

**Universität Stuttgart**

Institut für Nachrichtenvermittlung und Datenverarbeitung

Prof. Dr.-Ing. P. Kühn

**51. Bericht über verkehrstheoretische Arbeiten**

**VERKEHRSFLÜSSE IN ATM-NETZEN –  
MODELLIERUNG UND ANALYSE  
VON VERKEHRSQUELLEN UND  
QUELLFLUSSKONTROLLVERFAHREN**

von

**Erwin P. Rathgeb**

1991

© 1992 Institut für Nachrichtenvermittlung und Datenverarbeitung der Universität Stuttgart

Druck: E. Kurz & Co., Druckerei + Reprografie GmbH., Stuttgart

ISBN 3-922403-61-1



**University of Stuttgart**

Institute of Communications Switching and Data Technics

Prof. Dr.-Ing. P. Kühn

**51 th Report on Studies in Congestion Theory**

**TRAFFIC STREAMS IN ATM NETWORKS —  
MODELLING AND ANALYSIS  
OF TRAFFIC SOURCES  
AND POLICING MECHANISMS**

by

**Erwin P. Rathgeb**

1991



## Summary

To meet the increasing demand for broadband communications in the public area, a Broadband Integrated Services Digital Network (B-ISDN) will be introduced in the near future. The CCITT has decided that this B-ISDN will be based on the Asynchronous Transfer Mode (ATM) as defined in a set of recommendations currently under study.

All information in an ATM network is transferred in small, fixed size packets (called cells) by using the virtual connection concept. To share the network resources among the coexisting connections, the asynchronous time division multiplexing principle is used. This statistical multiplexing scheme provides the flexibility which is needed for a network intended to integrate the wide variety of existing and future services. On the other hand, it allows the different traffic streams to interfere with each other far more than in traditional circuit switched networks. This interference leads to variable cell delays within a connection (delay jitter) and – in the extreme – to cell losses due to momentary overload at a shared resource within the network.

The key problem is therefore to control the traffic load imposed on the network in such a way that a satisfactory quality of service can be provided to the users while utilizing the network resources sufficiently to allow an economical network operation. To solve this problem, a comprehensive set of interrelated traffic management functions working on different time scales has to be defined.

The main goal of this work was to investigate one of these functions, the so called “Policing” or “Usage Parameter Control” function, in detail. This function, which only exists in ATM networks, is located at the entrance of the network and its purpose is to enforce that a traffic source adheres to some statistical parameters negotiated at call setup during the whole active phase of a connection.

A significant part of the work has been devoted to the choice of proper models to characterize the traffic sources, because the model assumptions have considerable influence on the results obtained. Therefore, the most commonly used traffic source models have been reviewed and assessed with respect to their applicability.

**Chapter 1** provides an introductory overview about the current situation in the field of wide area communication networks and the possible evolution towards the B-ISDN.

**Chapter 2** focuses on architectural and operational aspects of broadband networks based on the Asynchronous Transfer Mode to provide some background information for the subsequent discussions. First, some of the requirements which have to be satisfied by a B-ISDN are described and the various concepts that have been proposed for its realization are discussed briefly. Then, the synchronous and the asynchronous time division multiplexing principles are introduced and used to highlight the basic characteristics of the Synchronous (STM) and

the Asynchronous Transfer Mode (ATM).

Section 2.4 provides an overview of the B-ISDN architecture and protocols as defined by CCITT. This overview reflects the state of the international standardization as of August 1990. Section 2.5 reviews some generic concepts for the architecture of an ATM switch. The traffic and congestion control functions including Call Acceptance Control (CAC), Policing (Usage Parameter Control, UPC) and Priority Control and their interrelations are discussed in section 2.6.

**Chapter 3** concentrates on modelling of traffic streams in ATM networks. In a more general part, various modelling aspects are described and a generic, hierarchical model for ATM traffic sources is developed. This model is used to show the different time scales which are relevant for the description of the traffic sources and to relate some of the effects determining the typical behavior of the various service types to these time scales.

