Steuerungsstrategien, die volle adaptive Steuerung und die Steuerung durch Signalprogramm-Wahl:

- Bei der vollen adaptiven Steuerung wird ein Referenzplan, der die vorgeschlagene Umlaufzeit und Freigabezeiteinteilung enthält, auf der Netzebene festgelegt. Auf der lokalen Ebene wird die Steuerung im gegebenen Rahmen über einen gleitenden Horizont fortgeschrieben. Mit Hilfe eines 'semi-heuristischen'³ Verfahrens wird die im Wirkungsmodell berechnete Zielfunktion⁴ optimiert, um das endgültige Signalprogramm zu bestimmen.
- Bei der Steuerung durch Signalprogramm-Auswahl werden Signalprogramme auf der Netzebene anhand der aktuellen Verkehrssituation aus der vorgegebenen Programmbibliothek gewählt. Diese Auswahl kann durch Befehle des Benutzers, einer Zeitplanung oder durch dynamische Erkennung des Szenarios erfolgen. Auf der lokalen Ebene wird die Steuerung auf Basis des gewählten Signalprogramms angepasst.

Hinsichtlich der Systemarchitektur zeigt UTOPIA/SPOT die Tendenz der Entwicklung einer adaptiven Netzsteuerung. Die makroskopische Netzsteuerung dient der Bestimmung des Rahmensignalprogramms. Die mikroskopische Knotenpunktsteuerung beschäftigt sich mit der Reaktion auf schnelle Ereignisse. Dadurch dass die Prognose der Fahrzeugankünfte durch die Information der benachbarten LSA ermittelt wird, erfolgt eine dezentrale Versatzzeit-Koordinierung im Netz. Dies setzt voraus, dass jeder Knotenpunkt durch SPOT gesteuert werden muss. Daher sind spezielle Infrastrukturen erforderlich [MAIER, 2007]. Weiterhin ist diese Versatzzeit-Optimierung sehr schwer kontrollierbar.

MOTION

Das Verfahren MOTION (Methode zur Optimierung der Lichtsignalsteuerung in online gesteuerten Netzen) [BUSCH UND KRUSE, 1993; BUSCH UND KRUSE, 2002] wurde Anfang der 90er Jahre erstmals vorgestellt und in Köln-Deutz eingesetzt. Seither wird das Verfahren weiterentwickelt und von der Siemens AG vertrieben.

³ Branch-and-Bound Algorithmus

⁴ Die Zielfunktion ist eine Linearkombination von Wartezeit, Staulänge, Abweichung des Signalplans von Referenzplan und Abweichung des Signalplans vom Signalplan des vorigen Intervalls.

MOTION nutzt ähnlich wie UTOPIA eine hierarchisch-dezentralisierte Systemarchitektur. Anstatt von zwei Ebenen wie in UTOPIA kommen drei Ebenen in MOTION zum Einsatz. Diese Systemarchitektur wird in Abbildung 2-11 dargestellt

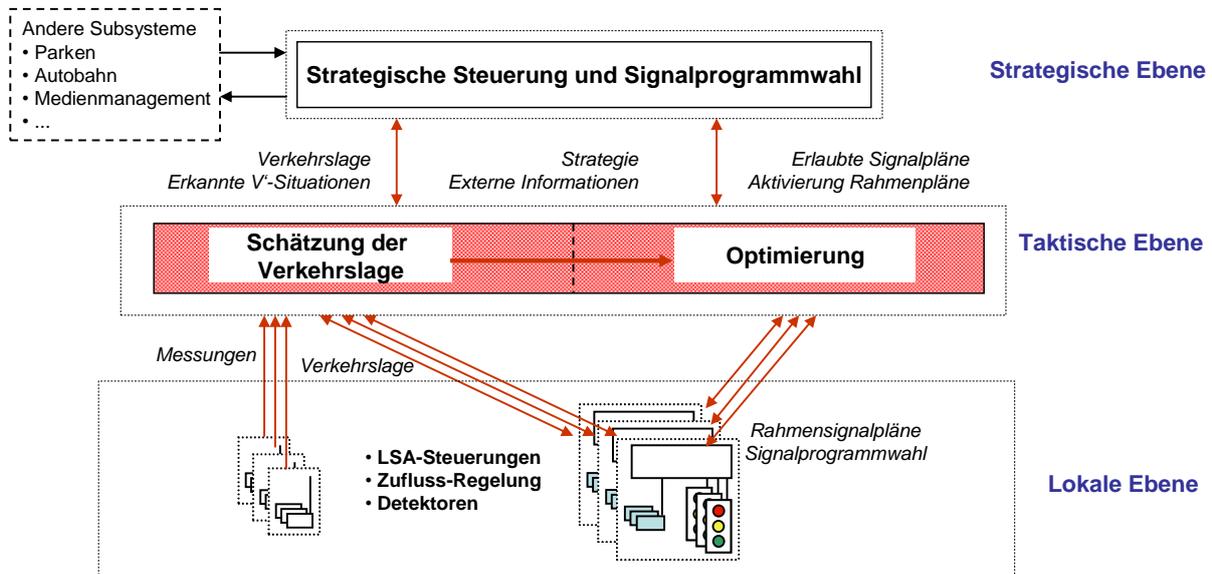


Abbildung 2-11: Die Systemarchitektur des Verfahrens MOTION [Mück, 2008]

Die strategische Steuerung dient der Verwirklichung der verkehrs- und umweltpolitischen Zielvorstellungen. Auf der taktischen Ebene wird die Verkehrslage mit Hilfe von Messungen ermittelt. Daraus können die Kenngrößen für die Wirkungsanalyse berechnet werden. Der Optimierungsprozess deckt den mittel- bis langfristigen Bereich (5-15 min) der verkehrsabhängigen Steuerung ab. MOTION gibt dabei den Knotenpunkten im Netz Rahmensignalprogramme vor, innerhalb derer die Flexibilität erhalten bleibt. Die Steuerung auf der lokalen Ebene betrifft vor allem die Reaktion auf kurzfristige Anforderungen im Verkehrsablauf. Anhand der Rahmensignalpläne wird die Phasendauer in dem erlaubten Bereich verkürzt bzw. verlängert.

Zur Einschätzung der Verkehrslage wurde in der älteren Version ein makroskopisches Modell verwendet. Auf Basis der Entropie wird gebietsweise eine Quell-Ziel-Matrix für ein 15 Minuten-Intervall geschätzt, mit der anschließend eine Verkehrsumlegung durchgeführt wird. Die daraus resultierenden Netzströme bzw. Routenbelastungen dienen der Optimierung der Signalprogramme und der Koordinierung. In der neuen Version MOTION MX [Mück, 2008] kommt eine lokale Quell-Ziel-Schätzung zum Einsatz. Mit Hilfe der Messwert-Korrelationen benachbarter Detektoren erfolgt eine dynamische Schätzung der Abbiegeraten in 1- bis 2-Minuten-Intervallen. Neben der Ermittlung der Verkehrsstärken gehören auch die Ermittlung der Staulängen [Mück, 2002] und die Berechnung von Warte- und Reisezeiten zu den Aufgaben der Verkehrslageschätzung. Diese Kenngrößen werden für Wirkungsanalysen verwendet.

Bei MOTION ist die Optimierung der LSA in drei Phasen aufgeteilt:

1. Die günstigsten Freigabezeiteinteilungen werden in einem ersten Schritt anhand der erlaubten Umlaufzeiten und Phaseneinteilungen ermittelt. Die Reservefrei-

gabezeiten werden auf Basis der Verkehrsstärken, der vorgegebenen Wunschauslastungen aller Signalgruppen und der berechneten Staulängen auf die Signalgruppen verteilt.

2. Für die Auswahl der Umlaufzeit wird für jede Umlaufzeit und Phasenfolge jeweils nur die Signalgruppe, deren Zielfunktionswert den höchsten Wert aufweist, herangezogen. Für die neue ausgewählte Umlaufzeit wird geprüft, ob die aktuelle verkehrliche Situation einen Wechsel in die neue Umlaufzeit rechtfertigt oder ob die bisherige Umlaufzeit beibehalten werden kann.
3. Anschließend wird die Koordinierung über die Gewichtung von Wartezeit und Halten optimiert. In der aktuellen Version stehen intern zur Zeit zwei Verfahren zur Verfügung:
 - Gleichzeitige Optimierung aller LSA unter Verwendung genetischer Algorithmen auf Basis eines mesoskopischen Verkehrsmodells.
 - Optimierung der Versatzeiten mit Hilfe eines deterministischen Optimierungsverfahrens auf Basis eines pulkorientierten Verkehrsmodells.

MOTION ist im Moment in 17 Städten in Betrieb bzw. im Aufbau. Die bisherigen Untersuchungen in Piräus, Kopenhagen und Graz [KRUSE, 2003] zeigen zufriedenstellende Ergebnisse. Hinsichtlich der Optimierungsmethode und dem Verkehrsmodell entspricht MOTION dem neuesten Stand der Technik. Nur der Einsatz von genetischen Algorithmen zur Versatzeitoptimierung in Münster [MÜCK, 2008] hat die Erwartung, dass eine klare Optimierung der dynamischen grünen Welle erzeugt wird, nicht erfüllt. MOTION ist nicht in der Lage, alle Optimierungsgrößen durch einen geschlossenen mathematischen Ansatz zu berechnen, da Umlaufzeit und Freigabezeiten von der Versatzeit getrennt werden. Weiterhin benötigt MOTION auch spezielle Siemens-Infrastrukturen.

2.4 Zwischenfazit

Im vorangegangenen Kapitel wurden zunächst die Grundlagen der Lichtsignalsteuerung und eine Übersicht der Steuerungsverfahren behandelt. Danach wurden die bekannten und etablierten Netzsteuerungsverfahren vorgestellt.

Die in Deutschland bzw. Europa etablierten Netzsteuerungsverfahren bauen auf einer hierarchisch-dezentralisierten Systemarchitektur auf. Auf der Netzebene werden mit Hilfe eines makroskopischen Verkehrsmodells die Rahmensignalprogramme optimiert. Die Steuerung auf der lokalen Ebene beschäftigt sich mit der Reaktion auf kurzfristige Ereignisse auf Basis der Rahmensignalprogramme.

Für die Optimierung der Rahmensignalprogramme werden verschiedene Methoden eingesetzt. Heuristische Methoden z. B. „genetische Algorithmen“ kommen in den letzten Jahren zum Einsatz. Künstliche Neuronale Netze (KNN) gehören zu den bedeutendsten heuristischen Verfahren und sind bislang noch nicht für netzweite Optimierung

der Lichtsignalprogramme verwendet worden. Andererseits ist der Zusammenhang zwischen dynamischer Verkehrsumlegung und netzweiter Lichtsignalsteuerung noch nicht untersucht.

Diese Arbeit hat das Ziel, ein neues auf KNN basiertes Lichtsignalsteuerungsverfahren zu entwickeln. Das neue Verfahren bietet die Möglichkeit, das Zusammenspiel der Verkehrsumlegung und der Signalsteuerung zu analysieren. Alle Signalsteuerungsparameter werden durch einen geschlossenen mathematischen Ansatz optimiert.

3 Rekurrente Neuronale Netze

Das in dieser Arbeit vorgestellte Verkehrsumlegungsmodell und das Verkehrssteuerungsmodell basieren auf einem spezifischen konnektionistischen Modell. Die nachfolgenden Ausführungen des Kapitels 3 basieren unter anderem auf den Veröffentlichungen von REY [2008] und PINEDA [1988, 1989, 1995]. Der Inhalt wird hier ausgerichtet auf die vorliegende Problemstellung in gekürzter Form zusammengefasst. Zunächst soll die Theorie der Künstlichen Neuronalen Netze (KNN) vorgestellt werden. Rekurrente Neuronale Netze (RNN = recurrent neural network) als ein spezielles Modell der KNN werden ausführlich erklärt. Ein besonderer Schwerpunkt wird auf einem speziellen Lernverfahren liegen, der sogenannten „rekurrenten Backpropagation“.

3.1 Das menschliche Gehirn und die Simulation

Das menschliche Gehirn besteht aus etwa 100 Milliarden in unterschiedlicher Weise untereinander verbundenen Neuronen. Wenn ein Neuron aktiviert wird, sendet es den benachbarten Neuronen einen elektrochemischen Reiz. Neuronen können auch gleichzeitig untereinander Informationen austauschen. Der erste Versuch, dieses menschliche Gehirn mit mathematischen Modellen zu simulieren, liegt 68 Jahre zurück [McCULLOCH, 1943]. Die künstlichen Neuronen eines KNNs aktivieren sich untereinander – ähnlich ihrem biologischen Vorbild - mit Hilfe von gerichteten Verbindungen. In Abbildung 3-1 sind das menschliche Neuron und eine KNN veranschaulicht.

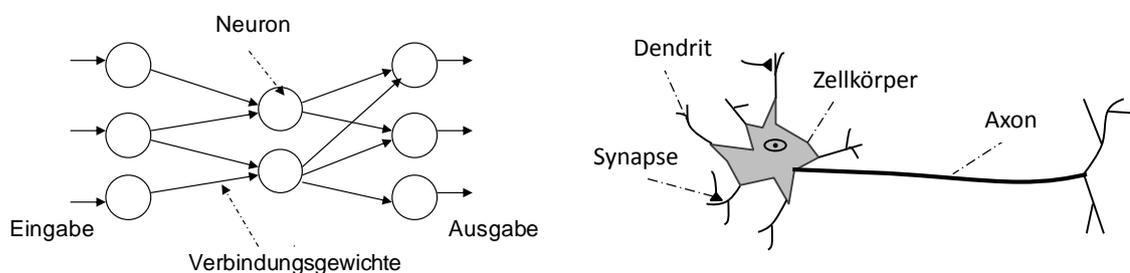


Abbildung 3-1: Künstliche Neuronale Netze und biologisches Vorbild

Um die neue Information zu speichern und auf wiederholte Information zu reagieren, wird das menschliche Gehirn trainiert. Selbiges Training benötigt auch ein neuronales Netz. Eine Vielzahl von Beispielen werden ihm vorgestellt, damit das Netz sich selbst

die nötigen internen Kenntnisse aufbauen kann, um die geforderte Aufgabe zu erfüllen. Mathematisch gesehen bedeutet ein Netz zu trainieren, ihm eine Menge Input und/oder Output (X/Y) Paare zu liefern und dafür zu sorgen, dass das Netz automatisch die Werte der Verbindungsgewichte W , in denen das Wissen des Neuronalen Netzes gespeichert ist, herausfindet. Die Grundidee des neuronalen Lernens ist, dass die Verbindungen der Elemente durch jedes gelieferte Beispiel verstärkt und optimiert werden und gleichzeitig aktiv sind.

Die Lernverfahren sind grundsätzlich in zwei Gruppen unterteilbar: Überwachtes Lernen und unüberwachtes Lernen. Beim überwachten Lernen werden dem neuronalen Netz sowohl Eingabemuster als auch Ausgabemuster zur Verfügung gestellt. Durch Vergleich von Soll- und Ist-Ausgabe, die das Netz in seinem aktuellen Zustand produziert, erfolgt eine Änderung der Netzkonfiguration. Bekannte Verfahren des überwachten Lernens sind die „Delta-Regel“ und das später ausführlich vorgestellte „Backpropagation“ (BP). Das unüberwachte Lernen findet nur statt, wenn das Eingabemuster bereits vorhanden ist. Es geht in diesem Fall darum, die gegebenen Beispiele zu organisieren und einzuordnen und dabei Beziehungen und Regelmäßigkeiten zu entdecken. „Kompetitives Lernen“ ist eine häufig eingesetzte unüberwachte Lernregel.

Neuronale Netze lassen sich nach unterschiedlichen Kriterien klassifizieren. Die Aspekte der Klassifizierung können folgende Fragen betreffen [REY ET AL, 2008]:

- Existenz von Hidden-Units vorhanden oder nicht
- Überwachtes oder unüberwachtes Lernen
- Existenz von Rückkopplungen von Neuronen zu anderen Neuronen derselben oder einer vorangegangenen Schicht
- Anwendungsbereich des Netzwerkes: Assoziativspeicher, Mustererkennung, Vorhersage, allgemeine Abbildung und Optimierung

Die dritte Frage bezieht sich auf die Architektur der Neuronalen Netze, dazu können die Netzwerke in zwei Klasse gruppiert werden: Feedforward-Netzwerke und Rekurrente Neuronale Netzwerke. Feedforward-Netzwerke, mit einer einzigen Schicht oder mehreren Schichten von Neuronen, verarbeiten die Eingabesignale vorwärts gerichtet. In Abbildung 3-1 ist beispielsweise ein mehrschichtiges neuronales Feedforward-Netz vorgestellt.

3.2 Grundlage des Rekurrenten Neuronalen Netzes

Im Gegensatz zur Feedforward-Netzen haben Rekurrente Neuronale Netze Feedback-Verbindung, wobei die Ausgaben einiger Neuronen an die Eingabe anderer Neuronen zurückgereicht werden, um wiederholte Berechnungen der Signale auszuführen. Unter anderem sind auch Self-Feedback-Verbindungen erlaubt. Abbildung 3-2 zeigt ein allgemeines RNN. Wegen dieser Rückkopplung eines Neurons sind die RNN im Allgemeinen nicht gedächtnislos. Sie besitzen einen internen Zustand und weisen ein dy-

namisches Verhalten auf. RNN können Abbildungen aufweisen, bei denen es um Funktion von Zeit und/oder Raum handelt.

Im Vergleich zu den konventionellen Feedforward-Netzwerken werden auch wesentliche Veränderungen hinsichtlich der Arbeitsweise und des Lernprozesses bewirkt. Die Schwierigkeit der RNN sind der Stabilitätsnachweis und die Konstruktion geeigneter Lernverfahren. Andererseits bringt diese Komplexität der Netzwerke auch Vorteile. Sie bieten eine gesteigerte Rechenleistung. Damit werden die Anwendungsbereiche erweitert, z.B. auf die Simulation und nichtlineare Steuerungssysteme.

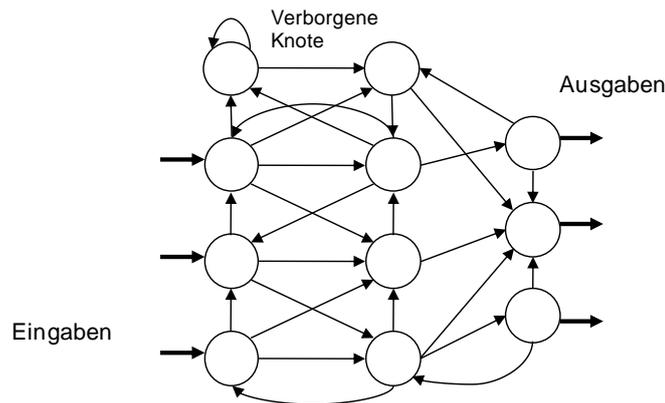


Abbildung 3-2: Ein allgemeines rekurrentes neuronales Netz

Das Verhalten von RNN wurde von PINEDA [1988, 1989,1995] beschrieben und von mehreren anderen Forschern untersucht, unter anderem ZIPSER [1989], WILLIAMS ET AL. [1990]. Gleichzeitig hat die Forschung die Übertragung des klassischen Lernverfahrens Backpropagation auf rekurrente Netze realisiert. Dadurch werden die Einsatzmöglichkeiten der RNN erweitert. Bis heute ist das Interesse an den RNN groß. Im Bereich der Verkehrstechnik werden RNN für Optimierung und Prognose seit den 90er Jahren eingesetzt, z.B. für die dynamische Verkehrsumlegung [MATHIAS, 1999], die Prognose der Reisezeit auf Autobahnen [VAN LINT, J. ET AL., 2005] und in Stadtstraßennetzen [LIU ET AL., 2007], sowie für die Stauerkennung auf der Autobahn [VAN LINT, H., 2007].

Im weiteren Verlauf dieser Arbeit wird das sogenannte rekurrente Backpropagation-Lernverfahren eine wesentliche Rolle bei der Optimierung der Verkehrssteuerung spielen. Deshalb ist es notwendig, dass dieses Lernverfahren ausführlich beschrieben wird.

3.3 Rekurrente Backpropagation-Lernverfahren

Das in dieser Arbeit verwendete rekurrente Backpropagation-Lernverfahren (RBP) wurde aus der klassischen Backpropagation hergeleitet. Deshalb wird im folgenden Abschnitt zunächst der Backpropagation-Algorithmus für Feedforward-Netze vorgestellt, anschließend wird das rekurrente Backpropagation-Lernverfahren für die RNN ausgeführt.

3.3.1 Backpropagation für Feedforward-Netze

Mehrere Forscher [WERBOS, 1974], [PARK, 1985] entdeckten unabhängig voneinander die Backpropagation Lernmethode. RUMMELHART UND HINTON [1986] haben das Verfahren in eine praktikable Prozedur eingebunden.

Das Backpropagation-Lernverfahren kann auf alle mehrschichtigen Feedforward-Netze angewendet werden, die differenzierbare Aktivierungsfunktionen und überwachtetes Lernen einsetzen. Es handelt sich um eine Optimierungsprozedur, die auf dem Gradientenabstieg basiert. Die Verbindungsgewichte werden so angepasst, dass Systemfehler oder Zielfunktionen reduziert werden. Die Backpropagation leitet sich von der Methode ab, wie diese Verbindungsgewichte zu korrigieren.

Während der Lernphase werden dem Netzwerk in einer bestimmten Reihenfolge Eingabemuster präsentiert. Jedes Muster wird Schicht für Schicht nach vorne weitergereicht, bis ein Ausgabemuster berechnet werden kann. Diese berechneten Ausgabewerte werden mit gewünschten Ausgaben oder Zielausgaben verglichen, woraus sich Fehlerwerte ableiten lassen. Diese Fehler werden als Eingaben für Feedback-Verknüpfungen verwendet, welche dann rückwärts Schicht für Schicht die Verbindungswerte anpassen.

Abbildung 3-3 zeigt ein Feedforward-Netzwerk und die Backpropagation desselben Netzes. Für dieses Beispiel wird die mathematische Beschreibung des Algorithmus vorgestellt. Dieses Beispielsystem hat eine Eingabeschicht I , die vollständig mit der verborgenen Schicht J verbunden ist, und eine Ausgabeschicht K . Die Verbindungsgewichte zwischen der Eingabeschicht-Einheit i und der Einheit j der verborgenen Schicht werden geschrieben als v_{ji} . w_{kj} bezeichnet die Verbindungsgewichte zwischen Einheit j der verborgenen Schicht und der Ausgabeschicht-Einheit k . Das Eingabemuster wird Schicht für Schicht vorwärts weitergereicht und das Netzwerk berechnet für ein gegebenes Muster (x^p, t^p) die Ausgabe:

$$z_k^p = \sigma(h_k^p) = \sigma\left(\sum_j w_{kj} y_j^p\right) = \sigma\left(\sum_j w_{kj} \sigma(h_j^p)\right) = \sigma\left(\sum_{i,j} w_{kj} \sigma(v_{ji} x_i^p)\right) \quad (3-1)$$

Hier ist σ eine Aktivierungsfunktion. In den meisten Modellen wird eine sigmoide Aktivierungsfunktion verwendet. Häufig wird die sigmoide Funktion $\sigma(x) = 1 / (1 + \exp(-\beta x))$ gewählt, da die Ableitung dieser Funktion besonders einfach ist ($\sigma'(x) = \beta \sigma(x)(1 - \sigma(x))$).

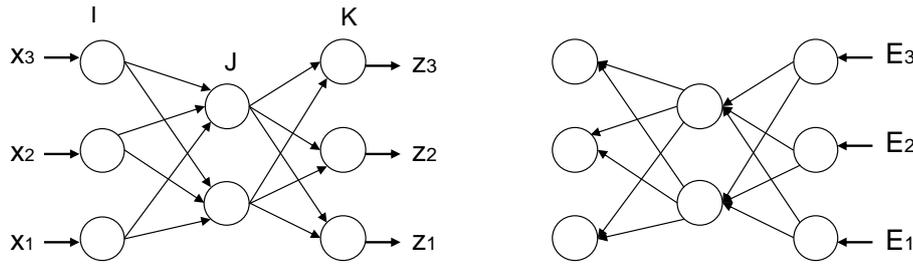


Abbildung 3-3: Ein mehrschichtiges Feedforward-Netzwerk und seine Backpropagation

Der Systemfehler (Gl. 3-2) für ein Muster p über alle Ausgabeneinheiten wird durch die Summe der quadrierten Abweichungen von gewünschter und tatsächlicher Ausgabe festgelegt.

$$E^p = \frac{1}{2} \sum_k (t_k^p - z_k^p)^2 \quad (3-2)$$

Wenn die Gl. 3-1 in 3-2 eingesetzt wird, ist E als eine Funktion der Gewichtungen v_{ji} , w_{kj} formuliert. Der Kern des Lernalgorithmus ist die Reduktion des gesamten Systemfehlers für alle Muster über eine entsprechende Anpassung der Gewichtungen. Einfachste und häufig verwendete Methode ist das Gradientenabstiegsverfahren:

$$\Delta w_{kj}^p = -\varepsilon \frac{\partial E^p}{\partial w_{kj}} \quad \text{und} \quad \Delta v_{ji}^p = -\varepsilon \frac{\partial E^p}{\partial v_{ji}} \quad (3-3)$$

Mit ε : eine positive Konstante und Parameter für den Lernumfang

Mit Hilfe der Methode der Kettenregel können die beiden Ableitungen nach der folgenden Formel berechnet werden:

$$\Delta w_{kj}^p = -\varepsilon \frac{\partial E^p}{\partial w_{kj}} = -\varepsilon \frac{\partial E^p}{\partial h_k^p} \frac{\partial h_k^p}{\partial w_{kj}} = \varepsilon \delta_k^p y_j^p \quad \text{mit} \quad \delta_k^p = (t_k^p - z_k^p) \sigma'(h_k^p)$$

Und

$$\Delta v_{ji}^p = -\varepsilon \frac{\partial E^p}{\partial v_{ji}} = -\varepsilon \frac{\partial E^p}{\partial h_j^p} \frac{\partial h_j^p}{\partial v_{ji}} = \varepsilon \delta_j^p x_i^p \quad \text{mit} \quad \delta_j^p = \sigma'(h_j^p) \sum_k \delta_k^p w_{kj}$$
(3-4)

Durch diese zwei Gleichungen können die Delta-Mengen δ_k^p für alle Neuronen, von den Ausgabenelementen zurück bis zu den Eingabeelementen, festgelegt werden. Mit Hilfe von δ_k^p werden alle Gewichtungen w_{kj} anhand von (3-5) aktualisiert.

$$w_{kj}^{neu} = w_{kj}^{alt} + \Delta w_{kj}^p = w_{kj}^{alt} + \varepsilon \delta_k^p y_j^p$$

Und

$$v_{ji}^{neu} = v_{ji}^{alt} + \Delta v_{ji}^p = v_{ji}^{alt} + \varepsilon \delta_j^p x_i^p$$
(3-5)

Diese Prozedur wird für jedes Musterpaar wiederholt, bis der Gesamtsystemfehler eine akzeptable Größe erreicht hat.

Gegenüber dem Vorteil des geringeren Rechenaufwands steht der Nachteil, dass dem Verfahren lediglich die lokale Umgebung (der Gradient) bekannt ist. Daraus resultieren Probleme (lokale Minima, flache Plateaus, direkte Oszillation und Verlassen guter Minima). Durch Initialisierung der Gewichte und Veränderung der Lernrate können diese Probleme teilweise gelöst werden.

3.3.2 Backpropagation für RNN

Aufgrund der weniger klaren Struktur im Vergleich zu einem Feedforward-Netzwerk ist die einfache schichtweise Fehlerpropagierung in rückwärtiger Richtung nicht mehr möglich, da jedes Neuron in RNN prinzipiell mit einem anderen beliebigen Neuron verbunden sein kann (Abbildung 3-2). Die Lernmethode der RNN fordert das Sammeln und Speichern von Fehlern über ganze Durchläufe, bevor die Verbindungsgewichte festgelegt werden.

RNN verfügt über eine Funktion von Zeit und/oder Raum. Die diskrete Zeitdynamik des RNN ist wie folgt definiert:

$$x_i(t + \Delta t) = \sigma \left(\sum_j w_{ij} x_j(t) + I_j \right) \quad (3-6)$$

Mit σ : eine sigmoid Funktion

I_j : externer Eingabewert

Vor den dynamischen Netzen soll hier zunächst das statische Modell betrachtet werden. Es wird in der Arbeit davon ausgegangen, dass das RNN sicher gegen einen Fixpunkt konvergiert und dies darüber hinaus sogar gewährleistet werden muss. Die Korrektur der Verbindungsgewichte kann durch Gradientenabstieg (Gl. 3-7) erfolgen. Dazu wird die Auswirkung der Dynamik ignoriert.

$$\Delta w_{ij} = -\varepsilon \frac{\partial E}{\partial w_{ij}} \quad (3-7)$$

Dabei ist die Optimierungsfunktion bzw. der Gesamtsystemfehler wie vorher eine Summe der Fehlerquadratrate der einzelnen Ausgabeinheiten:

$$E = \frac{1}{2} \sum_k E_k^2 \quad \text{mit } E_k = \begin{cases} d_k - x_k^* & \text{wenn } k \text{ eine Ausgabeinheit ist} \\ 0 & \text{sonst} \end{cases} \quad (3-8)$$

Als Besonderheit der Optimierungsfunktion wird hier erläutert, wie anstatt des konventionalen Fehlermaßes eine beliebige, in x stetig differenzierbare Funktion eingesetzt werden kann. In folgendem Kapitel werden bei dem später beschriebenen Verkehrsmo- dell andere Optimierungsfunktionen verwendet.

Die Gl. 3-7 kann mit Hilfe der Kettenregel wie folgt geschrieben werden (Die Details der Ableitung werden hier nicht erwähnt. In der Literatur [PINEDA 1988, 1989] wurde diese Verfahrensweise detailliert ausgeführt):

$$\Delta w_{ij} = -\varepsilon \sum_k E_k (L^{-1})_{ki} \sigma'(h_i^*) x_j^* \quad (3-9)$$

Und

$$L_{ki} = \delta_{ki} - \sigma'(h_k^*) w_{ki} \quad \text{mit} \quad \delta_{ki} = \begin{cases} 1 & \text{wenn } k = i \\ 0 & \text{sonst} \end{cases} \quad (3-10)$$

Gleichung 3-9 ist eine Lösung, die für die Aktualisierung der Gewichtungen verwendet werden kann. Mit Ausnahme der Matrixinvertierung sind alle beteiligten Größen in Gleichung 3-9 leicht zu berechnen. Um die relativ aufwendige Berechnung der Matrixinvertierung zu vermeiden, kann ein assoziiertes dynamisches System eingeführt werden. Zuerst wird die Gl. 3-9 in Vektor-Matrix-Notation geschrieben:

$$\nabla_w E = (y^* S) (x^*)^T \quad \text{mit } S \text{ ist die Diagonalmatrix } (S)_{ij} = \delta_{ij} \sigma'(h_i^*) \quad (3-11)$$

$$y^* := (L^{-1})^T e \quad \text{mit } e := (E_1, E_2, \dots, E_n) \quad (3-12)$$

Nachdem beide Seiten der Gl. 3-12 mit L^T multipliziert und umformiert sind, ergibt sich die etwas bekanntere Form, die *Fixpunktgleichung*:

$$y_k^* = \sum_i w_{ik} \sigma'(h_i^*) y_i^* + E_k \quad (3-13)$$

Analog zu Gl. 3-6 kann ein neues RNN, das sogenannte Fehlerpropagierungs-Netz, mit folgender Evolutionsgleichung definiert werden:

$$y_k(t + \Delta t) = y_k(t) + \sum_i w_{ik} \sigma'(h_i^*) y_i(t) + E_k \quad (3-14)$$

Die Topologie dieses Fehlerpropagierungs-Netzes (Abbildung 3-4) entspricht dem des originalen RNN (Abbildung 3-2). Nun werden die Verbindungsgewichte w_{ki} des originalen Netzes durch $w_{ik} \sigma'(h_i^*)$ in umgekehrter Richtung ersetzt. Die Fehler E_k des Originalnetzes entsprechen den Eingabewerten für das Fehlerpropagierungs-Netz.

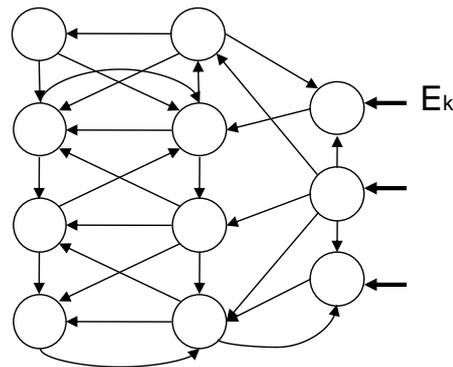


Abbildung 3-4: Das mit dem Originalnetz in Abbildung 3-2 korrespondierende Fehlerpropagations-Netz

Würde anstelle des Lernverfahrens mit RBP mit Relaxation eine Matrixinvertierung mit Standardmethoden durchgeführt, würde das einen Berechnungsaufwand der Ordnung N^3 (N = Anzahl der Neuronen) erfordern. Im Vergleich dazu benötigt RBP mit Relaxation eine Berechnungszeit der Ordnung N^2 . Dieser Vorteil kann für einen Einsatz mit großem Verkehrsnetz (z. B. Stadt München mit über 1,000 Lichtsignalanlagen) sehr attraktiv sein.

3.3.3 Backpropagation für zeitabhängiges RNN

Bisher wurde die rekurrente Backpropagation für die Fixpunkt-Form des RNN ausgeführt. Um ein zeitabhängiges RNN zu trainieren, wird eine Zeitdimension in die Berechnungen eingefügt. WILLIAMS UND ZIPSER [1989] entwickelten für ein zeitdiskretes RNN ein ständiges Lernen in Echtzeit und PEARLMUTTER [1989] macht für ein zeitkontinuierliches RNN die *zeitabhängige rekurrente Backpropagation*. Das zweite Modell wird im Verlauf dieser Arbeit verwendet, deshalb wird hier nur auf die zeitabhängige rekurrente Backpropagation eingegangen.

Die Dynamik des Netzes wird von kombinierten, nichtlinearen Differentialgleichungen bestimmt:

$$\tau_i \frac{dx_i(t)}{dt} = -x_i(t) + \sigma \left(\sum_j w_{ij} x_j \right) + I_i \quad \text{mit } \tau_i: \text{Relaxionszeiten} \quad (3-15)$$

Der Fehler in der Ausgabeeinheit muss für die Zeitvarianz beschrieben werden:

$$E_k(t) = \begin{cases} d_k(t) - x_k(t) & \text{wenn } d_k(t) \text{ zum Zeitpunkt } t \text{ definiert ist} \\ 0 & \text{sonst} \end{cases} \quad (3-16)$$

In die gesamte Optimierungsfunktion wird der quadratische Fehler aller Ausgabeeinheiten für Zeitintervall $[0, T]$ integriert. Das Verfahren hat zum Ziel, diese Optimierungsfunktion zu minimieren.

$$E = \frac{1}{2} \int_0^T \sum_k [E_k(t)] dt \quad (3-17)$$

Analog zur vorher vorgestellten rekurrenten Backpropagation (Fixpunkt) wird die für den Gradientenabstieg erforderliche Ableitung berechnet:

$$\Delta w_{ij} = \frac{\partial E}{\partial w_{ij}} = \frac{1}{\tau_i} \int_0^T \sum_i y_i(t) \sigma'(h_i^*(t)) x_j(t) dt \quad (3-18)$$

Wobei $y_i(t)$ die Lösung der folgenden Differentialgleichung ist:

$$\frac{dy_i(t)}{dt} = \frac{1}{\tau_y} y_i(t) + \sum_j w_{ij} \sigma'(h_i^*(t)) y_j(t) - E_i(t) \quad (3-19)$$

Ähnlich wie RBP für Fixpunkt-RNN steht die Gl. 3-19 für die Dynamik des Fehlerpropagierungs-Netzes. Zur Aktualisierung der Gewichtungen werden eine Vorwärts- und eine Rückwärtsphase durchgeführt. Als Startwert kann einfach $dx_i(t)/dt = 0$ angenommen werden, da davon ausgegangen werden kann, dass der Anfangszustand des Systems ein konstanter Zustand ist. In der Vorwärtsphase wird die Ausgabe $x_i(t)$ durch Integration des Originalnetzes nach Gl. 3-15 von $t = 0$ bis T ermittelt, dann wird $y_i(t)$ während der Rückwärtsphase durch Integration des Fehlerpropagierungs-Netzes nach Gl. 3-19 von $t = T$ bis 0 berechnet. Mit Hilfe der Gl. 3-18 kann die Änderung der Gewichtungen schließlich festgelegt werden.

4 Das RNN-Verkehrsmodell

Das Verkehrsmodell ist eines der wichtigsten Bausteine der modellbasierten Verkehrssteuerung. In der Arbeit von MATHIAS [1999] wurde ein auf Rekurrenten Neuronalen Netzen (RNN) basierendes Verkehrsmodell entwickelt. In den nachfolgenden Ausführungen des Kapitels 4 wird dieses Modell in gekürzter Form vorgestellt. Zunächst wird ein statisches Verkehrsmodell, das eine Schichtstruktur in Bezug auf die statischen Quell-Ziel-Matrizen besitzt, eingeführt. Das kooperierende Fixpunkt-Fehlerpropagation-Netzwerk dient der Festlegung der Abbiegeraten in dem Verkehrsmodell. Die Übertragung des statischen Modells auf ein dynamisches Modell wird anschließend beschrieben. Die zeitabhängige rekurrente Backpropagation, die als Routenwahlmodell dient, kommt hier zum Einsatz, um die definierten Zielfunktionen zu optimieren.

4.1 Das Simulations-Netzwerk

Das in dieser Arbeit entwickelte Verkehrssteuerungsmodell basiert auf vier rekurrenten Neuronalen Netzen. Das erste RNN (das Simulations-Netzwerk) simuliert das Verkehrsgeschehen, das zweite RNN (das Fehlerpropagierendes-Netzwerk) beschäftigt sich mit der Umlegung und die übrigen zwei RNN (die Fehlerpropagierenden-Netzwerke) sind zuständig für die Verkehrssteuerung. Anhand der Zielfunktion justieren die letzten drei RNN verschiedene Steuerungsparameter in dem Simulations-Netzwerk (siehe Abbildung 4-1).

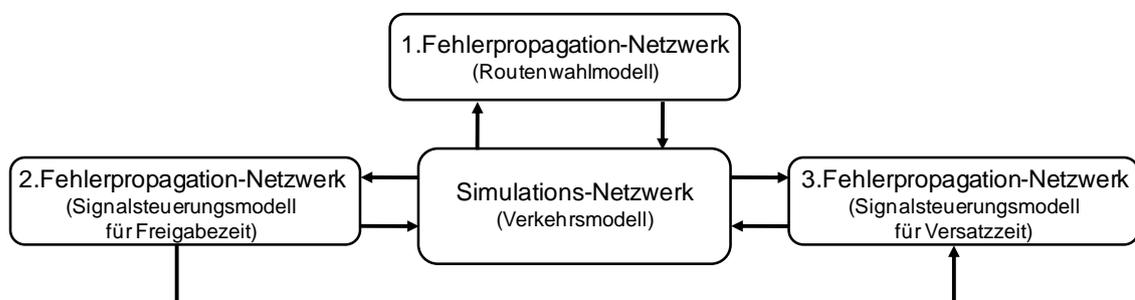


Abbildung 4-1: Die Grobstruktur des gesamten Modells

Die Ähnlichkeiten der Topologie und Dynamik von Verkehrsnetzen und RNN sind die Grundidee des neuen Ansatzes. Die Fahrzeuge im Verkehrsnetz entsprechen dem

Aktivitätspotential in Neuronalen Netzen, die Fortbewegung der Fahrzeuge im Straßennetz kann hauptsächlich durch die dichteabhängige Geschwindigkeit sowie Abbiegeraten bestimmt werden. Im Fall des RNN hängt die Systemdynamik von der Aktivitätsfunktion und den Verbindungsgewichten ab. Anhand dieser Analogie entwickelte MATHIAS [1999] ein auf RNN basiertes Verkehrsflussmodell, das in den kommenden Unterkapiteln ausführlich vorgestellt wird.

4.1.1 Topologie des Simulations-Netzes

Bevor ein rekurrentes neuronales Netz konstruiert wird, werden einige Definitionen für das Verkehrsnetz vorgestellt. Die eingeführten Formelzeichen werden weiterhin für das RNN-Modell übernommen. Ein Straßennetz besteht aus M Strecken und N Knoten. Q bzw. Z sei die Anzahl der Quellen bzw. Ziele. S ist die Menge der Strecken, K ist die Menge der Knoten, $O \subset S$ ist die Menge der Quellstrecken, $D \subset S$ ist die Menge der Zielstrecken, $A_j \subset S$ ist die Menge der am Knoten j beginnenden Strecken, $E_j \subset S$ ist die Menge der am Knoten j endenden Strecken. Abbildung 4-2 zeigt ein Straßennetz mit zwei Quellen und zwei Zielen, zwölf Strecken S_i und vier Knoten k_j .

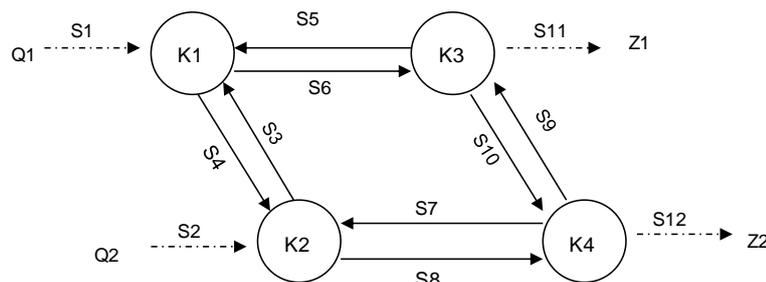


Abbildung 4-2: Ein einfaches Verkehrsnetzmuster

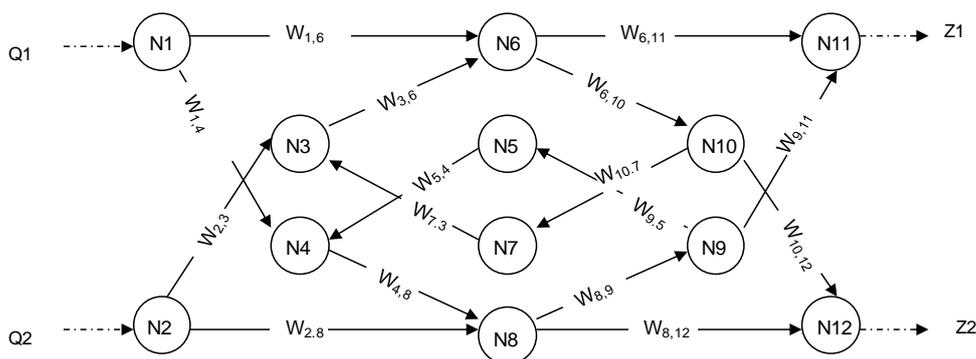


Abbildung 4-3: Das entsprechende Simulations-RNN

Die Nachbildung der Topologie des Straßennetzes auf dem Neuronalen Netz erfolgt wie folgt: Die Strecken des Straßennetzes werden mit den Neuronen des RNN identifiziert, die Nummerierung von Strecken und Neuronen sind dann auch identisch. Die Fahrbeziehungen bzw. die Verbindungen zwischen Strecken an den Knotenpunkten des Straßennetzes entsprechen den Verbindungsgewichten der Neuronen des RNN.

Die Abbildung 4-3 veranschaulicht das zum vorgestellten Straßennetz entsprechende Simulations-RNN.

Man kann erkennen, dass jedes Neuron mit jedem beliebigen Neuron verbunden werden kann, dieses Phänomen bestimmt die Eigenschaft des RNN. Aber tatsächlich sind die Neuronen nur mit wenig anderen Neuronen gekoppelt, da eine Strecke im Straßennetz in der Praxis maximal 5 Strecken als Nachfolger haben kann.

4.1.2 Dynamik des Simulations-Netzes

Aufgrund der Identifikation von Strecken und Neuronen kann der Verkehrsmenge $x_i(t)$ auf der Strecke i zum Zeitpunkt t der Aktivitätszustand des Neurons i entsprechen. Die Bewegung der Verkehrsmenge wird durch die Aktivierungsfunktion und Verbindungsgewichte w_{ik} des RNN beschrieben. Die von dem Zustand des Neurons bzw. der Verkehrsmenge der Strecken abhängige Aktivierungsfunktion besteht aus zwei Teilen. Einer ist der maximale mögliche Streckenabfluss, der andere ist die Zuflussrestriktion. Sie bestimmen, wie viele Fahrzeuge in einem bestimmten Zeitschritt eine Strecke verlassen und zur nächsten Strecke einfahren können. In RNN entspricht dies der Zustandsänderung des Neurons pro Zeitschritt. Die Verbindungsgewichte w_{ik} des RNN stellt die Abbiegeraten im Straßennetz dar. Tabelle 4-1 zeigt den Zusammenhang des RNN und des Straßennetzes im Überblick. In den folgenden Unterkapiteln werden die Verbindungsgewichte und die zwei Aktivierungsfunktionen in mathematischer Form beschrieben. Anschließend wird die dynamische Gleichung des Simulations-Netzwerks vorgestellt.

Straßennetz	Simulations-RNN
Strecke i	Neuron i
Abbiegebeziehung $w_{i,j}$	Verbindungsgewichte $w_{i,j}$
Fahrzeugmenge x_i auf Strecke i	Aktivitätszustand x_i des Neurons i
Netzzufluss q_i in Strecke i	Externer Input q_i des RNN
Netzabfluss z_i in Strecke i	Externer Output z_i des RNN
Möglicher Abfluss S_i aus Strecke i	Aktivierungsfunktion des Neurons i
Zuflussrestriktion r_i der Strecke i	Aktivierungsfunktion des Neurons i

Tabelle 4-1: Zusammenhang von Straßennetz und Simulations-RNN [MATHIAS, 1999]

Verbindungsgewichte

Die Verbindungsgewichte w_{ij} des Neurons i mit Neuron j sind im Straßennetz die Abbiegebeziehung einer Strecke i zu Strecke j . Ein Verbindungsgewicht $w_{ij}=0,6$ bedeutet, dass 60% der Fahrzeuge auf der Strecke i zur Strecke j abfließen. Abbildung 4-4 (links) veranschaulicht dieses einfache Beispiel.

Aufgrund des Verkehrsverhaltens müssen die Verbindungsgewichte folgende Bedingungen erfüllen:

$$w_{ij} \in [0,1], \forall i, j \quad \text{und} \quad \sum_j w_{ij} = 1, \forall i \quad (4-1)$$

Die Self-Feedback-Verbindungen des Neurons werden nicht durch die Gewichte beschrieben, sondern später in der dynamischen Gleichung des Simulations-Netzes eingebaut. Daher gilt: $w_{ii} \equiv 0, \forall i \in S$.

Aktivierungsfunktion – möglicher Abfluss

Der maximal mögliche Abfluss S_i definiert die Verkehrsmenge der Strecke i , die während eines Zeitintervalls τ_x abfließen könnte, wenn die nachfolgenden Strecken die gesamte Verkehrsmenge in diesem Zeitintervall aufnehmen könnten.

Im Unterkapitel 3.3.1 wurde die Grundform der Aktivierungsfunktion vorgestellt. Aufgrund der einfachen Ableitung wird hier eine sigmoide Funktion eingesetzt:

$$s_i(x_i) = \tau_x C_i \sigma_i(x_i) \quad (4-2)$$

Wobei C_i die Kapazität der Strecke ist und $\sigma_i(x_i)$ folgende Form besitzt:

$$\sigma_i(x_i) = \frac{2}{1 + \exp(-ax_i)} - 1 \quad (4-3)$$

Wobei a ein frei wählbarer Parameter ist, a sollte so gewählt werden, dass die aus $S_i(x_i)$ hergeleitete Reisezeitfunktion ähnlich der aus $M/D/1$ Modell ermittelten Reisezeitfunktion [Wu, 1990] ist. Auf dieses Detail wird hier nicht näher eingegangen. Abbildung 4-4 Mitte zeigt die Gestalt dieser Funktion ($c=1800$ kfz/h, $\tau_x=1$ h, $a=4$).

Wegen der sigmoiden Funktion kann die Ableitung (Gl. 4-4) von s_i durch $\sigma_i(x_i)$ beschrieben werden. Dies vereinfacht die spätere Berechnung des Fehlerpropagierungsnetzwerkes.

$$s_i'(x_i) = \tau_x C_i \sigma_i'(x_i) = \tau_x C_i (1 - \sigma_i(x_i))(1 + \sigma_i(x_i))a/2 \quad (4-4)$$

An dieser Stelle sei erwähnt, dass die Kapazität der Strecke in der Funktion (Gl. 4-2) im Verlauf der Arbeit für die Verkehrssteuerung eine entscheidende Rolle spielen wird.

Aktivierungsfunktion – Zuflussrestriktion

Die Annahme, dass die nachfolgenden Strecken den gesamten Abfluss der Strecke in einem Zeitintervall aufnehmen können, kann nicht immer wahr sein. Daher wird eine

Zuflussrestriktion r_k definiert, die von der Verkehrsmenge x_k abhängt. Die Zuflussrestriktion r_k beschreibt, wie viel Prozent aller Fahrzeuge in einem Zeitintervall auf der Strecke k zufließen könnten. Diese Funktion realisiert die qualitative Korrelation zwischen einer Strecke und ihren verbundenen Strecken, daraus können Stauaufbau-Effekte im Netz nachgebildet werden.

Eine sigmoide Funktion wird zur Bildung der Zuflussrestriktion r_k , genutzt. Der Wert liegt in einer offenen Menge (0,1).

$$r_k(x_k) = 1 - \frac{1}{1 + \exp(b_k - c_k(x_k/x_{\max,k}))}, \quad \forall k \in S \tag{4-5}$$

Wobei b_k, c_k freie Parameter sind. Diese sollten so gewählt werden, dass r_k den verkehrstechnischen Zusammenhang widerspiegelt. $x_{\max,k}$ (=maximale Verkehrsdichte * Länge der Strecke) ist die maximale Verkehrsmenge auf der Strecke k . Rechts in Abbildung 4-4 ist die Form der Funktion ($b_k=32, c_k=34$) dargestellt.

Wie s_i besitzt die Ableitung von r_k eine geschlossene Form von r_k :

$$r'_k(x_k) = -(c_k/x_{\max,k})r_k(x_k)(1-r_k(x_k)), \quad \forall k \in S \tag{4-6}$$

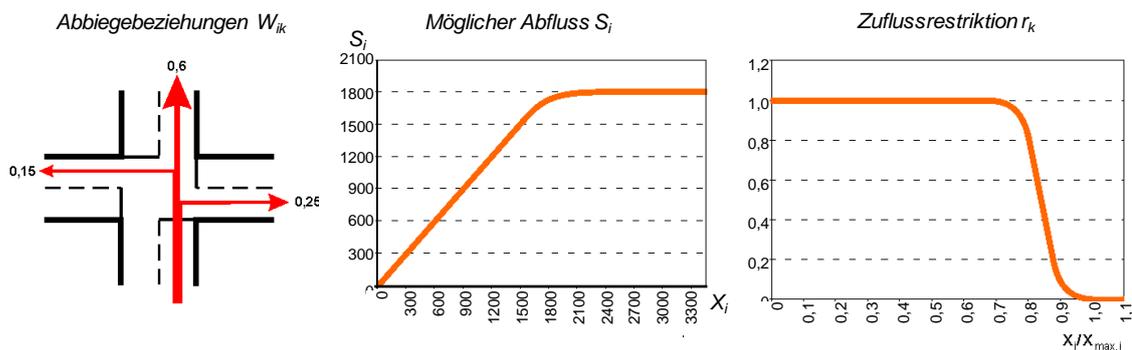


Abbildung 4-4: Drei Elementen für die dynamische Gleichung

Die dynamische Gleichung des Simulations-Netzes

Eine dynamische Gleichung des klassischen RNN enthält man aus Gleichung 3-6. Es wird auch eine diskrete zeitdynamische Gleichung für das Simulations-Netz definiert. Die zu einem Zeitpunkt auf einer Strecke bleibende Verkehrsmenge ist eine Funktion der zum vorherigen Zeitpunkt dort verbleibenden Verkehrsmenge, der während dieses Zeitintervalls auf die Strecke zufließenden Verkehrsmenge und der aus der Strecke abfließenden Verkehrsmenge. Die zeitdiskrete dynamische Gleichung des Netzes wird wie folgt definiert:

$$x_k(t + \tau_x) = x_k(t) + u_k(x_k(t)) - h_k(x_k(t)) + \tau_x q, \quad \forall k \in S \tag{4-7}$$

Wobei $u(x_k(t))$ der reale Streckenzufluss und $h(x_k(t))$ der reale Streckenabfluss sind. $\tau_x q$ ist die während des Zeitintervalls τ_x über die Quellstrecken ins Netz einfließende Verkehrsmenge. Der reale Streckenzufluss (Gl. 4-8) bzw. Streckenabfluss (Gl. 4-9) hängt von der Abbiegebeziehung w_{ij} und der vorher definierten Aktivierungsfunktionen: möglicher Abfluss und Zuflussrestriktion, ab.

$$u_k(x_k(t)) = r_k(x_k(t)) \sum_{i \in E(k)} w_{ik} s_i(x_i(t)), \quad \forall k \in S \quad (4-8)$$

$$h_k(x_k(t)) = s_k(x_k(t)) \sum_{i \in A(k)} w_{ki} r_i(x_i(t)), \quad \forall k \in S \quad (4-9)$$

r_k und s_k können in einer Kombination geschrieben werden:

$$a_{ik}(x) := s_i(x_i) r_k(x_k), \quad \forall i, k \in S \quad (4-10)$$

Durch Einsetzen kann die dynamische Gleichung 4-7 wie folgt umgeformt werden:

$$x_k(t + \tau_x) = x_k(t) + \sum_{i \in E(k)} w_{ik} a_{ik}(t) - \sum_{i \in A(k)} w_{ki} a_{ki}(t) + \tau_x q, \quad \forall k \in S \quad (4-11)$$

Gl. 4-11 zeigt, dass die Dynamik durch die Produkte der zwei verschiedenen sigmoiden Funktionen r_k und s_k , beschrieben ist. Sie unterscheidet sich von der in Gl. 3-6 definierte konventionellen RNN-Dynamik. Dieses Netzwerk wird Π - Σ -RNN genannt. Abbildung 4-5 zeigt die funktionale Struktur des Netzwerks in einzelnen Blöcken.

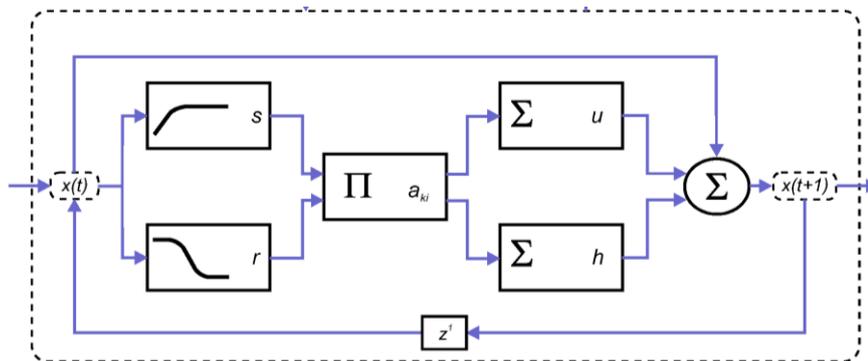


Abbildung 4-5: Flussdiagramm für Rekurrent Π - Σ -Netz [MATHIAS,2003]

4.1.3 Das Schichten-Modell des Simulations-Netzes

Das bisher eingeführte Verkehrsmodell berücksichtigt Quell- und Zielverkehrsmengen, aber keine Quell-Ziel-Matrizen. Ohne diese Matrizen kann das Verkehrsgeschehen nicht richtig simuliert werden. Im Folgenden wird das RNN-Simulations-Modell überarbeitet, damit die Quell-Ziel-Information in dem Modell integriert wird. Die Grundidee ist eine Strukturierung des Simulations-Netzes in einzelne Schichten. Jede Schicht model-

liert die Verkehrsflüsse gleichen Ziels, aber beliebiger Quelle. D.h. die Anzahl der Schichten ist gleich wie die Anzahl der Ziele. Das Beispielnetz mit zwei Zielen von Abbildung 4-3 kann in zwei Schichten aufgeteilt werden (siehe auch Abbildung 4-6)

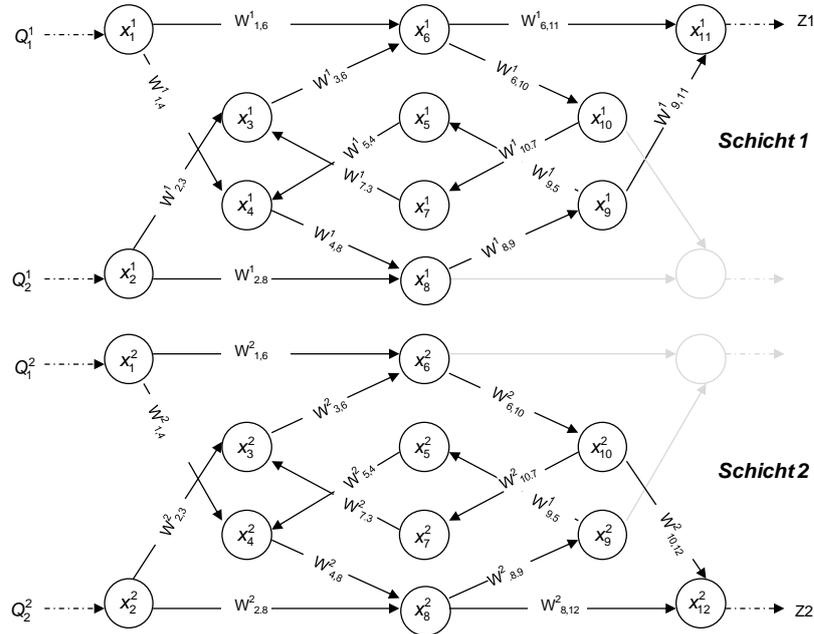


Abbildung 4-6: Ein zwei schichtiges Simulations-Netz mit zwei Zielen

Analog zu Gl. 4-11 werden die M - Z dynamischen Gleichungen des Systems definiert:

$$x_k^d(t + \tau_x) = x_k^d(t) + \sum_{i \in E(k)} w_{ik}^d a_{ik}^d(t) - \sum_{i \in A(k)} w_{ki}^d a_{ki}^d(t) + \tau_x q^d, \quad d = 1, 2 \dots Z, \quad \forall k \in S \quad (4-12)$$

Hier ist $x_k^d(t)$ der Teilfluss der Schicht d ($\sum_{d=1}^Z x_k^d(t) = x_k(t)$), w_{ik}^d ist die Abbiegebeziehung der Schicht d ($\sum_k w_{ik}^d = 1$). Es sollte beachtet werden, dass sowohl der mögliche Abfluss als auch die Zuflussrestriktion nicht abhängig von den Teilflüssen der Schicht, sondern von der gesamten Verkehrsmenge der Strecken sind. Deshalb sind die einzelnen Schichten des Netzes nicht unabhängig voneinander. Der Korrelationsterm $a_{ik}^d(t)$ kann wie folgt definiert werden:

$$a_{ik}^d(t) = s_i^d(x_i(t)) r_k(x_k(t)), \quad \forall i \in S, \quad \forall k \in S \quad (4-13)$$

Wobei

$$s_i^d(x_i(t)) = s_i(x_i(t)) \frac{x_i^d(t)}{x_i(t)}, \quad \forall i \in S \quad (4-14)$$

Das Schichtmodell schafft die Berücksichtigung der Quell-Ziel-Matrizen. Die Idee des Schichtmodells wird später in der Lichtsignalsteuerung bei der Koordinierung der einzelnen Lichtsignalanlagen eingesetzt.

4.2 Das Fehlerpropagierungs-Netz zur statischen Umlegung

Um die Verkehrsflüsse im Straßennetz zu modellieren, müssen die Quell-Ziel-Matrizen auf das Straßennetz umgelegt werden. Ein Fehlerpropagierungs-Netz kann als ein Routenwahlmodell dienen. Die Verbindungsgewichte bzw. die Abbiegebeziehungen in dem ersten Simulations-RNN werden anhand einer vorgegebenen Optimierungsfunktion angepasst, damit die Verkehrsmengen auf der Strecke festgelegt werden können. Die Abbildung eines Fehlerpropagierungs-Netzes wird auf Basis der in Kapitel 3 vorgestellten Rekurrenten Backpropagation aufgebaut und entsprechend dem Π - Σ -RNN erweitert.

In diesem Unterkapitel wird das statische Umlegungsverfahren eingeführt. Für das statische Modell sind zeitlich gleichbleibenden Randbedingungen vorgesehen, d.h. die Quell-Ziel-Matrizen, die Kapazitätswerte der Straßen werden für einen bestimmten Zeitraum konstant bleiben. Bevor die statische Umlegung implementiert wird, sind zunächst die im Verlauf der Arbeit immer betroffenen Optimierungsfunktionen im Detail zu beschreiben.

4.2.1 Optimierungsfunktion

Die Verkehrsmengen auf den Strecken im Gleichgewichtszustand werden durch die Abbiegeraten bzw. Gewichte bestimmt. Die benötigten Gewichtskorrekturen werden dazu mit Hilfe des Gradientenabstiegs von einer Optimierungsfunktion, die ein gewünschtes Ziel formuliert, berechnet.

Im Unterkapitel 3.3.2 wurde erwähnt, dass die Optimierungsfunktion des RNN eine beliebige, in x stetig differenzierbare Funktion sein kann. Dazu besitzt die zu minimierende Optimierungsfunktion die folgende Form:

$$E(x) = \sum_k E_k(x_k) \quad \forall k \in S \quad (4-15)$$

Grundsätzlich kann die Optimierungsfunktion eines Verkehrsmodells sich auf zwei Kategorien verteilen, das Systemoptimum und das Nutzeroptimum [WARDROP, 1952]. Für das vorhandene Simulationsmodell können sowohl das Systemoptimum als auch das Nutzeroptimum verwendet werden.

Da die Optimierungsfunktion später für die Verkehrssteuerung eingesetzt werden soll, welche ein Systemoptimum benutzt, wird im Folgenden nun das Systemoptimum vor-

gestellt. Im Verlauf der Arbeit wird auch für die Umlegung ausschließlich das Systemoptimum angewendet werden.

Optimierungsfunktion-Systemoptimum

Das Systemoptimum verfolgt das Ziel der Minimierung der Kosten im gesamten Verkehrsnetz. Das berücksichtigt in erster Linie das Interesse des Betreibers, durch geeignete Steuerungsmaßnahmen insgesamt möglichst minimale Reisezeiten zu ermöglichen. Dabei wird in Kauf genommen, dass der Nutzen für einzelne Nutzer erhöht werden könnte, dies aber zu Gunsten des Systemoptimums nicht getan wird. Mit dieser Sicht entspricht das Systemoptimum dem 2. Wardrop'sche Prinzip [WARDROP, 1952]: „Die durchschnittliche Fahrzeit (d. h. die Gesamtfahrzeit) ist minimal.“

Für das vorhandene Verkehrsmodell ist die Optimierungsfunktion wie folgend bestimmt:

$$E(x) = \sum_k E_k(x_k) = \sum_k \pi_k(x_k) f_k + \beta_k (x_k - \hat{x}_k)^2 \quad \forall k \in S \quad (4-16)$$

Hier entspricht $\pi_k(x_k)$ der Reisezeit auf der Strecke k und f_k dem Verkehrsfluss auf der Strecke k . Der Term $\beta_k (x_k - \hat{x}_k)^2$ bezieht sich auf die Messstellen. \hat{x}_k ist die gemessene Verkehrsmenge und β_k ist eine positive Konstante. Wenn eine Strecke keine Messstelle besitzt, ist β_k gleich null.

Ein in einer Strecke einfahrendes Fahrzeug braucht solange, um diese Strecke durchzureisen, bis alle vorher eingefahrenen Fahrzeuge diese Strecke verlassen haben. Daher kann die Reisezeit der Strecke beschrieben werden zu:

$$\pi_k(x_k) := \tau_x \frac{x_k}{h_k} \quad \forall k \in S \quad (4-17)$$

Hier ist h_k tatsächlicher Abfluss der Strecke k . Die Verkehrsstärke $f_k = h_k / \tau_x$ und Gl. 4-17 können in Gl. 4-16 eingesetzt werden, dadurch ist ein relativ einfacher Ausdruck der Optimierungsfunktion gegeben:

$$E(x) = \sum_k \left(\tau_x \frac{x_k}{h_k} \right) \left(\frac{h_k}{\tau_x} \right) + \beta_k (x_k - \hat{x}_k)^2 = \sum_k x_k + \beta_k (x_k - \hat{x}_k)^2 \quad \forall k \in S \quad (4-18)$$

Für dieses Kriterium ist die Optimierungsfunktion bzw. die Minimierung der Reisezeit im gesamten Verkehrsnetz identisch zur Minimierung der Gesamtzahl von Fahrzeugen im Netz.

4.2.2 Berechnung der Verbindungsgewichte

Ein Steuerungsparameter des Simulationsnetzes entspricht einem Verbindungsgewicht W des RNN bzw. einer Abbiegerate W des Straßennetzes. Dieser Parameter sollte so eingestellt sein, dass die festgelegte Optimierungsfunktion E minimiert wird.

Mit Hilfe des Gradientenabstiegsverfahrens und der Kettenregel kann die Gewichtskorrektur berechnet werden zu:

$$\Delta w_{pq}^e = -\varepsilon \sum_{k=1}^M \frac{\partial E_k}{\partial w_{pq}^e} = -\varepsilon \sum_{k=1}^M \sum_{d=1}^Z \frac{\partial E_k}{\partial x_k} \frac{\partial x_k^d}{\partial w_{pq}^e} = -\varepsilon \sum_{k=1}^M \sum_{d=1}^Z L_k \frac{\partial x_k^d}{\partial w_{pq}^e} \quad (4-19)$$

Hier besitzt L_k für in Gl. 4-18 definiertes Systemoptimum eine leicht zu berechnende Form:

$$L_k = 1 + 2\beta(x_k - \hat{x}_k) \quad (4-20)$$

Um die $\partial x_k^d / \partial w_{pq}^e$ ermitteln zu können, wird ein funktionaler Zusammenhang zwischen W und dem Aktivitätszustand x benötigt. Dieser Zusammenhang wurde schon in Gl. 4-12 erläutert. Eine partielle Differentiation der Gl. 4-12 führt zu

$$\begin{aligned} \frac{\partial x_k^d(t + \tau_x)}{\partial w_{pq}^e} &= \frac{\partial x_k^d(t)}{\partial w_{pq}^e} + \delta_{kq} \delta_{de} a_{pq}^e(t) + \sum_i w_{ik}^d \frac{\partial a_{ik}^d(t)}{\partial w_{pq}^e(t)} \\ &\quad - \delta_{kp} \delta_{de} a_{pq}^e(t) - \sum_i w_{ki}^d(t) \frac{\partial a_{ki}^d(t)}{\partial w_{pq}^e(t)} \end{aligned} \quad (4-21)$$

Für die Fixpunkte bzw. Gleichgewichtspunkte des Systems kann die Gl. 4-21 wie folgt formuliert werden:

$$\frac{\partial x_k^d}{\partial w_{pq}^e} = \frac{\partial x_k^d}{\partial w_{pq}^e} + \delta_{kq} \delta_{de} a_{pq}^e + \sum_i w_{ik}^d \frac{\partial a_{ik}^d}{\partial w_{pq}^e} - \delta_{kp} \delta_{de} a_{pq}^e - \sum_i w_{ki}^d(t) \frac{\partial a_{ki}^d}{\partial w_{pq}^e} \quad (4-22)$$

mit

$$\begin{aligned} \frac{\partial a_{ik}^d}{\partial w_{pq}^e} &= \frac{\partial s_i^d}{\partial w_{pq}^e} r_k + s_i^d \frac{\partial r_k}{\partial w_{pq}^e} \\ &= \left(\frac{s_i^d}{s_i^d} - \frac{1}{x_i} \right) s_i^d r_k \sum_{j=1}^Z \frac{\partial x_i^j}{\partial w_{pq}^e} + \frac{s_i^d}{x_i} r_k \frac{\partial x_i^d}{\partial w_{pq}^e} + s_i^d r_k \sum_{j=1}^Z \frac{\partial x_k^j}{\partial w_{pq}^e} \end{aligned} \quad (4-23)$$

Wenn man folgendes definiert,

$$R_{ik}^d = \begin{pmatrix} s_i' & 1 \\ s_i & x_i \end{pmatrix} s_i^d r_k'; \quad S_{ik} = \frac{s_i}{x_i} r; \quad T_{ik}^d =: s_i^d r_k' \quad (4-24)$$

ergibt sich daraus:

$$\frac{\partial a_{ik}^d}{\partial w_{pq}^e} = R_{ik}^d \sum_{j=1}^Z \frac{\partial x_i^j}{\partial w_{pq}^e} + S_{ik} \frac{\partial x_i^d}{\partial w_{pq}^e} + T_{ik}^d \sum_{j=1}^Z \frac{\partial x_k^j}{\partial w_{pq}^e} \quad (4-25)$$

Durch Einsetzen der Gl. 4-25 in Gl. 4-22 und Umsortieren hat man:

$$\sum_{i=1}^M \sum_{l=1}^Z \Phi_{kd,il} \frac{\partial x_i^l}{\partial w_{pq}^e} = \delta_{de} (\delta_{kp} - \delta_{kq}) a_{pq}^e \quad (4-26)$$

mit

$$\Phi_{kd,il} := (w_{ik}^d R_{ik}^d - w_{ki}^d T_{ki}^d) - \delta_{ik} \sum_j (w_{kj}^d R_{kj}^d - w_{jk}^d T_{jk}^d) + \delta_{dl} w_{ik}^d S_{ik} - \delta_{dl} \delta_{ik} \sum_j w_{kj}^d S_{kj} \quad (4-27)$$

Offenbar ist Gl. 4-26 ein lineares Gleichungssystem mit den Variablen $\partial x_i^l / \partial w_{pq}^e$. Aufgrund der Schichteneinteilung des Simulations-Netzwerks besitzt die Matrix $\Phi = (\Phi_{kd,il})$ eine natürliche Blockstruktur, die einen selbigen Aufbau wie die später ausführlich vorgestellte Blockmatrix V in Gl. 4-37 aufweist. Wenn davon ausgegangen wird, dass immer ein stabiler Gleichgewichtspunkt existiert, wird hier ein Iterationsverfahren aufgrund seines viel geringeren Rechenaufwandes verwendet.

Die Gl. 4-26 kann in Form der Gl. 4-27 invertiert werden:

$$\frac{\partial x_k^d}{\partial w_{pq}^e} = \sum_{i=1}^M \sum_{l=1}^Z (\Phi^{-1})_{kd,il} \delta_{le} (\delta_{ip} - \delta_{iq}) a_{pq}^e = \left((\Phi^{-1})_{kd,pe} - (\Phi^{-1})_{kd,qe} \right) a_{pq}^e \quad (4-28)$$

Um das Gleichungssystem (Gl. 4-26) mit Hilfe eines Fehlerpropagierungs-Netzwerks zu lösen, werden die folgenden neuen Variablen definiert [MATHIAS, 1999]:

$$y_i^l := \sum_{k=1}^M \sum_{d=1}^Z L_k (\Phi^{-1})_{kd,il} \quad (4-29)$$

und neuen Gewichten

$$v_{ik}^{ld} := \frac{1}{\alpha_k^d} \left[(w_{ki}^l R_{ki}^l - w_{ik}^l T_{ik}^l) - \delta_{ik} \sum_j (w_{kj}^l R_{kj}^l - w_{jk}^l T_{jk}^l) + \delta_{dl} w_{ki}^d S_{ki} \right] \quad (4-30)$$

mit $\alpha_k^d := \sum_j w_{kj}^d S_{kj}$

Dadurch ergibt sich

$$L_k / \alpha_k^d = \sum_{i=1}^M \sum_{l=1}^Z \Phi_{il,kd} y_i^l = \sum_{i=1}^M \sum_{l=1}^Z v_{ik}^{ld} y_i^l - y_k^d \quad (4-31)$$

bzw.

$$y_k^d = \sum_{i=1}^M \sum_{l=1}^Z v_{ik}^{ld} y_i^l - L_k / \alpha_k^d \quad (4-32)$$

Die Gleichung 4-32 wird jetzt als ein neues RNN definiert. Dieses neue RNN wird Fehlerpropagierungs-Netzwerk genannt. Das RNN hat $M \cdot Z$ lineare Neuronen und besitzt keine Schichtstruktur. Die Verbindungsgewichte v_{ik}^{ld} beziehen sich auf Verbindungen zwischen allen Neuronen des RNN. Seine zeitdiskrete dynamische Gleichung kann wie Gl. 4-33 beschrieben werden.

$$y_k^d(t + \tau_y) = \sum_{i=1}^M \sum_{l=1}^Z v_{ik}^{ld} y_i^l(t) - L_k / \alpha_k^d \quad (4-33)$$

Der Gleichgewichtszustand des Systems (Gl. 4-33) entspricht der Lösung des linearen Gleichungssystems (Gl. 4-32). Dann können die gesuchten Änderungen der Verbindungsgewichte mit Hilfe der Relaxation des Fehlerpropagierungs-Netzes berechnet werden, indem die Gl. 4-28 und 4-29 in Gl. 4-19 eingesetzt werden.

$$\Delta w_{pq}^e = -\varepsilon \sum_{k=1}^M \sum_{e=1}^Z L_k \frac{\partial x_k^d}{\partial w_{pq}^e} = \varepsilon (y_q^{*e} - y_p^{*e}) a_{pq}^e \quad (4-34)$$

Die Matrix $V = (v_{il,kd})$ besitzt zwar die Eigenschaft

$$\begin{aligned} \sum_{l=1}^Z \sum_{i=1}^M v_{ik}^{ld} &:= \frac{1}{\alpha_k^d} \left[\sum_{l=1}^Z \left(\sum_j (w_{kj}^l R_{kj}^l - w_{jk}^l T_{jk}^l) - \sum_j (w_{kj}^l R_{kj}^l - w_{jk}^l T_{jk}^l) \right) + \sum_{i=1}^M w_{ki}^d S_{ki} \right] \\ &= \frac{1}{\alpha_k^d} \left[0 + \sum_{i=1}^M w_{ki}^d S_{ki} \right] = 1 \end{aligned} \quad (4-35)$$

Die Maximumsnorm der Iterationsmatrix ist gleich dem maximalen Wert der Summe der Absolutwerte der Elemente jeder Reihe (Gl. 4-36).

$$\|V\|_{\infty} = \max_{e,k} \sum_{i=1}^M \sum_{l=1}^Z |v_{ik}^{le}| \quad (4-36)$$

Man kann es finden, dass nicht alle $v_{ik}^{ld} \geq 0$ sind, weshalb die Maximumsnorm der Iterationsmatrix größer als 1 ist. Das normale Iterationsverfahren zur Lösung des Gleichungssystems (4-32) ist nicht einsetzbar. Um dieses Problem zu lösen, wird die Matrix V zerlegt. Die Matrix V besitzt eine Blockstruktur.

$$V = \begin{bmatrix} V^{11} & \cdot & \cdot & \cdot & V^{1Z} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ V^{Z1} & \cdot & \cdot & \cdot & V^{ZZ} \end{bmatrix}, \quad (4-37)$$

mit

$$V^{le} = \begin{bmatrix} V_{11}^{le} & \cdot & \cdot & \cdot & V_{1M}^{le} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ V_{M1}^{le} & \cdot & \cdot & \cdot & V_{MM}^{le} \end{bmatrix}, \quad (4-38)$$

und kann aus den zwei Teilblockmatrizen zusammengesetzt werden:

$$V = V_D + V_{LR} \quad (4-39)$$

mit der Diagonal-Block-Matrix

$$V_D = \begin{bmatrix} V^{11} & 0 & \cdot & \cdot & 0 \\ 0 & V^{22} & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & 0 \\ 0 & \cdot & \cdot & 0 & V^{ZZ} \end{bmatrix}, \quad (4-40)$$

und der Rest-Block-Matrix

$$V_{LR} = \begin{bmatrix} 0 & V^{12} & \cdot & \cdot & V^{1Z} \\ V^{21} & 0 & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot & V^{(Z-1)Z} \\ V^{Z1} & \cdot & \cdot & V^{Z(Z-1)} & 0 \end{bmatrix}. \quad (4-41)$$

Man kann aufgrund $v_{ik}^{ee} \geq 0$ den Betrag der Maximumsnorm von V_D festlegen mit

$$\|V_D\|_\infty = \|V^{ee}\|_\infty = \sum_{i=1}^M |V_{ik}^{ee}| = \sum_{i=1}^M \mathcal{G}_{ik}^{ee} = \frac{1}{\alpha_k^e} \left[0 + \sum_{i=1}^F w_{ki}^e S_{ki} \right] = 1, \forall k, e \quad (4-42)$$

Zudem sind die Diagonalmatrizen V^{ee} dominant [MATHIAS, 1999]. Es gilt

$$\sum_{l \neq e} \|V^{el}\|_\infty < \|V^{ee}\|_\infty = 1, \forall e \quad (4-43)$$

Diese zwei Block-Matrizen V_D und V_{LR} erfüllen die vorher genannte Voraussetzung, deshalb kann das bekannte Block-Jacobi-Verfahren zur Lösung des Gleichungssystems 4-32 verwendet werden. Gl. 4-33 kann in folgenden zwei Gleichungen dargestellt werden.

$$y_k^d(t + \tau_y) = \sum_{i=1}^M v_{ik}^{dd} y_i^d(t) - (b_k^d - z_k^d(t)), \quad \forall d, k \quad (4-44)$$

wobei $b_k^d := L_k / \alpha_k^d$.

$$\begin{aligned} z_k^d(t + \tau_y) &= \sum_{i=1}^M \sum_{l \neq d} v_{ik}^{ld} y_i^l(t + \tau_y) \\ &= \sum_{i=1}^M \sum_{l \neq d} v_{ik}^{ld} \left[\sum_{j=1}^M v_{ji}^{ll} y_j^l(t) - b_i^l \right] + \sum_{i=1}^M \sum_{l \neq d} v_{ik}^{ld} z_i^l(t) \\ &=: \theta_k^d + \sum_{i=1}^M \sum_{l \neq d} v_{ik}^{ld} z_i^l(t), \quad \forall d, k \end{aligned} \quad (4-45)$$

Diese zwei Gleichungen können als zwei miteinander kooperierende Fehlerpropagierungs-Netzwerke interpretiert werden. Diese zwei Fehlerpropagierungs-Netzwerke müssen solange kooperieren (siehe Abbildung 4-7), bis y konstant bleibt. Das gefundene y ist die Lösung des Gleichungssystems 4-32 und kann in Gl. 4-34 eingesetzt werden, damit die Änderungen der Verbindungsgewichte aktualisiert werden können.

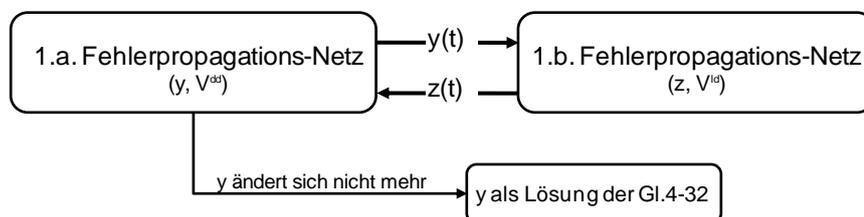


Abbildung 4-7: Die zwei kooperierenden Fehlerpropagierungs-Netzwerke

Das gesamte Verkehrsmodell besteht nun aus zwei Teilen, einem nichtlinearen, geschichteten RNN zur Simulation des Verkehrsnetzes und zwei zusammenarbeitenden, linearen Fehlerpropagierungs-Netzen zur Berechnung der Korrekturen für die Verbindungen.

dungsgewichte bzw. die Abbiegeraten des ersten Simulationsnetzes. Abbildung 4-8 zeigt der Struktur dieses RNN-Verkehrsmodells.

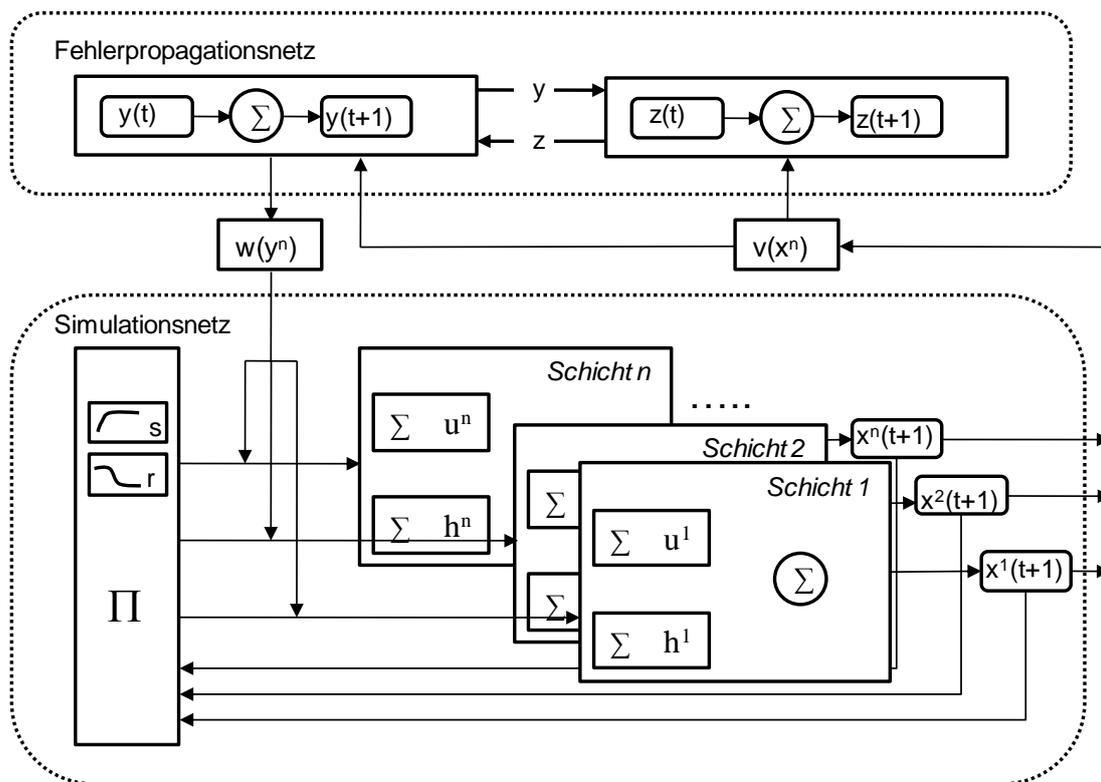


Abbildung 4-8: Das geschichtete nichtlineare Simulationsnetz und die zugehörigen zwei kooperierenden linearen Fehlerpropagations-Netze

Es sei hier darauf hingewiesen, dass die Berechnung der Gewichtskorrekturen im Falle des nicht vorgestellten Nutzeroptimums auch zwei Teile hat. Für den ersten Teil von L_k besitzt die Berechnung des Nutzeroptimums eine im Vergleich mit Gl. 4-20 etwas kompliziertere Form. Für den schwer zu berechnenden zweiten Teil von $\partial x_k^e / \partial w_{pq}^d$ bleibt die Gleichung unverändert für den Fall des Nutzeroptimums.

4.3 Das dynamische RNN-Umlegungsmodell

Das statische Verkehrsumlegungsmodell berücksichtigt bisher nur statische Quell-Ziel-Matrizen und Detektormesswerte. Das Simulations- und Fehlerpropagations-Netz kann zu einem Gleichgewichtszustand geführt und auch wechselseitig relaxiert werden. Im Gegensatz zu dem statischen Modell soll das dynamische Modell durch zeitlich veränderliche Quell-Ziel-Matrizen versorgt werden. Daher muss der Systemzustand in einer zeitlichen Entwicklung verfolgt werden.

Das dynamische RNN-Umlegungsmodell basiert auf dem bereits vorgestellten statischen Modell, im Gegensatz zum Fixpunktbetrieb des statischen Modells arbeitet das

dynamische Modell im Parallelbetrieb. Die zeitliche Entwicklung des Verkehrssystems kann durch permanentes Update der Abbiegeraten nachgebildet werden. Das Fehlerpropagierungs-Netz wird durch die in Unterkapitel 3.3.3 beschriebene Backpropagation für zeitabhängige RNN hergeleitet.

4.3.1 Erweitertes dynamisches Simulations-Modell

Im Gegensatz zum statischen Modell können sich die Quell-Ziel-Beziehungen im dynamischen Fall kontinuierlich ändern. Sie sind eine Funktion der Zeit: $q(t)$ und $z(t)$. In der Praxis ist die Erhebung der Daten sehr aufwendig, daher wird der Analysezeitraum in Abschnitte Δt unterteilt, die Längen liegen meist in der Größenordnung von 15 min bis 60 min [PLOSS, 1993].

Die Dynamik des dynamischen Simulationsnetzes soll diese zeitliche Änderung der Quell-Ziel-Beziehungen widerspiegeln. Es ist auch die einzige Änderung im Vergleich zum in Unterkapitel 4.1.3 definierten dynamischen Gesetz des statischen Modells. Man hat anhand des Schichtmodells insgesamt $M \cdot Z$ Gleichungen:

$$x_k^d(t + \tau_x) = x_k^d(t) + \sum_{i \in E(k)} w_{ik}^d a_{ik}^d(t) - \sum_{i \in A(k)} w_{ki}^d a_{ki}^d(t) + \tau_x q_k^d(t) \quad (4-46)$$

mit den zeitabhängigen und geschichteten Netzzuflüssen

$$q_k^d(t) := \begin{cases} q_k(t) \mu_k^d(t) & , k \in O \\ 0 & , \text{ansonst} \end{cases} \quad (4-47)$$

Hier ist $\mu_k^d(t) \in [0,1]$ die Quell (k)-Ziel (d)-Beziehung.

4.3.2 Optimierungsfunktion des dynamischen Modells

Die Optimierungsfunktion des dynamischen Modells ist auf das bekannte augenblickliche System-Optimum (ADSO) und auf das augenblickliche Nutzer-Optimum (ADNO) verteilt [HEYDECKER ET AL., 2006]. Diese beiden Kriterien beziehen sich auf den momentanen Systemzustand und berücksichtigen keine zukünftige Streckenreisezeitinformation. Im Falle des statischen Modells wurde nun das Systemoptimum beschrieben. Für den Fall des Systemoptimums gibt es fast keine Unterschiede zwischen statischem und dynamischem Modell. Die Optimierungsfunktion gemäß ADSO kann man wie folgt beschreiben:

$$\begin{aligned} E(x) &= \sum_k E_k(x_k) = \sum_k \pi_k(x_k(t)) f_k(x_k(t)) + \beta_k (x_k(t) - \hat{x}_k(t))^2 \\ &= \sum_k x_k(t) + \beta_k (x_k(t) - \hat{x}_k(t))^2 \quad \forall k \in S \end{aligned} \quad (4-48)$$

Die Optimierungsfunktion anhand ADNO ist deutlich komplizierter, hier wird dieser Aspekt nicht im Detail erläutert, da für die Verkehrssteuerung das Systemoptimum eingesetzt wird.