Then, two alternative methods to characterize traffic streams are discussed. The first, commonly used method describes the cell stream by using the moments or – more general – probability distribution functions of the cell interarrival times. The other method uses the counting process, i.e. the number of arrivals in a given time interval, and derived measures like the Index of Dispersion for Counts (IDC). Explicit recursive convolutional algorithms to obtain the distribution of the counting process for both continuous time and discrete time renewal processes are derived. These algorithms establish the basis for some of the analytical evaluations in chapter 5.

Section 3.3 provides a survey of traffic source models commonly used in ATM traffic studies. It covers, in addition to standard continuous and discrete time distributions, also the “sporadic” (On/Off, Burst/Silence) source, the Markov Modulated Poisson Process (MMPP), the Generally Modulated Deterministic Process (GMDP) and an autoregressive Markovian model for packet video. For some of the models, including the sporadic source model, explicit formulae are given for the distribution of the counting process.

Section 3.4 shows some examples for the calculation of the counting process distribution and of the IDC. These examples have been used to assess the numerical accuracy of the different algorithms.

**Chapter 4** deals with performance models and analysis techniques applicable to ATM traffic studies. Section 4.1 gives a brief overview of the simulation tool which has been implemented in the course of this work and discusses the limitations of traditional simulation methods with respect to ATM studies.

The main part of this chapter is devoted to a survey of the single server delay/loss systems with deterministic service time which arise as models for system components like statistical multiplexers and also as models for some policing mechanisms. For the simpler models,

explicit solutions are given for the loss probabilities while for the more complex models only the basic solution methods are described. Special emphasis has been put on discrete time models, for which effects like the simultaneity of arrival events and service completion events are studied in detail. Explicit algorithms for the numerical solution of the respective systems are also presented.

Section 4.3 contains quantitative comparisons of different models which lead to the conclusion that the sporadic source model is well suited for the subsequent studies.

**Chapter 5** describes the comparative performance study of several policing mechanisms that has been carried out. Following a discussion of the requirements and quality measures which have been defined for the mechanisms implementing the policing function, the assumptions for the comparison are listed in section 5.1.

Section 5.2 provides a description of the five mechanisms included in the comparison, namely the Leaky Bucket, the Jumping Window, the Triggered Jumping Window, the Moving Window and the Exponentially Weighted Moving Average (EWMA) mechanisms. The resulting performance models which have been used to assess these mechanisms are also mentioned. For the two Jumping Window mechanisms, explicit formulae are derived which allow to calculate the probability that a violation of the controlled parameters is indicated.

The qualitative influence of the parameters of the mechanisms on the probability for a violation indication is summarized in section 5.3. Also, a new dimensioning approach is described which provides an additional degree of freedom by introducing the overdimensioning factor  $C$ .

Section 5.4 gives the qualitative results of the comparison with respect to the following criteria:

- Exact control of the mean bit rate,
- overdimensioning of the mechanisms,
- dynamic behavior of the mechanisms,
- influence of the traffic source model on the results and
- worst case traffic accepted by the mechanisms.

The comparison shows that the Leaky Bucket and the EWMA mechanisms are more promising than the other three window mechanisms. However, none of the mechanisms allows to control the mean bit rates with the accuracy that would be required if the long term mean bit rate would be used as key traffic control parameter.

Based on these results, two new and improved mechanisms are suggested in section 5.5. These mechanisms, namely a Leaky Bucket mechanism with variable counter limit and a

mechanism with adaptive counter, are compared with the Leaky Bucket mechanism to show their superior performance.

Chapter 6 summarizes the results of the study and emphasizes the basic problems of the policing concept. One of these problems is that a policing mechanism has to estimate long term statistical characteristics with very high accuracy based on relatively short samples of the cell arrival process. Such conflicting requirements lead to a clearly non-optimal behavior of the mechanisms. When dimensioning the mechanisms, a compromise involving tradeoffs between

- the probability for unjustified discarding of cells from connections adhering to the parameters negotiated at call setup,
- the amount of real overload that is not detected,
- the dynamic reaction time of the mechanisms,
- the accepted worst case traffic and
- the implementation complexity

has to be found taking into consideration the requirements and characteristics of the different services.