4.3.3 Das Fehlerpropagierungs-Netz und das Betriebsmodell

Analog zum statischen Modell kann das Fehlerpropagierungs-Netz mit Hilfe des Gradientenabstiegsverfahrens hergeleitet werden. Die partielle Ableitung der Systemoptimierungsfunktion nach den Verbindungsgewichten wird wie folgt ermittelt:

$$\Delta w_{pq}^e = -\varepsilon \sum_{k=1}^M \frac{\partial E_k}{\partial w_{pq}^e} = -\varepsilon \sum_{k=1}^M \sum_{d=1}^Z \frac{\partial E_k}{\partial x_k} \frac{\partial x_k^d}{\partial w_{pq}^e} = -\varepsilon \sum_{k=1}^M \sum_{d=1}^Z L_k \frac{\partial x_k^d}{\partial w_{pq}^e} \quad (4-49)$$

mit

$$L_k = 1 + 2\beta(x_k(t) - \hat{x}_k(t)) \quad (4-50)$$

Man kann erkennen, dass diese Ergebnisse identisch zum statischen Fall (Gl. 4-19, 4-20) sind. Die Berechnung der Anteile $\partial x_k^d / \partial w_{pq}^e$ sind die Aufgabe des Fehlerpropagierungs-Netzes, dessen Herleitung bereits aus Kapitel 4.2.2 bekannt ist. Der große Unterschied zum statischen Fall ist das Betriebsmodell. Aufgrund des zeitlich veränderlichen Verkehrszustands kann insbesondere das Simulations-RNN nicht relaxiert werden. Daher ist statt des Fixpunktbetriebs der Parallelbetrieb erforderlich.

Das gesamte System fängt mit einem Gleichgewichtszustand an, danach arbeitet das Modell im Parallelmodus, der durch einen Zyklus beschrieben werden kann. Ein ganzzahliges Paar $[z_s, z_f]$ soll für einen Zyklus festgelegt werden. Als ein Zyklus gilt, wenn das Simulations-RNN für z_s diskrete Zeitschritte τ_x und das Fehlerpropagierungs-RNN für z_f mal diskret Zeitschritte τ_y abwechselnd arbeiten. Die Abbildung 4-9 veranschaulicht den Überblick der beiden Betriebsmodelle.

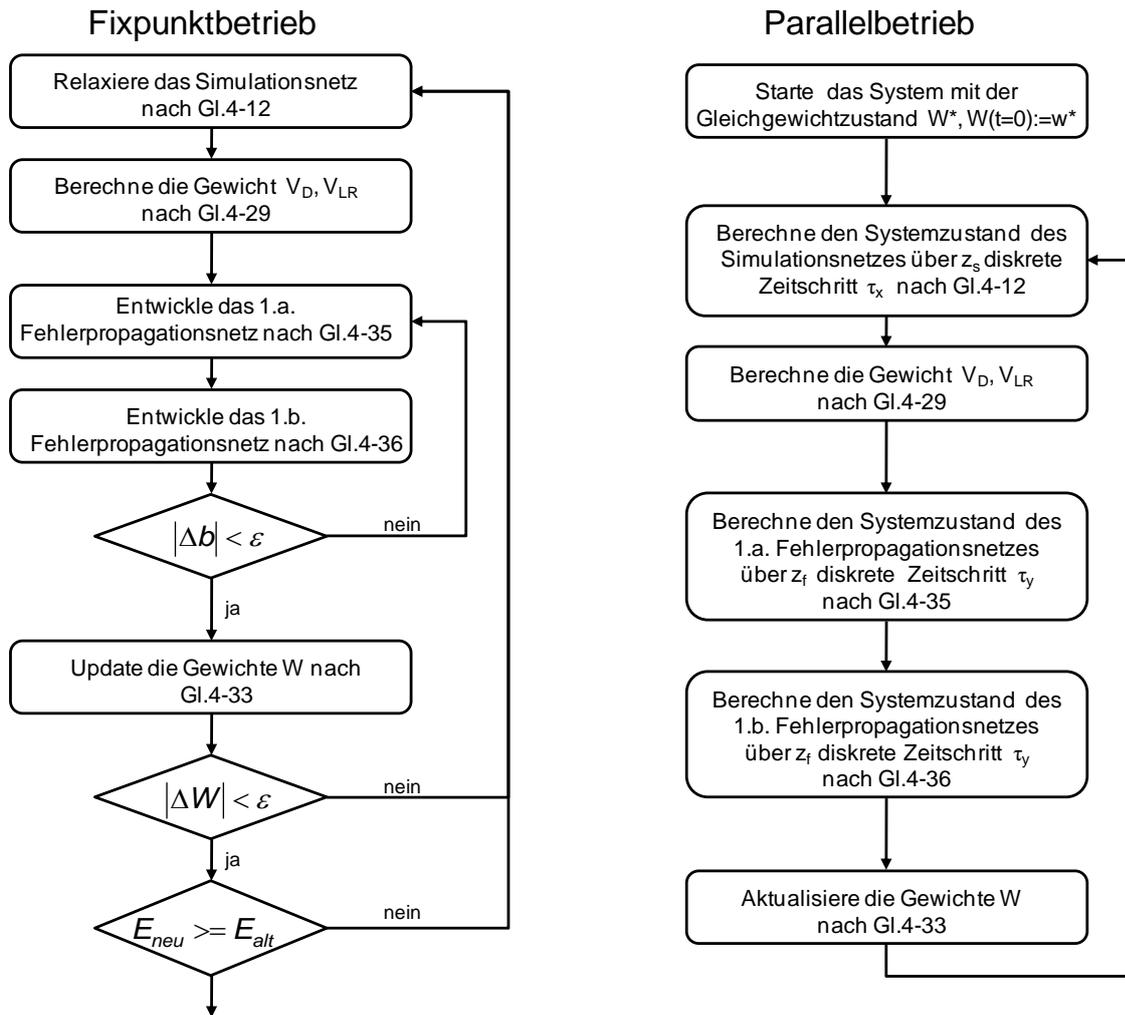


Abbildung 4-9: Die beide Betriebsmodelle von statischem Modell und dynamischem Modell

4.4 Zwischenfazit

In diesem Kapitel wurde ein auf RNN basiertes Verkehrsmodell vorgestellt, das als Grundlage eines netzweiten modellbasierten Verkehrssteuerungsverfahrens dient.

Insgesamt entstand so ein Verkehrsmodell, das aus zwei kooperierenden RNN besteht: Eins ist das Simulations-Netzwerk, das das reale Verkehrsnetz nachbildet. Das andere ist ein Fehlerpropagierungs-Netzwerk, das gewisse Steuerungsparameter bzw. die Abbiegeraten des Simulations-Netzwerks einstellt.

Das Simulations-Netzwerk ist ein nichtlineares System und besitzt eine Schicht-Struktur. Auf diese Weise kann die Integration der Verkehrsnachfrage implementiert werden. Die Aktivierungsfunktion setzt sich aus dem möglichen Abfluss und den Zuflussrestriktionen zusammen, die maßgeblich die Dynamik bestimmen. Der mögliche

Abfluss ist auch eine Funktion über Kapazitätswerte, die eine sehr große Rolle bei der Steuerung von LSA, wie in den nächsten Kapiteln beschrieben, spielen.

Entsprechend einer vorgegebenen Systemoptimierungsfunktion stellt ein lineares Fehlerpropagierungs-Netzwerk die Gewichte bzw. die Abbiegeraten des Simulations-Netzwerks ein. Die Herleitung des Fehlerpropagierungs-Netzwerks geht letztlich auf das häufig benutzte Gradientenabstiegsverfahren zurück.

Aufgrund der dynamischen Verkehrsnachfrage kommt der Parallelbetrieb bei dynamischer Umlegung zum Einsatz, da die Fixpunkte der RNN dann nicht mehr abgesichert werden können.

Das in diesem Kapitel vorgestellte Verkehrsmodell liefert die Grundgrößen, hauptsächlich die Verkehrsstärke jeder Strecke, für die Berechnung der Wirkungsgrößen und die Ermittlung und Optimierung der Signalprogramme. Im nächsten Kapitel sollen diese Aufgaben, die der wichtigste Teil eines modellbasierten netzweiten Verkehrssteuerungsverfahrens ist, ausführlich behandelt werden.

5 RNN-Verkehrssteuerungsmodell für Freigabezeitanteiloptimierung

Hauptgegenstand dieser Arbeit ist die modellbasierte Lichtsignalsteuerung in einem Straßennetz. Das in Kapitel 4 entwickelte Verkehrsmodell basiert auf einem RNN-Simulationsnetz und einem RNN-Fehlerpropagierungs-Netz. Das zweite RNN bzw. Fehlerpropagierungs-Netz hat das Ziel, die Abbiegeraten bzw. die Routenwahl in dem Simulationsnetz festzulegen. Die dynamische Gleichung des Simulationsnetzes enthält fest vorgegebene Kapazitätswerte, die von den aktuellen Signalprogrammen der Knoten abhängig sind. Darüber hinaus bietet es die Möglichkeit, die Signalprogrammparameter als neue Steuerungsparameter aufzufassen, die durch weitere neu entwickelte Fehlerpropagierungs-Netze justiert werden.

Dieses Kapitel soll sich nun auf einzelne Knotenpunkte der LSA-Steuerung konzentrieren. Für die lokale Steuerung wird hauptsächlich der Freigabezeitanteil optimiert. Man kann den quantitativen Zusammenhang zwischen Kapazitätswerten und Freigabezeitanteil in verschiedenen Fällen bestimmen. Wenn die Umlaufzeit vorher festgelegt wird, kann die optimale Freigabezeitverteilung bzw. der optimale Kapazitätswert aller auf einen Knoten zuführenden Links durch zweites Fehlerpropagierungsnetz auf Basis des Verkehrsmodells (das Simulationsnetz und erstes Fehlerpropagierungsnetz) ermittelt werden. Obwohl die LSA-Steuerung in diesem Kapitel nur am Einzelknoten betrachtet wird, bezieht sich die Optimierungsfunktion auf das gesamte Netz. Die Form der Optimierungsfunktion der Verkehrssteuerung ist nicht identisch mit der Umlegung, die im Unterkapitel 5.2 vorgestellt wird. Nach der Herleitung des zweiten Fehlerpropagierungs-Netzwerkes und dem Update der Freigabezeitoptimierung wird die Besonderheit des Betriebsmodells beschreiben.

5.1 Freigabezeitverteilung als Steuerungsparameter

Die im Kapitel 4.1.2 vorgestellte Dynamik des Simulationsnetzes ist durch den maximal möglichen Abfluss, die Zuflussbeschränkung und die Abbiegeraten beschrieben. Der maximal mögliche Abfluss bestimmt die Verkehrsmenge, die während eines Zeitintervalls abfließen könnte, wenn die Zuflussbeschränkungen aller weiterführenden Strecken gleich eins sind. Anhand der Form der Aktivierungsfunktion von RNN und War-

temodell der Verkehrstechnik wird der maximal mögliche Abfluss in Form der Gl. 5-1 definiert.

$$s_i(x_i) = \tau_x C_i \left(\frac{2}{1 + \exp(-ax_i)} - 1 \right) \quad (5-1)$$

Im bisher beschriebenen Verkehrsmodell wurde die in der Gl. 5-1 enthaltene Kapazität als fest vorgegebener Wert eingesetzt. Die Kapazität eines Fahrstreifens an einem signalgesteuerten Knoten, aus dem die Fahrzeuge unbehindert abfließen können, hängt von den aktuell laufenden Signalprogrammen ab. Im Fall verkehrsabhängiger Steuerungen oder adaptiver Steuerungen (z.B. MOTION, BALANCE) werden die Kapazitätswerte durch die Rahmensignalprogramme bestimmt.

Die Strecken, deren Kapazitäten abhängig vom Signalprogramm sind, sind im HBS [FGSV, 2001] als die Fahrstreifen einer Signalzufahrt definiert. Die Links im Knotenpunktbereich, die Zufahrt- und Ausfahrtstreifen verbinden (siehe Abbildung 5-1), werden in dieser Arbeit als Verbindungsstrecken im Knotenpunkt definiert. Man kann leicht erkennen, dass die Kapazitätswerte der Verbindungsstrecken auch vom Signalprogramm abhängen. Der Grund der Einführung der Verbindungsstrecken im Knotenpunkt ist, dass dies die Gesamtlängen der von dem Signalprogramm abhängigen Strecken verlängert. $F \subset S$ ist als die Menge solcher Strecken definiert. In dieser Arbeit werden die Zusammenhänge zwischen den Kapazitätswerten der Verbindungsstrecken und Signalprogrammen ausschließlich anhand HBS [FGSV, 2001] definiert.

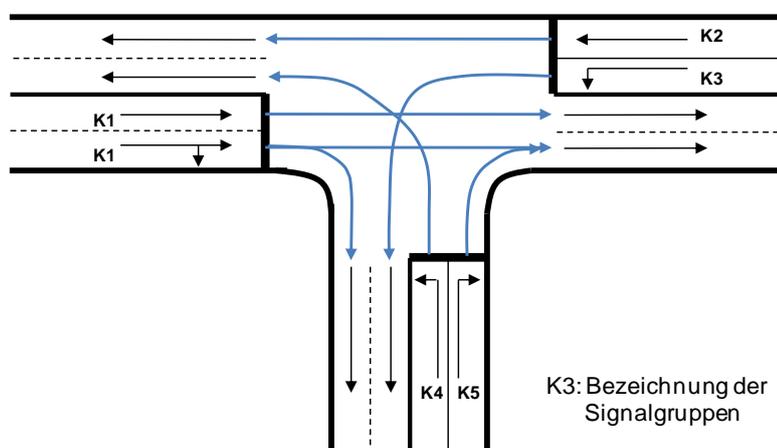


Abbildung 5-1: Beispiel der Verbindungsstrecken im Knotenpunkt

5.1.1 Kapazitätswerte des Geradeausfahrers

Der Zusammenhang zwischen Kapazitätswerten und Signalprogrammen soll für verschiedene Fälle ermittelt werden. Im Falle des Geradeausfahrers werden die Kapazitätswerte über den Freigabezeitanteil wie folgt definiert:

$$C_i = \frac{tg_i}{tu_i} S_i \quad i \in F \quad (5-2)$$

Hier ist S_i die Sättigungsverkehrsstärke der Strecke i , tg_i ist die Freigabezeit der Strecke i bzw. der dem Fahrstrom zugehörigen Signalgruppe, tu_i ist die Umlaufzeit der Strecke i zugehörigen Knotenpunkte. Die Kapazitätswerte sind proportional zum Anteil der Freigabezeit an der gesamten Umlaufzeit. Die Sättigungsverkehrsstärke S_i ist ein vom aktuellen Verkehrszustand unabhängiger konstanter Wert, der durch die Sättigungsverkehrsstärke unter Standardbedingungen und Angleichungsfaktoren berechnet wird. Die Angleichungsfaktoren sind hauptsächlich von straßenbaulichen Gegebenheiten bestimmt. Dadurch sind die Kapazitätswerte eindeutig durch die Freigabezeitanteile des Signalprogramms bestimmt. Die Ableitung, die später beim Gradientenverfahren verwendet wird, kann man berechnen mit:

$$\frac{\partial C_i}{\partial tg_i} = \frac{1}{tu_i} S_i \quad i \in F \quad (5-3)$$

5.1.2 Kapazitätswerte des Abbiegers

Für den Abbieger ist die Berechnung nicht gleich wie für den Geradeausfahrer, da der Kapazitätswert der bedingt verträglichen Abbieger nicht nur vom Freigabezeitanteil sondern auch von anderen Faktoren z.B. Fußgänger, Gegenverkehr, etc. bestimmt wird. In diesem Unterkapitel soll hierfür ein Berechnungsschema angeführt werden.

Wenn die Abbieger in einer eigenen Phase konfliktfrei geführt werden, sollen die Kapazitätswerte dieser Ströme wie Geradeausfahrer behandelt werden. Im diesem Fall wird lediglich der Abbiegeradius zur Bestimmung der Angleichungsfaktoren beachtet. Bei bedingt verträglichen Abbiegern hingegen muss auf die Berechnung der Kapazitätswerte gesondert eingegangen werden.

Bedingt verträgliche Rechtsabbieger

Bedingt verträgliche Rechtsabbieger sind gemäß StVO auf den parallel gerichteten Fußgängerstrom wartepflichtig. Erst nach dem von Fußgängern blockierten Freigabezeitanteil können die Rechtsabbieger abfließen. Um die Kapazitätswerte der bedingt verträglichen Rechtsabbieger zu ermitteln, muss zunächst die Blockierungszeit der Fußgänger berechnet werden. Über die Anzahl der Fußgänger pro Umlauf P kann die Blockierungszeit t_{fuss} nach Gl. 5-4 kalkuliert werden.

$$t_{fuss} = \frac{P}{0.024P + 0.48} - t_{vor} \quad (5-4)$$

mit $t_{vor} =$ Zeitvorsprung für Fußgänger 1 bis 2 s

Bei schwachem Fußgängerverkehr kann man mit einer Blockierungszeit von $t_{fuss}=4$ bis 8 s rechnen. In dieser Arbeit wird t_{fuss} als 6s angenommen, wenn die Fußgängerverkehrsstärke nicht bekannt ist. Daraus bestimmt sich die fußgängerfreie Freigabezeit tg_u zu

$$tg_{Fu} = \max \begin{cases} t_g - t_{fuss} - N_A t_B \\ 0 \end{cases} \quad (5-5)$$

mit

tg_{Fu}	=	fußgängerfreie Freigabezeit für Rechtsabbieger
N_A	=	Anzahl der Aufstellplätze zwischen Haltelinie und Fußgängerfurt
t_B	=	Zeitbedarfswert für einen Rechtsabbieger

Der Kapazitätswert für die Rechtsabbieger kann nach folgender Gleichung bestimmt werden:

$$C_{RA} = \min \begin{cases} S \cdot tg_{Fu} / t_U + N_A \cdot 3600 / t_U \\ S \cdot tg / t_U \end{cases} \quad (5-6)$$

mit t_U = Umlaufzeit.

Weil N_A von den straßenbaulichen Gegebenheiten abhängt, ist der Kapazitätswert des Rechtsabbiegers eine eindeutige Funktion des Signalprogramms bzw. der Freigabezeit und der Umlaufzeit. Die Ableitung lässt sich auch in folgender geschlossener Form ausdrücken:

$$\frac{\partial C_{RA,i}}{\partial tg_i} = \frac{1}{t_U} S_i \quad \forall i \in F \quad (5-7)$$

Im amerikanischen Regelwerk HCM 2000 [TRB, 2000] wird eine andere Formel zur Bestimmung von t_{fuss} verwendet, aber die Formel für die darauf folgende Kapazitätsberechnung ist identisch und liefert auch zahlenmäßig fast die gleichen Werte [Wu, 2003].

Bedingt verträgliche Linksabbieger

Die Kapazität für bedingt verträgliche Linksabbieger besteht aus zwei Teilen, der erste Teil ermittelt die Linksabbieger, die während ausreichender Zeitlücken den Gegenverkehrsstrom durchsetzen können. HBS 2001 geht von der Annahme aus, dass sich der Gegenverkehr über die Gesamtlänge der Freigabezeit gleichmäßig verteilt. Nach der Theorie der Grenzzeitlücken wird der erste Teil der Kapazitätswerte berechnet. Der zweite Teil bestimmt die Kapazität durch Phasenwechseln. Daraus ergibt sich der Kapazitätswert der bedingten verträglichen Linksabbieger zu:

$$C_{LA} = C_D + C_{PW} \quad (5-8)$$

mit

$$C_D = \frac{3600 \cdot t_g / t_U - q \cdot t_c}{t_f} e^{-q(t_0 - t_c) / (3600 \cdot t_g / t_U - q \cdot t_c)} \quad (5-9)$$

und

$$C_{PW} = N_A \cdot 3600 / t_U \quad (5-10)$$

Hier ist N_A der gleiche Parameter wie im Fall der Rechtsabbieger, $t_c=1,8$ s, $t_f=3,0$ s und $t_0=4,2$ s sind feste Parameter der Zeitlücke und q ist die Stärke des Gegenstroms. Das später vorgestellte Gesamtbetriebsmodell läuft im Parallelbetrieb, das heißt, dass die Stärke aller Ströme bekannt ist und als Grundlage für die Verkehrssteuerung dient. Daher kann q als bekannter Wert betrachtet werden. Daraus kann man erkennen, dass der Kapazitätswert hier auch eine eindeutige Funktion abhängig von Signalprogramm bzw. Freigabezeit und Umlaufzeit ist.

Die Ableitung für spätere Anwendung lässt sich auch in folgender Form ausdrücken:

$$\frac{\partial C_{LA,i}}{\partial t g_i} = e^{-q(t_0 - t_c) / (3600 \cdot t_g / t_U - q \cdot t_c)} \left(\frac{3600}{t_U t_f} + \frac{3600 q (t_0 - t_c)}{(3600 t_g / t_U - q \cdot t_c) t_U t_f} \right) \quad \forall i \in F \quad (5-11)$$

Im HCM 2000 wird die reale Ganglinie des Gegenverkehrs berücksichtigt, weshalb die Kapazität C_D eine ganz andere Form besitzt. Die Kapazität durch Phasenwechsel wird im Gegensatz zum HBS nicht addiert, sondern als Mindestkapazität eingesetzt. Auf die Details wird hier nicht weiter eingegangen. Das Ergebnis jedoch zeigt, dass der Kapazitätswert nur von der Freigabezeit und der Umlaufzeit abhängig ist.

5.1.3 Festlegung der Umlaufzeit

Die Umlaufzeit ist die Summe von maßgebenden Freigabezeiten und dazwischen liegenden erforderlichen Zwischenzeiten. Die erforderlichen Zwischenzeiten sind für den Entwurf des aktuellen Signalprogramms ein konstanter Wert. Die Leistungsfähigkeit der signalgesteuerten Knoten entspricht der Summe der Kapazitätswerte der Strecken. Man kann leicht beweisen, dass die Leistungsfähigkeit der Knoten eine monoton steigende Funktion über die Umlaufzeit ist. Das heißt, je länger die Umlaufzeit ist, desto größer ist die Leistungsfähigkeit der Knotenpunkte. Die Umlaufzeit ist als neuer Steuerungsparameter des Fehlerpropagierungsnetzes nicht sinnvoll, da die Umlaufzeiten verschiedener Knotenpunkte nicht in einer mathematischen Beziehung dargestellt werden können. Ohne die mathematische Beziehung wird die Umlaufzeit aller Knoten einen Grenzwert erreichen, damit den Kapazitäten aller Knoten ein Maximalwert zugewiesen werden kann.

Im Gegensatz zur Umlaufzeit kann der Freigabezeitanteil als Parameter netzweiter Verkehrssteuerungen aufgefasst werden, weil die Freigabezeitanteile der einzelnen Phasen bzw. der einzelnen Strecken nicht voneinander mathematisch unabhängig sind.

Bevor der Freigabezeitanteil als neuer Parameter eingeführt wird, sollen die Umlaufzeiten mit konventionellen Methoden festgelegt werden.

Zur Berechnung der Umlaufzeit wurden bereits zahlreiche Verfahren entwickelt, denen die Warteschlangentheorie zugrunde liegt. Anhand der Berücksichtigung der Einflussgröße unterscheidet man die Ermittlung der Umlaufzeit am Einzelknotenpunkt von der Ermittlung der optimalen Umlaufzeit für die Knotenpunkte eines Straßenzuges oder Straßennetzes. HBS 2001 und HCM 2000 stellen jeweils die zur Ermittlung der Umlaufzeit am Einzelknotenpunkt gebräuchlichen Verfahren vor. Im HBS 2001 werden sowohl die wartzeitoptimale Umlaufzeit (Gl. 5-12) als auch die Mindestumlaufzeit (Gl. 5-13) erläutert.

$$t_{U,Opt} = \frac{1,5 \cdot T_Z + 5}{1 - \sum_{i=1}^p q_{maßg,i} / S_i} \quad (5-12)$$

$$t_{U,Min} = \frac{T_Z}{1 - \sum_{i=1}^p q_{maßg,i} / (g_i \cdot S_i)} \quad (5-13)$$

mit	T_Z	=	Summe der maßgebenden Zwischenzeit
	p	=	Anzahl der Phasen
	$q_{maßg,i}$	=	maßgebende Fahrstreifenverkehrsstärke der Phase i
	S_i	=	zugehörige Sättigungsverkehrsstärke für den maßgebenden Fahrstreifen der Phase i
	g_i	=	Sättigungsgrad für den maßgebenden Fahrstreifen der betreffenden Phase.

Im HCM 2000 wird die gleiche Formel wie Gl. 5-13 verwendet, daher berechnet das HCM 2000 nur die minimale Umlaufzeit mit einem vorgegebenen Sättigungsgrad. Die Mindestumlaufzeit nach Gl. 5-13 kommt auch in dieser Arbeit zum Einsatz.

Zur Bestimmung der Umlaufzeit mehrerer Knotenpunkte in einem Straßennetz kann man entweder einen einheitlichen Wert oder unterschiedliche Werte verwenden. Das HBS enthält ebenso wie das HCM keinen Verfahrensvorschlag zur Festlegung unterschiedlicher Umlaufzeiten. BOLTZE [1988] entwickelte ein Optimierungsverfahren zur Berechnung unterschiedlicher Umlaufzeiten in einem Straßennetz, in dem alle Verkehrsteilnehmergruppen und alle relevanten Zielgrößen für die Kraftfahrzeuge angemessen berücksichtigt werden. Mit Hilfe der Methode der Verbesserungsheuristiken wird dabei die Umlaufzeit optimiert. Da sich die Optimierungsmethode dieser Arbeit auf KNN konzentrieren und der Verfahrensvorschlag von BOLTZE nicht für diese Art der Optimierung einsetzbar ist, wird die Umlaufzeitberechnung ohne dem Verfahrensvorschlag von BOLTZE durchgeführt.

In Deutschland ist die Anwendung der einheitlichen Umlaufzeit weit verbreitet. In den RiLSA [1992] wird im Kapitel „Grüne Welle“ der quantitative Zusammenhang zwischen

Umlaufzeit für die Knotenpunkte eines Straßenzuges, Progressionsgeschwindigkeit und Teilpunktastand beschrieben (Gl. 5-14).

$$t_U = \frac{x}{V_{Pr, Richtung}} + \frac{x}{V_{Pr, Gegenrichtung}} \quad (5-14)$$

mit x = Abstand zwischen benachbarten Knotenpunkten
 $V_{Pr, Richtung}$ = Progressionsgeschwindigkeit in Hinrichtung
 $V_{Pr, Gegenrichtung}$ = Progressionsgeschwindigkeit in Gegenrichtung

Das bis heute bekannteste Berechnungsverfahren zur Bestimmung einer einheitlichen Umlaufzeit für Knotenpunkte eines Straßennetzes wurde von ALLSOP [1972] beschrieben. In der Regel kommt diese einheitliche Umlaufzeit aus der minimal erforderlichen Umlaufzeit für den am stärksten ausgelasteten Knotenpunkt in dem Straßennetz. Anhand dieses Vorgehens optimiert beispielsweise die Netzsteuerung SCOOT die Umlaufzeit von Teilgebieten [FRIEDRICH, 2002]. Die vorliegende Arbeit verwendet auch diesen Ansatz für die Ermittlung der einheitlichen Umlaufzeit.

Um die Umlaufzeit zu bestimmen, sollen zuerst die Phaseneinteilungen für die Knotenpunkte festgelegt werden. Die Phaseneinteilung wird in dieser Arbeit nicht zeitabhängig optimiert, sondern bleibt während der Optimierung anderer Parameter des Signalprogramms konstant. Aus dem Verkehrsmodell kann man die für die Berechnung der Mindestumlaufzeit nach Gl. 5-13 erforderlichen Verkehrsstärken erhalten. Mit der vorgegebenen Phaseneinteilung für jeden Knotenpunkt, die maßgebende Fahrstreifenverkehrsstärke der jeweiligen Phase und dem festgelegten maximalen Sättigungsgrad von 90% errechnet sich die Mindestumlaufzeit jedes Knotenpunkts. Darüber hinaus sollte die Umlaufzeit gemäß RiLSA zwischen 30 s und 120 s liegen. Im Rahmen dieses unteren und oberen Grenzwerts kann die maximale Mindestumlaufzeit für einen einzelnen Knotenpunkt als die einheitliche Umlaufzeit für das gesamte Netz gewählt werden.

Die Ermittlung der optimalen einheitlichen Umlaufzeit erfolgt in der Regel im Abstand von 180 s. Der Grund liegt darin, dass in dem parallelen Betriebsmodell ein Zyklus durch den gewählten Zyklusparameter 90 s beträgt. Daher kann die Umlaufzeit jeden zweiten Zyklus bzw. alle 180 s angepasst werden. Diese Realisierung wird im späteren Kapitel 6 ausführlich erklärt.

Nach der Festlegung der Umlaufzeit kann man erkennen, dass die Kapazitätswerte (Gl. 5-2, 5-6, 5-8) durch die Freigabezeit eindeutig abgebildet werden. Das ist der Grund, warum die Freigabezeitverteilung als Steuerungsparameter zur Entwicklung eines neuen Fehlerpropagierungsnetzes aufgefasst werden kann. Da Freigabezeiten der feindlichen Zufahrt im Signalprogramm disjunkt sind, verursacht die Vergrößerung der Freigabezeit einer Phase in der Regel die Verkleinerung der Freigabezeit anderer Phasen des Knotenpunkts. Dieser Zusammenhang wird für die Berechnung der Freigabezeitverteilung im späteren Kapitel herangezogen. Vor der Ermittlung der Freigabezeitverteilung wird zunächst auf die erforderliche Optimierungsfunktion für die dynamische Verkehrssteuerung eingegangen.

5.2 Optimierungsfunktion für Freigabezeitoptimum

Im Unterkapitel 3.3.2 wurde erläutert, dass die Optimierungsfunktion des RNN eine beliebige, in x stetig differenzierbare Funktion sein muss. Im Unterkapitel 4.2.1 wurde zudem eine Systemoptimierungsfunktion für die Umlegung vorgestellt. Im folgenden Abschnitt wird darauf aufbauend eine neue Optimierungsfunktion für das Freigabezeitoptimum abgeleitet. Zunächst wird die Optimierungsfunktion entsprechend der Zielvorstellung formuliert. Da die Variablen in der Optimierungsfunktion des Steuerungsverfahrens als Entscheidungskriterien für die Optimierung dienen, werden anschließend die ausgewählten Variablen (Wartezeiten) mit Hilfe eines Wartemodells bestimmt.

5.2.1 Formulierung der Optimierungsfunktion

Im Hinblick auf das zweite Wardrop'sche Prinzip wird bei der Optimierungsfunktion zwischen nutzer- und systemoptimaler Lösung unterschieden. Das Nutzeroptimierungsmodell hat das Ziel, das tatsächliche Verhalten des Verkehrsteilnehmers und des Verkehrsgeschehens eines Straßennetzes zu beschreiben. Statt des Systems werden die Belange der jeweils Betroffenen berücksichtigt. So kann beispielsweise die mittlere Wartezeit in allen Zufahrten minimiert werden, z.B.:

$$\min! E = \sum_i c_i \quad (5-15)$$

Hier ist c_i der verallgemeinerte durchschnittliche Kostenwert für den Link i . Da das Nutzeroptimierungsmodell zur Abbildung der realen Verkehrssituation führt, ist es meist nicht in der Lage, die Ziele des Systembetreibers (z.B. Bündelung des Verkehrs auf Hauptstraßen) widerzuspiegeln.

Der Betreiber des Verkehrssystems hat im Gegensatz zum „egoistischen“ Verhalten der Verkehrsteilnehmer das Ziel, die Systemkosten des Straßennetzes zu minimieren. Bei der Systemoptimierung

$$\min! E = \sum_{i \in F} c_i \cdot f_i \quad \text{mit } f_i: \text{Verkehrsstärke} \quad (5-16)$$

werden die Hauptrichtungen der Signalprogramme so definiert, dass die auf eine Kreuzung zufahrenden Verkehrsteilnehmer (Nutzer) so viel wie möglich Freigabezeit erhalten. Im Gegensatz dazu müssen in den Zufahrten der Nebenrichtung größere Wartezeiten in Kauf genommen werden. Dieser Effekt entspricht dem vorher genannten Ziel der Stadtplanung. Aber eine systemoptimale Steuerung birgt die Gefahr, dass das Interesse des gesamten Verkehrssystems nicht die Akzeptanz des Verkehrsteilnehmers findet.

Aus diesem Grund sollte eine Optimierungsfunktion gesucht werden, die die Vorzüge beider Kriterien vereinen kann. Das heißt, dass sowohl die Belange einzelner Verkehrsteilnehmer als auch die Interessen des Systembetreibers berücksichtigt werden. FRIEDRICH ET AL. [2002] stellen einen Ansatz für diese Anforderung vor:

$$\min! E = \sum_{i \in F} \alpha_i \cdot c_i \cdot f_i \quad \text{mit } \alpha_i: \text{Gewicht der Streckengruppen} \quad (5-17)$$

Diese Optimierungsfunktion wird strategisches Systemoptimum genannt. Die Gewichte können durch die Dimension der Straßenkategorie bestimmt werden. So können z.B. die Hauptverkehrsstraße und die Erschließungsstraße in einer Straßenkategorie definiert werden. Über diese Erweiterung der Optimierungsfunktion kann eine suboptimale Lösung des Gesamtsystems realisiert werden, damit die Grundsätze der in Deutschland verfolgten Strategien des ‚Stadtfahrplans‘ gewährleistet sind.

Für die verallgemeinerten durchschnittlichen Kosten existieren mehrere mögliche Wirkungskriterien. Dabei sind folgende Wirkungskomponenten von den in Kapitel 2 vorgestellten adaptiven Steuerungsverfahren für die Optimierungsfunktion angewendet worden:

- Wartezeiten des mIV und des ÖPNV
- Staulänge des mIV , die direkt von den Wartezeiten abhängig sind
- Anzahl der Halte des mIV und des ÖPNV

Da die Lichtsignalsteuerung in dieser Arbeit nur den mIV berücksichtigen soll und das Verkehrsmodell nicht alle Wirkungsgrößen berechnen kann, soll die Optimierungsfunktion die Wartezeit als das wesentliche Kriterium enthalten. Damit ergibt sich die strategische Systemoptimierungsfunktion

$$\min! E = \sum_{i \in F} \alpha_i \cdot w_i^{mIV} \cdot f_i \quad (5-18)$$

5.2.2 Berechnung der Wartezeit

Die signalgruppenbezogenen Wartezeiten gelten als wichtigstes Kriterium für die Optimierungsfunktionen zur Freigabezeitberechnung. Die Wartezeiten können ausgehend von den Warteprozessen mit Modellen der Warteschlangentheorie bestimmt werden. Die Warteschlangentheorie unterscheidet zwischen den drei Komponenten Ankunftsprozess, Bedienungssystem und Warteordnung, deren Eigenschaften die Wartemodelle klassifizieren.

Deterministische Wartemodelle –D/D/1

Die einfachste und auch sehr häufig eingesetzte Methode zur Analyse der Wartezeiten ist die Annahme deterministischer Zu- und Abflussgesetze. Bei der deterministischen Betrachtungsweise setzen sich die Wartezeiten an Lichtsignalanlagen aus zwei Teilen

zusammen: Eine ist durch die Servicezeiten der LSA verursacht, die andere ist die Folge einer Überlastung ($r > 1$). Der Fall der Überlastung wird in dieser Arbeit nicht berücksichtigt, daher können die Wartezeiten ausschließlich durch die Servicezeiten der LSA verursacht werden. CLAYTON [1941] berechnet die mittleren individuellen Wartezeiten wie folgt:

$$w(r) = \frac{t_U/tg}{t_U/tg - \rho} \cdot \frac{(t_U - tg)^2}{2t_U} \quad \text{mit} \quad \rho = \frac{f}{s \cdot tg/t_U} \quad (5-19)$$

Stochastische Wartemodelle –M/D/1

Die stochastischen Wartemodelle setzen Ankunfts- und Serviceraten voraus, die über den Betrachtungszeitraum konstant sind, jedoch um ihre Mittelwerte schwanken können. Das Wartemodell M/D/1 geht von Poisson verteilten Ankünften und konstanten Servicezeiten aus. Die Wartezeiten werden in zwei Teilprozesse unterteilt. Der erste Teil wird durch die effektive Sperrzeit der LSA verursacht, die daraus sich ergebenden Wartezeiten sind identisch zum Fall des deterministischen Modells. Zusätzliche Wartezeiten entstehen in einem anderen Teilprozess durch stochastische Schwankungen der Verkehrsstärke q der Ankunft gegenüber der Servicerate.

WEBSTER [1958] bestimmt für einen Poisson verteilten Zufluss die durchschnittliche Wartezeit eines Fahrzeuges mit Gl. 5-20:

$$w(\rho) = \frac{t_U/tg}{t_U/tg - \rho} \cdot \frac{(t_U - tg)^2}{2t_U} + \frac{\rho}{2q(1-\rho)} - 0,65 \left(\frac{tg}{q^2} \right)^{\frac{1}{3}} \rho^{\left(2 + \frac{5tg}{t_u} \right)} \quad (5-20)$$

Anstelle von Gl. 5-20 kann in guter Näherung vereinfacht folgende Gleichung [WEBSTER 1958] in der Praxis verwendet werden:

$$w(\rho) = 0.9 \left[\frac{t_U/tg}{t_U/tg - \rho} \cdot \frac{(t_U - tg)^2}{2t_U} + \frac{\rho^2}{2q(1-\rho)} \right] \quad (5-21)$$

5.2.3 Dynamische Optimierungsfunktion

In Kapitel 4.3.2 verwendet die Verkehrsumlegung ein augenblickliches dynamisches System-Optimum (ADSO). Anhand der vorherigen Formulierung der Optimierungsfunktion soll hier ein augenblickliches strategisches System-Optimum festgelegt werden. Durch Einsetzen von Gl. 5-21 in Gl. 5-18 erhält man

$$E(x) = \sum_{k \in F} \alpha_k f_k(x_k) 0.9 \left[\frac{t_U/tg_k}{t_U/tg_k - \rho_k} \cdot \frac{(t_U - tg_k)^2}{2t_U} + \frac{\rho_k^2}{2f_k(1-\rho_k)} \right] \quad (5-22)$$

wobei ρ_k sich wie folgt bestimmen lässt:

$$\rho_k = \frac{f_k(x_k)}{S_k \cdot t g_k / t_U} \quad (5-23)$$

Mit der Definition einer neuen Variablen

$$g_k = t g_k / t_U \quad \text{und} \quad b_k = f_k / S_k \quad (5-24)$$

und Umsortieren kann man die Gl. 5-22 in folgender Form schreiben:

$$E(x(t)) = \sum_{k \in F} \alpha_k f_k(x_k) 0.9 \left[\frac{t_U \cdot (1 - g_k)^2}{2(1 - b_k)} + \frac{\rho_k^2}{2f_k(x_k) \cdot (1 - \rho_k)} \right]. \quad (5-25)$$

Die Gl. 5-25 ist die sogenannte strategische Systemoptimierungsfunktion, die das Ziel hat, ausgehend von den augenblicklich vorgefundenen Verhältnissen den Systemzustand so einzustellen, dass die Summe der gewichteten Wartezeiten minimiert wird. Wie der Begriff „strategisch“ schon andeutet, orientiert sich das Systemoptimum nicht an einem fixierten Gleichgewichtszustand, sondern an dem augenblicklichen Systemstatus.

Im Falle der Umlegung basiert das Fehlerpropagierungs-Netzwerk auf einem Gradientenabstieg über der x -dimensionalen Potentialfläche der vorgegebenen Optimierungsfunktion. Die durch Gl. 5-25 definierte Potentialfläche ist unter anderem von der veränderlichen O-D-Matrix abhängig, daher ist diese Potentialfläche auch nicht konstant und bewegt sich ständig. Aus diesem Grund sind die Minima bzw. die Gleichgewichtspunkten der Optimierungsfunktion in Bewegung. Das neu entwickelte Fehlerpropagierungs-Netzwerk versucht durch permanentes Nachführen der Steuerungsparameter bzw. der Freigabezeitverteilung, dass die vorgegebene Optimierungsfunktion dem augenblicklichen Optimum folgt. Das bedeutet auch, einer optimalen zeitlichen Trajektorie zu folgen. Die optimale zeitliche Trajektorie ist der Weg im x -dimensionalen Raum, der für jeden Zeitpunkt die Optimierungsfunktion minimiert. Wenn die Rahmenbedingungen bzw. die Netzzuflüsse über einen Zeitraum unverändert bleiben, kann der Gleichgewichtspunkt bzw. das Minimum des Systems erreicht werden. Die Abbildung 5-2 veranschaulicht ein schematisches Beispiel über die Bewegung einer Potentialfläche und die zeitlich optimale Trajektorie.

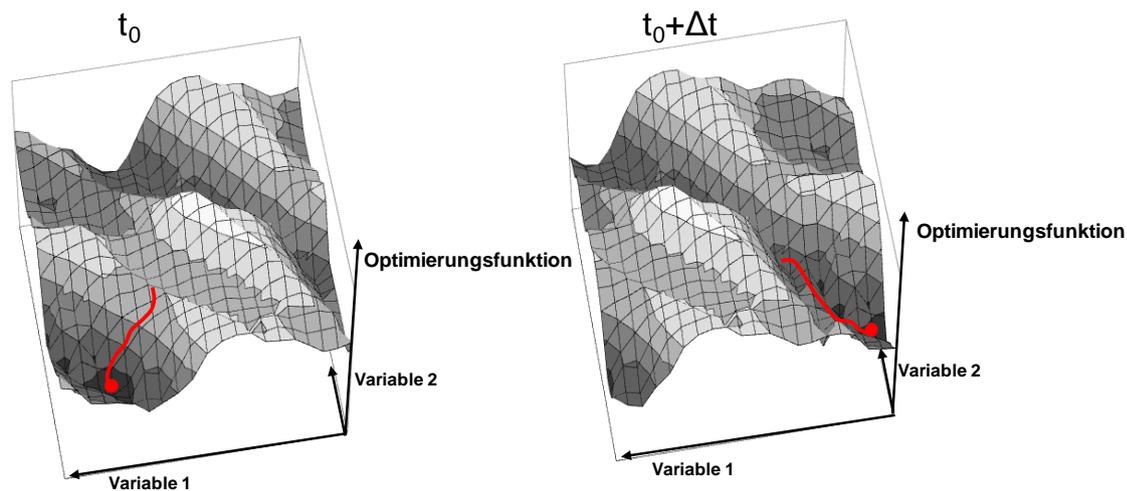


Abbildung 5-2: Die Bewegung der Potentialflächen und die optimale Trajektorie

Hier stellt sich jedoch die Frage, ob der sich entlang der optimalen Trajektorie bewegte Systemzustand in jeder kleinen Umgebung des optimalen Zustands, mit Hilfe des Fehlerpropagierungs-Netztes mitgeführt werden kann. Diese Frage konnte mathematisch bisher nicht beantwortet werden, allerdings liefern die bis jetzt durchgeführten numerischen Experimente, sowohl der Umlegung [MATHIAS, 1999] als auch der späteren Lichtsignalsteuerung, die im Rahmen dieser Arbeit durchgeführt werden, zufriedenstellende Ergebnisse, denn bei den numerischen Experimenten wird der Systemstatus bzw. die Optimierungsfunktion immer in die Nähe des optimalen Zustands bzw. des Gleichgewichtspunkts geführt.

5.3 Erweiterungen des RNN-Modells zur Optimierung der Freigabezeitverteilung

Im Kapitel 5.2 wurden die Freigabezeitanteile bzw. die Kapazitätswerte als neue Menge von Steuerungsparameter des Simulations-Netzwerks interpretiert. Die Freigabezeitanteile sollen so eingestellt werden, dass die festgelegte Optimierungsfunktion (Gl. 5-25) minimiert wird. Im folgenden Abschnitt wird ein neues Fehlerpropagierungs-Netzwerk konstruiert, um die Änderung der Freigabezeit zu bestimmen. Auf Basis der Randbedingung der Freigabezeit werden die Freigabezeitverteilungen aktualisiert.

5.3.1 Berechnung der Freigabezeitkorrekturen

Zur netzweiten Optimierung der Freigabezeitverteilung bzw. der Kapazitätswerte wird ein neues Fehlerpropagierungs-Netzwerk entwickelt. Der Ansatz zur Lösung der Aufgabe ist, die Kapazitätswerte als neue Steuerungsparameter des Simulations-Netzwerks zu verwenden. Sie bestimmen über den maximal möglichen Abfluss die Dynamik des Simulations-Netzwerks.