Another problem is that for most services the relevant parameters like the mean bit rate (and even the peak bit rate) will not be known with sufficient accuracy at call setup. Therefore, effective policing would be impossible for those services even if nearly optimal policing mechanisms were available.

The main conclusion of this discussion is that the policing concept as such has severe limitations, and that there is only limited benefit from increasing the complexity of the mechanisms implementing the policing function. Therefore, additional service specific mechanisms are necessary to improve the effectiveness of the policing concept and consequently of the entire traffic management. For services with less severe real time constraints, mechanisms like traffic shaping or additional protocols to control the source characteristics may be applied successfully. For interactive services, adaptive coding schemes restricting the sources appropriately while degrading the quality only very gracefully are a possible solution.

With measures like these, it should be possible to design, dimension and operate networks based on ATM such that they are economically competitive and still offer the extended flexibility which is required for a future save and universal broadband network.



# Inhaltsverzeichnis

<b>Abkürzungen und Formelzeichen</b>	<b>4</b>
<b>1 Einleitung</b>	<b>10</b>
1.1 Evolution zum integrierten Breitband-Kommunikationsnetz . . . . .	10
1.2 Übersicht über die Arbeit . . . . .	13
<b>2 Architektur und Betrieb von Breitbandnetzen auf der Basis von ATM</b>	<b>15</b>
2.1 Anforderungen an ein integriertes Breitband-Kommunikationsnetz . . . . .	15
2.2 Prinzipien für ein integriertes Breitband-Kommunikationsnetz . . . . .	16
2.3 Zeitmultiplexverfahren und Verbindungskonzepte . . . . .	17
2.3.1 Synchrones Zeitmultiplex . . . . .	17
2.3.2 Asynchrones Zeitmultiplex . . . . .	20
2.4 Konzepte für das Breitband-ISDN . . . . .	23
2.4.1 Dienste und Dienstkonzept . . . . .	23
2.4.2 Prinzipielle Eigenschaften von ATM . . . . .	26
2.4.3 Basis-Architekturmodell . . . . .	26
2.4.4 Bezugsconfiguration . . . . .	27
2.4.5 Protokollmodell . . . . .	29
2.4.6 Zellformat . . . . .	34
2.4.6.1 Format des Informationsfeldes . . . . .	34
2.4.6.2 Format und Funktionen des Kopffeldes . . . . .	35
2.4.7 Benutzer/Netzschnittstelle . . . . .	37
2.4.8 Signalisierung . . . . .	38
2.5 Prinzipielle Architektur einer ATM-Vermittlungsstelle . . . . .	39
2.6 Betrieb von ATM-Breitbandnetzen . . . . .	43
2.6.1 Verbindungsannahme und Wegesuche . . . . .	43
2.6.2 Quellfußkontrolle während der Verbindung (Policing) . . . . .	45
2.6.3 Dienstgüteklassen und Prioritäten . . . . .	47
2.6.4 Zusammenhang zwischen den Verkehrssteuerungsmechanismen . . . . .	48