Das Simulationsnetz kann unverändert aus dem Verkehrsmodell übernommen werden. Das bekannte Gesetz der Dynamik bleibt wie folgt:

$$x_k^d(t + \tau_x) = x_k^d(t) + \sum_{i \in E(k)} w_{ik}^d a_{ik}^d(t) - \sum_{i \in A(k)} w_{ki}^d a_{ki}^d(t) + \tau_x q_k^d(t) \quad \forall k \in S \quad (5-26)$$

Diese Gleichung hängt von den Freigabezeiten über den maximalen möglichen Abfluss ab.

$$s_i^d(x_i) = s_i(x_i, C_i) \frac{x_i^d}{x_i} = s_i(x_i, C_i(tg_i)) \frac{x_i^d}{x_i} \quad \forall i \in F \quad (5-27)$$

Anhand des parallelen Betriebs kann man davon ausgehen, dass die Abbiegeraten unter Verwendung des ersten Fehlerpropagierungs-Netzes festgehalten werden. Es heißt auch, dass der Verkehrsfluss jeder Strecke im Netz bestimmt ist und hier im Bereich der Verkehrssteuerung als unveränderlich angenommen werden kann.

Zur Lösung der Korrekturen der Kapazitätswerte bzw. der Freigabezeitanteile wird analog zu den Abbiegeraten wieder das Gradientenabstiegsverfahren herangezogen. Die Änderung der Freigabezeit Δtg_k ist proportional zur Ableitung der Optimierungsfunktion E nach dieser Freigabezeit.

$$\begin{aligned} \Delta tg_t &= -\varepsilon \sum_{k=1}^F \frac{\partial E_k}{\partial tg_t} \\ &= -\varepsilon \sum_{t=1}^F \sum_{k=1}^F \left(\begin{aligned} &\frac{\partial E_k}{\partial tg_t} + \frac{\partial E_k}{\partial g_k} \frac{\partial g_k}{\partial tg_t} + \frac{\partial E_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial tg_t} + \frac{\partial E_k}{\partial f_k} \sum_e^Z \frac{\partial f_k}{\partial x_k^e} \frac{\partial x_k^e}{\partial c_t} \frac{\partial c_t}{\partial tg_t} \\ &+ \frac{\partial E_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial f_k} \sum_e^Z \frac{\partial f_k}{\partial x_k^e} \frac{\partial x_k^e}{\partial c_t} \frac{\partial c_t}{\partial tg_t} + \frac{\partial E_k}{\partial b_k} \frac{\partial b_k}{\partial f_k} \sum_e^Z \frac{\partial f_k}{\partial x_k^e} \frac{\partial x_k^e}{\partial c_t} \frac{\partial c_t}{\partial tg_t} \end{aligned} \right) \end{aligned} \quad (5-28)$$

Mit der Definitionen von neuen Variablen x

$$\begin{aligned}
A_{kt} &:= \frac{\partial E_k}{\partial t g_t} = 0 \\
B_{kt} &:= \frac{\partial E_k}{\partial g_k} \frac{\partial g_k}{\partial t g_t} = \alpha_k f_k 0.9 \frac{2(g_k - 1)}{2(1 - b_k)} \\
C_{kt} &:= \frac{\partial E_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial t g_t} = \alpha_k 0.9 \frac{1}{2} \frac{2\rho_k - \rho_k^2 - b_k t_U}{(1 - \rho_k)^2 t g_t^2} \\
D_{kt}^e &:= \frac{\partial E_k}{\partial f_k} \frac{\partial f_k}{\partial x_k^e} \frac{\partial c_t}{\partial t g_t} = \alpha_k 0.9 \left(\frac{t_U \cdot (1 - g_k)^2}{2(1 - b_k)} \right) \frac{s_k'}{\tau_x} \frac{\partial c_t}{\partial t g_t} \\
E_{kt}^e &:= \frac{\partial E_k}{\partial \rho_k} \frac{\partial \rho_k}{\partial f_k} \frac{\partial f_k}{\partial x_k^e} \frac{\partial c_t}{\partial t g_t} = \alpha_k 0.9 \frac{1}{2} \frac{2\rho_k - \rho_k^2}{(1 - \rho_k)^2} \frac{1}{S_k \cdot t g_k / t_U} \frac{s_k'}{\tau_x} \frac{\partial c_t}{\partial t g_t} \\
F_{kt}^e &:= \frac{\partial E_k}{\partial b_k} \frac{\partial b_k}{\partial f_k} \frac{\partial f_k}{\partial x_k^e} \frac{\partial c_t}{\partial t g_t} = \alpha_k f_k 0.9 \frac{t_U \cdot (1 - g_k)^2}{2(1 - b_k)^2} \frac{1}{S_k} \frac{s_k'}{\tau_x} \frac{\partial c_t}{\partial t g_t}
\end{aligned} \tag{5-29}$$

kann Gl. 5-28 wie folgt vereinfacht werden:

$$\begin{aligned}
\Delta t g_t &= -\varepsilon \sum_{t=1}^F \sum_{k=1}^F \left(A_{kt} + B_{kt} + C_{kt} + \sum_{e=1}^Z D_{kt}^e \frac{\partial x_k^e}{\partial c_t} + \sum_{e=1}^Z E_{kt}^e \frac{\partial x_k^e}{\partial c_t} + \sum_{e=1}^Z F_{kt}^e \frac{\partial x_k^e}{\partial c_t} \right) \\
&:= -\varepsilon \sum_{t=1}^F \sum_{k=1}^F \left(ABC_{kt} + \sum_{e=1}^Z \frac{\partial x_k^e}{\partial c_t} DEF_{kt}^e \right)
\end{aligned} \tag{5-30}$$

Die Komponenten A_{kt} , B_{kt} , C_{kt} , D_{kt} , E_{kt} und F_{kt} bzw. ABC_{kt} und DEF_{kt} sind in geschlossener Form erhältlich, $\partial c_k / \partial t g_t$ kann durch die in Kapitel 5.1 ermittelten Ableitungen der Kapazitätswerte (Gl. 5-3, 5-7, 5-11) entsprechend der dort unterschiedenen Fälle eingesetzt werden. Im Gegensatz dazu kann man die partiellen Ableitungen $\partial x_k^e / \partial c_t$ nicht so leicht berechnen. Sie können aber über die Differentiation der dynamischen Gleichung (Gl. 5-26) bestimmt werden. Dies führt zu

$$\frac{\partial x_k^e}{\partial c_t} = \frac{\partial x_k^e}{\partial c_t} + \sum_i w_{ik}^e \frac{\partial a_{ik}^e}{\partial c_t} - \sum_i w_{ki}^e \frac{\partial a_{ki}^e}{\partial c_t} \tag{5-31}$$

mit

$$\begin{aligned}
\frac{\partial a_{ij}^e}{\partial c_t} &= \frac{\partial s_i^e}{\partial c_t} r_j + \frac{\partial r_j}{\partial c_t} s_i^e \\
&= \delta_{it} \frac{\partial s_i}{\partial c_t} \frac{x_i^e}{x_i} r_j + \frac{s_i}{x_i} r_j \frac{\partial x_i^e}{\partial c_t} + \left(\frac{\partial s_i}{\partial x_i} - \frac{s_i}{x_i} \right) \frac{x_i^e}{x_i} r_j \sum_{l=1}^Z \frac{\partial x_i^l}{\partial c_t} + s_i^e r_j \sum_{l=1}^Z \frac{\partial x_j^l}{\partial c_t} \\
&=: \delta_{it} H_{ij}^e + I_{ij} \frac{\partial x_i^e}{\partial c_t} + J_{ij}^e \sum_{l=1}^Z \frac{\partial x_i^l}{\partial c_t} + K_{ij}^e \sum_{l=1}^Z \frac{\partial x_j^l}{\partial c_t}
\end{aligned} \tag{5-32}$$

Durch Einsetzen von Gl 5-32 in Gl. 5-31 und anschließendes Umsortieren erhält man

$$HW_{kt}^e := \sum_j \left(w_{kj}^e \delta_{kt} H_{kj}^e - w_{jk}^e \delta_{jt} H_{jk}^e \right) = \sum_{i=1}^M \sum_{l=1}^Z \Phi_{ke,il} \frac{\partial x_i^l}{\partial c_t} \tag{5-33}$$

mit

$$\Phi_{ke,il} := \left(w_{ik}^e J_{ik}^e - w_{ki}^e K_{ki}^e \right) - \delta_{ik} \sum_j \left(w_{kj}^e J_{kj}^e - w_{jk}^e K_{jk}^e \right) + \delta_{el} w_{ik}^e I_{ik} - \delta_{el} \delta_{ik} \sum_j w_{kj}^e I_{kj} \tag{5-34}$$

Die Gl. 5-33 kann auch in folgender Form invertiert werden:

$$\frac{\partial x_k^e}{\partial c_t} = \sum_{i=1}^F \sum_{l=1}^Z \Phi_{ke,il}^{-1} HW_{it}^l \tag{5-35}$$

Analog der Vorgehensweise von Kapitel 4.2.2 werden hier folgenden neuen Variablen definiert:

$$y_i^l := \sum_{k=1}^F \sum_{e=1}^Z \Phi_{ke,il}^{-1} DEF_{kt}^e \tag{5-36}$$

und

$$g_{ik}^{le} := \frac{1}{\gamma_k^e} \left[\left(w_{ki}^e J_{ki}^e - w_{ik}^e K_{ik}^e \right) - \delta_{ik} \sum_j \left(w_{kj}^e J_{kj}^e - w_{jk}^e K_{jk}^e \right) + \delta_{el} w_{ki}^e I_{ki} \right] \tag{5-37}$$

Dadurch kann man die dynamische Gleichung bekommen.

$$DEF_{kt}^e = \sum_{i=1}^F \sum_{l=1}^Z \Phi_{il,ke} y_i^l = \gamma_k^e \left(\sum_{i=1}^F \sum_{l=1}^Z g_{ik}^{le} y_i^l - y_k^e \right) \tag{5-38}$$

bzw.

$$y_k^e = \sum_{i=1}^F \sum_{l=1}^Z g_{ik}^{le} y_i^l - \frac{DEF_{kt}^e}{\gamma_k^e} \tag{5-39}$$

wobei

$$\gamma_k^e := \sum_j w_{kj}^e I_{kj} \quad (5-40)$$

In Form einer zeitdiskreten Evolutionsgleichung kann Gl. 5-39 wie folgt umformuliert werden.

$$y_k^e(t + \tau_y) = \sum_{i=1}^F \sum_{l=1}^Z g_{ik}^{le} y_i^l(t) - \frac{DEF_{kt}^e}{\gamma_k^e} \quad (5-41)$$

Das neue dynamische System (Gl. 5-41) definiert das zweite Fehlerpropagierungs-Netzwerk, das aus $F \cdot Z$ linearen Neuronen besteht. Es besitzt keine Schichtstruktur, da die Verbindungsgewichte des Fehlerpropagierungs-Netzwerks g_{ik}^{le} zwischen allen Neuronen definiert sind.

Zur Lösung des durch Gl. 5-39 definierten linearen Gleichungssystems soll zunächst überprüft werden, ob ein normales Iterationsverfahren verwendbar ist. Das Iterationsverfahren konvergiert, wenn der Spektralradius der Iterationsmatrix kleiner als 1 oder die Maximumsnorm der Iterationsmatrix nicht größer als 1 ist. Die Maximumsnorm der Iterationsmatrix ist gleich dem maximalen Wert der Summe der Absolutwerte der Elemente jeder Reihe (Gl. 5-42).

$$\|\Theta\|_{\infty} = \max_{e,k} \sum_{i=1}^F \sum_{l=1}^Z |g_{ik}^{le}| \quad (5-42)$$

Die Matrix Θ besitzt die gute Eigenschaft

$$\begin{aligned} \sum_{i=1}^F \sum_{l=1}^Z g_{ik}^{le} &= \frac{1}{\gamma_k^e} \left[\sum_{l=1}^Z \left(\sum_{i=1}^F (w_{ki}^e J_{ki}^e - w_{ik}^e K_{ik}^e) \right) - \sum_{i=1}^F (w_{ki}^e J_{ki}^e - w_{ik}^e K_{ik}^e) \right] + \sum_{i=1}^F w_{ki}^e I_{ki} \\ &= \frac{1}{\gamma_k^e} \left(0 + \sum_{i=1}^F w_{ki}^e I_{ki} \right) = 1 \end{aligned} \quad (5-43)$$

Man kann es finden, dass nicht alle $g_{ik}^{le} \geq 0$ sind. Deshalb ist die Maximumsnorm der Iterationsmatrix größer als 1.

$$\|\Theta\|_{\infty} = \max_{e,k} \sum_{i=1}^F \sum_{l=1}^Z |g_{ik}^{le}| > 1. \quad (5-44)$$

Der Einsatz des normalen Iterationsverfahrens ist nicht möglich, da die Konvergenz für beliebige Startwerte nicht strikt gewährleistet werden kann. Dennoch kann aufgrund der speziellen Eigenschaft der Iterationsmatrix ein Fehlerpropagierungs-Netzwerk konstruiert und das Problem des Iterationsverfahrens gelöst werden.

Die Iterationsmatrix Θ besitzt eine Blockstruktur,

$$\Theta = \begin{bmatrix} \Theta^{11} & . & . & . & \Theta^{1Z} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ \Theta^{Z1} & . & . & . & \Theta^{ZZ} \end{bmatrix}, \quad (5-45)$$

mit

$$\Theta^{le} = \begin{bmatrix} \mathcal{G}_{11}^{le} & . & . & . & \mathcal{G}_{1F}^{le} \\ . & & & & . \\ . & & & & . \\ . & & & & . \\ \mathcal{G}_{F1}^{le} & . & . & . & \mathcal{G}_{FF}^{le} \end{bmatrix}, \quad (5-46)$$

und kann aus den zwei Teilblockmatrizen zusammengesetzt werden:

$$\Theta = \Theta_{Diagonal} + \Theta_{Link+Rechts} \quad (5-47)$$

mit der Diagonal-Block-Matrix

$$\Theta_D = \begin{bmatrix} \Theta^{11} & 0 & . & . & 0 \\ 0 & \Theta^{22} & & & . \\ . & & . & & . \\ . & & & . & 0 \\ 0 & . & . & 0 & \Theta^{ZZ} \end{bmatrix}, \quad (5-48)$$

und der Rest-Block-Matrix

$$\Theta_{LR} = \begin{bmatrix} 0 & \Theta^{12} & . & . & \Theta^{1Z} \\ \Theta^{21} & 0 & & & . \\ . & . & . & & . \\ . & . & & . & \Theta^{(Z-1)Z} \\ \Theta^{Z1} & . & . & \Theta^{Z(Z-1)} & 0 \end{bmatrix}. \quad (5-49)$$

Man kann aufgrund $\mathcal{G}_{ik}^{ee} \geq 0$ den Betrag der Maximumsnorm von Θ_D festlegen mit

$$\|\Theta_D\|_{\infty} = \|\Theta^{ee}\|_{\infty} = \max_k \sum_{i=1}^F |\mathcal{G}_{ik}^{ee}| = \max_k \sum_{i=1}^F \mathcal{G}_{ik}^{ee} = \max_k \frac{1}{\gamma_k^e} \left[0 + \sum_{i=1}^F w_{ki}^e I_{ki} \right] = 1. \quad (5-50)$$

Analog zu Kapitel 4.2.2 sind die Diagonalmatrizen Θ_D hier auch dominant, daher gilt:

$$\|\Theta_{LR}\|_{\infty} = \sum_{l \neq e} \|\Theta^{le}\|_{\infty} < \|\Theta^{ee}\|_{\infty} = 1. \quad (5-51)$$

Nach den Festlegungen aus (Gl. 5-50) und (Gl. 5-51) sind die Diagonal-Block-Matrix und die Rest-Block-Matrix für das klassische Iterationsverfahren einsetzbar. Zur Lösung des Gleichungssystems (Gl. 5-41) geht das Verfahren nun letztlich auf das klassische Block-Jacobi-Verfahren zurück. Das Gleichungssystem kann in folgender Schreibweise dargestellt werden.

$$b - \Theta_{LR}y = \Theta_D y \quad (5-52)$$

mit

$$b := \left(\frac{DEF_1^1}{\gamma_1^1}, \dots, \frac{DEF_F^1}{\gamma_F^1}, \dots, \frac{DEF_1^Z}{\gamma_1^Z}, \dots, \frac{DEF_F^Z}{\gamma_F^Z} \right)^T \quad (5-53)$$

Die Komponentendarstellung der Gl. 5-52 ist

$$\sum_{i=1}^F g_{ik}^{ee} y_i^e = b_k^e - \sum_{i=1}^F \sum_{l \neq e} g_{ik}^{le} y_i^l. \quad (5-54)$$

Die Gl. 5-54 ermöglicht die Konstruktion von neuen Fehlerpropagierungs-Netzwerken. Zwei neu entwickelte, miteinander kooperierende Fehlerpropagierungs-Netzwerke können die Gleichung interpretieren. Das erste Fehlerpropagierungs-Netzwerk wird durch die linke Seite der Gleichung definiert, das eine Verbindungsgewichtsmatrix Θ_D besitzt. Die dynamische Gleichung des 1. Fehlerpropagierungs-Netzwerks lautet

$$y_k^e(t + \tau_y) = \sum_{i=1}^M g_{ik}^{ee} y_i^e(t) - (b_k^e - u_k^e(t)), \quad \forall e, k. \quad (5-55)$$

Das zweite Fehlerpropagierungs-Netzwerk kann durch die folgende lineare Operation abgebildet werden.

$$\begin{aligned}
u_k^e(t + \tau_y) &= \sum_{i=1}^M \sum_{l \neq e}^Z g_{ik}^{le} y_i^l(t + \tau_y) = \sum_{i=1}^M \sum_{l \neq e}^Z g_{ik}^{le} \left(\sum_i^M g_{ik}^{le} y_i^l(t) - (b_i^l - u_i^l(t)) \right) \\
&= \sum_{i=1}^M \sum_{l \neq e}^Z g_{ik}^{le} u_i^l(t) + \sum_{i=1}^M \sum_{l \neq e}^Z g_{ik}^{le} \left(\sum_{i=1}^M g_{ik}^{le} y_i^l(t) - b_i^l \right) \\
&=: \sum_{i=1}^M \sum_{l \neq e}^Z g_{ik}^{le} u_i^l(t) + \phi_k^e \quad \forall e, k.
\end{aligned} \tag{5-56}$$

Um die Lösung des Gleichungssystems (Gl. 5-54) zu liefern, arbeiten die zwei kooperierenden Fehlerpropagierungs-Netzwerke zusammen. Das erste Netzwerk löst das lineare Gleichungssystem iterativ, gleichzeitig korrigiert das zweite Netzwerk einen Teil $u_k^e(t)$ der rechten Seite von dem Gleichungssystem permanent. Abbildung 5-3 veranschaulicht diese Zusammenarbeit der beiden Fehlerpropagierungs-Netzwerke.

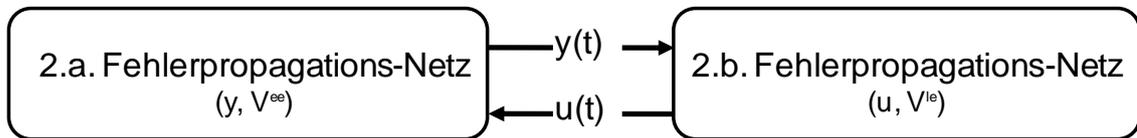


Abbildung 5-3: Die beiden zusammenarbeitenden, eine Lösung der Gl. 5-54 liefernden, Fehlerpropagierungs-Netzwerke

Die gefundene augenblickliche Lösung des ersten Systems (Gl. 5-55) ist auch das gesuchte y von (Gl. 5-54). Die Freigabezeitkorrektur Δtg lässt sich mit Hilfe von y berechnen. Durch das Einsetzen von Gl. 5-35 in Gl. 5-30 ergibt sich schließlich:

$$\begin{aligned}
\Delta tg_t &= -\varepsilon \sum_{t=1}^F \sum_{k=1}^F ABC_{kt} - \varepsilon \sum_{t=1}^F \sum_{k=1}^F \sum_{e=1}^Z \frac{\partial \chi_k^e}{\partial C_t} DEF_{kt}^e \\
&= -\varepsilon \sum_{t=1}^F \sum_{k=1}^F ABC_{kt} - \varepsilon \sum_{i=1}^F \sum_{e=1}^Z DEF_{kt}^e \sum_{k=1}^F \sum_{l=1}^Z \Phi_{ke,il}^{-1} HW_{it}^l \\
&=: -\varepsilon \sum_{t=1}^F \sum_{k=1}^F ABC_{kt} - \varepsilon \sum_{i=1}^F \sum_{l=1}^Z HW_{it}^l \hat{y}_i^l
\end{aligned} \tag{5-57}$$

5.3.2 Update der Freigabezeitverteilung

Bisher wurde die Randbedingung der Freigabezeit noch nicht berücksichtigt, dass die Freigabezeiten voneinander abhängig sind. Denn sonst würden die Freigabezeiten der zugehörigen Signalgruppen einfach maximal, das heißt bis zur Umlaufzeit vergrößert werden.

In Kapitel 5.1.3 wurde bereits erläutert, dass die Umlaufzeit und die Phaseneinteilungen für die Freigabezeitoptimierung schon vorher festgelegt werden. Das heißt, der maßgebende Verkehrsstrom in jeder Signalphase kann bestimmt werden. Die Differenz zwischen der Umlaufzeit und der Summe der Zwischenzeiten zwischen den maßgebenden Verkehrsströmen bleibt während der Optimierung konstant. Die in Kapitel 5.1 eingeführte Verbindungsstrecke, auf die der maßgebende Verkehrsstrom fließt, wird als maßgebende Verbindungsstrecke definiert.

Für die folgenden Betrachtungen sei ein Knotenpunkt $n \in K$ mit den maßgebenden Verbindungsstrecken $s \in F(n)$ fest gewählt. Wenn die Signalphase für Knotenpunkt n als Ph_n , die maßgebenden Verbindungsstrecken für die zugehörige Signalphase Ph_n als VS_{Ph_n} und die Freigabezeiten für die maßgebende Verbindungsstrecke als tg_{p_n} definiert wird,

$$Ph_n := \{Ph_{n1}, Ph_{n2} \dots Ph_{na}\}; \quad VS_{Ph_n} := \{s_{n1}, s_{n2} \dots s_{nb}\}; \quad tg_{p_n} := \{tg_{s_{n1}}, tg_{s_{n2}} \dots tg_{s_{nb}}\}^T \quad (5-58)$$

gilt die Randbedingung

$$\sum_{Anz_Phase} \left(\frac{tg_{p_n}}{tu - \sum Zwischenzeit} \right) = \sum_{Anz_Phase} \left(\frac{tg_{p_n}}{Summe_tg} \right) = \left\| \frac{tg_{p_n}}{Summe_tg} \right\|_1 = 1 \quad (5-59)$$

Aufgrund der Bedingung (Gl. 5-59) liegt $tg_{p_n} := \{tg_{s_{n1}}, tg_{s_{n2}} \dots tg_{s_{nb}}\}^T$ auf der positiven Kugelfläche bezüglich der 1-Norm mit fixem Radius $G = tu - \sum massg.Zwischenzeiten$ $S_a := \{x \in R^a; \|x\|_1 = G \text{ und } x_k > tg_{\min}, k = 1, \dots, a\}$, Der Mindestbetrag von x entspricht der Mindestfreigabezeit, die gemäß Definition die kleinste, unabhängig von der Verkehrsstärke zu gewährende Freigabezeit ist. In der Steuerung kann sowohl die statische als auch die dynamische Mindestfreigabezeit berücksichtigt werden. In der vorliegenden Arbeit wird eine Mindestfreigabezeit von 10 Sekunden als einheitlich statisch angenommen.

Zur Betrachtung der Randbedingung (Gl. 5-59) muss die Optimierungsfunktion E über die a -dimensionale Kugelfläche S_a differenziert werden, was auch heißt, dass der ermittelte Gradientenvektor Δtg_{p_n} im Tangentialraum $T_{S_a}(tg_{p_n})$ von tg_{p_n} an S_a liegt. Über die differenzierbaren Mannigfaltigkeiten können Funktionen abgeleitet werden, indem man für die Mannigfaltigkeit eine Parameterdarstellung konstruiert. Eine mögliche Parameterdarstellung für tg_{p_n} ist gegeben durch

$$tg_{p_n} = \left\{ \begin{array}{l} tg_{s_{n1}}(\theta_1, \theta_2, \dots, \theta_{a-1}) \\ tg_{s_{n2}}(\theta_1, \theta_2, \dots, \theta_{a-1}) \\ \dots \\ \dots \\ tg_{s_{na-1}}(\theta_1, \theta_2, \dots, \theta_{a-1}) \\ tg_{s_{na}}(\theta_1, \theta_2, \dots, \theta_{a-1}) \end{array} \right\} := \left\{ \begin{array}{l} (1-\theta_1) \cdot (1-\theta_2) \dots (1-\theta_{a-1}) G \\ \theta_1 \cdot (1-\theta_2) \dots (1-\theta_{a-1}) G \\ \dots \\ \dots \\ \theta_{a-2} (1-\theta_{a-1}) G \\ \theta_{a-1} G \end{array} \right\} \quad (5-60)$$

Hier sind die Parameter $\theta_1, \theta_2, \dots, \theta_{a-1} \in \left[\frac{tg_{\min}}{G}; 1 - \frac{a \cdot tg_{\min}}{G} \right]$. Statt der Freigabezeitkorrektur werden nun die Parameterkorrekturen bestimmt. Anhand der Kettenregel ergibt sich

$$\begin{aligned} \Delta\theta_p &= -\varepsilon \sum_i \frac{\partial E_i}{\partial \theta_p} = -\varepsilon \sum_i \sum_{t=1}^a \frac{\partial E_i}{\partial tg_t} \frac{\partial tg_t}{\partial \theta_p} = \sum_{t=1}^a \frac{\partial tg_t}{\partial \theta_p} \left[-\varepsilon \sum_i \frac{\partial E_i}{\partial tg_t} \right] \\ &= \sum_{t=1}^a \frac{\partial tg_t}{\partial \theta_p} \left[-\varepsilon \sum_i \frac{\partial E_i}{\partial tg_t} \right] \end{aligned} \quad (5-61)$$

Der erste Summenteil $\partial tg_t / \partial \theta_p$ der rechten Seite von (Gl. 5-61) ist die partielle Ableitung der Mannigfaltigkeit, die sich über (Gl. 5-60) berechnen lässt. Der hintere Teil in der eckigen Klammer von (Gl. 5-61) ist identisch mit der Form (Gl. 5-28). Daher kann die Herleitung des Fehlerpropagierungs-Netzwerks aus dem letzten Unterkapitel übernommen werden. Entsprechend ergibt sich mit (Gl. 5-57) für die Berechnung der $\Delta\theta_k$:

$$\Delta\theta_p = -\sum_{t=1}^a \frac{\partial tg_t}{\partial \theta_p} \varepsilon \left[\sum_{k=1}^F \sum_{l=1}^F ABC_{kt} + \sum_{t=1}^F \sum_{e=1}^Z DEF_{kt}^e \sum_{s=1}^F \sum_{e=1}^Z HW_{st}^l y_s^l \right] \quad (5-62)$$

Aus den berechneten $\Delta\theta_p$ erhält man wieder die benötigten maßgebenden Freigabezeitkorrekturen durch einfache Beziehungen, die durch (Gl. 5-60) festgelegt sind.

$$tg_k = tg_k(f(\theta_1 + \Delta\theta_1), f(\theta_2 + \Delta\theta_2), \dots, f(\theta_a + \Delta\theta_a)) \quad (5-63)$$

Die Funktion $f(\theta)$ lässt sich wie folgt darstellen:

$$f(\theta) := \begin{cases} \frac{tg_{\min}}{\text{Summe_tg}}, & \text{wenn } \theta < \frac{tg_{\min}}{\text{Summe_tg}} \\ \theta, & \text{wenn } \theta \in \left[\frac{tg_{\min}}{\text{Summe_tg}}; 1 - \frac{a \cdot tg_{\min}}{\text{Summe_tg}} \right] \\ 1 - \frac{a \cdot tg_{\min}}{\text{Summe_tg}}, & \text{wenn } \theta > 1 - \frac{a \cdot tg_{\min}}{\text{Summe_tg}} \end{cases} \quad (5-64)$$

Schließlich kann das Update der Freigabezeiten von allen Verbindungsstrecken über die aktualisierte maßgebende Freigabezeit (Gl. 5-65) mit Hilfe von vorher festgelegter Phaseneinteilung und Zwischenzeit durchgeführt werden.

5.4 Das Betriebsmodell

Das Betriebsmodell der Steuerung einzelner LSA basiert auf dem in Kapitel 4 vorgestellten Modell, da das für die Freigabezeitoptimierung zuständige Fehlerpropagierungs-Netzwerk mit dem Simulationsnetzwerk kooperieren muss. In dem nächsten Kapitel wird die Koordinierung der einzelnen LSA bzw. die Versatzzeitoptimierung eingeführt, weshalb dann näher auf das gesamte Betriebsmodell der LSA-Steuerung eingegangen werden kann. Im folgenden Abschnitt betrachten wir daher nur die Arbeitsweise der Kooperation des zweiten Fehlerpropagierungs-Netzwerks mit dem bestehenden Verkehrsmodell.

Umlegungsmodell und Steuerungsmodell im Parallelmodus

Das im Kapitel 4 ausführlich vorgestellte Verkehrsmodell arbeitet mit zwei RNN: Ein Simulationsnetzwerk beschreibt den Verkehrszustand, das Fehlerpropagierungs-Netzwerk bestimmt die Abbiegungsbeziehung. Das Steuerungsmodell für die Freigabezeitoptimierung soll auf Basis des Verkehrsmodells erweitert werden. Das zweite Fehlerpropagierungs-Netzwerk wurde mit Hilfe von Systemzustandsinformationen konstruiert, damit die Freigabezeitanteile so eingestellt werden, dass sich der Wert der gegebenen Optimierungsfunktion (Gl. 5-22) nicht weiter verringert.

Das Gesamtmodell setzt sich zusammen aus den drei kooperierenden RNN: dem Simulations-Netzwerk und den zwei Fehlerpropagierungs-Netzwerken, wobei die zwei Fehlerpropagierungs-Netzwerke aus je zwei Sub-Netzwerken bestehen. Die Abbildung 5-4 verdeutlicht diese gesamte Struktur.

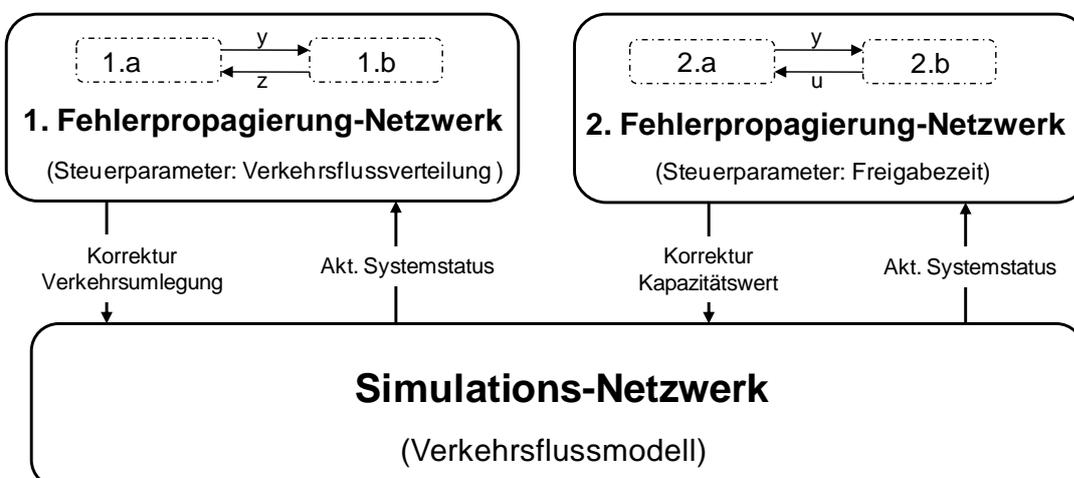


Abbildung 5-4: Das Gesamtsystem mit Umlegungsmodell und Steuerungsmodell

Die Zusammenarbeit der drei RNN kann grundsätzlich in zwei unterschiedlichen Betriebsmodi erfolgen. Bei dem ersten, dem *Fixpunktbetrieb* arbeitet jedes der drei RNN mit den Fixpunkten des jeweils anderen, deshalb müssen die drei RNN wechselweise relaxiert werden. Die Daten, die unter den RNN ausgetauscht werden, sind die Gleichgewichtszustände bzw. Fixpunkte der jeweiligen RNN. Aber der Verkehrszustand befindet sich mit einer dynamisch veränderlichen Nachfrage nur selten in einem Gleichgewichtszustand, weshalb das Simulations-Netzwerk meist nicht einfach einem Fixpunkt zustrebt, sondern in ständiger Bewegung bleibt.

Um diesem Problem zu begegnen, laufen beim Parallelbetrieb alle drei RNN permanent. Im Gegensatz zum Fixpunktbetrieb werden aktuelle Statusdaten (nicht die Gleichgewichtszustände) unter den drei RNN nach bestimmten Zeittakten ausgetauscht. In der Regel beschleunigt dies auch die Konvergenz des Systems, aber es beinhaltet auch die theoretische Gefahr der Instabilität (siehe hierzu auch Kapitel 5.2.3). Die Auswahl der Zeittakte spielt eine wichtige Rolle bei der Verfolgung der Trajektorie. Dieser Effekt wird in Kapitel 6.4 ausführlich erklärt. Die im Rahmen dieser Arbeit durchgeführten numerischen Experimente im Parallelbetrieb mit dynamischen Netzzuflüssen zeigten allerdings in keinem Fall negative Effekte.

Bei Bearbeitungsstart sollte das Gesamtsystem mit einem optimalen Systemzustand bzw. einem Gleichgewichtszustand beginnen, das heißt, dass das Verkehrsnetz zum Startzeitpunkt ($t=0$) konstante geringe Netzzuflüsse hat. Im anschließenden Betrieb können die Netzzuflüsse dann zeitlich veränderlich sein. Um die Arbeitsweise des Parallelbetriebs zu beschreiben, werden drei ganze Zahlen (C_s ; C_U ; C_G) definiert. Wenn das Simulations-RNN für C_s diskrete Zeitschritte τ_x , das für die Umlegung zuständige Fehlerpropagierungs-RNN für C_U diskrete Zeitschritte τ_y und das für die Grünzeitoptimierung zuständige Fehlerpropagierungs-RNN für C_G diskrete Zeitschritte τ_z abwechselnd arbeiten, ist es ein Zyklus. Die Abbildung 5-5 veranschaulicht die Zusammenhänge.

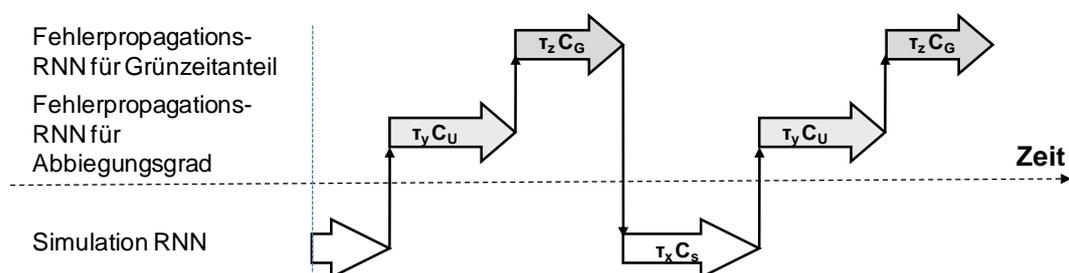


Abbildung 5-5: Die Arbeitsweise bei Parallelbetrieb

Die Zyklusparameter C_s , C_U und C_G haben besondere Bedeutung bei der Verfolgung der optimalen Trajektorie. Im Falle zu großer Zyklusparameter kann sich der Systemzustand zu weit von der optimalen Trajektorie entfernen und nicht mehr dorthin zurückfinden (divergieren). Andererseits steigt mit kleinen Werten der Zyklusparameter der Rechenaufwand.

5.5 Zwischenfazit

Im Kapitel 5 wurde ein neues modellbasiertes LSA-Steuerungsverfahren auf Basis von RNN für einzelne Knotenpunkte entwickelt. Hauptsächlich werden die Freigabezeitverteilungen mit Hilfe eines Fehlerpropagierungs-Netzwerks optimiert.

Auf Basis des Verkehrsmodells kann die Beziehung von Freigabezeitanteil und Kapazitätswerten anhand der Verträglichkeit von Verkehrsströmen bestimmt werden. Die Freigabezeitanteile können so als neuer Steuerungsparameter des bestehenden RNN-Modells aufgefasst werden. Die Herleitung des neuen Fehlerpropagierungs-Netzwerks zur Freigabezeitoptimierung verwendet wie bei der Verkehrsumlegung das Gradientenabstiegsverfahren. Schließlich berechnen zwei Sub-Fehlerpropagierungs-Netzwerke iterativ die Korrektur der Freigabezeit. Unter gewissen Randbedingungen kann mit der Korrektur der Freigabezeit die Optimierung der Freigabezeitverteilung ermittelt werden.

Obwohl die LSA-Steuerung in diesem Kapitel nur bezogen auf einzelne Knoten betrachtet wird, bezieht sich die Optimierungsfunktion auf das gesamte Netz. Durch die Optimierung der einzelnen Knotensteuerungen erhöht sich die Leistungsfähigkeit des gesamten Straßennetzes, womit die Summe der Wartezeiten eines Verkehrsstroms minimiert wird.

Die in diesem Kapitel optimierte Freigabezeitverteilung einzelner Knotenpunkte und die netzweit einheitliche Umlaufzeit dienen als Grundgrößen für die Koordinierung der LSA, die im nächsten Kapitel behandelt wird.

6 RNN-Verkehrssteuerungsmodell für Versatzzeitoptimierung

Eine wichtige und auch komplexe Aufgabe einer netzweiten modellbasierten Verkehrssteuerung ist die netzweite Koordinierung. Im Kapitel 5 wurde die Freigabezeitverteilung einzelner Knotenpunkte unter der Voraussetzung optimiert, dass die Umlaufzeit und die Phaseneinteilung schon festgelegt sind. Die einzige noch nicht betrachtete Variable der LSA-Steuerung ist die Versatzzeit. Das folgende Kapitel soll sich mit der Versatzzeitoptimierung beschäftigen.

Um die Versatzzeit als Steuerungsparameter des bestehenden RNN-Modells aufzufassen, werden zusätzlich die Kapazitätswerte benötigt. Mit Hilfe der Wartezeit wird der Zusammenhang zwischen Versatzzeit und der sogenannten Streckenleistungsfähigkeit bestimmt. Auf Basis einer bekannten Optimierungsfunktion wird analog zum Kapitel 5 ein neues Fehlerpropagierungs-Netzwerk entwickelt. Nach der Optimierung der Versatzzeit werden die gesamte Systemarchitektur und das Betriebsmodell der LSA-Steuerung beschrieben.

6.1 Versatzzeit als Steuerungsparameter

Um die Versatzzeit mit Hilfe des RNN zu optimieren, wird die Versatzzeit in die dynamische Gleichung (Gl. 4-12) des Simulations-Netzwerks integriert, so dass die Versatzzeit als neuer Steuerungsparameter des bestehenden RNN-Modells aufgefasst werden kann. Wie bei der Freigabezeitoptimierung sollen die Kapazitätswerte die Rolle einer Brücke übernehmen. Das Problem besteht darin, einen Zusammenhang zwischen der Kapazität und der Versatzzeit zu finden. In diesem Unterkapitel soll diese Frage geklärt werden.

6.1.1 Grunddefinitionen und Konzept

Die Kapazität der sogenannten Verbindungsstrecken ist eine Funktion des Freigabezeitanteils. Die Kapazität der Strecken zwischen zwei lichtsignalgesteuerten Knotenpunkten ist während der Freigabezeitoptimierung noch konstant und vorgegeben. Die

Strecke k in der Abbildung 6-2 ist ein Beispiel dieser Streckenkategorie, deren Kapazität versatzzeitabhängig sein kann.

Vor der Einführung einer versatzzeitabhängigen Streckenleistungsfähigkeit wird hier die Definition der relativen Versatzzeit vorgestellt. Eine relative Versatzzeit ist die Zeitdifferenz zwischen dem Freigabezeitbeginn zweier aufeinanderfolgender Signalgruppen auf einer Strecke. Die Abbildung 6-1 veranschaulicht diese Definition.

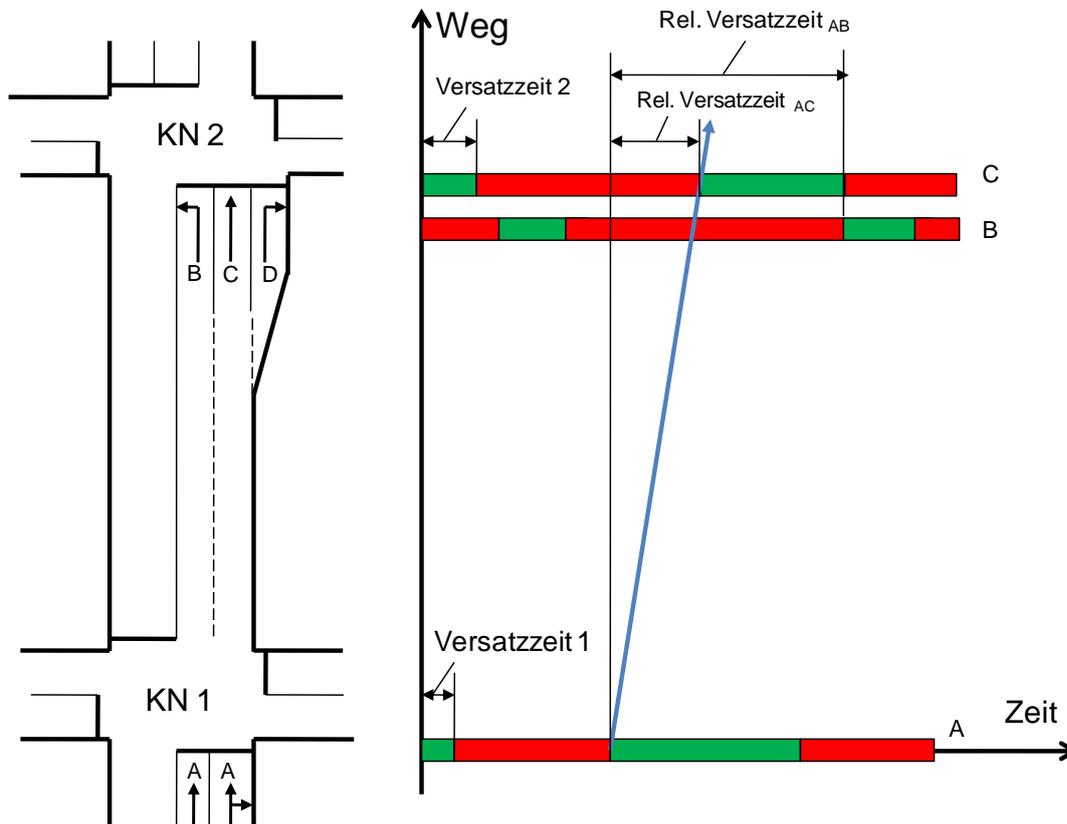


Abbildung 6-1: Zeit-Weg-Diagramm zur Definition der relativen Versatzzeiten

Die Beziehung zwischen den relativen und absoluten Versatzzeiten kann man leicht aus der Abbildung 6-1 ablesen, die Umrechnung ergibt sich wie folgt:

$$VZ_1 + t_{rot_A} = VZ_2 + t_{rot_C} - rel.VZ_{AC} . \quad (6-1)$$

Die Abhängigkeit von verschiedenen relativen Versatzzeiten für dieses Beispiel lässt sich ebenfalls ohne Schwierigkeit wie folgt berechnen:

$$rel.VZ_{AC} + t_{gruen_C} = rel.VZ_{AB} . \quad (6-2)$$

Die ermittelte Beziehung wird in Kapitel 6.3 zum Einsatz kommen. Dieses Unterkapitel konzentriert sich auf den Zusammenhang zwischen Versatzzeit und Kapazität. Eigentlich ist die im HBS oder in den RiISA definierte Streckenkapazität theoretisch unabhängig von der Versatzzeit. Sie ist der Quotient der Freigabezeit durch den Zeitbedarf.