<b>3</b>	<b>Modellierung von Verkehrsströmen in ATM-Netzen</b>	<b>50</b>
3.1	Modellierungsaspekte	50
3.2	Charakterisierung von Verkehrsströmen	53
3.2.1	Zwischenankunftsabstände	53
3.2.2	Zählprozeß	55
3.2.2.1	Verteilung	55
3.2.2.2	Index of Dispersion for Counts	59
3.2.3	Autokovarianz	60
3.3	Verwendete Modelle für Verkehrsströme	60
3.3.1	Erneuerungsprozesse	60
3.3.1.1	Zeitkontinuierliche Erneuerungsprozesse	61
3.3.1.1.1	Negativ-exponentielle Verteilung (M)	61
3.3.1.1.2	Deterministische Verteilung (D)	62
3.3.1.1.3	Erlang-k-Verteilung ( $E_k$ )	62
3.3.1.1.4	Hyperexponentielle Verteilung 2. Ordnung ( $H_2$ )	63
3.3.1.2	Zeitdiskrete Erneuerungsprozesse	64
3.3.1.2.1	Bernoulli-Verteilung (BER)	65
3.3.1.2.2	Verschobene Geometrische Verteilung ( $GEO_m$ )	65
3.3.1.2.3	Binomial-Verteilung (BIN)	67
3.3.1.2.4	Negativ-Binomial-Verteilung (NEGBIN)	67
3.3.1.2.5	Poisson-Verteilung (POIS)	68
3.3.2	Modell für eine sporadische Quelle	68
3.3.3	Markoff-modulierter Poisson-Prozeß (MMPP)	71
3.3.4	Allgemein modulierter deterministischer Prozeß (GMDP)	72
3.3.5	Autoregressives Modell für paketierte Videoübertragung	73
3.3.6	Andere Modelle für ATM-Verkehrsströme	75
3.4	Beispiele für die Berechnung des IDC	76
<b>4</b>	<b>Systemmodelle und Analysemethoden</b>	<b>82</b>
4.1	Simulation	82
4.1.1	Prinzip der zeittreuen Simulation	82
4.1.2	Simulationsbibliothek	83
4.1.3	Anwendbarkeit simulativer Methoden	87
4.2	Analytische Modelle und Methoden	89
4.2.1	Systeme mit zeitkontinuierlichem Ankunftsprozeß	90
4.2.1.1	M/D/1-s	90
4.2.1.2	M/D/n-0	92
4.2.1.3	Systeme mit allgemeinen Markoffschen Ankunftsprozessen	92
4.2.1.4	UAS-Modell	93

4.2.2	Systeme mit zeitdiskretem Ankunftsprozeß . . . . .	94
4.2.2.1	Gleichzeitigkeit von Ankunft und Bedienende . . . . .	94
4.2.2.2	$GEO_1/D/1-s$ . . . . .	95
4.2.2.3	$GI^{[X]}/D/1-s$ . . . . .	97
4.2.2.4	$DMAP/D/1-s$ . . . . .	104
4.2.3	Andere Modelle . . . . .	105
4.3	Beispiele für die Berechnung der Verlustwahrscheinlichkeit . . . . .	105
<b>5</b>	<b>Leistungsuntersuchung von Quellflußkontrollmechanismen</b> . . . . .	<b>111</b>
5.1	Anforderungen und Voraussetzungen . . . . .	111
5.2	Beschreibung und Modellierung der Mechanismen . . . . .	114
5.2.1	Leaky Bucket . . . . .	114
5.2.2	Jumping Window . . . . .	115
5.2.3	Triggered Jumping Window . . . . .	117
5.2.4	Moving Window . . . . .	119
5.2.5	Exponentially Weighted Moving Average . . . . .	120
5.3	Dimensionierung der Mechanismen . . . . .	122
5.4	Untersuchung und Vergleich der Verfahren . . . . .	124
5.4.1	Exakte Überwachung der mittleren Zellrate . . . . .	124
5.4.2	Überdimensionierung der Verfahren . . . . .	125
5.4.3	Dynamisches Verhalten der Verfahren . . . . .	128
5.4.4	Einfluß der Quellcharakteristik . . . . .	130
5.4.5	Im ungünstigsten Fall zugelassener Verkehr . . . . .	132
5.5	Vorschläge für verbesserte Mechanismen . . . . .	134
5.5.1	Adaptive Counter . . . . .	134
5.5.2	Leaky Bucket mit variabler Zählergrenze . . . . .	137
5.5.3	Vergleich mit dem LB-Verfahren . . . . .	138
<b>6</b>	<b>Zusammenfassung und Ausblick</b> . . . . .	<b>141</b>
	<b>Literaturverzeichnis</b> . . . . .	<b>146</b>

## Abkürzungen

AAL	ATM Adaptation Layer
AC	Adaptive Counter
ADPCM	Adaptive Differential Pulse Code Modulation
AF	Arrival First
ANSI	American National Standards Institute
ATD	Asynchronous Time Division Multiplexing
ATM	Asynchronous Transfer Mode
B-ISDN	Broadband Integrated Services Digital Network
BER	Bernoulli-Verteilung
BIN	Binomial-Verteilung
BP	Blocking Policy
BVN	Breitband-Vorläufernetz
CBR	Constant Bitrate
CAC	Call Acceptance Control
CCITT	Comité Consultatif International Téléphonique et Télégraphique
CLP	Cell Loss Priority
CNET	Centre National d'Etudes des Télécommunications
CPN	Customer Premises Network
CS	Circuit Switching (Durchschaltvermittlung) oder
CS	Convergence Sublayer
D	Deterministische Verteilung
DCT	Discrete Cosine Transform
DF	Departure First
DMAP	Discrete Markovian Arrival Process
DPCM	Differential Pulse Code Modulation
DQDB	Distributed Queue Dual Bus
EF	Wahrscheinlichkeitserzeugende Funktion
$E_k$	Erlang-k-Verteilung
ET	Exchange Termination
EWMA	Exponentially Weighted Moving Average
FIFO	First In First Out
FCS	Fast Circuit Switching
FFT	Fast Fourier Transform
FPS	Fast Packet Switching
G	General (allgemeiner Zufallsprozeß)
GASTA	Geometric Arrivals See Time Averages
$GEO_m$	Um m Elemente verschobene geometrische Verteilung

GFC	Generic Flow Control
GI	General Independent (allgemeiner unabhängiger Zufallsprozeß)
GMDP	Generally Modulated Deterministic Process
H <sub>2</sub>	Hyperexponentielle Verteilung zweiter Ordnung
HDTV	High Definition Television
HEC	Header Error Control
IBCN	Integrated Broadband Communication Network
IDC	Index of Dispersion for Counts
IEEE	Institute of Electrical and Electronics Engineers
IPP	Interrupted Poisson Process
ISDN	Integrated Services Digital Network
ISO	International Organization for Standardization
ITG	Informationstechnische Gesellschaft
JW	Jumping Window
LAN	Local Area Network
LB	Leaky Bucket
LB-VL	Leaky Bucket - Variable Limit
LCN	Logical Channel Number
LEX	Local Exchange
LFC	Local Functional Capabilities
LLC	Logical Link Control
LST	Laplace-Stieltjes Transformation
LT	Laplace Transformation
M	Markoffsche (negativ-exponentielle) Verteilung
MAN	Metropolitan Area Network
MAP	Markovian Arrival Process
MMPP	Markov Modulated Poisson Process
MRCS	Multi Rate Circuit Switching (Mehrkanal-Durchschaltevermittlung)
MW	Moving Window
N	Neuts-Prozeß
NA	Network Adaptor
NEGBIN	Negativ-Binomial-Verteilung
NNI	Network Node Interface
NT	Network Termination
OAM	Operation and Maintenance
OSI	Open System Interconnection
PBX	Private Branch Exchange (Nebenstellenanlage)
PCM	Pulse Code Modulation
PM	Physical Medium Sublayer

POIS	Poisson-Verteilung
PS	Packet Switching (Paketvermittlung)
PTI	Payload Type Identifier
QOS	Quality of Service
RACE	Research and Development in Advanced Communications in Europe
SAR	Segmentation and Reassembly Sublayer
SDH	Synchronous Digital Hierarchy
SMDS	Switched Multi-Megabit Data Service
SONET	Synchronous Optical Network
SPP	Switched Poisson Process
STD	Synchronous Time Division Multiplexing
STM	Synchronous Transfer Mode
STS	Synchronous Transport Signal
SVC	Signalling Virtual Channel
SVCI	Signalling Virtual Channel Identifier
TA	Terminal Adaptor
TASI	Time Assignment Speech Interpolation
TC	Transmission Convergence Sublayer
TE	Terminal Equipment
TEI	Terminal Endpoint Identifier
TJW	Triggered Jumping Window
UAS	Uniform Arrival and Service
UNI	User-Network Interface
VBR	Variable Bitrate
VC	Virtual Channel
VCI	Virtual Channel Identifier
VP	Virtual Path
VPI	Virtual Path Identifier
WPT	Wideband Packet Technology
ZT	Z-Transformation
ZV	Zufallsvariable