Wu [1990] unterscheidet bezüglich der Kapazität eines LSA-gesteuerten Knotenpunkts zwischen der Grenzkapazität und der praktischen Kapazität. Die Grenzkapazität ent-

spricht der in Kapitel 5 eingesetzten, im HBS definierten Kapazität, bei der der Auslastungsgrad $x = 1$ ist. In diesem Fall treten sehr lange Wartezeiten auf; im extremen Fall unendlich lange Wartezeiten. Im Gegensatz zur Grenzkapazität ist die praktische Kapazität so festgelegt, dass eine vorgegebene mittlere Wartezeit eingehalten wird. Damit ist die Kapazität eine Funktion der Wartezeit.

Das Konzept zur Erstellung der Beziehung zwischen Kapazität und Versatzzeit ist, die Wartezeit als ein Bindeglied zu verwenden. Die Abhängigkeit von Versatzzeit und Wartezeit kann für verschiedene Fälle ohne große Schwierigkeit festgelegt werden. Anschließend kann der Zusammenhang zwischen Kapazität und Wartezeit mit Hilfe eines Warteschlangemodells bestimmt werden.

Um die im Kapitel 5 verwendeten Kapazitätswerte zu unterscheiden, wird die versatzzeitabhängige Streckenleistungsfähigkeit eingeführt. Die Versatzzeit bezieht sich auf die relative Versatzzeit. Wie in Abbildung 6-2 dargestellt, sind $VZ[h][j]$ bzw. $VZ[2][5]$, $VZ[2][6]$, $VZ[3][5]$, $VZ[3][6]$ usw. die für die Strecke k relevanten relativen Versatzzeiten.

Um die Wartezeit des Verkehrsstroms auf der Strecke k in Abhängigkeit von der Versatzzeit zu ermitteln, wird der Verkehrsstrom auf der Strecke k anhand der für die Strecke k relevanten relativen Versatzzeiten in einem Teilstrom-Modell aufgeteilt. Für das Beispiel in Abbildung 6-2 ist der Verkehrsstrom auf der Strecke k in die Teilströme $TS[2][5]$, $TS[2][6]$, $TS[3][5]$, $TS[3][6]$ usw. aufgeteilt. Die mittlere Wartezeit des Verkehrsstroms auf der Strecke k ist die nach Verkehrsstärke gewichtete mittlere Wartezeit aller Teilströme, die durch die relativen Versatzzeiten bestimmt werden können.

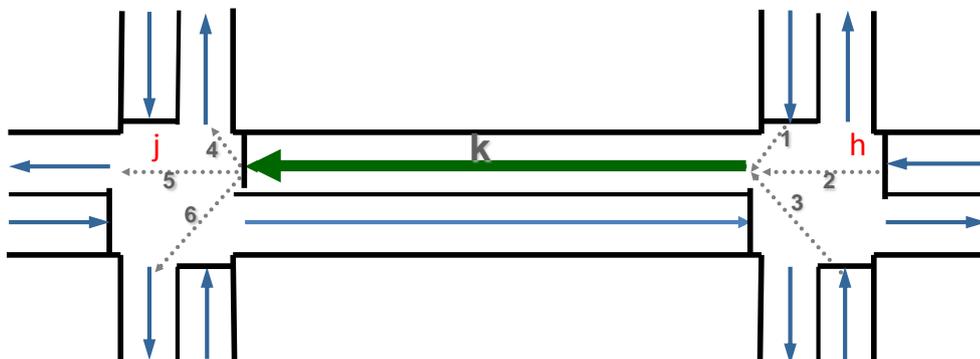


Abbildung 6-2: Beispiel zur Definition versatzzeitabhängiger Streckenleistungsfähigkeit

In den kommenden Unterkapiteln werden zuerst die beiden Beziehungen zwischen Versatzzeit und Wartezeit, sowie zwischen Wartezeit und Streckenleistungsfähigkeit hergestellt. Danach soll die dynamische Gleichung des Simulationsnetzwerks für die Versatzzeitoptimierung aktualisiert werden.

6.1.2 Versatzzeit und Wartezeit

Über die Versatzzeit werden die Freigabezeiten benachbarter LSA koordiniert. Ziel der Koordinierung ist es, eine sogenannte Grüne Welle zu schaffen, und so die Wartezeiten und Halte der Fahrzeuge zu minimieren. Im Idealfall können alle Fahrzeuge bei einer bestimmten Geschwindigkeit mehrere Knotenpunkte ohne Halt passieren.

Ausgangsgrößen für Koordinierung bzw. Versatzzeitoptimierung sind die Progressionsgeschwindigkeit und die Distanz zwischen zwei Knotenpunkten. Im Rahmen der Arbeit wird die Progressionsgeschwindigkeit mit 40 km/h als konstant angenommen. Die aus den beiden Ausgangsgrößen zu bestimmende Reisezeit stellt eine optimale Versatzzeit eines betrachteten Verkehrsstroms dar, falls keine behindernden Fahrzeuge vor dem ankommenden Verkehrsstrom stehen. Anderenfalls muss die Freigabezeit der zweiten LSA früher als anhand der Reisezeit ermittelt eingeschaltet werden, damit die behindernden Fahrzeuge vor der Ankunft des betrachteten Verkehrsstroms abfließen können. Abbildung 6-3 veranschaulicht diese beiden Fälle.

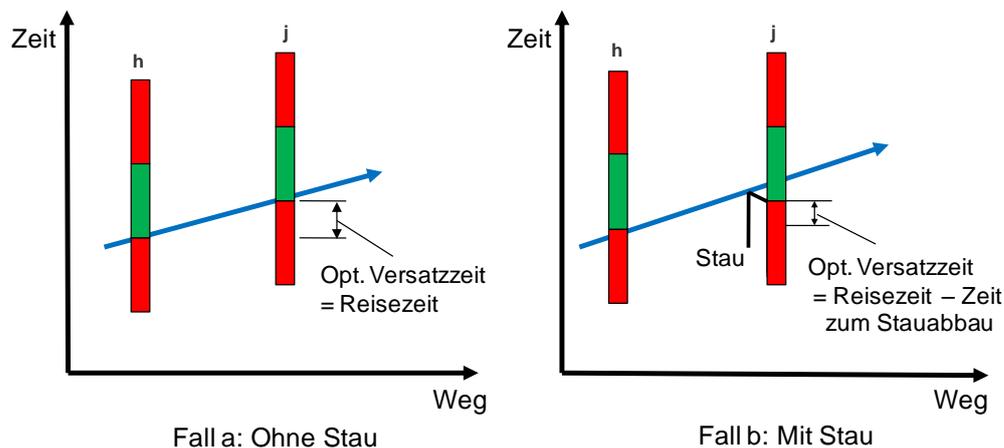


Abbildung 6-3: Zeit-Weg-Diagramm zur Definition der optimalen Versatzzeit

Es sei hier darauf hingewiesen, dass der Auslastungsgrad im Rahmen der Arbeit nicht größer als 1 sein soll. Das heißt, bei deterministischer Betrachtungsweise treten ausschließlich unmittelbar von der LSA verursachte Warteschlangen und Wartezeiten auf.

Die Staulänge, die von dem von h nach j fließenden Verkehrsstrom verursacht ist, kann durch die zugehörige Verkehrsstärke f und die Sperrzeit T_{Sperr} bestimmt werden:

$$l_{Stau} = f \cdot T_{Sperr} \quad (6-3)$$

Die Verkehrsstärke und die Sperrzeit sind durch das erste und zweite Fehlerpropagationsnetzwerk bekannt. Damit sind die optimalen Versatzzeiten für beide Fälle berechenbar. Die optimale Versatzzeit ist ein fixer Parameter in der unten vorgestellten Wartezeitfunktion über die Versatzzeit.

Zur Bestimmung der versatzzeitabhängigen Wartezeit wird das Verkehrsflussmodell auf der Strecke genutzt. Dieses betrachtet kein einzelnes Fahrzeug und wird durch ein deterministisches durch ein Zweibereichsmodell charakterisiert (Abbildung 6-4).

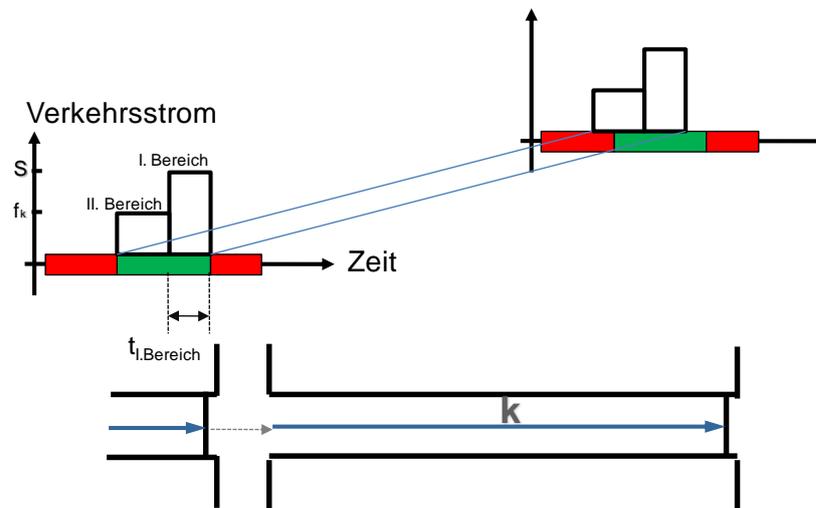


Abbildung 6-4: Zweibereichsmodell des Verkehrsmodells auf der Strecke

Der erste Bereich entspricht der Sättigungsverkehrstärke bzw. der Servicerate, der zweite Bereich steht für die zu diesem Zeitpunkt vorhandene Verkehrsstärke. Aus der Gleichung

$$t_{I.Bereich} \cdot S + (t_{gruen} - t_{I.Bereich}) \cdot f_k = t_{Umlauf} \cdot f_k \quad (6-4)$$

ergibt sich die Zeitdauer des ersten Bereichs:

$$t_{I.Bereich} = \frac{(t_{Umlauf} - t_{gruen}) \cdot f_k}{S - f_k} \quad (6-5)$$

Die Bestimmung der mathematischen Zusammenhänge von Wartezeit und Versatzeit kann mit Hilfe des bekannten Polygon-Verfahrens erfolgen. Die Arbeitsweise wird in Abbildung 6-5 anhand eines Beispiels dargestellt.

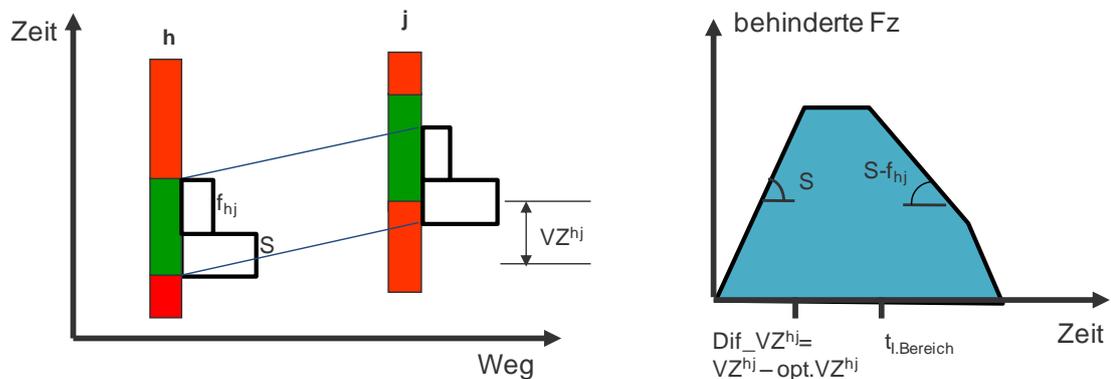


Abbildung 6-5: Beispiel zur Bestimmung der versatzzeitabhängigen Wartezeit

Die mittlere Wartezeit pro Fahrzeug ist der Quotient aus der Polygonfläche durch die Gesamtzahl der abfließenden Fahrzeuge:

$$w = \frac{\left(2 \cdot t_{I,Bereich} + Dif_{-}VZ^{hj} \cdot S / (S - f_{hj}) - Dif_{-}VZ^{hj}\right) Dif_{-}VZ^{hj} \cdot S}{f_{hj} \cdot t_{Umlauf} \cdot 2} \quad (6-6)$$

Es sei angemerkt, dass das oben dargestellte Beispiel die Bedingung:

$$\begin{cases} opt.VZ^{hj} \leq VZ^{hj} \leq t_{sperr,j} + opt.VZ^{hj} \\ t_{I,Bereich} \geq VZ^{hj} - opt.VZ^{hj} \end{cases} \quad (6-7)$$

erfüllt. Unter verschiedenen Bedingungen ist die Wartezeitfunktion (Gl. 6-6) auch unterschiedlich. Die ausführliche, vollständige Berechnung wird im Anhang dargestellt.

6.1.3 Streckenleistungsfähigkeit und Wartezeit

Die Leistungsfähigkeit der Strecke zwischen zwei benachbarten Knotenpunkten, ist prinzipiell nur vom Freigabezeitanteil abhängig. Wenn man den Verkehr innerhalb eines Umlaufs als stationär bzw. deterministisch betrachtet, ist die mittlere Wartezeit pro Fahrzeug für jeden Umlauf wie folgt (Auf die Herleitung wird hier nicht eingegangen, es wird auf [BUSCH, 2007] verwiesen.):

$$w = \frac{t_U (1 - \lambda)^2}{2 \left(1 - \lambda \frac{f}{c}\right)} \quad \text{mit } \lambda = \frac{t_g}{t_U} \quad (6-8)$$

Wenn man

$$\lambda = \frac{t_g}{t_U} = \frac{c}{S} \quad (6-9)$$

in Gl. 6-8 einsetzt, ergibt sich

$$w = t_U \left(1 - \frac{c}{S}\right)^2 / 2 \left(1 - \frac{f}{S}\right). \quad (6-10)$$

Man löst die Gl. 6-10 nach der Streckenleistungsfähigkeit c und erhält

$$c = \left(1 - \sqrt{\frac{w \cdot 2(1 - f/S)}{t_U}}\right) \cdot S. \quad (6-11)$$

Wie leicht zu erkennen ist, erreicht die Streckenleistungsfähigkeit die Sättigungsverkehrsstärke, wenn die Wartezeit gleich Null ist. Oder anders ausgedrückt: Im Fall einer idealen Koordinierung bzw. Grünen Welle hat die Streckenleistungsfähigkeit den maximalen Wert.

Im anderen Extremfall, wenn die Streckenleistungsfähigkeit gleich Null ist, kann mit Gl. 6-10 die maximale Wartezeit bestimmt werden.

$$w_{\max} = \frac{t_U}{2(1-f/S)} \quad (6-12)$$

Die Umlaufzeit t_U , die Verkehrsstärke x und die Sättigungsverkehrsstärke s sind bereits bekannten Variablen, die maximale Wartezeit (Gl. 6-12) ist auch eine bestimmte Größe für die Versatzozeitoptimierung. Durch Einsetzen von Gl. 6-12 in Gl. 6-11 erhält man die wartezeitabhängige Funktion der Streckenleistungsfähigkeit (Gl. 6-13). Abbildung 6-6 stellt den Zusammenhang grafisch dar.

$$c = \left(1 - \sqrt{\frac{w}{w_{\max}}}\right) \cdot S \quad (6-13)$$

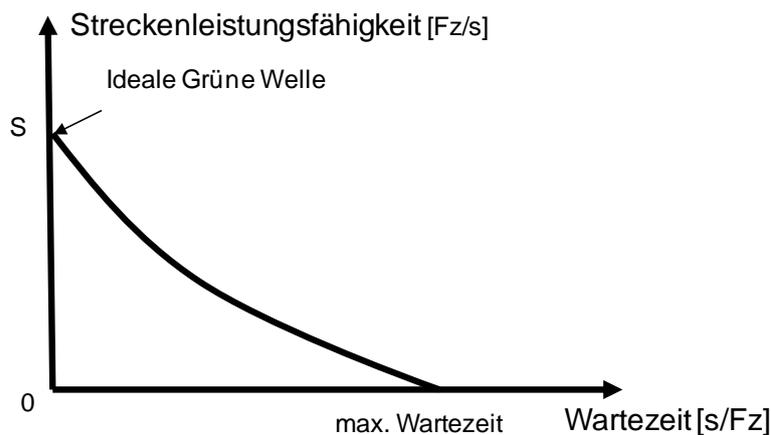


Abbildung 6-6: Die wartezeitabhängige Funktion von Streckenleistungsfähigkeit

Zur Erklärung der beiden Extremfälle der Funktion (Gl. 6-13) betrachtet man den Verkehrsstrom zwischen benachbarten Knotenpunkten innerhalb eines sogenannten Grünbands. Im Falle der idealen grünen Welle können alle von dem ersten Knotenpunkt kommenden Fahrzeuge den zweiten Knotenpunkt ohne Halt passieren. Aufgrund des vollständigen Abflusses wird der maximale Wert der Streckenleistungsfähigkeit erreicht. Müssen im Gegensatz dazu alle Fahrzeuge vor der Haltlinie des zweiten Knotenpunkts warten, so fließt kein einziges Fahrzeug innerhalb dieses Grünbands ab. Die Streckenleistungsfähigkeit wird gleich Null gesetzt.

Damit sind die mathematischen Zusammenhänge von Versatzozeit und Wartezeit (Gl. 6-6), sowie Wartezeit und Streckenleistungsfähigkeit (Gl. 6-13) ermittelt. Durch einfaches

Einsetzen erhält man die Beziehung zwischen Streckenleistungsfähigkeit und Versatzzeit (Gl. 6-14).

$$Lst_k^{hj} = f(VZ^{hj}) = \left(1 - \sqrt{\frac{w^{hj}}{w_{\max}^{hj}}}\right) \cdot S_k = \left(1 - \sqrt{\frac{w^{hj}(VZ^{hj})}{w_{\max}^{hj}}}\right) \cdot S_k \quad (6-14)$$

Es sei angemerkt, dass der absolute Wert der versatzzeitabhängigen Streckenleistungsfähigkeit einer einzelnen Strecke für die Versatzzeitoptimierung nicht maßgebend ist. Vielmehr ist der relative Zusammenhang der Streckenleistungsfähigkeit einer einzelnen Strecke für die Einstellung der Versatzzeit entscheidend. Aus dieser Sicht ist die Annahme der Funktion einer Streckenleistungsfähigkeit sinnvoll. Die Versatzzeit soll so eingestellt werden, dass die Streckenleistungsfähigkeit einer einzelnen Strecke optimal verteilt wird. Dadurch wird die vorgegebene Optimierungsfunktion minimiert.

6.2 Optimierungsfunktion für Versatzzeitoptimum

Die Optimierungsfunktion für die Freigabezeitoptimierung ist eine auf der Wartezeit basierende strategische Systemoptimierungsfunktion. Das Kriterium für die Erstellung einer Optimierungsfunktion wurde in Kap. 3 beschrieben. Es besagt, dass die Optimierungsfunktion des RNN eine beliebige, in x stetig differenzierbare Funktion sein muss.

Mit Hilfe des Warteschlangenmodells $M/D/1$ wurde die Optimierungsfunktion von Kap. 5 über die Wartezeit erstellt, denn die von der Freigabezeit abhängige Wartezeitfunktion (Gl. 5-21) ist in x stetig differenzierbar. Leider ist die von der Versatzzeit abhängige Wartezeitfunktion (z.B. Gl. 6-6 und Anhang) eine abschnittsweise definierte Funktion, weshalb die stetige Differenzierbarkeit nicht gewährleistet ist. Aus diesem Grund ist eine wartezeitabhängige Optimierungsfunktion für die Versatzzeitoptimierung nicht anwendbar. Die für das Verkehrsumlegungsmodell verwendete in x stetig differenzierbare Optimierungsfunktion kann auch hier zum Einsatz kommen.

Die für die Verkehrsumlegung verwendete Systemoptimierungsfunktion (Gl. 4-18) beschreibt die Summe der Reisezeiten in dem gesamten Netz. Die Reisezeit im Stadtstraßennetz setzt sich aus zwei Komponenten zusammen. Ein Teil ist die Reisezeit der freien Strecke, der andere Teil steht für die Wartezeit vor dem signalisierten Knotenpunkt. Die Reisezeit (Gl. 4-17) ist eine Funktion über den möglichen Abfluss bzw. die Kapazität. Da die Kapazität und die Streckenleistungsfähigkeit über die LSA bestimmt werden können, erhält die Reisezeit (Gl. 4-17) die Information von den LSA.

Aus dieser Sicht wird die im Kap. 4 eingesetzte Optimierungsfunktion ebenfalls für die Versatzzeitoptimierung zum Einsatz kommen, wie in Gleichung 6-15 beschrieben.

$$E(x) = \sum_k \pi_k(x_k) f_k + \beta_k (x_k - \hat{x}_k)^2 = \sum_k x_k + \beta_k (x_k - \hat{x}_k)^2 \quad \forall k \in S \quad (6-15)$$

Die gegebene Optimierungsfunktion E definiert eine Potentialfläche über den Zustand des Simulationsnetzwerks, entlang der die Versatzzeiten justiert werden können. Dabei soll E so eingestellt werden, dass die Gesamtreisezeit möglichst kurz ist.

6.3 Erweiterungen des RNN-Modells zur Optimierung der Versatzzeit

Für die netzweite Optimierung der Versatzzeiten soll das in Kapitel 5 beschriebene RNN-Modell erweitert werden. Der Ansatz zur Lösung der Aufgabe ist, die Versatzzeit bzw. die Streckenleistungsfähigkeit als dritten Steuerungsparameter zu verwenden. Im Unterschied zur Kapazität für die Freigabezeitoptimierung soll die Streckenleistungsfähigkeit schichtweise betrachtet werden. Daher muss die dynamische Gleichung des Simulationsnetzwerks entsprechend verändert werden. Dies wird im folgenden Abschnitt beschrieben.

Analog zum Verfahren im Kapitel 5 wird das dritte Fehlerpropagierungsnetzwerk konstruiert, wodurch man die Änderung der Versatzzeit so bestimmen kann, dass die vorgegebene Optimierungsfunktion minimiert wird.

6.3.1 Dynamische Gleichung der Versatzzeitoptimierung

Die dynamische Gleichung des Simulations-Netzwerks variiert in Abhängigkeit von den Streckenleistungsfähigkeiten einer einzelnen Schicht bzw. den Versatzzeiten. Es soll davon ausgegangen werden, dass unter Verwendung des Umlegungsmodells und des Freigabezeitoptimierungsmodells bereits ein augenblicklicher optimaler Satz von Abbiegeraten und Freigabezeitverteilungen bestimmt wurde. Diese Abbiegeraten und Freigabezeitverteilungen werden während der Versatzzeitoptimierung beibehalten.

Aufgrund des durch die relativen Versatzzeiten definierten Teilstrom-Modells wird der maximal mögliche Abfluss für jeden Teilstrom beschrieben. Für das Beispiel in Abbildung 6-2 kann der maximal mögliche Abfluss für einen Teilstrom auf der Strecke k durch

$$s_k^{d-hj} = \tau_x \cdot Lst_k^{hj} \cdot \sigma_k(x_k) \frac{x_k^{d-hj}}{x_k} \quad \forall k \in V \quad (6-16)$$

definiert werden. $V \subset S$ ist die Menge der Strecken, die für die Versatzzeitoptimierung relevant sind. Lst_k^{hj} ist die Streckenleistungsfähigkeit und d kommt aus dem Ziel-Schicht-Modell.

Die dynamische Gleichung basiert auf in Kapitel 4 eingeführten dynamischen Gesetzen (Gl. 4-12). Das bekannte Gesetz der Dynamik bleibt wie folgt:

$$x_k^d(t + \tau_x) = x_k^d(t) + \sum_{i \in E(k)} w_{ik}^d a_{ik}^d(t) - \sum_{i \in A(k)} w_{ki}^d a_{ki}^d(t) + \tau_x q_k^d(t) \quad \forall k \in S \quad (6-17)$$

Diese Gleichung hängt von den Versatzzeiten über den maximalen möglichen Abfluss ab.

$$\begin{aligned} s_k^d &= \sum_h \sum_j s_k^{d-hj} = \tau_x \cdot \sum_h \sum_j Lst_k^{hj} \cdot \sigma_k(x_k) \sum_h \sum_j \frac{x_k^{d-hj}}{x_k} \\ &= \tau_x \cdot \sum_h \sum_j Lst_k^{hj} \cdot \sigma_k(x_k) \frac{x_k^d}{x_k} = \tau_x \sigma_k(x_k) \frac{x_k^d}{x_k} \sum_h \sum_j Lst_k^{hj} (VZ_{hj}) \quad \forall k \in V \end{aligned} \quad (6-18)$$

6.3.2 Berechnung der Versatzzeitkorrekturen

Auf Basis der mathematischen Zusammenhänge zwischen Versatzzeit und Streckenleistungsfähigkeit ist die Aufgabe der Versatzzeitoptimierung die Ermittlung der optimalen Streckenleistungsfähigkeit. Die Versatzzeiten werden optimiert bezüglich der vorgegebenen Optimierungsfunktion, indem die Streckenleistungsfähigkeiten netzweit optimal verteilt werden.

Zur Lösung dieser Optimierungsaufgabe wird wieder die Gradientenabstiegsmethode herangezogen. Der Korrekturwert der Versatzzeit VZ_{hj} , ΔVZ_{hj} , ist proportional zur Ableitung der Optimierungsfunktion (Gl. 6-15) nach VZ_{hj} :

$$\begin{aligned} \Delta VZ_{hj} &= -\varepsilon \sum_k \frac{\partial E_k}{\partial VZ_{hj}} = -\varepsilon \sum_{t=1}^V \frac{\partial Lst_t^{hj}}{\partial VZ_{hj}} \sum_k \frac{\partial E_k}{\partial x_k} \frac{\partial x_k}{\partial Lst_t^{hj}} = -\varepsilon \sum_{t=1}^V \frac{\partial Lst_t^{hj}}{\partial VZ_{hj}} \sum_{k=1}^V \frac{\partial E_k}{\partial x_k} \sum_{e=1}^Z \frac{\partial x_k^e}{\partial Lst_t^{hj}} \\ &=: -\varepsilon \sum_{t=1}^V J_t \sum_{k=1}^V \sum_{e=1}^Z L_k \frac{\partial x_k^e}{\partial Lst_t^{hj}} \end{aligned} \quad (6-19)$$

Die Komponente $J_t = \partial Lst_t^{hj} / \partial VZ_{hj}$ ist leicht berechenbar, die Komponente $L_k = \partial E_k / \partial x_k$ ist identisch mit der Berechnung nach Gl. 4-20. Nicht direkt verfügbar sind dagegen die partiellen Ableitungen $\partial x_k^e / \partial Lst_t^{hj}$, die aber über die spezielle dynamische Gleichung (Gl. 6-18) in Verbindung stehen. Zur Berechnung der partiellen Ableitung wird daher ein drittes Fehlerpropagierungs-Netzwerk konstruiert. Zuerst soll die Gleichung nach der Streckenleistungsfähigkeit differenziert werden.

$$\frac{\partial x_k^e}{\partial Lst_t^{hj}} = \frac{\partial x_k^e}{\partial Lst_t^{hj}} + \sum_i w_{ik}^e \frac{\partial a_{ik}^e}{\partial Lst_t^{hj}} - \sum_i w_{ki}^e \frac{\partial a_{ki}^e}{\partial Lst_t^{hj}} \quad (6-20)$$

Wenn man den möglichen Abfluss der Strecke in der folgenden Form (Gl. 6-21) schreibt,

$$s_m^e = \sum_h \sum_j s_m^{e-hj} \quad \forall m \in V \quad (6-21)$$

kann die Komponente $\partial a_{ik}^e / \partial LS_t^{hj}$ von Gl. 6-20 wie folgt berechnet werden

$$\begin{aligned} \frac{\partial a_{ki}^e}{\partial LS_t^{hj}} &= \frac{\partial S_k^e}{\partial LS_t^{hj}} r_i + \frac{\partial r_i}{\partial LS_t^{hj}} S_k^e \\ &= \delta_{kt} \frac{\partial S_k^e}{\partial LS_t^{hj}} r_i \frac{x_k^d}{x_k} + \left(\frac{s_k'}{s_k} - \frac{1}{x_k} \right) s_k^e r_i \sum_{l=1}^Z \frac{\partial x_k^l}{\partial LS_t^{hj}} + \frac{s_k}{x_k} r_i \frac{\partial x_k^e}{\partial LS_t^{hj}} + s_k^e r_i \sum_{l=1}^Z \frac{\partial x_i^l}{\partial LS_t^{hj}} \quad (6-22) \\ &=: \delta_{kt} A_{ki}^e + B_{ki} \frac{\partial x_k^e}{\partial LS_t^{hj}} + C_{ki}^e \sum_{l=1}^Z \frac{\partial x_k^l}{\partial LS_t^{hj}} + D_{ki}^e \sum_{l=1}^Z \frac{\partial x_i^l}{\partial LS_t^{hj}} \end{aligned}$$

Durch das Einsetzen von Gl. 6-22 in Gl. 6-20 und Umsortieren erhält man

$$H_{kt}^e := \sum_j (\delta_{kt} A_{kj}^e - \delta_{jt} A_{jk}^e) = \sum_{i=1}^M \sum_{l=1}^Z \Omega_{ke,il} \frac{\partial x_i^l}{\partial LS_t^{hj}} \quad (6-23)$$

oder

$$\frac{\partial x_k^l}{\partial LS_t^{hj}} = \sum_{i=1}^M \sum_{l=1}^Z \Omega_{ke,il}^{-1} H_{it}^e \quad (6-24)$$

mit

$$\Omega_{ke,il} := (w_{ik}^e C_{ik}^e - w_{ki}^e D_{ki}^e) - \delta_{ik} \sum_j (w_{kj}^e C_{kj}^e - w_{jk}^e D_{jk}^e) + \delta_{el} w_{ik}^e B_{ik} - \delta_{el} \delta_{ik} \sum_j w_{kj}^e B_{kj} \quad (6-25)$$

Analog der Vorgehensweise von Kapitel 5.3.2 wird hier eine neue Variable definiert

$$y_i^l := \sum_{k=1}^M \sum_{e=1}^Z J_k \Omega_{ke,il}^{-1} \quad (6-26)$$

dadurch kann ein neues lineares dynamisches System (Gl. 6-27) erstellt werden.

$$y_k^e(t + \tau_y) = \sum_{i=1}^M \sum_{l=1}^Z y_i^l(t) \mu_{ik}^{le} - \frac{J_k}{\alpha_k^e} \quad (6-27)$$

mit

$$\mu_{ik}^{le} = \frac{1}{\alpha_k^e} \left[\left(w_{ik}^l C_{ik}^l - w_{ik}^l D_{ik}^l \right) - \delta_{ik} \sum_j \left(w_{kj}^l C_{kj}^l - w_{jk}^l D_{jk}^l \right) + \delta_{el} w_{ki}^e B_{ki} \right] \quad (6-28)$$

und

$$\alpha_k^e = \sum_j w_{kj}^e B_{kj} \quad . \quad (6-29)$$

Gl. 6-27 definiert hier ein neues bzw. das dritte Fehlerpropagierungs-Netzwerk, das V-Z linearen Neuronen besitzt. Weil die Verbindungsgewichte μ_{ik}^{le} zwischen beliebigen Neuronen definiert sind, ist das Netzwerk auch ein typisches RNN.

Das System hat wie das in Kapitel 5.3.2 beschriebene System ebenfalls eine Maximumsnorm $\|U\|_\infty = \max_{e,k} \sum_i^F \sum_l^Z |\mu_{ik}^{le}| > 1$, daher ist die durch die Matrix U definierte Abbildung nicht kontrahierend. Analog der Vorgehensweise von Kapitel 5.3.2 kann auch dieses Problem dadurch gelöst werden, dass zwei kooperierende Fehlerpropagierungs-Netzwerke die Lösung bestimmen. Aufgrund der Analogie zu Kapitel 5.3.2 soll hier die Konstruktion nicht ausführlich sondern nur kurz skizziert werden. Mit Hilfe des Block-Jacobi-Verfahrens kann die Gl. 6-27 wie folgt umschrieben werden:

$$\sum_{i=1}^M \mu_{ik}^{ee} y_i^e = \frac{J_k}{\alpha_k^e} - \sum_{i=1}^M \sum_{l \neq e}^Z \mu_{ik}^{le} y_i^l \quad . \quad (6-30)$$

Die Darstellungsform (Gl. 6-30) lässt sich als Zusammenspiel zweier linearer Systeme deuten. Die zwei dynamischen Gleichungen können wie folgt beschrieben werden.

$$y_k^e(t + \tau_x) = \sum_{i=1}^M \mu_{ik}^{ee} y_i^e(t) - \left(\frac{J_k}{\alpha_k^e} - z_k^e(t) \right) \quad . \quad (6-31)$$

und

$$\begin{aligned} z_k^e(t + \tau_x) &= \sum_{i=1}^M \sum_{l \neq e}^Z \mu_{ik}^{le} y_i^l(t + \tau_x) = \sum_{i=1}^M \sum_{l \neq e}^Z \mu_{ik}^{le} \left[\sum_{i=1}^M \mu_{ik}^{le} y_i^l(t) - \left(\frac{J_k}{\alpha_k^e} - z_k^l(t) \right) \right] \\ &= \sum_{i=1}^M \sum_{l \neq e}^Z \mu_{ik}^{le} \left[\sum_{i=1}^M \mu_{ik}^{le} y_i^l(t) - \frac{J_k}{\alpha_k^e} \right] + \sum_{i=1}^M \sum_{l \neq e}^Z \mu_{ik}^{le} z_k^l(t) \\ &=: \eta_{ke} + \sum_{i=1}^M \sum_{l \neq e}^Z \mu_{ik}^{le} z_k^l(t) \end{aligned} \quad (6-32)$$

Die Lösung der ersten Gl. 6-31 ist auch die gesuchte Lösung von Gl. 6-27. Die Versatzeitkorrektur ΔVZ_{hj} kann man leicht mit Hilfe der gefundenen Lösung berechnen. Durch anschließendes Einsetzen von Gl. 6-24 in Gl. 6-19 ergibt sich:

$$\begin{aligned}
 \Delta VZ_t^{hj} &= -\varepsilon \sum_{t=1}^V J_t \sum_k L_k \sum_{e=1}^Z \sum_{i=1}^M \sum_{l=1}^Z \Omega_{ke,il}^{-1} H_{it}^l \\
 &= -\varepsilon \sum_{t=1}^V J_t \sum_{i=1}^M \sum_{l=1}^Z H_{it}^l \sum_{k=1}^M \sum_{e=1}^Z L_k \Omega_{ke,il}^{-1} \\
 &= -\varepsilon \sum_{t=1}^V J_t \sum_{i=1}^M \sum_{l=1}^Z H_{it}^l y_i^l
 \end{aligned} \tag{6-33}$$

6.3.3 Update der Versatzeit

Bisher wurde nur die Korrektur der Versatzeit nach der vorgegebenen Optimierungsfunktion bestimmt. Die Randbedingungen der Versatzeit wurden noch nicht betrachtet. Aufgrund der Abhängigkeit der relativen Versatzeiten ist das Verfahren hier anders als bei der Berechnung der optimalen Freigabezeit. Während der Definition der relativen Versatzeiten im Kap. 6.1 wurden diese Abhängigkeiten (Gl. 6-2) schon vorgestellt. Die Zusammenhänge von relativen und absoluten Versatzeiten (Gl. 6-1) sind ebenfalls definiert.

Um die Versatzeiten netzweit zu optimieren, wird zunächst die absolute Versatzeit zwischen zwei Knotenpunkten anhand der relativen Versatzeiten festgestellt. Abbildung 6-7 veranschaulicht Beispielstrecken (s und t) zur Bestimmung der maßgebenden relativen Versatzeit.

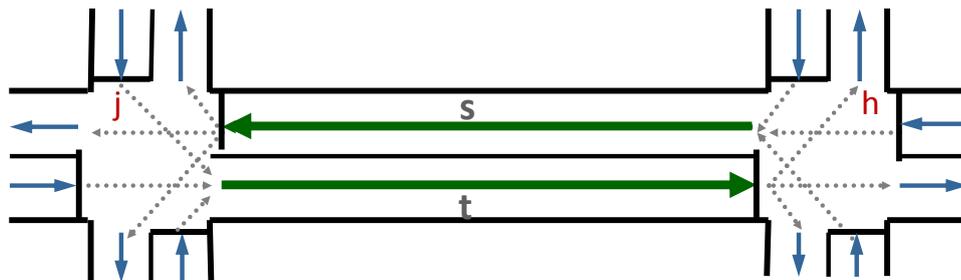


Abbildung 6-7: Beispielstrecken zum Update der Versatzeit

Für die folgenden Betrachtungen seien zwei Strecken s, t (Abbildung 6-7) für Hin- und Rückrichtung $s, t \in V$ fest gewählt. Wenn man die verbundenen Knotenpunkte als H und J , A_j und B_h für die Strecken von Knotenpunkte H und J definiert, gilt

$$\begin{aligned}
 A_j &:= \{A_1, A_2, \dots, A_a\}; & B_h &:= \{B_1, B_2, \dots, B_b\}; \\
 VZ_{s-t}^{hj-jh} &:= \{VZ_s^{11}, VZ_s^{12}, VZ_t^{44}, VZ_t^{45}, \dots, VZ_s^{ab}, VZ_t^{ba}\}^T
 \end{aligned} \tag{6-34}$$

Die maßgebende Korrektur der relativen Versatzzeit wird aus dem maximalen Absolutbetrag aller relativen Versatzzeiten ermittelt. Gleichzeitig können die Strecken bestimmt werden, die für die relative Versatzzeitkorrektur maßgebend sind:

$$ma\beta g.\Delta VZ_{s-t}^{mn} = \begin{cases} \max |\Delta VZ_{s-t}^{mn}| = \|\Delta VZ_{s-t}^{hj-jh}\|_{\infty} & \text{wenn } \Delta VZ_{s-t}^{mn} \geq 0 \\ -\max |\Delta VZ_{s-t}^{mn}| = -\|\Delta VZ_{s-t}^{hj-jh}\|_{\infty} & \text{wenn } \Delta VZ_{s-t}^{mn} < 0 \end{cases} \quad (6-35)$$

Schließlich können alle relativen Versatzzeiten wie folgt berechnet werden:

$$VZ_{s-t}^{hj-jh} := \begin{cases} VZ_{s-t}^{hj-jh} + ma\beta g.\Delta VZ_{s-t}^{mn} + tu, & \text{wenn } VZ_{s-t}^{hj-jh} + ma\beta g.\Delta VZ_{s-t}^{mn} < 0 \\ VZ_{s-t}^{hj-jh} + ma\beta g.\Delta VZ_{s-t}^{mn}, & \text{wenn } VZ_{s-t}^{hj-jh} + ma\beta g.\Delta VZ_{s-t}^{mn} \in [0; tu] \\ VZ_{s-t}^{hj-jh} + ma\beta g.\Delta VZ_{s-t}^{mn} - tu, & \text{wenn } VZ_{s-t}^{hj-jh} + ma\beta g.\Delta VZ_{s-t}^{mn} > tu \end{cases} \quad (6-36)$$

Durch den Zusammenhang (Gl. 6-1) zwischen der ermittelten relativen Versatzzeit (Gl. 6-36) und der absoluten Versatzzeit kann man leicht die absolute Versatzzeit von Knotenpunkt B erhalten, wenn die Versatzzeit von Knotenpunkt A schon festgelegt ist.

Nach der Festlegung der absoluten Versatzzeit zwischen benachbarten Knotenpunkten können die Versatzzeiten im gesamten Netz bestimmt werden. Abbildung 6-8 stellt ein einfaches Netz mit vier Knotenpunkten dar. Man kann zuerst einen Knotenpunkt z.B. Knotenpunkt 1 als Ausgangspunkt setzen. Danach kann die absolute Versatzzeit der Knotenpunkte 2 und 3 berechnet werden. Um die absolute Versatzzeit von Knotenpunkt 4 zu bestimmen, sollen die relativen Versatzzeiten zwischen K3 und K4 bzw. K2 und K4 während der Ermittlung von maßgebenden Versatzzeiten (Gl. 6-35) nicht getrennt betrachtet werden. Das heißt, dass die maßgebende Korrektur der relativen Versatzzeit aus dem maximalen Absolutbetrag aller relativen Versatzzeiten von K3 und K4 bzw. K2 und K4 ermittelt wird. Mit Hilfe der ermittelten maßgebenden relativen Versatzzeit wird schließlich das Update der absoluten Versatzzeit von Knotenpunkt 4 durchgeführt.

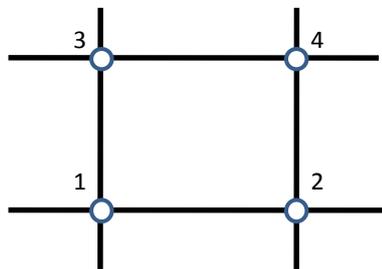


Abbildung 6-8: Beispielnetz zur Update der Versatzzeit

6.4 Das gesamte Betriebsmodell der LSA-Steuerung

Das Gesamtmodell der modellbasierten LSA-Steuerung ergibt sich aus dem Zusammenwirken von vier RNN: dem Simulations-Netzwerk (Inklusive eines Sub-Simulations-Netzwerk, das für Versatzzeitoptimierung zuständig ist) und den drei Fehlerpropagierungs-Netzwerken. Das Simulations-Netzwerk und das erste Fehlerpropagierungs-Netzwerk dienen als Verkehrsmodell, das die Grundgrößen für das Wirkungsmodell liefert. Das zweite und dritte Fehlerpropagierungs-Netzwerk optimieren die Signalpläne bzw. den Freigabezeitanteil und die Versatzzeit. Die Umlaufzeitoptimierung soll durch eine separate Komponente erfolgen. In Abbildung 6-9 wird der Ablauf der Optimierung dargestellt.

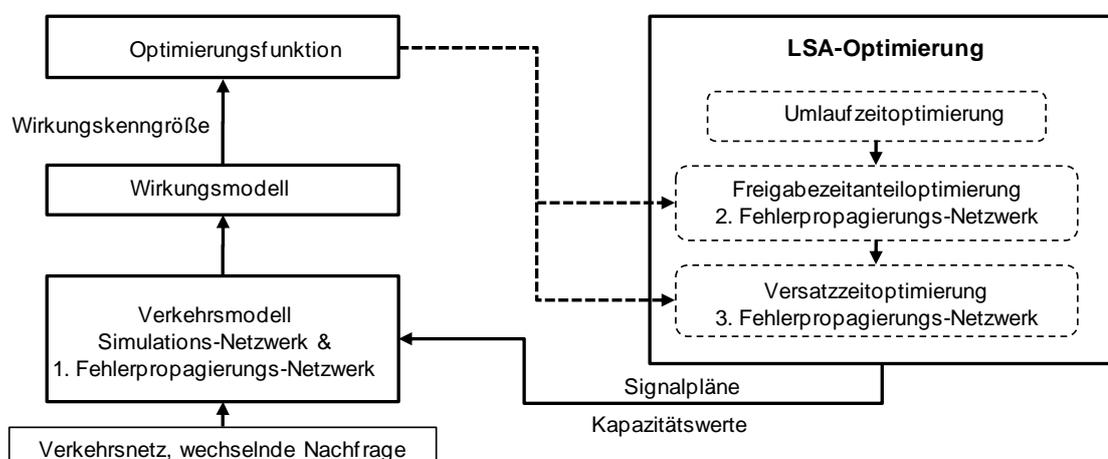


Abbildung 6-9: Ablauf der modellbasierten LSA-Optimierung

Es sei noch darauf hingewiesen, dass die zwei Fehlerpropagierungs-Netzwerke zur LSA-Optimierung nach verschiedenen Optimierungsfunktionen arbeiten. Die Zusammenarbeit der vier RNN muss solange dauern, bis sich der Wert der Optimierungsfunktionen nicht weiter verringert. Dazu können die RNN entweder wechselweise relaxiert werden (Fixpunktbetrieb), oder parallel laufen (Parallelbetrieb). Aufgrund einer dynamisch veränderlichen Nachfrage steht insbesondere das Simulations-Netzwerk nun selten in einem Gleichgewichtszustand, weshalb Parallelbetrieb zum Einsatz kommen wird.

Analog der Vorstellung in Kapitel 5.4 soll der Parallelbetrieb des gesamten Systems mit einem optimalen Systemzustand bzw. einem Gleichgewichtszustand beginnen, im anschließenden Betrieb läuft der Parallelbetrieb. Vier ganze Zahlen (C_s ; C_U ; C_G ; C_V) werden definiert, damit der Parallelbetrieb als Zyklus beschrieben werden kann. Der Zyklus legt fest, dass zunächst das Simulations-RNN für eine Zeitdauer von C_s mal die Anzahl diskreter Zeitschritte τ_x , dann das erste Fehlerpropagierungs-RNN für C_U mal die Anzahl diskreter Zeitschritte τ_y , dann das zweite Fehlerpropagierungs-RNN für C_G mal die Anzahl diskreter Zeitschritte τ_z und abschließend das dritte Fehlerpropagierungs-RNN für C_V mal die Anzahl diskreter Zeitschritte τ_v abgearbeitet werden.

Die Zyklusparameter C_S , C_U , C_G und C_V spielen bei der Verfolgung der Trajektorie (Kap. 5.2.3) eine wichtige Rolle. Die Verhältnisse zwischen C_U , C_G , C_V und C_S beschreiben, wie häufig die Steuerungsparameter des Simulations-Netzwerks bzw. die Abbiegeraten, die Freigabezeitanteile und die Versatzzeit korrigiert werden und wie genau diese Korrektur ist. Im Fall eines zu großen Parameterwerts C_S entfernt sich der Systemzustand eventuell zu weit von der optimalen Trajektorie und läuft dabei Gefahr ganz abzudriften. Hingegen werden kleinere Parameterwerte C_U , C_G und C_V einen größeren Rechenaufwand verursachen.

Es ist zu beachten, dass die Steuerungsparameter C_G und C_V bzw. die Lichtsignalsteuerungsparameter kleiner als der Parameter C_U sein sollen, denn in der Realität reagiert die Verkehrsumlegung nicht verzögerungsfrei auf die Veränderung der Verkehrsnachfrage. Im Gegensatz dazu muss sich die Lichtsignalsteuerung sich deutlich schneller auf die Veränderung einstellen.

Man muss auch hinzufügen, dass die Konvergenzgeschwindigkeit bezüglich der Simulationszeit unabhängig von den Werten der Zyklusparameter ist. Aber die Zeitschritte τ_x , τ_y , τ_z und τ_v sowie die Zyklusparameter C_S , C_U , C_G und C_V haben einen Einfluss auf die Arbeitsweise des Modells. Bei der Erprobung dieses Steuerungsverfahrens im kommenden Kapitel werden die Zeitschritte τ_x , τ_y , τ_z und τ_v einheitlich auf 2 s und die Variablen C_S gleich 5, C_U gleich 20, C_G und C_V gleich 10 festgelegt. Mit diesen Werten wurden bei allen Experimenten gute und stabile Ergebnisse erzielt, weshalb keine weitere Optimierung der Parameter notwendig ist.

Auf Basis des Zyklus kann die Umlaufzeitoptimierung im Parallelbetrieb leicht integriert werden. Wie in Kapitel 5.1.3 erwähnt, erfolgt die Ermittlung der optimalen einheitlichen Umlaufzeit in der Regel im Abstand von 180 s. Wenn alle Parameter wie beschrieben festgelegt werden, dann wird die Umlaufzeit im Abstand von 2 Zyklen nach Gl. 5-13 neu berechnet. Die Abbildung 6-10 veranschaulicht die Zusammenhänge.

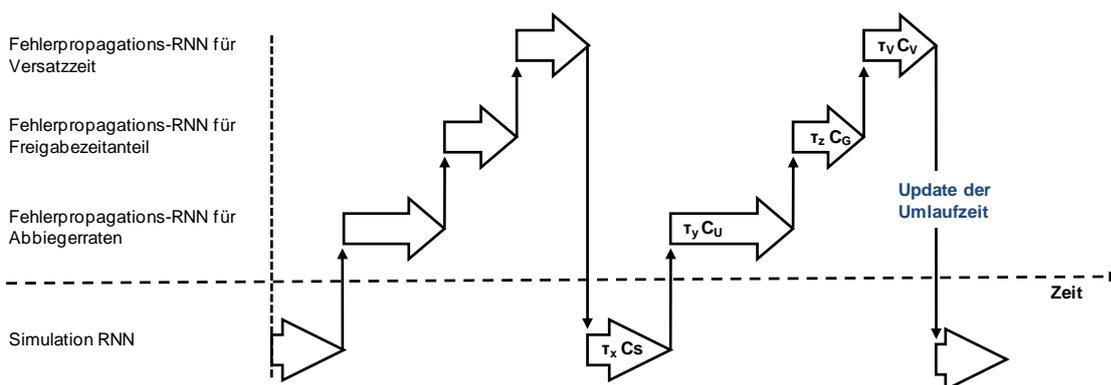


Abbildung 6-10: Arbeitsweise der vier RNN und der Update der Umlaufzeit

6.5 Zwischenfazit

Im Kapitel 6 wurde die Koordinierung von LSA auf Basis des RNN-Modells beschrieben. Die Versatzzeiten werden dabei mit Hilfe eines neuen Fehlerpropagierungsnetzwerks optimiert.

Mit Hilfe der Wartezeit kann die mathematische Beziehung zwischen der relativen Versatzzeit und der Streckenleistungsfähigkeit bestimmt werden. Dies ermöglicht, dass die Versatzzeit als neuer bzw. als dritter Steuerungsparameter des bestehenden RNN-Modells aufgefasst werden kann.

Aufgrund der mathematischen Anforderung der Optimierungsfunktion von RNN unterscheidet sich die für die Versatzzeitoptimierung gewählte Optimierungsfunktion stark von der Funktion für die Freigabezeitoptimierung, die sich auf die Reisezeit aller Fahrzeuge im gesamten Netz bezieht.

Analog zur Freigabezeitoptimierung wird das Gradientenabstiegsverfahren zur Konstruktion des neuen Fehlerpropagierungs-Netzwerks zur Versatzzeitoptimierung verwendet. Schließlich berechnen zwei Sub-Fehlerpropagierungs-Netzwerke iterativ die Korrektur der Versatzzeit. Mit Hilfe der Randbedingungen von relativen und absoluten Versatzzeiten kann die Optimierung der Versatzzeit erfolgen.

Im Betrieb des gesamten Systems laufen die vier RNN parallel, wobei die Netzzustände der vier RNN permanent bzw. in Zeittakten ausgetauscht werden. Das Update der Umlaufzeitoptimierung wird im Parallelbetrieb nach der Zykluslänge integriert.

7 Erprobung und Bewertung des Steuerungsverfahrens

Im vorangegangenen Kapitel wurde ein neues Verfahren für die netzweite LSA-Steuerung in städtischen Straßennetzen entwickelt. Um dieses neue Steuerungsverfahren zu bewerten und die Ergebnisse zu überprüfen, wird das Verfahren, zusammen mit einer Visualisierungsoberfläche, als PC-Programm NEURA-NESC (Neural net wide traffic signal control) in C++ umgesetzt. Es werden numerische Experimente durchgeführt, um die Funktionstüchtigkeit des Modells zu verifizieren.

Um ein möglichst objektives Bild des Nutzens von NEURA-NESC zu erhalten, wird das Verfahren in einer Simulationsstudie untersucht. Ein virtuelles Netz mit acht Knotenpunkten wird als Testnetz herangezogen. Referenzfall in der Simulationsstudie ist eine Festzeitsteuerung. Es werden die charakteristischen Kennungen des Netzes (Anzahl der Halte, Wartezeit und gesamte Reisezeit) für unterschiedliche Nachfragemuster aufgeführt und miteinander verglichen.

7.1 Programm NEURA-NESC

Zur Durchführung des im folgenden Kapitel vorgestellten Experiments wurde das Steuerungsverfahren als PC-Programm unter dem Namen NEURA-NESC umgesetzt. Das Programm NEURA-NESC wurde auf Basis von NEURO-DYN-NEMO [MATHIAS, 1999] entwickelt. Es besteht aus einer Bedien- und Visualisierungsoberfläche (Visual Basic) und dem gekapselten, in C++ implementierten RNN-Modell (getrenntes OCX-Control). In Abbildung 7-1 ist die Bedienoberfläche dargestellt.

Alle benötigten Daten des Verkehrsmodells z. B. die Netz-Topologie, Quell-Ziel-Matrizen, die Phaseneinteilung und weitere Modelparameter werden über spezielle ASCII-Dateien eingelesen. Deshalb wurde die Phaseneinteilung auch in diesen ASCII-

Dateien beschrieben. Die Ausgabe erfolgt wahlweise grafisch⁵ im Bildschirmfenster oder als ASCII-Datei, die optimierte Freigabezeit und Versatzzeit enthält. Über spezielle Debug-Variablen wird festgelegt, was genau in die ASCII-Datei ausgegeben werden soll. Diese Datei ist die Grundlage für die weitere Auswertung.

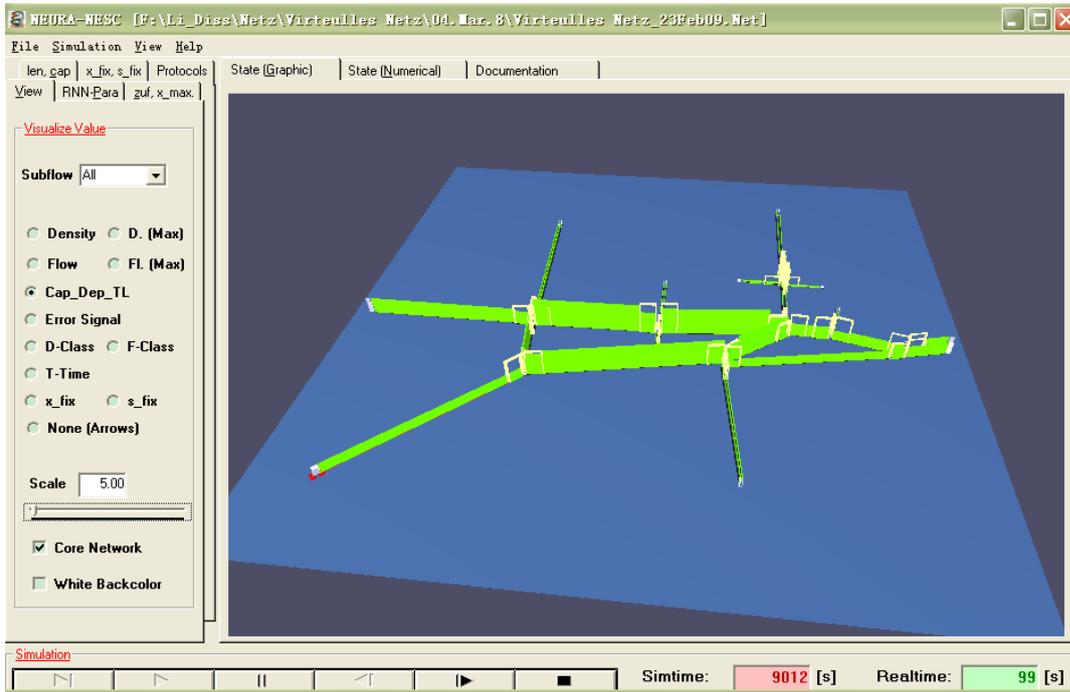


Abbildung 7-1: Bedienoberfläche des Programms NEURA-NESC

Das implementierte RNN-Steuerungsmodell kann sowohl eine statische als auch eine dynamische Steuerung durchführen. Im Fall einer dynamischen Steuerung müssen in der Input-ASCII Datei die entsprechenden Werte (Quell-Ziel-Matrizen, Netzzuflüsse) als Zeitreihen definiert werden. Weiterhin müssen die Zeitskalen τ der vier RNN vorgegeben werden. Auch der in dem Gradientenabstiegsverfahren eingesetzte Faktor ε muss angegeben werden. Abbildung 7-2 gibt einen schematischen Überblick zu den einzelnen Dateien und Funktionen und zu den internen Schnittstellen.

⁵ Die vom LSA-Programm abhängigen Streckenkapazitätswerte können direkt in der Grafik über die jeweilige Strichstärke verfolgt werden.

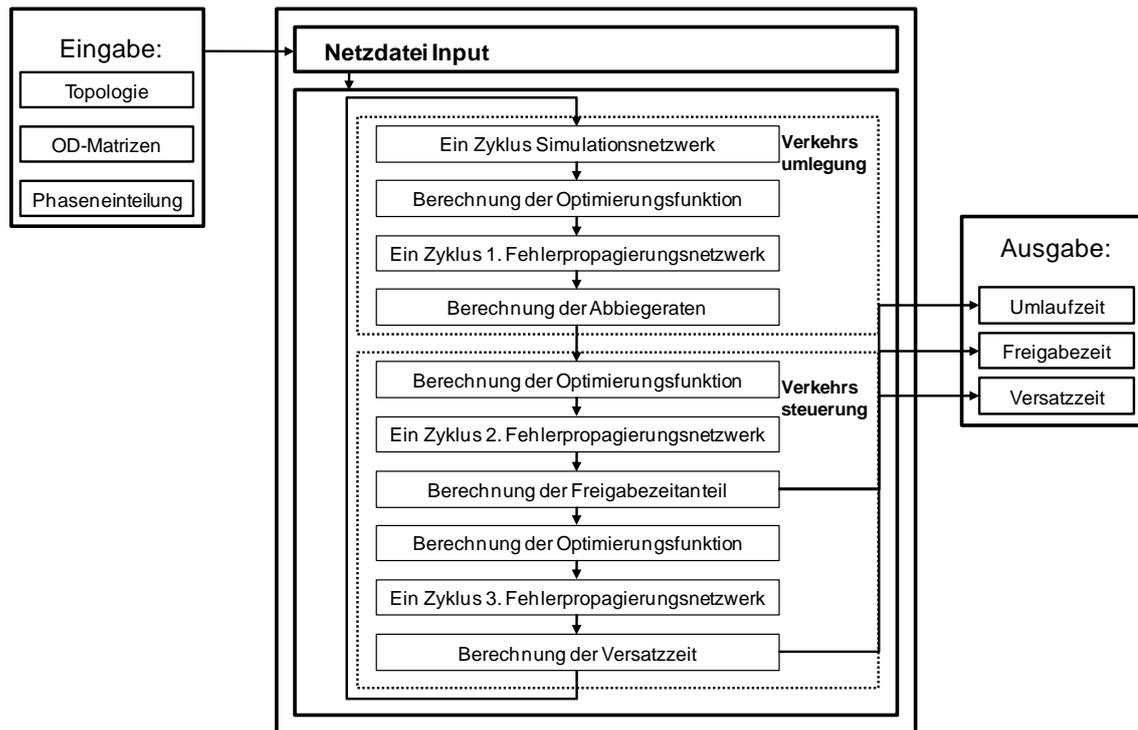


Abbildung 7-2: Funktionsdiagramm des Programms NEURA-NESC

Zur Überprüfung der Funktionstüchtigkeit des Programms wurden numerische Experimente durchgeführt. Ein isolierter Knotenpunkt und zwei koordinierte Knotenpunkte wurden bei den Experimenten eingesetzt. Die beiden Untersuchungen liefern plausible und befriedigende Ergebnisse [LI, 2008], wodurch die Nutzung des Programms für ein komplexes Straßennetz ermöglicht wird. Dieses Experiment wird im kommenden Unterkapitel ausführlich dokumentiert.

7.2 Das Testnetz

Für die Laboruntersuchung wurde ein virtuelles Netz mit acht Knotenpunkten und elf Zielen ausgewählt. Im Rahmen des vom Bundesministerium für Verkehr, Bau und Stadtentwicklung geförderten Forschungsprojekts AMONES [LÜSSMANN, 2008] wurde die verkehrliche Wirkung modellbasierter Netzsteuerungsverfahren in einem virtuellen Testfeld untersucht. Dieses virtuelle Testnetz (siehe Abbildung 7-3 und Abbildung 7-4) wird in dieser Arbeit für die Auswertung des Verfahrens NEURA-NESC eingesetzt.

Die gesamte Reisezeit des individuellen Verkehrs im innenörtlichen Straßennetz besteht aus zwei Teilen, nämlich aus der Reisezeit auf freien Strecken und aus der Verlustzeit an Knotenpunkten. Im Gegensatz zur Verlustzeit an Knotenpunkten ist die Reisezeit auf freien Strecken in einem Straßennetz ein konstanter Wert. Das virtuelle Testnetz enthält acht lichtsignalgesteuerte Knotenpunkte.

Die nachfolgende Simulation wird zeigen, dass der Anteil der Verlustzeit an festzeitgesteuerten Knotenpunkten an der gesamten Reisezeit ca. 25% beträgt. Wenn die gesamte Reisezeit im Netz minimiert werden kann, folgt daraus, dass die durch die LSA verursachte Verlustzeit deutlich reduziert werden muss. Aus diesem Grund kommt die Gesamtreisezeit als die Optimierungsfunktion in der Laboruntersuchung sowohl für die Freigabezeitoptimierung als auch für die Versatzzeitoptimierung zum Einsatz. Die in Kapitel 5 vorgestellte Wartezeitfunktion wird in der Laboruntersuchung nicht für die Freigabezeitoptimierung verwendet. Das spätere Simulationsergebnis wird zeigen, dass die Reduzierung der Gesamtreisezeit der Verringerung der Verlustzeit an Knotenpunkten entspricht. Es zeigt sich, dass die Anwendung der Gesamtreisezeit als einheitliche Optimierungsfunktion plausibel ist.

Zur Simulation der Pulkeffekte in der Einfahrt eines Straßennetzes werden fünf zusätzliche Knotenpunkte bei den Haupteinfahrten des Testnetzes eingebaut. Die LSA dieser fünf Knotenpunkte sollen nicht durch NEURA-NESC optimiert sondern nur mit einer Festzeitsteuerung betrieben werden. Diese fünf Knotenpunkte sind in der Abbildung 7-3 als 1a, 2a, 6a, 7a und 8a dargestellt.

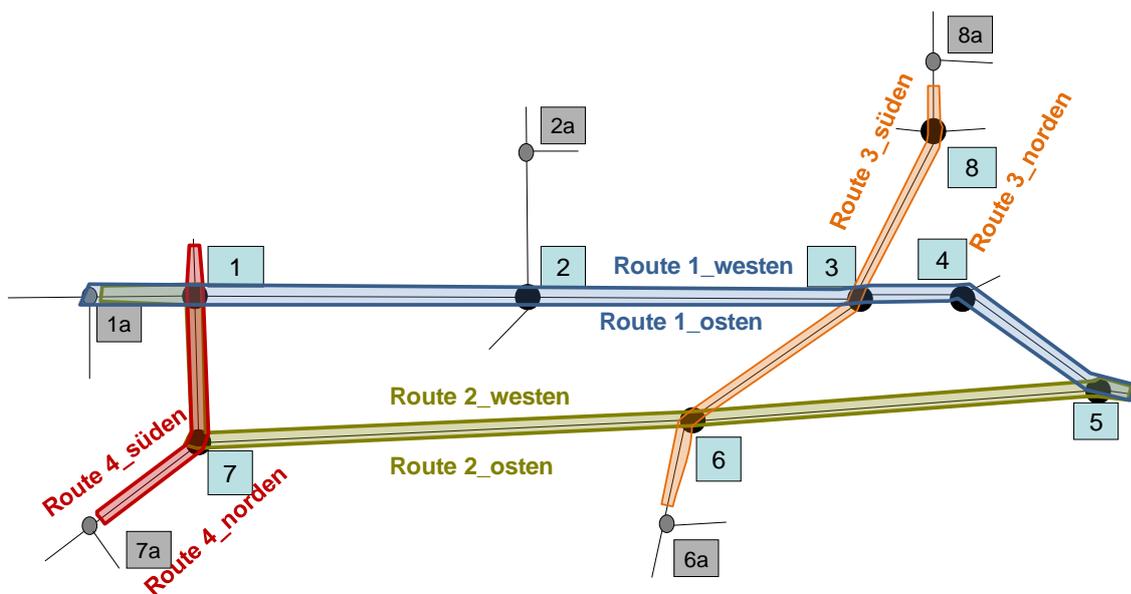


Abbildung 7-3: Übersichtplan des gesamten virtuellen Netzes

Zusätzlich werden die Umweltwirkungen des Steuerungsverfahrens betrachtet. Die CO₂-Emissionen werden für verschiedene Fälle durch das Emissionsmodul des Programms VISSIM ermittelt und bewertet.

7.3.2 Aufbau der Simulationsstudie

Um die Festzeitsteuerungsverfahren und NEURA-NESC mit Hilfe der mikroskopischen Simulation zu bewerten und zu vergleichen, wurde die Simulation über eine Schnittstelle mit den Steuerungsverfahren verbunden werden (Abbildung 7-5).

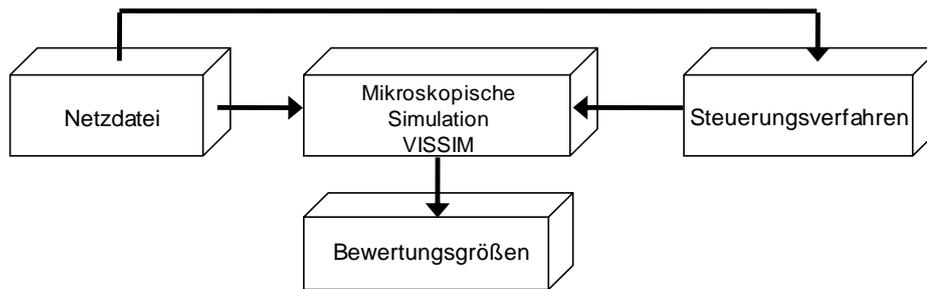


Abbildung 7-5: Allgemeines Schema des Versuchsaufbau für die Simulationsuntersuchung

Das Programm NEURA-NESC optimiert gleichzeitig sowohl die Verkehrsumlegung als auch die LSA-Steuerung, wohingegen die Festzeitsteuerung auf Basis einer vorgegebenen Umlegung mit Hilfe des Programms SITRAFFIC⁶ berechnet wird. Der vorgegebene Umlegungsgrad wird ohne Berücksichtigung der LSA-Programme bestimmt. Um Festzeitsteuerungsverfahren und NEURA-NESC-Steuerungsverfahren mit gleichen Randbedingungen zu vergleichen, wurden in der Simulationsstudie die vier Szenarien

1. Festzeitsteuerung mit vorgegebener Umlegung,
2. NEURA-NESC mit ausgeschalteter Umlegungsfunktion,
3. Festzeitsteuerung mit optimierter Umlegung sowie
4. NEURA-NESC

untersucht.

Szenario 1 basiert auf einer Festzeitsteuerung und einer vorgegebenen Umlegung. Bei Szenario 2 wird nur der NEURA-NESC Teil der Verkehrssteuerungsoptimierung (Abbildung 7-2) aktiv, die Verkehrsumlegung ist entsprechend der in Szenario 1 ermittelten Werte fixiert. Im Gegensatz zu Szenario 2 wird bei Szenario 3 nur die Verkehrsumlegung von NEURA-NESC optimiert, der Teil der Verkehrssteuerung von NEURA-NESC wird von einem Festzeitprogramm ersetzt. Schließlich werden bei Szenario 4 alle Funktionen von NEURA-NESC eingeschaltet, um die Umlegung und die Steuerung zu optimieren. Szenario 3 und 4 stellen ideale Fälle dar, bei der jeder Fahrer nach einer Rou-

⁶ Verkehrsingenieurarbeitsplatz für Lichtsignalsteuerung, Produkt von Siemens AG

tenempfehlung fährt, die anhand der LSA-Steuerung optimiert ist. In Abbildung 7-6 ist die Zusammensetzung der vier Versuchsaufbauten veranschaulicht.

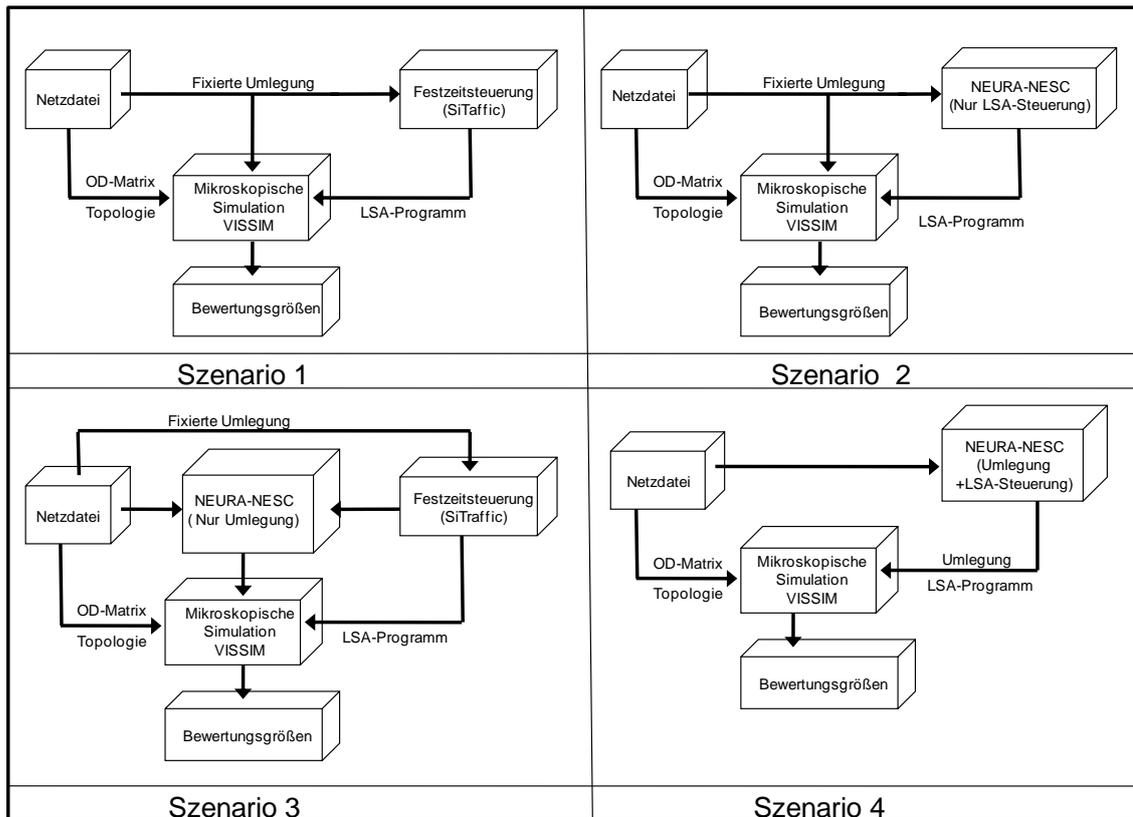


Abbildung 7-6: Zusammensetzung der Versuchsaufbauten für verschiedenen Szenarien

Sowohl die von SITRAFFIC erzeugten Festzeitsteuerungen als auch die von NEURA-NESC entwickelten Netzsteuerungen basieren auf einer vorgegebenen Phaseneinteilung. In der Simulationsstudie wird eine Zweiphasensteuerung für alle Knotenpunkte als Basis herangezogen. Im Gegensatz dazu werden die anderen Variablen wie Umlaufzeit, Freigabezeit und Versatzzeit von SITRAFFIC und NEURA-NESC ermittelt bzw. optimiert.

7.3.3 Durchführung der Simulationsstudie

Die Simulationsuntersuchungen basieren auf vorgegebenen Quell-Ziel-Matrizen, die an einem normalen Werktag von 06:00 Uhr bis 22:30 Uhr alle relevanten Verkehrsstärken und Verkehrsbeziehungen in 15 Min enthalten. Für Szenario 1 und 2 werden auch die nicht zeitabhängigen Routen vorgegeben. In Abbildung 7-7 werden die Ganglinien der elf Zuflüsse vorgestellt. Die Quell-Ziel-Matrix wird in Anhang B dargestellt.

Die Festzeitsteuerung beinhaltet die folgenden Programme:

- Morgenprogramm von 06:00 Uhr bis 09:00 Uhr
- Tagesprogramm von 09:00 Uhr bis 16:00 Uhr
- Abendprogramm von 16:00 Uhr bis 22:30 Uhr

NEURA-NESC entwickelt im Gegensatz dazu jede 15 min neue Signalprogramme, um die Steuerung der zeitabhängigen Nachfrage anzupassen.

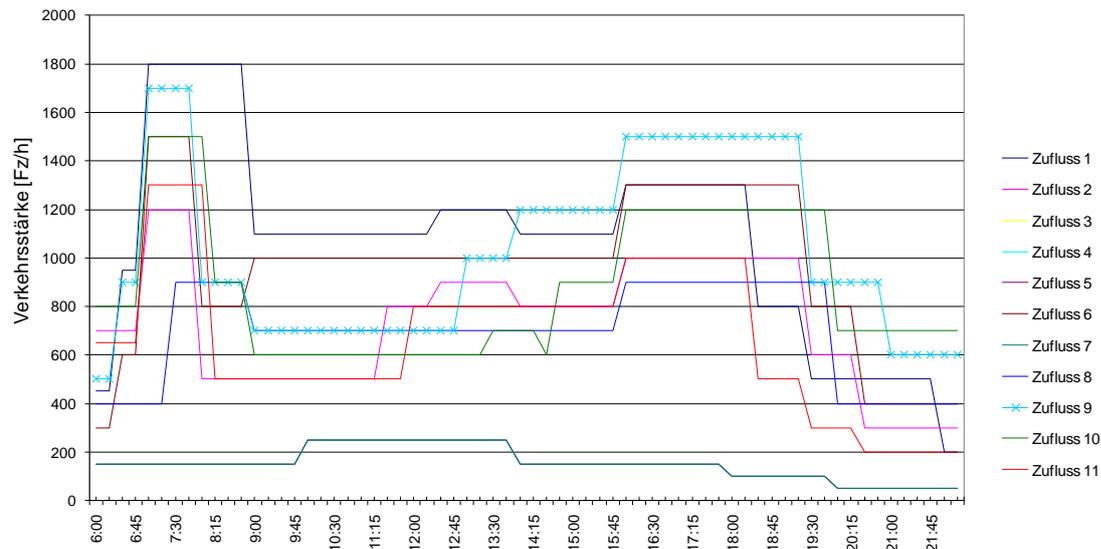


Abbildung 7-7: Tagesganglinien der Zuflüsse

Ein einziger Simulationslauf kann ein völlig falsches Ergebnis liefern [FGSV, 2006]. Um dies zu vermeiden, werden in der Simulationsstudie dieser Arbeit mehrere Simulationsverläufe mit unterschiedlichen Startzufallszahlen durchgeführt. Die erforderliche Anzahl an Simulationsläufen wird nach einer Voruntersuchung für die Morgenspitze berechnet. Die ausführliche Berechnung dazu wird in Anhang C vorgestellt. Die Anzahl der notwendigen Wiederholungen wird aufgrund der Berechnungsergebnisse mit zehn festgelegt. In den folgenden Unterkapiteln werden die Mittelwerte der zehn Simulationsverläufe für die Ergebnisbewertung verwendet.

Vor der Wirkungsanalyse der Simulationsstudie wird zuerst qualitativ überprüft, ob die von NEURA-NESC erzeugten Signalpläne die Freigabezeiten sinnvoll an die wechselnden Verkehrsverhältnisse anpassen. In den folgenden Abbildungen ist für die zwei maßgebenden Verkehrsströme die Anpassung der Freigabezeit [s] an die Verkehrsstärke 2 [Fz] im Szenario dargestellt. Bei Knotenpunkt 3 am Morgen handelt es sich dabei um die Geradeausfahrer der Westzufahrt und die Geradeausfahrer der Nordzufahrt, am Abend sind das die Geradeausfahrer der Ostzufahrt und die Geradeausfahrer der Nordzufahrt. Da in Szenario 2 auch die Umlaufzeit von NEURA-NESC optimiert wird, gibt es keine einheitliche Umlaufzeit für das Gesamtnetz. Um die Anpassung der Freigabezeit an die Verkehrsstärke besser darzustellen, sind sie in der Darstellung auf eine Umlaufzeit von 90 s genormt.

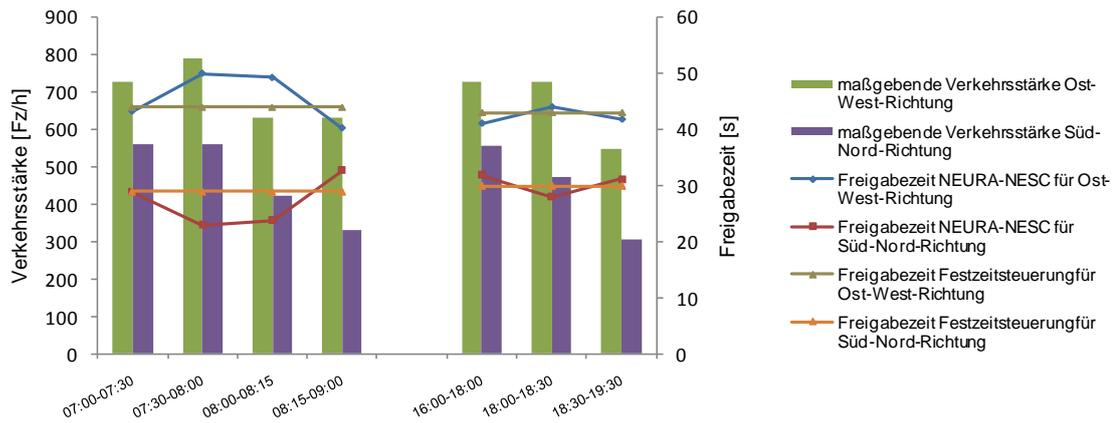


Abbildung 7-8: Anpassung der Freigabezeit an die Verkehrsstärke

Die Abbildung 7-8 zeigt, dass sich die von NEURA-NESC ermittelten Freigabezeiten im Vergleich zu denen der Festzeitsteuerung bei den betrachteten Strömen an die Verkehrsnachfrage anpasst. Im kommenden Unterkapitel wird analysiert, ob NERA-NESC in der Lage ist, Freigabezeit und Versatzzeit netzweit anzupassen, die Ergebnisse der Simulationsstudie werden auch vorgestellt.

7.3.4 Ergebnisse der Simulationsstudie

In der Simulationsstudie wurden die mittlere Verlustzeit, die Gesamtreisezeit, die CO₂-Emission und die Routenreisezeit ermittelt. Der Beobachtungszeitraum wurde von 06:00 Uhr bis 22:30 Uhr gewählt. In den Abbildungen 7-9, 7-10 und 7-11 sind die Tagesverläufe der drei erstgenannten Bewertungskriterien veranschaulicht. In Abbildung 7-12, 7-14 und 7-16 sind die relativen Veränderungen der ersten drei Bewertungskriterien in den vier Szenarien für Morgen (06:00 Uhr bis 09:00 Uhr), Tag (09:00 Uhr bis 16:00 Uhr) und Abend (16:00 Uhr bis 22:30 Uhr) vorgestellt. Die Abbildungen 7-13, 7-15 und 7-17 veranschaulichen die Vergleiche der Reisezeit entlang der acht Routen.

In der Simulation ergaben sich folgende Ergebnisse:

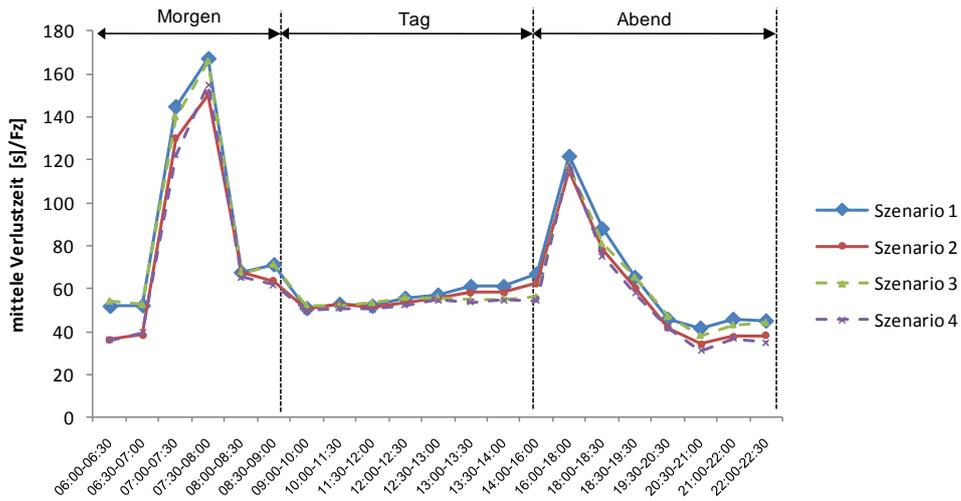


Abbildung 7-9: Mittlere Verlustzeit pro Fahrzeug

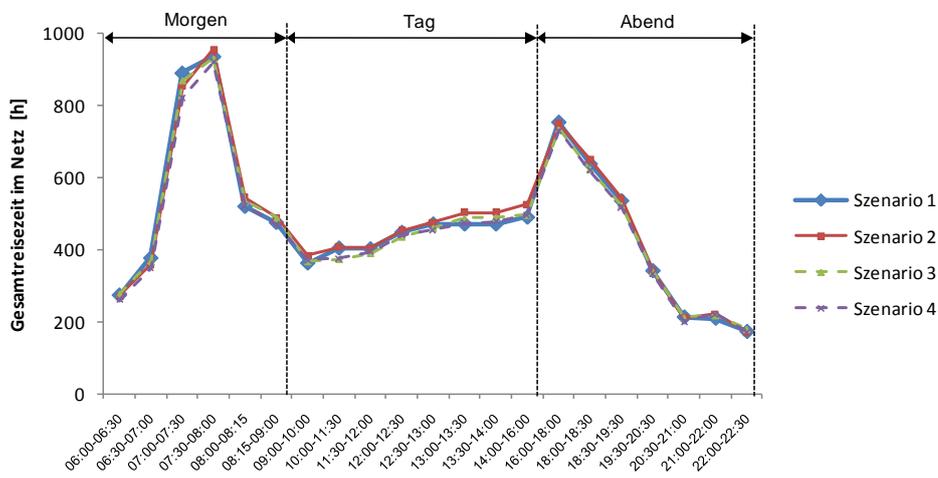


Abbildung 7-10: Gesamtreisezeit im Netz

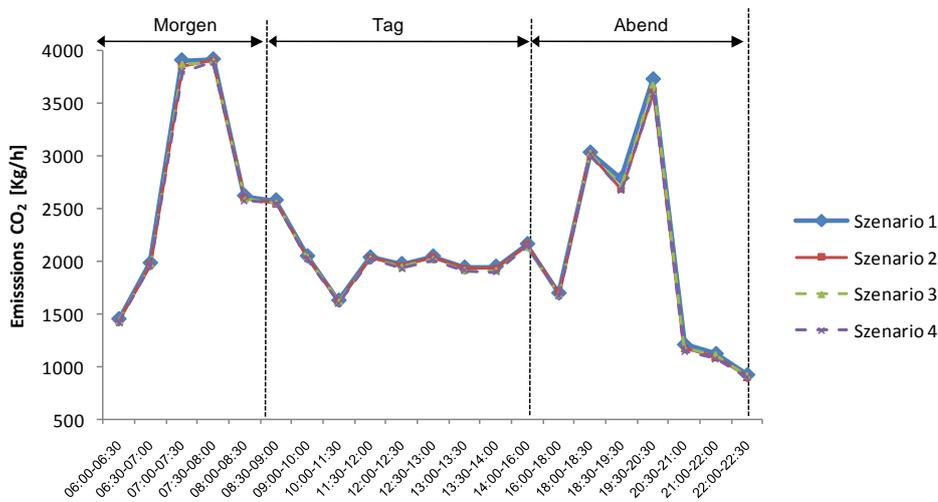


Abbildung 7-11: CO₂-Emission

Morgens (06:00Uhr bis 09:00 Uhr)

- Nach dem t-Test sind die Mittelwertdifferenzen der Verlustzeiten von Szenario 2 zum Szenario 1 auf dem 10%-Niveau signifikant. Im Vergleich zu Szenario 1 konnten bei Szenario 2 13,1 % der mittleren Verlustzeit eingespart werden. Von 06:00 Uhr bis 07:00 Uhr werden fast 22 % der Verlustzeit eingespart. Während der Morgenspitze von 07:00 Uhr bis 08:00 Uhr verringert sich die Verlustzeit mit Hilfe der NEURA-NESC-Steuerung um 10,8%. Hier soll erwähnt werden, dass von 06:00 Uhr bis 07:00 Uhr die Umlaufzeiten von NEURA-NESC kürzer als die der Festzeitsteuerung sind, wohingegen die Umlaufzeit während des Zeitraums von 07:00 Uhr bis 08:00 Uhr länger als die der Festzeitsteuerung ist. Das ist ein Grund, dass die Verbesserung von 06:00 Uhr bis 07:00 Uhr deutlich höher ist.
- Für die Gesamtreisezeit der Fahrzeuge unterscheidet sich zwischen Szenario 1 und Szenario 2 signifikant (10 %-Niveau). Die Verbesserung ist im Vergleich zur Wartezeit deutlich geringer: Die Gesamtreisezeit bei Szenario 2 ist 4,2 % kürzer als bei Szenario 1.
- Signifikant (10 %-Niveau) sind die Verkürzungen der Reisezeit auf der Route 1 in Richtung Osten zwischen Szenario 1 (456 s) und Szenario 2 (400 s), auf der Route 3 in Richtung Norden zwischen Szenario 1 (360 s) und Szenario 2 (302 s), auf der Route 4 in Richtung Norden zwischen Szenario 1 (381,1 s) und Szenario 2 (331 s) und auf der Route 4 in Richtung Süden zwischen Szenario 1 (228 s) und Szenario 2 (195 s). Dagegen sind die Reisezeiten auf den schwach belasteten Routen z. B. Route 2_osten und Route 2_westen in Szenario 2 länger als in Szenario 1. Die Reisezeit auf der Route 1 in Richtung Westen und auf der Route 3 in Richtung Süden in Szenario 1 und Szenario 2 bleiben fast unverändert.
- Obwohl die mittlere Verlustzeit und Gesamtreisezeit in Szenario 2 viel geringer als in Szenario 1 sind, ist die Reduzierung der CO₂-Emissionen nicht signifikant (10%-Niveau). Die Ursache liegt darin, dass sich die Anzahl der Halte nicht reduziert. Die häufigen Brems- und Beschleunigungsvorgänge gleichen die Wirkung der Verringerung der Verlustzeit aus.
- Die zusätzlich auf Basis der Signalsteuerung optimierte Umlegung verringert auch die mittlere Verlustzeit und die Gesamtreisezeit (10%-Niveau signifikant). Szenario 3 und Szenario 4 haben jeweils 5,2 % bzw. 2,8 % geringere Verlustzeiten im Vergleich zu Szenario 1 und Szenario 2. Bei der Gesamtreisezeit liegt die Verbesserung jeweils bei 2,3 % und 1,7 %. Bei den CO₂-Emissionen ist die Reduzierung nicht signifikant (10%-Niveau).
- Die Verkürzungen der Reisezeit sind auf Route 2 in beiden Richtungen und auf Route 4 in Richtung Norden zwischen Szenario 3 und Szenario 1 signifikant (10 %-Niveau). Die hauptsächlichsten Veränderungen der unter Berücksichtigung der Signalsteuerung optimierten Umlegung sind die Fahrten von 1 nach 9 und von

9 nach 1. Im Vergleich zu Szenario 1 befahren bei Szenario 3 11,0 % mehr Fahrzeuge die Route 1 in Richtung Osten und 8,2 % mehr Fahrzeuge die Route 1 in Richtung Westen. Dadurch wird die Route 2 in beiden Richtungen entlastet, was zu einer Reduktion der Reisezeit führt. Zusätzlich profitiert die Route 4 in Richtung Norden von der Entlastung der Route 2.