## Formelzeichen

Die typische Verwendung einiger Formelzeichen soll am Beispiel des Ankunftsabstandes dargestellt werden:

$t$	Variable für die Zeit, zeitkontinuierlich
$k$	Variable für die Zeit, zeitdiskret, normiert auf die Basiszeiteinheit
$A$	Zufallsvariable (ZV) für den Ankunftsabstand
$A(t)$	Verteilungsfunktion, zeitkontinuierlich
$A(k)$	Verteilungsfunktion, zeitdiskret
$a(t)$	Verteilungsdichtefunktion
$a_k$	Verteilung
$a_{LST}(s)$	Laplace-Stieltjes Transformierte
$a_{EF}(z)$	Wahrscheinlichkeitserzeugende Funktion
$a_{ZT}(z)$	Z-Transformierte
$E[A]$	Mittelwert oder erstes gewöhnliches Moment
$E[A^2]$	zweites gewöhnliches Moment
$VAR[A]$	Varianz
$c_a$	Variationskoeffizient
$\lambda = 1/E[A]$	Ankunftsrate

Weitere Formelzeichen:

$[R]$	Kleinste ganze Zahl größer oder gleich $R$
$\lceil R \rceil$	Größte ganze Zahl kleiner oder gleich $R$
$s(t - t_0) = \begin{cases} 0 & \text{für } t < t_0 \\ 1 & \text{für } t \geq t_0 \end{cases}$	Sprungfunktion
$\delta(t - t_0) = 0$ für $t \neq t_0$	Dirac-Funktion
$f_{LT}(s)$	Laplace-Transformierte
$\mathcal{LT}\{\dots\}$	Laplace-Transformation
$\mathcal{LT}^{-1}\{\dots\}$	Inverse Laplace-Transformation
$\mathcal{ZT}\{\dots\}$	Z-Transformation
$\mathcal{ZT}^{-1}\{\dots\}$	Inverse Z-Transformation
$A_{OL}$	Verkehrsangebot (engl. Offered Load)
$A^*$	ZV für die Vorwärtsrekurrenzzeit
$A_n$	ZV für den Ankunftsabstand zwischen der $n$ -ten und der $(n + 1)$ -ten Gruppe

$B$	Burstiness-Faktor
$C$	Überdimensionierungsfaktor bei Quellflußkontrollverfahren
$D$	Deterministische Bedienzeit
$D_A$	Dekrementierintervall in der Aktivitätsphase beim AC-Verfahren
$D_P$	Dekrementierintervall in der Pausenphase beim AC-Verfahren
$d_i$	Zellabstand in der Phase $i$ beim GMDP
$E[X]$	Mittlere Anzahl von Zellen pro Aktivitätsphase bei der sporadischen Quelle
$E[X_i]$	Mittlere Anzahl von Zellen in der Phase $i$ beim GMDP
$E[S]$	Mittlere Pausendauer bei der sporadischen Quelle und beim GMDP
$F_A$	Faktor für das Dekrementierintervall in der Aktivitätsphase beim AC-Verfahren
$F_P$	Faktor für das Dekrementierintervall in der Pausenphase beim AC-Verfahren
$F^{(i)}$	ZV für die Zeitdauer bis zum $i$ -ten Ereignis
$IDC(t)$	Index of Dispersion for Counts (IDC)
$I(U_N)$	Anzahl freier Warteplätze in Abhängigkeit von der Restarbeit im System
$k_i$	Konstanten beim UAS-Modell
$N$	Zählergrenze bei Quellflußkontrollverfahren
$N_i$	Zählergrenze im $i$ -ten Intervall beim EWMA- und beim LB-VL-Verfahren
$n_i$	Wahrscheinlichkeit für $i$ Ankünfte während einer Bedienzeit
$N_t$	ZV für die Anzahl der Ankünfte in einem Intervall der Länge $t$ , zeitkontinuierlich
$N_k$	ZV für die Anzahl der Ankünfte in einem Intervall der Länge $k$ , zeitdiskret
$n_i(t)$	Wahrscheinlichkeit für $i$ Ankünfte in einem Intervall der Länge $t$ , zeitkontinuierlich
$n_i(k)$	Wahrscheinlichkeit für $i$ Ankünfte in einem Intervall der Länge $k$ , zeitdiskret
$p_i$	Zeitgemittelte Zustandswahrscheinlichkeiten
$p_{ij}$	Übergangswahrscheinlichkeiten für den GMDP
$p_v$	Verlustwahrscheinlichkeit
$r_i$	Rate für das Verlassen des Zustands $i$ beim MMPP