- Signifikant (10%-Niveau) sind die Verkürzungen der Reisezeit auf Route 2 in beiden Richtungen und auf Route 4 in beiden Richtungen zwischen Szenario 4 und Szenario 2. Im Vergleich zu Szenario 2 fahren in Szenario 4 7,3 % mehr Fahrzeuge von 1 nach 9 auf Route 1 in Richtung Osten, wodurch die Route 2 in Richtung Süden entlastet wird. Dadurch wird die Reisezeit auf Route 4 in Richtung Süden verkürzt. Die Verbesserungen der Reisezeiten auf Route 2 in Richtung Westen und auf Route 4 in Richtung Norden sind auf die Verlängerung der zugehörigen Freigabezeit am Knotenpunkt 1 zurückzuführen.

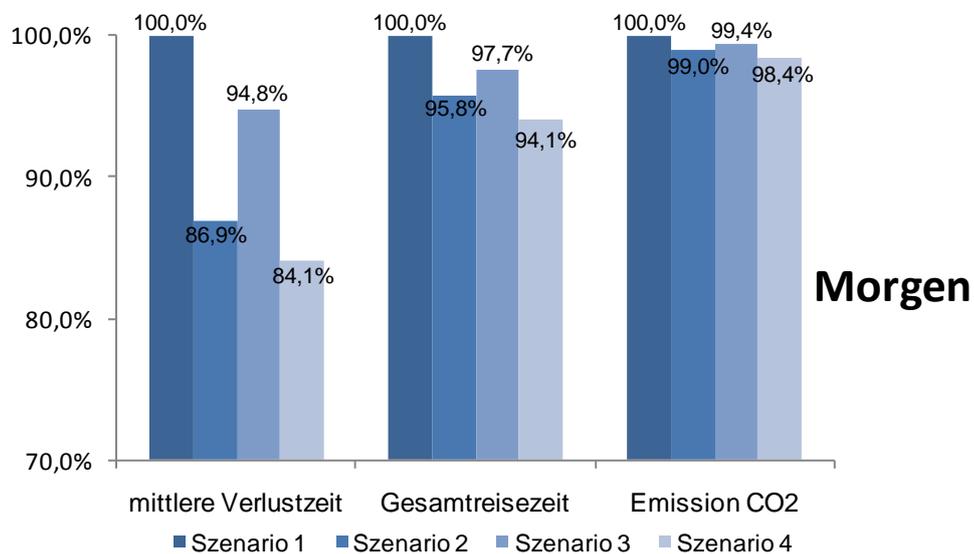


Abbildung 7-12: Relative Veränderung der mittleren Verlustzeit, Gesamtreisezeit und Emission CO₂ morgens

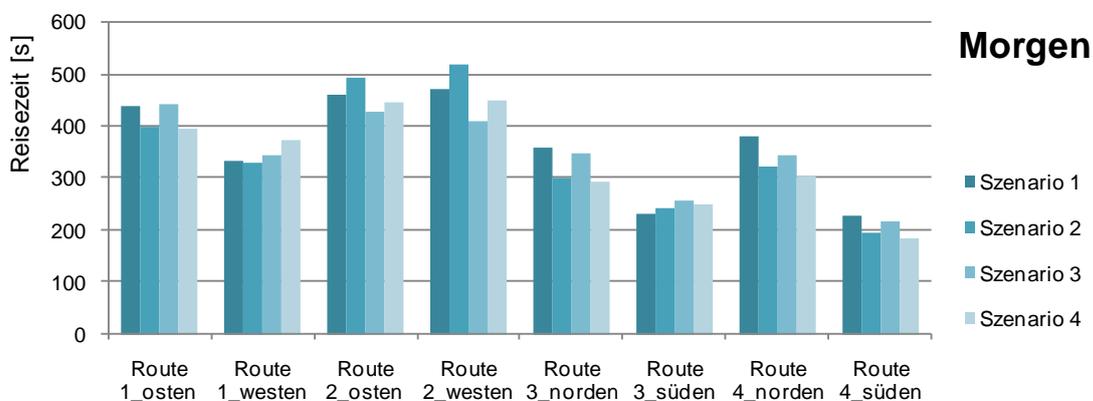


Abbildung 7-13: Vergleich der acht Routenreisezeiten morgens

Tag (09:00 Uhr bis 16:00 Uhr)

- Die Differenz der Mittelwerte der Verlustzeiten zwischen Szenario 1 und Szenario 2 ist signifikant (10 %-Niveau). Im Vergleich zu Szenario 1 konnten bei Szenario 2 2,6 % der mittleren Verlustzeit eingespart werden. Das ist deutlich weniger als morgens.
- Die Differenzen der Gesamtreisezeiten und der CO₂-Emissionen sind zwischen Szenario 2 und Szenario 1 nicht signifikant (10 %-Niveau).
- Signifikant (10 %-Niveau) sind die Verkürzungen der Reisezeiten auf der Route 1 in Richtung Osten zwischen Szenario 1 (347 s) und Szenario 2 (338 s) und auf Route 4 in Richtung Norden zwischen Szenario 1 (194 s) und Szenario 2 (180 s). Dagegen sind die Reisezeiten auf den schwach belasteten Routen z. B. Route 2_osten und Route 2_westen bei Szenario 2 länger als bei Szenario 1.
- Die zusätzlich auf Basis der Signalsteuerung optimierte Umlegung verringert auch die mittlere Verlustzeit und Gesamtreisezeit (10 %-Niveau signifikant). Szenario 3 und Szenario 4 hat jeweils 5,0 % und 4,0 % geringe Verlustzeiten im Vergleich zu Szenario 1 und Szenario 2. Bei der Gesamtreisezeit liegt die Verbesserung jeweils bei 3,6 % und 3,0 %. Die Veränderungen von Szenario 3 und 4 zu Szenario 1 und 2 sind tagsüber relativ größer als morgens. Die Ursache liegt darin, dass die Umlegung bei leicht belastetem Netz mehr Optimierungsraum hat.
- Die Verkürzungen der Reisezeiten sind auf der Route 1 in Richtung Osten, auf Route 2 in Richtung Westen und auf Route 4 in Richtung Norden zwischen Szenario 3 und Szenario 1 signifikant (10 %-Niveau). Die hauptsächlichsten Veränderungen der auf Basis der Signalsteuerung optimierten Umlegung sind die Fahrten von 1 nach 9 und von 9 nach 1. Im Vergleich zu Szenario 1 fahren bei Szenario 3 9,1 % mehr Fahrzeuge von 1 nach 9, auf Route 2 in Richtung Osten und 5,0 % mehr Fahrzeuge von 9 nach 1, auf Route 1 in Richtung Westen. Damit die Routen 2 in Richtung Westen entlastet werden. Dadurch wird die Reisezeit auf der Route 4 in Richtung Norden auch verkürzt.
- Signifikant (10 %-Niveau) sind die Verkürzungen der Reisezeiten auf der Route 1 in Richtung Osten, auf der Route 2 in Richtung Westen und auf der Route 4 in Richtung Norden zwischen Szenario 4 und Szenario 2. Im Vergleich zu Szenario 1 fahren in Szenario 3 4,0 % mehr Fahrzeuge von 1 nach 9, auf Route 2 in Richtung Osten und 2,9 % mehr Fahrzeug von 9 nach 1, auf Route 1 in Richtung Westen. Die Verbesserungen der Reisezzeiten auf Route 2 in Richtung Westen und auf Route 4 in Richtung Norden sind nicht nur von der Entlastung von Route 2 sondern auch von der Verlängerung der zugehörigen Freigabezeit am Knotenpunkt 1 beeinflusst.

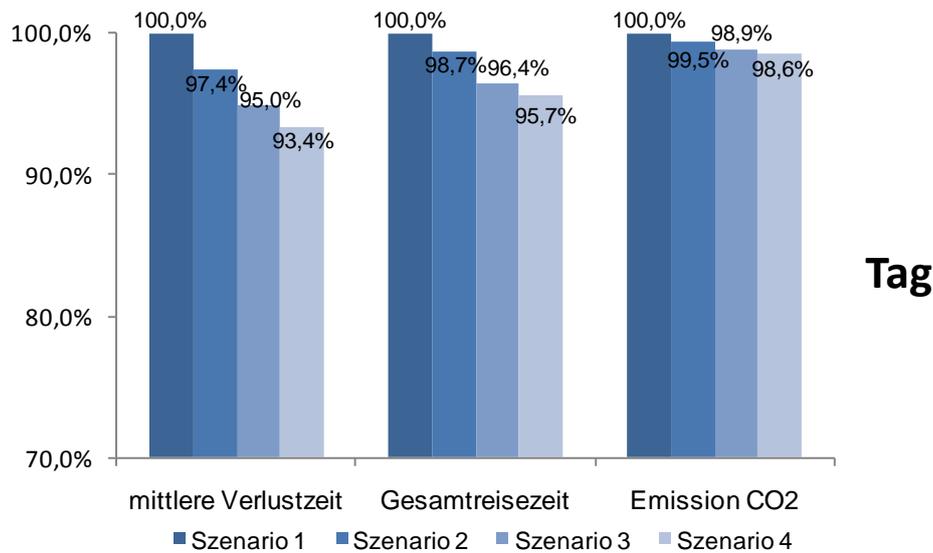


Abbildung 7-14: Relative Veränderung der mittleren Verlustzeit, Gesamtreisezeit und Emission CO₂ tagsüber

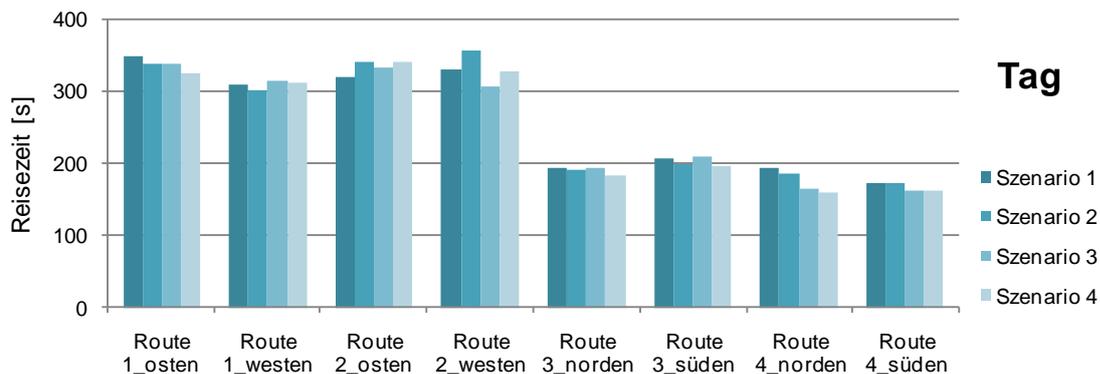


Abbildung 7-15: Vergleich der acht Routenreisezeiten tagsüber

Abend (16:00 Uhr bis 22:30 Uhr)

- Signifikant (10 %-Niveau) ist der Unterschied der Verlustzeit zwischen Szenario 1 und Szenario 2. Im Vergleich zu Szenario 1 ergeben sich bei Szenario 2 abends Verminderungen der mittleren Verlustzeit von 8,5 %. Während der Abendspitze von 18:00 Uhr bis 19:30 Uhr wird eine Einsparung der Verlustzeit durch die NEURA-NESC-Steuerung von 8,1 % erzielt.
- Die Differenz der Gesamtreisezeit von Szenario 2 und Szenario 1 ist signifikant (10 %-Niveau). Die Reduzierung der CO₂-Emissionen ist auch signifikant (10%-Niveau) und mit 2,1 % relativ größer.
- Signifikant (10 %-Niveau) sind die Verkürzungen der Reisezeiten auf der Route 1 in Richtung Osten zwischen Szenario 1 (448 s) und Szenario 2 (416 s), auf Route 3 in Richtung Norden zwischen Szenario 1 (232 s) und Szenario 2 (205 s) und auf Route 4 in Richtung Norden zwischen Szenario 1 (246 s) und Szenario 2 (216 s).

rio 2 (225 s). Dagegen sind die Reisezeiten auf den schwach belasteten Routen z. B. Route 2_osten und Route 2_westen bei Szenario 2 länger als bei Szenario 1.

- Die auch die Signalsteuerung berücksichtigende optimierte Umlegung verringert auch die mittlere Verlustzeit und die Gesamtreiszeit (10 %-Niveau signifikant). Szenario 3 und Szenario 4 haben jeweils 5,2 % und 1,5 % geringe Verlustzeit im Vergleich zum Szenario 1 und Szenario 2. Für die Gesamtreiszeit beträgt die Verbesserung jeweils um 1,7 % und 1,6 %. Die Reduzierung der CO₂-Emission ist nicht signifikant (10 %-Niveau). Es wird aus den Simulationsergebnissen deutlich, dass das Verbesserungspotential von Szenario 3 zu 1 viel größer als das von Szenario 4 zu 2 ist. Die Ursache für diesen Effekt liegt darin, dass das Netzsteuerungsverfahren Netzströme anpassen kann, so dass der Optimierungsraum für die Umlegung von NEURA-NESC-Verfahren eher geringer als der Festzeitsteuerung ist.
- Die Verkürzungen der Reisezeiten sind auf der Route 2 in Richtung Westen und auf der Route 4 in Richtung Norden zwischen Szenario 3 und Szenario 1 signifikant (10 %-Niveau). Im Vergleich zu Szenario 1 fahren bei Szenario 3 7,9 % mehr Fahrzeuge von 9 nach 1 fließen auf Route 1 in Richtung Westen. Damit wird die Route 2 in Richtung Westen entlastet.
- Signifikant (10 %-Niveau) sind die Verkürzungen der Reisezeiten auf Route 2 in beiden Richtungen zwischen Szenario 4 und Szenario 2. Im Vergleich zur Szenario 2 fahren bei Szenario 4 2,1 % mehr Fahrzeuge von 1 nach 9 auf Route 1 in Richtung Osten, wodurch die Route 2 in Richtung Osten entlastet wird. Die Verbesserungen der Reisezeiten auf Route 2 in Richtung Westen und auf Route 4 in Richtung Norden sind hauptsächlich durch die Verlängerung der zugehörigen Freigabezeit am Knotenpunkt 1 verursacht.

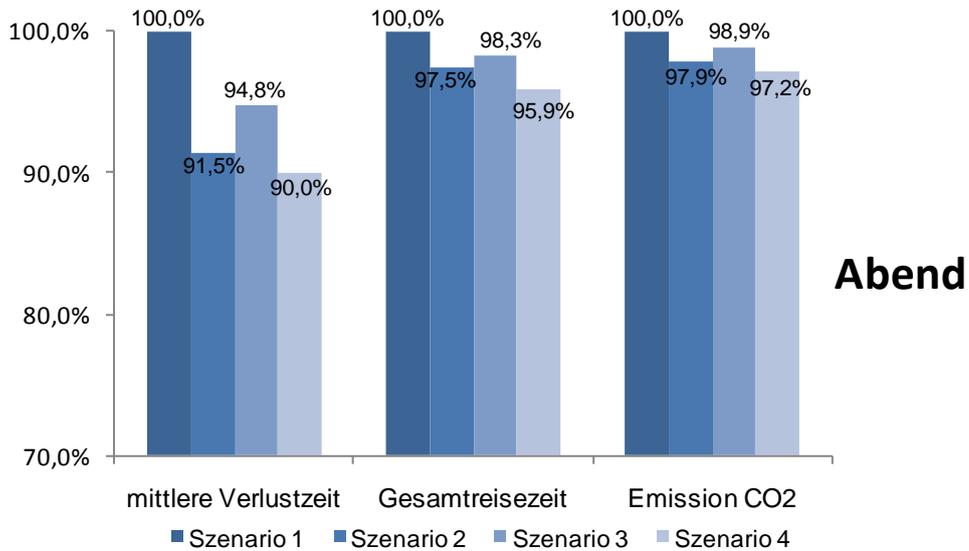


Abbildung 7-16: Relative Veränderung der mittleren Verlustzeit, Gesamtreisezeit und Emission CO₂ abends

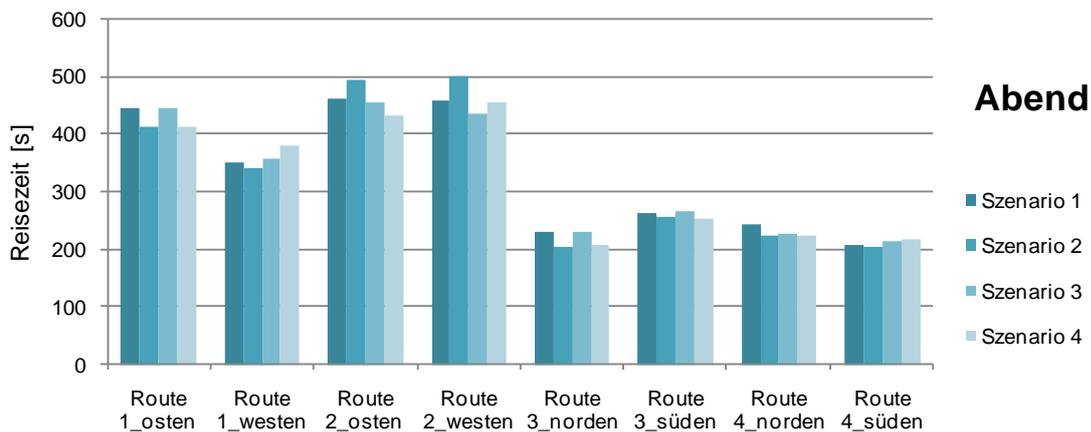


Abbildung 7-17: Vergleich der acht Routenreisezeiten abends

7.4 Zwischenfazit

In Kapitel 7 wurde zuerst das Programm NEURA-NESC vorgestellt, das die softwaretechnische Umsetzung des neuartigen Verfahrens realisiert. Um die Funktionstüchtigkeit und die Leistungsfähigkeit des Programms zu überprüfen, wurde NEURA-NESC mit einem virtuellen Netz getestet.

Der Referenzfall für die Bewertung ist eine tageszeitabhängige Festzeitsteuerung. Vier Szenarien wurden in der Simulationsstudie so definiert, dass die Optimierung der Umlegung für die Festzeitsteuerung und für NEURA-NESC zugleich betrachtet werden können

Drei Bewertungskriterien, nämlich die mittlere Verlustzeit, die Gesamtreisezeit und die CO₂-Emissionen, wurden für den Test betrachtet. Die Ergebnisse der Simulationsstudie zeigen, dass die mittlere Verlustzeit morgens und abends durch die NEURA-NESC-Steuerung um 13,1 % und 8,5 % verringert werden können. Für die Gesamtreisezeit und die CO₂-Emissionen sind die Reduzierungen nicht so deutlich. Unter Berücksichtigung einer aufgrund der geänderten LSA-Steuerung anderen Verkehrsverteilung können noch mehr Einsparungen von Verlustzeit, Gesamtreisezeit und CO₂-Emissionen erzielt werden. Die Berücksichtigung der veränderten Verkehrsnachfrageverteilung durch das Umlegungsverfahren wirkt sich dabei bei der Festzeitsteuerung stärker aus als bei der bereits mit NEURA-NESC optimierten Steuerung.

8 Fazit und Ausblick

8.1 Zusammenfassung

Die vorliegende Dissertation widmet sich der Entwicklung eines neuartigen Verfahrens zur netzweiten modelbasierten Lichtsignalsteuerung auf der Basis von Rekurrenten Neuronalen Netzen (RNN). Die zugrundeliegende Idee war dabei, die quantitative Abhängigkeit der Kapazitätswerte von Strecken und der Lichtsignalsteuerung auszunutzen. Über die direkte Abbildung von Topologie und Dynamik des Straßennetzes mit dem RNN gelang es, das Verkehrsmodell einzurichten. Mit Hilfe eines leistungsfähigen Lernalgorithmus „Recurrent Backpropagation“ werden die Lichtsignalsteuerungsparameter eingestellt.

Das Verkehrsmodell besteht aus zwei kooperierenden RNN, von denen das erste, das Simulations-Netzwerk, das reale Verkehrsgeschehen nachbildet. Die Definition der dynamischen Gleichung bezieht verschiedene Parameter, unter anderem Abbiegebeziehungen und Kapazität, ein. Das zweite RNN des Verkehrsmodells bzw. das Fehlerpropagierungs-Netzwerk hat die Aufgabe, den Parameter Abbiegebeziehung einzustellen.

Zur Optimierung der Lichtsignalprogramme wurde der Kapazitätswert für die Interpretation der Freigabezeit und der Versatzzeit als Steuerungsparameter verwendet, wofür zwei weitere Fehlerpropagierungs-Netzwerke konstruiert wurden. Dem dritten Fehlerpropagierungs-Netzwerk kommt die Aufgabe zu, die Freigabezeitanteile festzulegen. Die Umlaufzeit wird in der Netzebene einheitlich in Intervallen von 180 s so berechnet werden, dass die Auslastung der maßgebenden Knotenpunkte gerade 90 % erreicht. Der Zusammenhang zwischen Freigabezeitanteil und Kapazitätswert der Strecken in den Knotenpunkten wird in einer dynamischen Gleichung interpretiert. Die strategische Systemoptimierungsfunktion bezieht die Wartezeit ein und berücksichtigt die Unterschiede der Straßentypen. Die Berechnung der Freigabezeitkorrektur geht letztlich auf das bekannte Gradientenabstiegsverfahren zurück. Anhand gewisser Randbedingungen wird auf Basis der Freigabezeitkorrektur die Freigabezeitverteilung optimiert.

Das vierte Fehlerpropagierungs-Netzwerk beschäftigt sich mit der Optimierung der Versatzzeit. Die Streckenleistungsfähigkeiten der Strecken, die die zwei koordinierten Knotenpunkte verbindet, werden zur Definition der Versatzzeit als dritter Steuerungs-

parameter eingeführt. Mit Hilfe der Wartezeit kann die quantitative Abhängigkeit von Streckenleistungsfähigkeit und Versatzzeit bestimmt werden. Die Optimierungsfunktion für die Versatzzeitoptimierung bezieht die Gesamtreisezeit ein. Analog der Freigabezeitoptimierung wird die Versatzzeitkorrektur durch das vierte Fehlerpropagierungs-Netzwerk berechnet. Unter den Randbedingungen von relativen und absoluten Versatzzeiten erfolgt die Optimierung der Versatzzeit.

Die softwaretechnische Umsetzung wurde durch das Programm NEURA-NESC realisiert. Zur Bewertung des neuartigen netzweiten Steuerungsverfahrens wird für ein virtuelles Netzwerk mit 8 Knotenpunkten das Programm NEURA-NESC mit einer Festzeitsteuerung verglichen. Da NEURA-NESC sowohl die LSA-Programme als auch die Umlegung optimiert, werden vier Szenarien für die Simulationsstudie definiert, so dass auch berücksichtigt wird, ob eine Umlegungsoptimierung vorhanden ist. Als Bewertungskriterien kommen die mittlere Verlustzeit, die gesamte Reisezeit im Netz und CO₂-Emissionen zum Einsatz.

In der Simulationsstudie wurde nachgewiesen, dass morgens und abends durch die NEURA-NESC-Steuerung die Verbesserung der mittleren Verlustzeit von 13,1 % bzw. 8,5 % erzielt werden können. Die Verminderung der Gesamtreisezeit und der CO₂-Emissionen sind nicht so deutlich. Durch die Berücksichtigung der LSA-Steuerung in der Verkehrsumlegung wird auch eine bemerkenswerte Einsparung der Verlustzeit, der Gesamtreisezeit und der CO₂-Emissionen realisiert. Die Berücksichtigung der veränderten Verkehrsverteilung wirkt sich bei der Festzeitsteuerung stärker aus als bei der bereits mit NEURA-NESC optimierten Steuerung.

8.2 Weiterer Forschungsbedarf

Das in der Arbeit entwickelte Verfahren auf Basis kooperierender RNN ist neuartig. Die Funktionsfähigkeit und die Leistungsfähigkeit des Verfahrens wurden grundsätzlich nachgewiesen. Das Verfahren sollte in umfassenderen Tests weiter überprüft werden. Zusätzlich kann das Verfahren auch in einigen Bereichen weiterentwickelt werden. Deshalb seien diese Aspekte als weiterer Forschungsbedarf zusammengestellt:

1. Für die Überprüfung der Auswirkungen des Verfahrens in einer realen Testumgebung ist ein Feldtest notwendig. Damit werden all diejenigen Einflüsse überprüft, die aufgrund der Komplexität des realen Straßennetzes in der Simulationsstudie nicht ausreichend abgebildet werden können.
2. Das Verfahren wurde bis jetzt als Offline-Verfahren getestet. Durch die Erstellung einer Schnittstelle zwischen NEURA-NESC und den Lichtsignalanlagen kann das Verfahren auch Online zum Einsatz kommen. Zusätzlich wird eine online berechnete O-D-Matrix für bestimmte Zeitintervalle erforderlich sein.

3. Das in der Arbeit entwickelte Verfahren verwendet zwei verschiedene Optimierungsfunktionen, die jeweils Verlustzeit und Gesamtreisezeit einbeziehen. Aufgrund dessen ist die Optimierung auf diese beiden Kriterien beschränkt. Weitere Kriterien, wie z. B. CO₂-Emmission könnten über eine erweiterte Optimierungsfunktion einbezogen werden. Es ist darauf zu achten, dass die Optimierungsfunktion stetig und differenzierbar ist.
4. Die Berücksichtigung des ÖPNV kann mit Hilfe der Schichtstruktur von Simulationsnetzwerken theoretisch realisiert werden. Prinzipiell können die Steuerungsparameter verschiedener Schichten über unterschiedliche Optimierungsfunktionen eingestellt werden. Damit eröffnet sich eine Möglichkeit, priorisierte Routen zu berücksichtigen.
5. Für die Stabilität des für die Lichtsignalsteuerung zuständigen Fehlerpropagierungs-Netzwerks fehlt bisher ein analytischer Beweis. Beim parallelen Betrieb der RNN stellt sich auch die Frage, unter welchen Bedingungen (Zeitskala, Zykluslänge etc.) der Systemzustand durch das Fehlerpropagierungs-Netzwerk stabil nachgeführt werden kann.
6. Ist das für Innerorts entwickelte Signalsteuerungsverfahren auch für die Verkehrssteuerung auf Autobahnen einsetzbar? Hier soll vorerst die Modellierung des Autobahnnetzes auf Basis von RNN erklärt werden. Welche dynamische Gleichung und welche Optimierungsfunktion z. B. für die Harmonisierung der Geschwindigkeit kann hier verwendet werden?

Literaturverzeichnis

ALMASRI, E.; FRIEDRICH, B. (2005):

Online offset optimisation in urban networks based on cell transmission model
Proc. of the 5th European Congress on Intelligent Transport Systems and Services,
Hannover.

ALMASRI, E. (2006):

A New Offset Optimization Method for Signalized Road Networks.
Veröffentlichung des Instituts für Verkehrswirtschaft, Straßenwesen und Städtebau,
Universität Hannover.

ALMEIDA, L.B. (1988):

Backpropagation in Perceptrons with Feedback. In Neural Computers, eds. Eckmiller
und von der Malsburg, Berlin: Springer-Verlag, S. 199-208.

ALLSOP, R.E. (1972):

Delay at a fixed Time Traffic-Signal: Theoretical Analysis.
In Transportation Science 6, Nr.3, S. 260-305.

BRETHERTON R.D; RAI,G.I. (1982):

The use of SCOOT in low flow conditions.
Traffic Engineering & Control, December.

BINNING, J. (2008):

TRANSYT: Current Versions and known problems
TRANSYT User Group, London.

BOHLINGER,M. (2006):

Koordinierte Lichtsignalsteuerung
OptiV– Entscheidungs- und Optimierungsmethoden im Verkehr, Hannover.

BOLTZE, M. (1988):

Optimierung von Umlaufzeiten in der Lichtsignalsteuerung für Straßennetze.
Dissertation von Fachbereich Wasser und Verkehr der TH Darmstadt.

BOLTZE, M.; FRIEDRICH, B. (2007):

Innovation in der Lichtsignalsteuerung – Die Neufassung der Richtlinien für Lichtsig-

nalanlagen (RiLSA).

Straßenverkehrstechnik, Heft 4.

BRAUN, R.; WEICHENMEIER, F. (2005):

Automatische Offline-Optimierung der lichtsignaltechnischen Koordinierung des mIV im städtischen Netz unter Verwendung genetischer Algorithmen.

Straßenverkehrstechnik, Heft 5.

BRAUN, R. (2006):

Projektsteckbrief TRAVOLUTION.

Lehrstuhl für Verkehrstechnik, TU München.

BRAUN, R.; KEMPER, C. (2008):

GALOP – Online, Ein Genetischer Algorithmus zur netzweiten Online-Optimierung der Lichtsignalsteuerung.

HEUREKA- Tagungsband, Stuttgart.

BUSCH, F.; KRUSE, G. (1993);

MOTION – Ein neues Verfahren für die städtische Lichtsignalsteuerung in Netzen.

HEUREKA- Tagungsband, Karlsruhe.

BUSCH, F. ; KRUSE, G. (2002):

MOTION for SITRAFFIC – Optimierung der Lichtsignalsteuerung im Einsatz.

HEUREKA- Tagungsband, Karlsruhe.

BUSCH, F. (2007):

Vorlesungsskript Verkehrstechnik A.

Lehrstuhl für Verkehrstechnik, TU München.

BUSCH, F. (2011):

Verkehrssteuerung mit Lichtsignalanlagen -verkehrsabhängige Steuerung in Netzen.

Summer School Verkehr, Verkehrsbeeinfluss in Netzen

CEYLAN, H.; BELL, M.G.H. (2004):

Traffic Signal optimization based on genetic algorithm approach, including driver's Routing.

Traffic Research Report B 38, S. 329-342.

CEYLAN, H.; BELL, M.G.H. (2005):

Genetic Algorithm solution for the stochastic equilibrium transportation networks under congestion.

Traffic Research Report B 39, S. 169-185.

CHILUKURI, B.R.; PERRIN, J.; MARTIN, P. T. (2004):

SCOOT and Incidents: Performance Evaluation in Simulated Environment.

Transportation Research Record 1867, National Research Council, Washington, D.C., S. 224-232.

- CLAYTON, A.J:H. (1941);
Road traffic calculations.
J. Inst. Civil Engineers.16 Page 247-284
- DONATI, F.; MAURO, V.; RONCOLINI, G.; VALLAURI, M. (1984);
A Hierarchical Decentralised traffic Light Control System.
The first realisation: 'Progeto Torin'. IFAC 9 th World Congress, Vol II, 11G/A-1.
- FAGHRI, A.; HUA, J. (1997):
Development of a Neural Traffic Signal Control System.
Delaware Transportation Institute, DTI Report 103, May.
- FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN (1992):
Richtlinien für Lichtsignalanlagen – RiLSA. Köln.
- FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN (2001):
Handbuch für die Bemessung von Straßenverkehrsanlagen. Köln.
- FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN (2003):
Richtlinien für Lichtsignalanlagen – RiLSA Teilfortbeschreibung. Köln.
- FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN (2006):
Hinweise zur mikroskopischen Verkehrsflusssimulation. Köln.
- FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN (2008):
OptiV– Entscheidungs- und Optimierungsmethoden im Verkehr.
CD-Veröffentlichung der Ergebnisse des vom BMBF/BMWi geförderten Projekts OptiV
durch den Arbeitskreis „Entscheidungs- und Optimierungsmethoden“, Köln.
- FORSCHUNGSGESELLSCHAFT FÜR STRAßEN- UND VERKEHRSWESEN (2010):
Richtlinien für Lichtsignalanlagen – RiLSA. Köln.
- FRIEDRICH, B. (1999):
Ein verkehrsadaptives Verfahren zur Steuerung von Lichtsignalanlagen.
Schriftenreihe des Fachgebiets Verkehrstechnik und Verkehrsplanung, TU München.
- FRIEDRICH, B. (2002):
Verkehrsadaptive Steuerung von Lichtsignalanlagen - Ein Überblick.
Festschrift zum Ehrenkolloquium für Univ.- Prof. Dr./UCB Hartmut Keller 01.03.2002 an
der TU München. Veröffentlichung des Fachgebiets Verkehrstechnik und Verkehrspla-
nung, TU München.
- FRIEDRICH, B.; SCHÜTTE, C.; KESSLER, D. (2002):
Strategisches Verkehrsmanagement - Eine konsistente Theorie und ihre Umsetzung.
Straßenverkehrstechnik – Heft 4.

GARBEN, M.; HECK, H.M.; HOTOP, R.; KELLER, H.; MEIßNER, J.D.; SAHLING, B.M.; STOTTMEISTER V. (1988):
SIGMA: Ein Optimierungsverfahren zur koordinierten Lichtsignalsteuerung.
Straßenverkehrstechnik – Heft 4.

GARTNER, N.H. (1982):
Demand-responsive decentralised urban traffic control.
US Department of Transportation DOT/RSPA/DPB.

GEVAS SOFTWARE GMBH; PTV AG (2008):
CROSSIG Standard Handbuch.

GIFFIN, J. (2005):
Traffic System Presentation.
Sydney, Australia.

HALE, D.K. (2005)
Traffic Network Study Tool, Transyt-7F, United States version.
Mc Trans Center, University of Florida.

HEYDECKER, B.G.; ADDISON, J.D. (2006)
Analysis of dynamic traffic assignment.
Proceedings of the 1st international symposium on dynamic traffic assignment, Leeds, UK.

HOLROYD, J.; HILLER, J.A. (1971):
The Glasgow Experiment: PLIDENT and after. RRL Report.

HÜFTLE, M. (2006):
Heuristiken "Ameisen, Neuronen und Vererbung".
OptiV– Entscheidungs- und Optimierungsmethoden im Verkehr, Hannvor.

HUNT, P.B; ROBERTSON, D.I.; BERTHERTON, R.D.; WINTON, R.I. (1981):
SCOOT – a traffic responsive method of coordinating signals.
TRRL Laboratory Report 1014.

HUNT, P.B; ROBERTSON, D.I.; BERTHERTON, R.D.; WINTON, R.I. (1981):
The SCOOT online traffic signal optimization technique.
Traffic Engineering & Control, April, pp.190-192.

KELLER; BOESEFELDT; GRABEN; HECK; HOTOP; SAHLING; STOTTMEISTER (1982):
Forschungsgesellschaft für Straßen- und Verkehrswesen.
Hinweise zur Anwendung von Entscheidungs- und Optimierungsmethoden im Verkehrswesen FGSV Schriftreihe, Nr. 113.

KELLER, H.; FRIEDRICH, B.; WULFFIUS, H.; WOLTERS, W.; MERTZ, J.; SCHÜLER, T. (2002):
Verkehrsadaptive Netzsteuerung von Lichtsignalanlagen.
Straßenverkehrstechnik, Heft 7.

KRUSE, G. (2003):
MOTION Netzsteuerung-Optimierung der Lichtsignalsteuerung im Einsatz.
Steuerung kommunaler Verkehrsnetze: Was können und was kosten die neuen Ver-
fahren? Was setzen sie voraus? Stuttgart.

KURZAK, H. (1974):
Untersuchung zur verkehrsabhängigen Signalsteuerung von Stadtstraßenknotenpunk-
ten mit Hilfe der Simulation.
Schriftenreihe des Lehrstuhls Verkehrs- und Stadtplanung der TU München.

LI, Y.; MATHIAS, P. (2008):
Net wide traffic signal control with recurrent neural network.
15th World Congress on ITS, New York City, 16.-20. Nov. 2008

LIU, H.; VAN LINT, H.; VAN ZUYLEN, H. (2007):
Neural-Network-Based Traffic Flow Model for Urban Arterial Travel Time Prediction.
TRB 86th Annual Meeting Compendium of Papers CD-ROM.

LOWRIE, P.R. (1980):
The Sydney Coordinated Adaptive Traffic System – Principles, Methodology, Algo-
rithmus.
Proceedings International Conference on Road Traffic Signalling (IEE), London.

LÜßMANN, J. (2008):
Anwendung und Analyse modellbasierter Netzsteuerungsverfahren in städtischen
Straßennetzen.
Projektsteckbrief, Lehrstuhl für Verkehrstechnik, TU München.

MAIER, P. (2007):
Verkehrsadaptive Netzsteuerungen: Vergleich und Bewertung verschiedener Verfah-
ren.
BALANCE-Roadshow, München.

MATHIAS, P. (1999):
Statische und dynamische Verkehrsumlegung mit Rekurrenten Neuronalen Netzen.
Shaker Verlag (ISBN 3-8265-6720-X), Herzogenrath.

MATHIAS, P. (1999):
Static and dynamic traffic assignment with recurrent neural networks.
The 14th International Symposium on Transportation and Traffic Theory (ISTTT), Jeru-
salem, Israel.

- MATHIAS, P. (2003):
Integrated Dynamic Traffic Assignment and Control for Road Networks that are regarded as Recurrent Neural Networks.
Workshop of Traffic Modeling, Barcelona.
- MAURO, V.; DI TARRANTO, C. (1989):
UTOPIA. Proc. Of the IFAC Symposium, Paris.
- MAURO, V; DI TARRANTO, C. (1990):
"UTOPIA." Control, Computers, Communications in Transportation.
Selected Papers from the IFAC Symposium, Pergamon Press, S. 245 – 252.
- MCCULLOCH, W.; PITTS, W. (1943):
A logical calculus of the ideas immanent in nervous activity.
Bulletin of mathematical Biophysics.
- MERTZ, J. (2001):
Ein mikroskopisches Verfahren zur verkehrsadaptiven Knotenpunktsteuerung mit Vorrang des öffentlichen Verkehrs.
Veröffentlichung des Fachgebiets Verkehrstechnik und Verkehrsplanung, TU München.
- MÜCK, J. (2002):
Schätzverfahren für den Verkehrszustand an Lichtsignalanlagen unter Verwendung haltliniennaher Detektoren.
HEUREKA - Tagungsband, Karlsruhe.
- MÜCK, J. (2008):
Schätz- und Optimierungsverfahren in der Adaptiven Netzsteuerung SITRAFFIC Motion MX.
HEUREKA - Tagungsband, Stuttgart.
- MÖLLER, K. (1987):
Signalgruppeorientiertes Modell zur Optimierung von Festzeitprogrammen an Einzelknotenpunkten.
Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe, Heft 37.
- NHUN, H.; HESS, M. (2006)
Verkehrsgeographie
Schöningh, Paderborn.
- PARKER D.B. (1985):
Learning Logic.
Technical Report TR-47, Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology.

- PEARL, J. (1984):
Heuristics: intelligent Search Strategies for Computer Problem Solving.
Addison-Wesley, Reading Massachusetts.
- PEARLMUTTER, B.A. (1989):
Learning State Space Trajectories in Recurrent Neural Networks.
International Joint Conference on Neural Networks. IEEE. Vol. II, 365-372. New York.
- PINEDA, F.J. (1988):
Dynamics and architectures for Neural Computation.
Journal of Complexity, Band 4, S. 213-215.
- PINEDA, F.J. (1989):
Recurrent Back-Propagation and the Dynamical Approach to Adaptive Neural Computation.
Neural Computation 1, S. 161-172.
- PINEDA, F.J. (1995):
Backpropagation: theory, architectures, and applications.
IEEE Expert: Intelligent Systems and Their Applications, Vol. 10, no. 2, S. 99-135.
- PLOSS, G. (1993):
Ein dynamisches Verfahren zur Schätzung von Verkehrsbeziehungen aus Querschnittszählungen.
Schriftenreihe des Fachgebiets Verkehrstechnik, TU München.
- REY, G.; WENDER, K. (2008):
Neuronale Netze - Eine Einführung in die Grundlagen, Anwendungen und Datenauswertung.
Verlage Hans Huber, Bern.
- ROBERTSON, D. I. (1969):
TRANSYT, A Traffic Network Study Tool.
Road Research Laboratory, RRL Report LR 253.
- ROHDE, J. (2006):
Bildung von Gleichgewichtszuständen in Verkehrsnetzen nach Nutzer- oder Systemoptimum.
OptiV- Entscheidungs- und Optimierungsmethoden im Verkehr, Hannover.
- RUMMELHART D.E., HINTON G.E., WILLIAMS R.J. (1986):
Learning Representations by Back-Propagation Errors.
Nature 323, S. 533-536.
- SARAF, R. (1994):
Adaptive Traffic Control using Neural Networks.
Dissertation, Department of Civil and Environmental Engineering, Vanderbilt University.

- SCHNABEL, W.; LÄTZSCH, L. ; LOHSE, D. (2006)
Grundlagen der Straßenverkehrstechnik und der Verkehrsplanung, Band1: Verkehrstechnik.
Beuth-Verlag, Berlin.
- SCOOT (2008):
Advice Leaflet 1: The "SCOOT" Urban Traffic Control System,
www.scoot-utc.com, Stand 20.5.2008.
- SIMS, A. G.; DOBINSON, K. W. (1980):
The Sydney Coordinated Adaptive Traffic (SCATS) System: Philosophy and Benefits.
IEEE Transactions on Vehicular Technology, Volume 29(2), S. 130-137.
- SIEMENS AG I&S ITS (2008):
Motion MX 4.0 Handbuch, München.
- STEVANOVIC, A.Z., MARTIN, P.T. (2007):
SCOOT and Coordinated Actuated Traffic Control Evaluated through Microsimulation.
Presented at 87th TRB Annual Meeting and recommended for publication in TRR,
Journal of the Transportation Research Board.
- TRB, TRANSPORT RESEARCH BOARD (2000):
Highway Capacity manual (HCM)
National Research Council, Waschington, D.C.
- TULLY, I. M. (1977):
Synthesis of sequences for traffic signal controllers using techniques of the theory of graph.
OUEL Report 1189/77. University of Oxford.
- UTOPIA (2006):
UTOPIA ,Urban Traffic Control: Overview.
SWARCO Traffic Management .
- VAN LINT, H. (2007):
Identifying congestion patterns with state space neural networks.
Advanced OR and AI Methods in Transportation S. 736-741.
- VAN LINT, J.; HOOGENDOORN, S. (2005):
Accurate freeway travel time prediction with state-space neural networks under missing data
Transport Research Part C13 S. 347-369
- WARDROP, J.G. (1952):
Some theoretical aspects of road traffic research.
Proceedings, Institution of Civil Engineers II(1), S. 325-378.

- WEBSTER, F.V. (1958):
Traffic Signal Settings.
Road Research Technical Paper No.39, Road Research Laboratory, London.
- WERBOS, P. (1974):
Beyond Regression: New Tools for prediction and Analysis in Behavioral Sciences.
Dissertation, Harvard University.
- WILLIAMS, R.J.; ZIPSER, D. (1989):
A Learning Algorithm for Continually Running Fully Recurrent Neural Networks.
Neural Computation, Band 1, S. 270-280.
- WILLIAMS, R.J.; PENG, J. (1990):
An efficient gradient-based algorithms for online training of recurrent network trajectories.
Neural Computation, Band 2, S. 490-501.
- WU, N. (1990):
Wartezeit und Leistungsfähigkeit von Lichtsignalanlagen unter Berücksichtigung von Instationarität und Teilgebundenheit des Verkehrs.
Schriftenreihe, Lehrstuhl für Verkehrswesen, Ruhr-Universität Bochum.
- WU, N. (2003):
Bemessung und Bewertung von Lichtsignalanlagen – Vergleich der neuen Regelwerke HCM 2000 und HBS 2001.
Straßenverkehrstechnik, Heft 12.
- ZHANG, J. (2008):
TRANSYT: The CTM for TRANSYT.
TRANSYT User Group, London.
- ZIPSER, D. (1989):
A Subgrouping Strategy that Reduces Complexity and Speeds Up Learning in Recurrent Networks.
Neural Computation, Band 1, S. 552-558.

Abbildungsverzeichnis

Abbildung 2-1:	Modellknotenpunkt und die zugehörigen Signalgruppen	5
Abbildung 2-2:	Verriegelungsmatrix des Modellknotenpunkts.....	6
Abbildung 2-3:	Zwischenzeitenmatrix des Modellknotenpunkts	6
Abbildung 2-4:	Mögliche Phaseneinteilung des Modellknotenpunkts	7
Abbildung 2-5:	Signalprogramm des Modellknotenpunkts	8
Abbildung 2-6:	Einteilung der Steuerungsverfahren.....	10
Abbildung 2-7:	Aufbau der modellbasierte Steuerungen [RiLSA, 2010]	12
Abbildung 2-8:	Überblick der modellbasierten LSA-Steuerungsverfahren [BUSCH, 2011]	13
Abbildung 2-9:	Ablaufdiagramm des Steuerungsverfahrens TRANSYT	14
Abbildung 2-10:	Prinzipien des Verkehrsmodells SCOOT [SCOOT, 2008]	16
Abbildung 2-11:	Die Systemarchitektur des Verfahrens MOTION [Mück, 2008]	18
Abbildung 3-1:	Künstliche Neuronale Netze und biologisches Vorbild.....	21
Abbildung 3-2:	Ein allgemeines rekurrentes neuronales Netz.....	23
Abbildung 3-3:	Ein mehrschichtiges Feedforward-Netzwerk und seine Backpropagation	25
Abbildung 3-4:	Das mit dem Originalnetz in Abbildung 3-2 korrespondierende Fehlerpropagierungs-Netz.....	28
Abbildung 4-1:	Die Grobstruktur des gesamten Modells	31
Abbildung 4-2:	Ein einfaches Verkehrsnetzmuster.....	32
Abbildung 4-3:	Das entsprechende Simulations-RNN.....	32
Abbildung 4-4:	Drei Elementen für die dynamische Gleichung	35
Abbildung 4-5:	Flussdiagramm für Rekurrent Π - Σ -Netz [MATHIAS,2003]	36
Abbildung 4-6:	Ein zwei schichtiges Simulations-Netz mit zwei Zielen	37
Abbildung 4-7:	Die zwei kooperierenden Fehlerpropagierungs-Netzwerke	44
Abbildung 4-8:	Das geschichtete nichtlineare Simulationsnetz und die zugehörigen zwei kooperierenden linearen Fehlerpropagierungs-Netze.....	45
Abbildung 4-9:	Die beide Betriebsmodelle von statischem Modell und dynamischem Modell	48
Abbildung 5-1:	Beispiel der Verbindungsstrecken im Knotenpunkt.....	52
Abbildung 5-2:	Die Bewegung der Potentialflächen und die optimale Trajektorie	62
Abbildung 5-3:	Die beiden zusammenarbeitenden, eine Lösung der Gl. 5-54 liefernden, Fehlerpropagierungs-Netzwerke	69
Abbildung 5-4:	Das Gesamtsystem mit Umlegungsmodell und Steuerungsmodell	72
Abbildung 5-5:	Die Arbeitsweise bei Parallelbetrieb.....	73
Abbildung 6-1:	Zeit-Weg-Diagramm zur Definition der relativen Versatzzeiten	76
Abbildung 6-2:	Beispiel zur Definition versatzzeitabhängiger Streckenleistungsfähigkeit	77
Abbildung 6-3:	Zeit-Weg-Diagramm zur Definition der optimalen Versatzzeit	78
Abbildung 6-4:	Zweibereichsmodell des Verkehrsmodells auf der Strecke	79

Abbildung 6-5:	Beispiel zur Bestimmung der versatzzeitabhängigen Wartezeit	79
Abbildung 6-6:	Die wartzeitabhängige Funktion von Streckenleistungsfähigkeit	81
Abbildung 6-7:	Beispielstrecken zum Update der Versatzzeit	87
Abbildung 6-8:	Beispielnetz zur Update der Versatzzeit.....	88
Abbildung 6-9:	Ablauf der modellbasierten LSA-Optimierung	89
Abbildung 6-10:	Arbeitsweise der vier RNN und der Update der Umlaufzeit	90
Abbildung 7-1:	Bedienoberfläche des Programms NEURA-NESC	94
Abbildung 7-2:	Funktionsdiagramm des Programms NEURA-NESC.....	95
Abbildung 7-3:	Übersichtplan des gesamten virtuellen Netzes	96
Abbildung 7-4:	Übersichtplan der Knotenpunkte des virtuellen Netzes.....	97
Abbildung 7-5:	Allgemeines Schema des Versuchsaufbau für die Simulationsuntersuchung	98
Abbildung 7-6:	Zusammensetzung der Versuchsaufbauten für verschiedenen Szenarien.....	99
Abbildung 7-7:	Tagesganglinien der Zuflüsse.....	100
Abbildung 7-8:	Anpassung der Freigabezeit an die Verkehrsstärke.....	101
Abbildung 7-9:	Mittlere Verlustzeit pro Fahrzeug.....	102
Abbildung 7-10:	Gesamtreisezeit im Netz.....	102
Abbildung 7-11:	CO ₂ -Emission.....	102
Abbildung 7-12:	Relative Veränderung der mittleren Verlustzeit, Gesamtreisezeit und Emission CO ₂ morgens	104
Abbildung 7-13:	Vergleich der acht Routenreisezeiten morgens.....	104
Abbildung 7-14:	Relative Veränderung der mittleren Verlustzeit, Gesamtreisezeit und Emission CO ₂ tagsüber	106
Abbildung 7-15:	Vergleich der acht Routenreisezeiten tagsüber.....	106
Abbildung 7-16:	Relative Veränderung der mittleren Verlustzeit, Gesamtreisezeit und Emission CO ₂ abends	108
Abbildung 7-17:	Vergleich der acht Routenreisezeiten abends.....	108

Tabellenverzeichnis

Tabelle 4-1: Zusammenhang von Straßennetz und Simulations-RNN [MATHIAS,1999]	33
--	----

Symbolverzeichnis

Neuronale Netze:

C_S	Zyklusparameter des Simulations-Netzwerks
C_U	Zyklusparameter des 1. Fehlerpropagierungs-Netzwerks
C_G	Zyklusparameter des 2. Fehlerpropagierungs-Netzwerks
C_V	Zyklusparameter des 3. Fehlerpropagierungs-Netzwerks
E	Optimierungsfunktion
t_x	Zeitintervall des Simulations-Netzwerks
t_y	Zeitintervall des 1. Fehlerpropagierungs-Netzwerks
t_z	Zeitintervall des 2. Fehlerpropagierungs-Netzwerks
t_v	Zeitintervall des 3. Fehlerpropagierungs-Netzwerks
$V = (v_{ij})$	Gewichte-Matrix des Fehlerpropagierungs-Netzwerks
$W = (w_{ij})$	Gewichte-Matrix des Simulations-Netzwerks
$x(t)$	Statusvariable des Simulations-Netzwerks
$y(t)$	Statusvariable des Fehlerpropagierungs-Netzwerks
$z(t)$	Statusvariable des Fehlerpropagierungs-Netzwerks

Topologie des Verkehrsnetzes:

A_i	Menge der Strecken, die von Knoten i ausgehen
D	Menge der Zielstrecken
E_i	Menge der Strecken, die in Knoten i enden
F	Menge der Strecken, deren Kapazitäten von der LSA abhängig sind
K	Menge der Knoten eines Netzes
M	Anzahl der Strecken eines Netzes
N	Anzahl der Knoten eines Netzes
O	Menge der Quellstrecken
Q	Anzahl der Quellen
S	Menge der Strecken eines Netzes
V	Menge der Strecken, deren Leistungsfähigkeiten von Versatzzeit abhängig sind
Z	Anzahl der Ziele

Verkehrstechnik:

a_{ij}	Produkt von s_i und r_j
C_i	Kapazität der Strecke i
f_i	Verkehrsfluss auf der Strecke i
h_i	tatsächlicher Abfluss der Strecke i
Lst_k^{hj}	Leistungsfähigkeit der Strecke k , die Strecke h und Strecke j verbindet
p	Anzahl der Phasen
q_i	Netzzufluss in Quellstrecke i
$q_{maßg,i}$	maßgebende Fahrstreifenverkehrsstärke in der Phase i
s_i	möglicher maximaler Abfluss der Strecke i
S_i	Sättigungsverkehrsstärke der Strecke i
r_i	Zuflussrestriktion der Strecke i
tg_i	Freigabezeit der Strecke i
tg_{erf}	erforderliche Freigabezeit
$t_{sperr,i}$	Sperrzeit der Strecke i
Δtg_i	Veränderung der Freigabezeit der Strecke i
tu_i	Umlaufzeit des Signalprogramms von Knoten i oder Umlaufzeit der der Strecke i zugehörigen Knotenpunkte
tu_{erf}	erforderliche Umlaufzeit
tu_{gew}	gewählte Umlaufzeit
$t_{z,i}$	maßgebende Zwischenzeit für einen Phasenwechsel
u_i	tatsächlicher Zufluss der Strecke i
v_i	mittlere Geschwindigkeit auf der Strecke i
VZ_i	absolute Versatzzeit von Knoten i
$rel.VZ_{AB}$	relative Versatzzeit zwischen Fahrstreifen A und Fahrstreifen B
VZ_{AB}	relative Versatzzeit zwischen Fahrstreifen A und Fahrstreifen B
ΔVZ_{AB}	Veränderung der relativen Versatzzeit zwischen Fahrstreifen A und Fahrstreifen B
$opt. VZ_{AB}$	optimale relative Versatzzeit zwischen Fahrstreifen A und Fahrstreifen B
Dif_VZ_{AB}	Differenz zwischen relativer Versatzzeit VZ_{AB} und optimaler relativer Versatzzeit $opt. VZ_{AB}$
w_i	mittlere Verlustzeit in der Warteschlange auf der Strecke i
w_{ij}	Anteil des Verkehrsflusses der Strecke i mit Abbiegerate zu Strecke j
x_i	Verkehrsmenge (Anzahl Fahrzeuge) auf der Strecke i
\hat{x}_i	gemessene Verkehrsmenge auf der Strecke i
z_i	Netzabfluss aus Zielstrecke i
π_i	mittlere Reisezeit auf der Strecke i
τ	Zeitintervall

Anhang A – Berechnungen der Zusammenhänge von Versatzzeit und mittlerer Wartezeit

Die vollständigen Berechnungen der Zusammenhänge von Versatzzeit und mittlerer Wartezeit pro Fahrzeug werden in diesem Anhang dargestellt. Die Wartezeitfunktionen sollen anhand von Bedingungen in drei Gruppen (Gl. A-1) ermittelt werden.

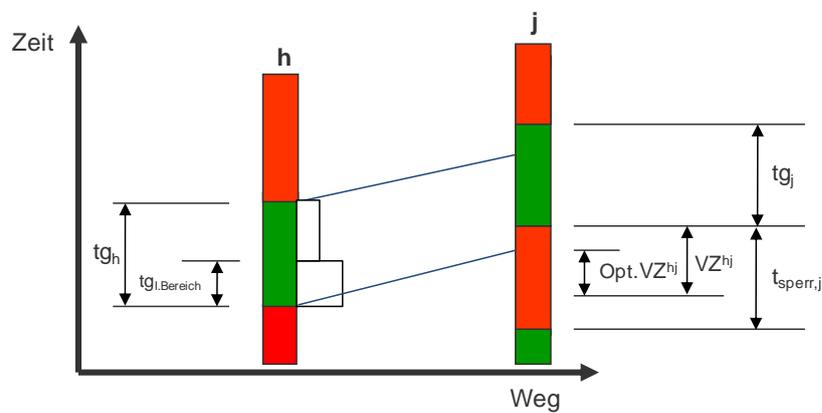


Abbildung A-1: Parameter zur Berechnung der Wartezeitfunktionen

$$\left\{ \begin{array}{l} \text{Gruppe 1: } VZ^{hj} > t_{sperr,j} + opt.VZ^{hj} \\ \text{Gruppe 2: } opt.VZ^{hj} \leq VZ^{hj} \leq t_{sperr,j} + opt.VZ^{hj} \\ \text{Gruppe 3: } VZ^{hj} < opt.VZ^{hj} \end{array} \right. \quad (A-1)$$

Gruppe 1:

Unter der Bedingung aus A-1 wird die Gruppe 1 zusätzlich in drei Fälle unterteilt.

Fall 1a:

$$VZ^{hj} > tg_h + opt.VZ^{hj} + t_{sperr,j} \quad (A-2)$$

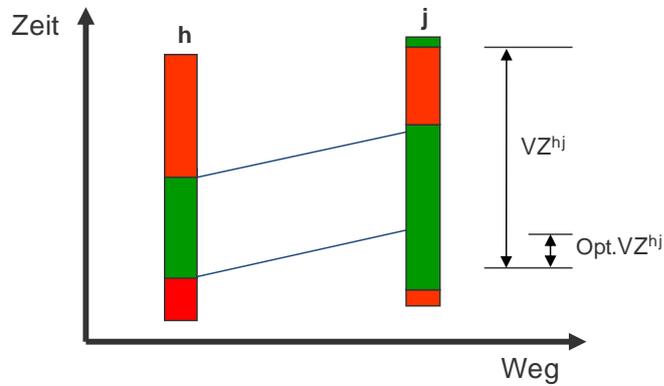


Abbildung A-2: Zeit-Weg-Diagramm zur Berechnung des Falls 1a

$$w = 0 \quad (A-3)$$

Fall 1b:

$$tg_h + opt.VZ^{hj} \leq VZ^{hj} \leq tg_h + opt.VZ^{hj} + t_{sperr,j} \quad (A-4)$$

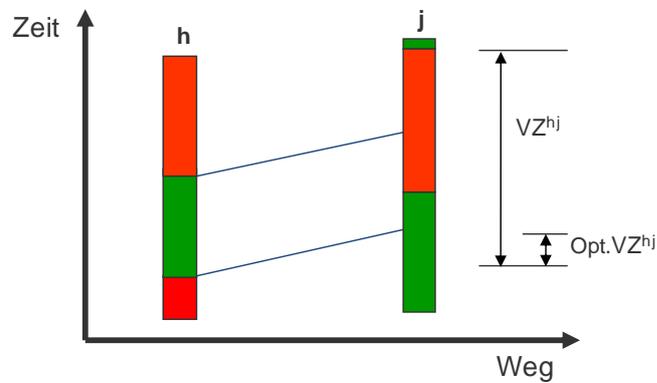


Abbildung A-3: Zeit-Weg-Diagramm zur Berechnung des Falls 1b

In Hinsicht des Bereichsmodells (Abbildung 6-4) kann man die Wartezeitfunktionen (Gl. A-5 und A-7) des Falls 1b in zwei Situationen (Abbildung A-4 und Abbildung A-5) ermitteln.

Fall 1b-I: $VZ^{hj} \leq tg_{I.Bereich} + opt.VZ^{hj} + t_{sperr,j}$

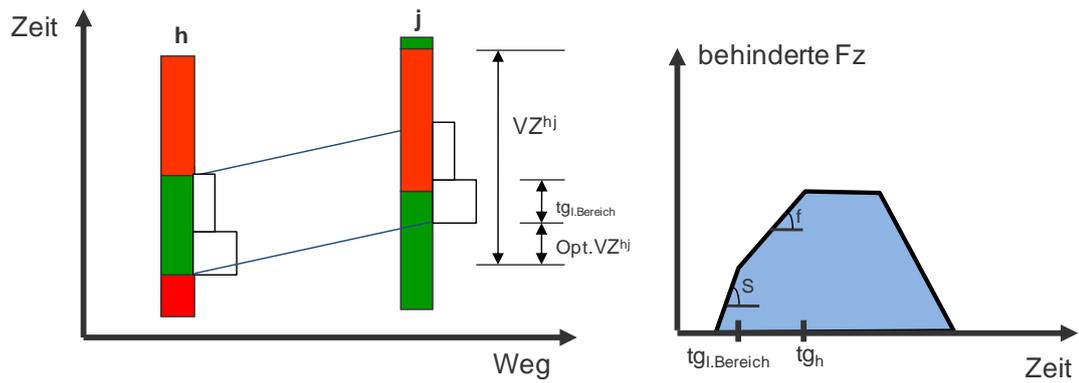


Abbildung A-4: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 1b-I

$$w = \frac{(t_{II.Bereich} + tg_{-d}) \cdot S \cdot t_{I-d} + f_{hj} \cdot t_{II.Bereich}^2 + 2(tg - tg_{-d}) \cdot Sum_Fz + Sum_Fz^2 / S}{2f_{hj} \cdot t_{Umlauf}} \quad (A-5)$$

mit

$$\begin{cases} tg_{II.Bereich} := tg - tg_{I.Bereich} \\ tg_{-d} := tg_h - VZ^{hj} + opt.VZ^{hj} + t_{sperr,j} \\ t_{I-d} := tg_{I.Bereich} - VZ^{hj} + opt.VZ^{hj} + t_{sperr,j} \\ (Sum_Fz := S \cdot t_{I-d} + f_{hj} \cdot tg_{II.Bereich}) \end{cases} \quad (A-6)$$

Fall 1b-II: $VZ^{hj} > tg_{I.Bereich} + opt.VZ^{hj} + t_{sperr,j}$

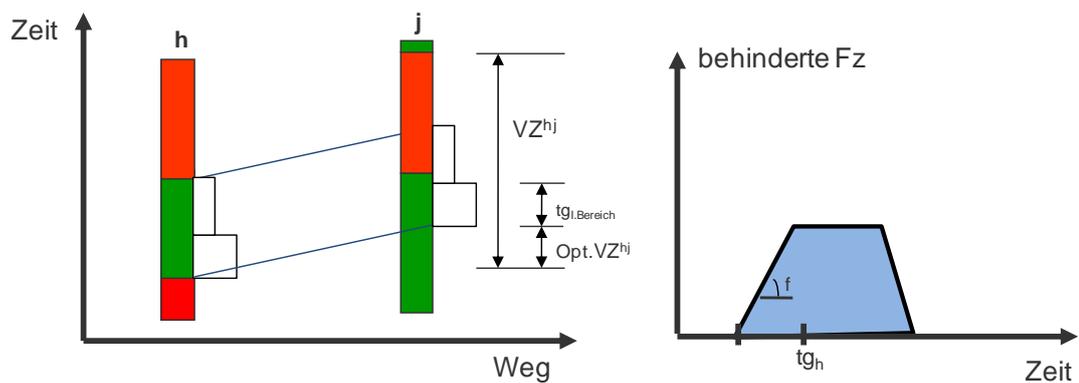


Abbildung A-5: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 1b-II

$$w = \frac{tg_{-d}^2 \cdot f + 2(tg - tg_{-d}) \cdot Sum_Fz + Sum_Fz^2 / S}{2f_{hj} \cdot t_{Umlauf}} \quad (A-7)$$

mit

$$\begin{cases} tg_d := tg_h - VZ^{hj} + opt.VZ^{hj} + t_{sperr,j} \\ Sum_Fz := f_{hj} \cdot tg_d \end{cases} \quad (A-8)$$

Fall 1c:

$$VZ^{hj} < tg_h + opt.VZ^{hj} \quad (A-9)$$

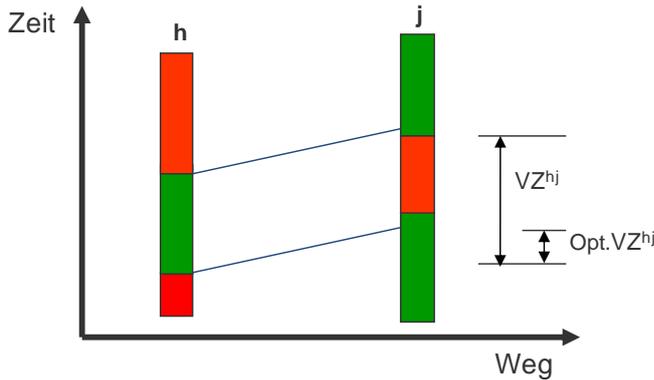


Abbildung A-6: Zeit-Weg-Diagramm des Falls 1c

In Hinsicht des Bereichsmodells (Abbildung 6-4) kann man die Wartezeitfunktionen (Gl. A-10 und A-12) des Falls1b in zwei Situationen (Abbildung A-7 und Abbildung A-8) ermitteln.

Fall 1c-I: $VZ^{hj} \leq tg_{I.Bereich} + opt.VZ^{hj} + t_{sperr,j}$

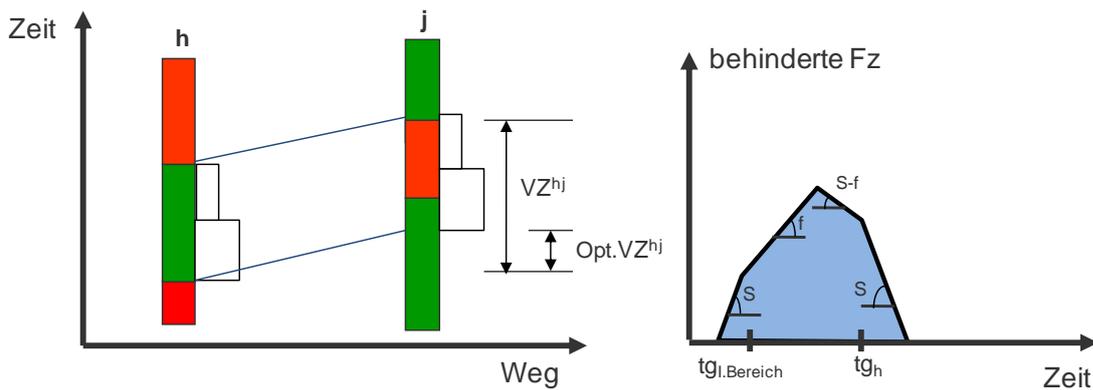


Abbildung A-7: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 1c-I

$$w = \frac{tg_d^2 \cdot S + (t_{sperr,j} - tg_d)^2 \cdot f + (2Sum_Fz - (S-f) \cdot tg_d)^2 \cdot tg_d + (Sum_Fz - (S-f) \cdot tg_d)^2 / S}{2f_{hj} \cdot t_{Umlauf}} \quad (A-10)$$

mit

$$\begin{cases} tg_d := tg_h - VZ^{hj} + opt.VZ^{hj} \\ t_l_d := tg_{l.Bereich} - VZ^{hj} + opt.VZ^{hj} + t_{sperr,j} \\ Sum_Fz := S \cdot t_l_d + f_{hj} \cdot (t_{sperr,j} - t_l_d) \end{cases} \quad (A-11)$$

Fall 1c-II: $VZ^{hj} > tg_{l.Bereich} + opt.VZ^{hj} + t_{sperr,j}$

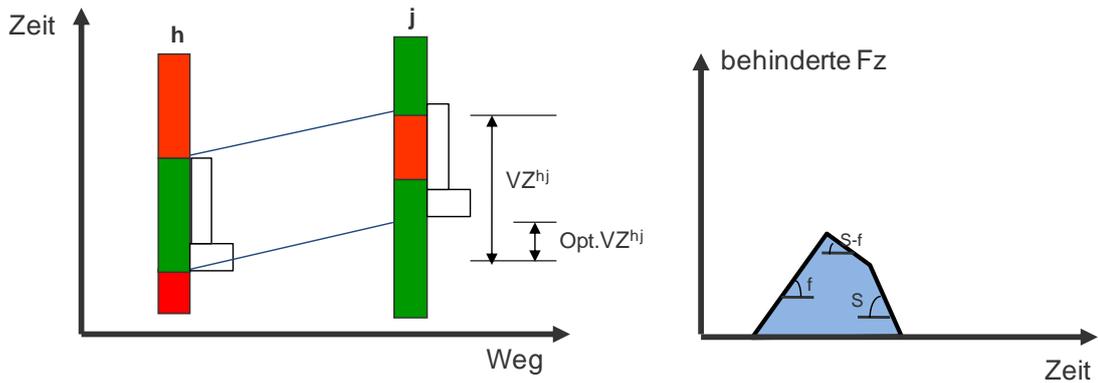


Abbildung A- 8: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 1c-II

$$w = \frac{t_{sperr,h}^2 \cdot f + (2Sum_Fz - (S-f) \cdot tg_d)^2 \cdot tg_d + (Sum_Fz - (S-f) \cdot tg_d)^2 / S}{2f_{hj} \cdot t_{Umlauf}} \quad (A-12)$$

mit

$$\begin{cases} tg_d := tg_h - VZ^{hj} + opt.VZ^{hj} \\ Sum_Fz := f_{hj} \cdot t_{sperr,j} \end{cases} \quad (A-13)$$

Gruppe 2:

Analog zur Gruppe 1 wird Gruppe 2 auch weithin in drei Fälle unterteilt. Die Bedingung aus Gl. A-1 ist die Grundlage für alle folgenden Fälle.

Fall 2a:

$$VZ^{hj} > tg_h + opt.VZ^{hj} \quad (A-14)$$

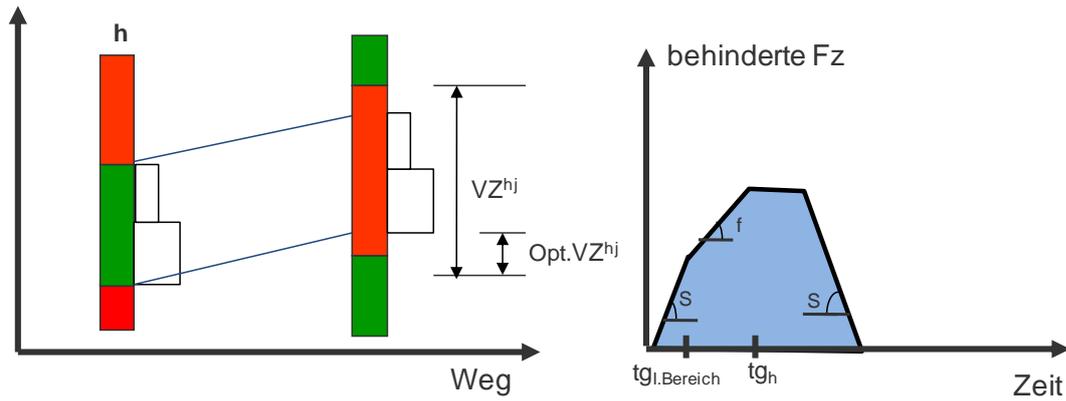


Abbildung A-9: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 2a

$$w = \frac{(tg_h + tg_{II.Bereich}) \cdot tg_{I.Bereich} \cdot S + tg_{II.Bereich}^2 \cdot f + Sum_Fz \cdot (2tg_d + Sum_Fz / S)}{2f_{hj} \cdot t_{Umlauf}} \quad (A-15)$$

mit

$$\begin{cases} tg_{II.Bereich} := tg_h - tg_{I.Bereich} \\ Sum_Fz := S \cdot tg_{I.Bereich} + f_{hj} \cdot tg_{II.Bereich} \\ tg_d = VZ^{hj} - opt.VZ^{hj} - tg_h \end{cases} \quad (A-16)$$

Fall 2b:

$$VZ^{hj} \leq tg_h + opt.VZ^{hj} \quad (A-17)$$

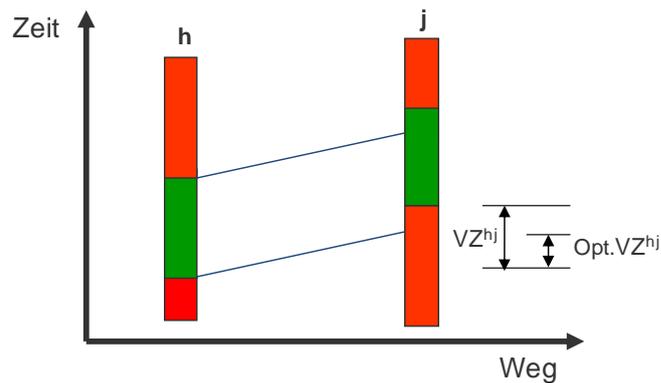


Abbildung A-10: Zeit-Weg-Diagramm des Falls 2b

In Hinsicht des Bereichsmodells kann man die Wartezeitfunktionen (Gl. A-18 und A-20) des Falls 2b in zwei Situationen (Abbildung A-11 und Abbildung A-12) ermitteln.

Fall 2b-I: $VZ^{hj} \leq tg_{I.Bereich} + opt.VZ^{hj}$

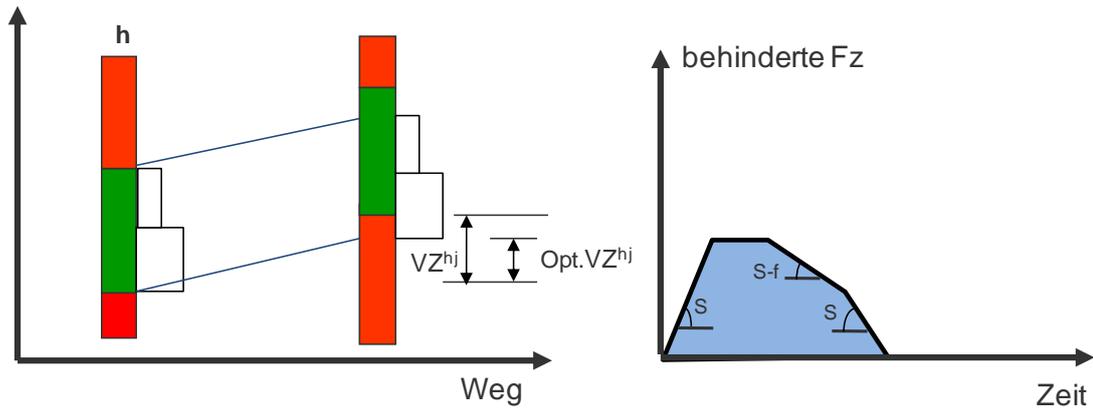


Abbildung A-11: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 2b-I

$$w = \frac{(2tg_{I.Berei.} - tg_{I_d}) \cdot Su - Fz + (2Su - Fz - tg_{II.Bere.} \cdot (s - f)) \cdot tg_{II.Bere.} + (Su - Fz - tg_{II.Bere.} \cdot (s - f))^2 / S}{2f_{hj} \cdot t_{Umlauf}} \quad (A-18)$$

mit

$$\begin{cases} tg_{II.Bereich} := tg_h - tg_{I.Bereich} \\ tg_{I_d} := tg_{I.Bereich} - VZ^{hj} + opt.VZ^{hj} \\ Su - Fz := S \cdot tg_{I.Bereich} \end{cases} \quad (A-19)$$

Fall 2b-II: $VZ^{hj} > tg_{I.Bereich} + opt.VZ^{hj}$

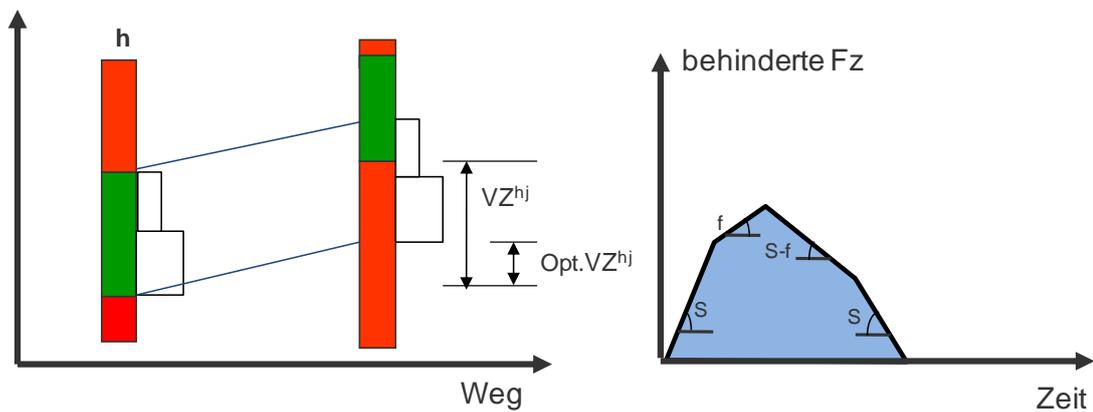


Abbildung A-12: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 2b-II

$$w = \frac{tg_{I.Berei.}^2 \cdot S + (tg_{I.Berei.} \cdot S + Su - Fz) \cdot tg_{I_d} + (2Su - Fz - (tg_{II.Bere.} - tg_{I_d}) \cdot (s - f)) \cdot (tg_{II.Bere.} - tg_{I_d})}{2f_{hj} \cdot t_{Umlauf}} + \frac{(Su - Fz - (tg_{II.Bere.} - tg_{I_d}) \cdot (s - f))^2 / S}{2f_{hj} \cdot t_{Umlauf}} \quad (A-20)$$

mit

$$\begin{cases} tg_{II.Bereich} := tg_h - tg_{I.Bereich} \\ tg_{I_d} := VZ^{hj} - opt.VZ^{hj} - tg_{I.Bereich} \\ Su_Fz := S \cdot tg_{I.Bereich} + tg_{I_d} \cdot f \end{cases} \quad (A-21)$$

Gruppe 3:

Gruppe 3 kann man auf Basis der Bedingung (Gl. A-1) in weitere zwei Fälle unterteilen.

Fall 3a:

$$VZ^{hj} > tg_j - tg_h - opt.VZ^{hj} \quad (A-22)$$

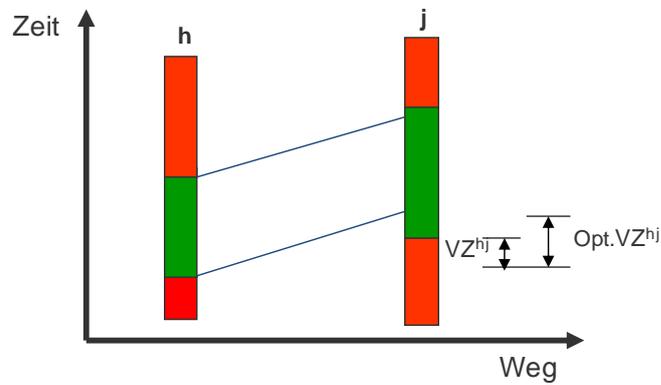


Abbildung A-13: Zeit-Weg-Diagramm zur Berechnung des Falls 3a

$$w = 0 \quad (A-23)$$

Fall 3b:

$$VZ^{hj} < tg_j - tg_h - opt.VZ^{hj} \quad (A-24)$$

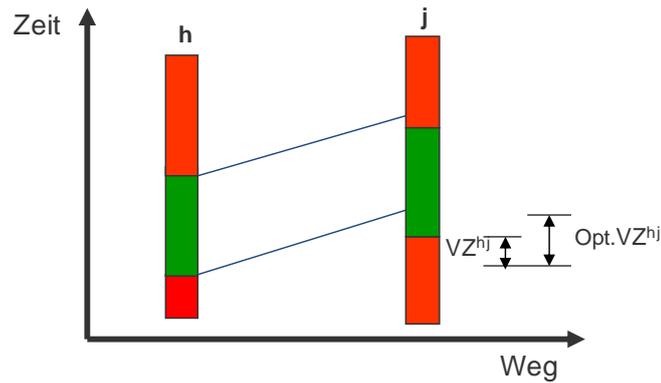


Abbildung A-14: Zeit-Weg-Diagramm des Falls 3b

In Hinsicht des Bereichsmodells kann man die Wartezeitfunktionen (Gl. A-25 und A-27) des Falls 3b in zwei Situationen (Abbildung A-15 und Abbildung A-16) ermitteln.

$$\text{Fall 3b-I: } VZ^{hj} \leq tg_{I.Bereich} + opt.VZ^{hj} - tg_j$$

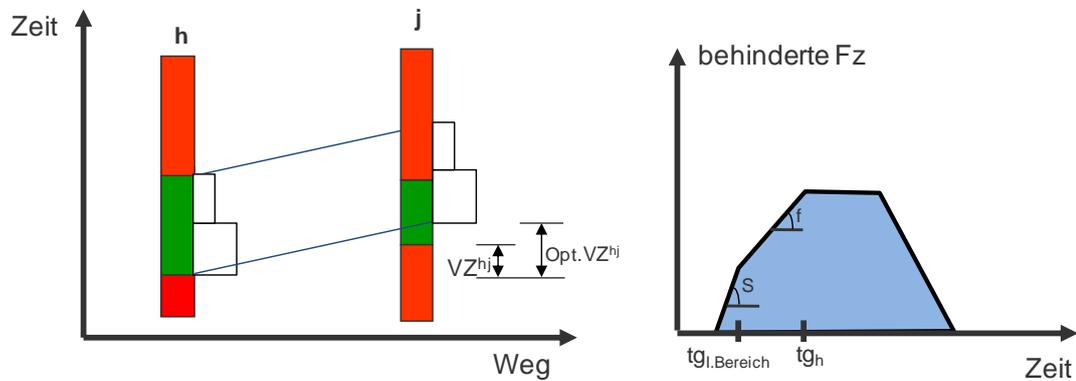


Abbildung A-15: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 3b-I

$$w = \frac{(t_{II.Bereich} + tg - d) \cdot S \cdot t_{I-d} + f_{hj} \cdot t_{II.Bereich}^2 + 2(tg - tg - d) \cdot Sum_Fz + Sum_Fz^2 / S}{2f_{hj} \cdot t_{Umlauf}} \quad (\text{A-25})$$

mit

$$\begin{cases} tg_{II.Bereich} := tg - tg_{I.Bereich} \\ tg - d := tg_h - VZ^{hj} + opt.VZ^{hj} - tg_j \\ t_{I-d} := tg_{I.Bereich} - VZ^{hj} + opt.VZ^{hj} - tg_j \\ Sum_Fz := S \cdot t_{I-d} + f_{hj} \cdot tg_{II.Bereich} \end{cases} \quad (\text{A-26})$$

Fall 3b-II: $VZ^{hj} > tg_{l.Bereich} + opt.VZ^{hj} - tg_j$

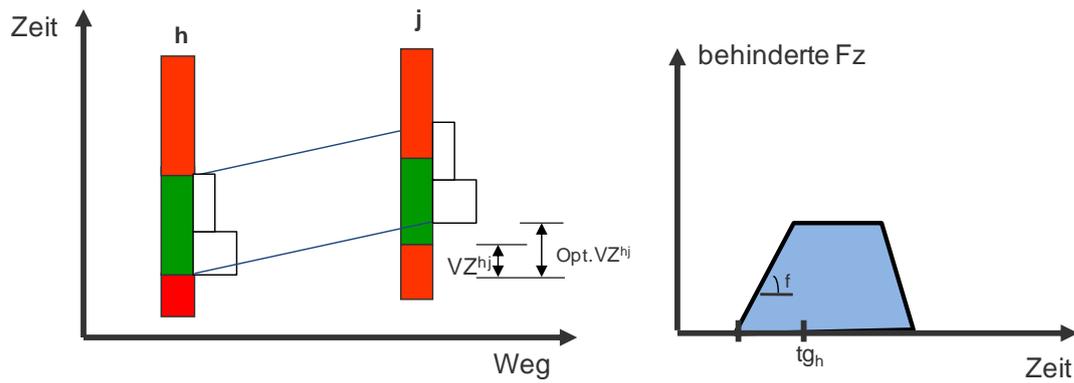


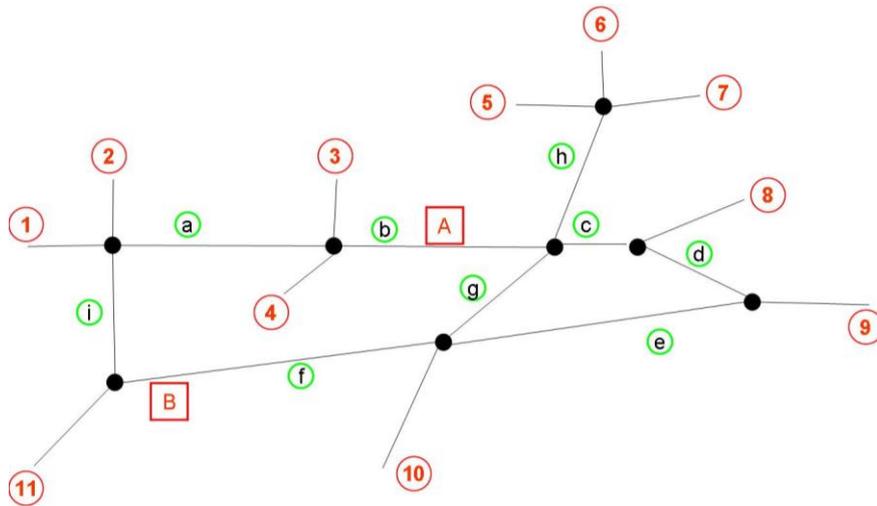
Abbildung A-16: Bestimmung der versatzzeitabhängigen Wartezeit bei Fall 3b-II

$$w = \frac{tg_d^2 \cdot f + 2(tg - tg_d) \cdot Sum_Fz + Sum_Fz^2 / S}{2f_{hj} \cdot t_{Umlauf}} \quad (A-27)$$

mit

$$\begin{cases} tg_d := tg_h - VZ^{hj} + opt.VZ^{hj} + tg_j \\ Sum_Fz := f_{hj} \cdot tg_d \end{cases} \quad (A-28)$$

Anhang B – Quell-Ziel-Matrix des virtuellen Straßennetzes



Nach	Von										
	1	2	3	4	5	6	7	8	9	10	11
1 über A									36	10	10
1 über B		20	20	20	12	12	12	40	6	5	
2 über A									4	3	25
2 über B	15		5	7	2	5	8	5	2	3	
3	2	1		30	0	0	0	2	2	1	2
4	2	1	30		2	2	2	2	2	1	2
5	1	1	0	0		10	20	0	2	2	0
6	6	1	0	6	20		30	8	12	32	12
7	1	1	0	4	25	10		0	2	2	0
8	7	5	10	3	3	3	0		4	5	6
9 über A	40	15	15	16	15	10	10	18		24	28
9 über B	6	7	0	0							
10 über A	10	8	10	7	16	36	12	10	10		15
10 über B	2	10	0	0							
11 über A	8	30	10	1	5	12	6	5	18	12	
11 über B				6				10			
	100	100	100	100	100	100	100	100	100	100	100

Abbildung A-17: Quell-Ziel-Matrix des virtuellen Straßennetzes

Anhang C – Anzahl der Simulationsläufe

Die Anzahl der Simulationsläufe kann nach folgender Gleichung geschätzt werden:

$$n \geq \frac{t(\alpha, n-1)^2 \cdot s^2}{e_r^2 \bar{x}^2} \quad (\text{A-29})$$

mit

- $t(\alpha, n-1)$ = Wert aus der Student-Verteilung
- s = Standardabweichung der untersuchten Kenngröße aus der Voruntersuchung
- e_r = relative Abweichung
- \bar{x} = Mittelwerte aus der Voruntersuchung.

Aus einer Voruntersuchung (Stichprobenumfang $n = 10$) für Szenario 2 während der Morgenspitze (07:00 Uhr bis 08:30 Uhr) sind folgende Ergebnisse bekannt:

Probe	Mittlere Verlustzeit (s)	Gesamte Reisezeit (h)	Route 1_Ost (s)	Route 1_West (s)
1	151,7	478,288	337,6	325,6
2	141,65	460,988	357,3	315,1
3	144	457,1	344,4	316,1
4	145,8	468,02	378,8	318,9
5	128,17	458,716	359,2	322,2
6	148,785	471,3	331,1	323,6
7	140	467,8	356	330,9
8	138	463	335,1	309,3
9	137,248	457,98	353,2	315,5
10	141,84	456,5	348,2	325,6
Mittelwert	141,719	463,96	350,09	320,28
Abweichung s	6,597	7,196	14,067	6,443

Als zweckmäßige Genauigkeitsanforderungen werden festgelegt:

$$\alpha = 0,95 \quad \text{und} \quad e_r = 3\% \quad (\text{A-30})$$

Gemäß Gl. A-29 ist

	Verlustzeit	Reisezeit	Route 1_Ost	Route 1_West
$n \geq$	8,1	0,9	6,0	1,5

Die erforderliche Anzahl der Simulationsläufe ist kleiner als zehn. Für die Simulationstests ist deshalb eine Anzahl von zehn Simulationsläufen ausreichend.