$s$	Anzahl der Warteplätze (ohne Bedieneinheit)
$T$	Intervalldauer bei Quellflußkontrollverfahren auf der Basis von Zeitfenstern
$T_B$	Ausgleichsintervall beim AC-Verfahren
$U_n$	ZV für die Restarbeit unmittelbar vor der Ankunft der $n$ -ten Gruppe
$U_n^+$	ZV für die Restarbeit unmittelbar nach der Ankunft der $n$ -ten Gruppe
$X_A$	Maximale Anzahl von schnellen Dekrementierungen pro Intervall beim AC-Verfahren
$X_{max}$	Anzahl von Zellen pro Aktivitätsphase des „Worst Case“-Verkehrs beim LB-Verfahren
$X_n$	ZV für die Anzahl von Zellen in der $n$ -ten Gruppe
$X_P$	Maximale Anzahl von langsamen Dekrementierungen pro Intervall beim AC-Verfahren
$\alpha$	Rate für das Verlassen des aktiven Zustands bei der sporadischen Quelle
$\beta$	Rate für das Verlassen des Pausenzustands bei der sporadischen Quelle
$\gamma$	Dämpfungsfaktor beim EWMA- und beim LB-VL-Verfahren
$\Delta$	Minimaler Zellabstand bei der sporadischen Quelle
$\lambda_i$	Ankunftsrate im Zustand $i$ beim MMPP
$\lambda_p$	Kontrollierte mittlere Zellrate bei den Quellflußkontrollverfahren
$\pi_i$	Zustandswahrscheinlichkeiten zum Bedienende

# Kapitel 1

## Einleitung

### 1.1 Evolution zum integrierten Breitband-Kommunikationsnetz

Die heutige Situation im Bereich der öffentlichen Kommunikationsnetze ist gekennzeichnet durch vollkommen getrennte Netze für Sprach-, Daten-, Text- und, in eingeschränktem Umfang, Bewegtbildübermittlung. Dabei kommen digitale und – vor allem im Teilnehmeranschlußbereich des Telefonnetzes – analoge Techniken zum Einsatz. Die hauptsächlich verwendeten Vermittlungsprinzipien sind die Durchschaltvermittlung auf der Basis der synchronen Zeitmultiplextechnik sowie die Paketvermittlung auf der Basis der asynchronen Paketmultiplextechnik für die Datenkommunikation. Das öffentliche Telefonnetz wird zur Zeit digitalisiert, da die Komponenten für die digitale Übertragung und Vermittlung wie z.B. integrierte Schaltungen, Mikrocomputer und optische Komponenten in den letzten Jahren eine stürmische Entwicklung hin zu größerer Leistungsfähigkeit erlebt haben und inzwischen den analogen Komponenten in Wirtschaftlichkeit und Leistung weit überlegen sind. Ein weiterer Vorteil der entstehenden vollen digitalen Netze ist die einheitliche Darstellung sämtlicher Informationen in digitaler Form, wodurch eine Integration der getrennten und oft unwirtschaftlichen dienstspezifischen Netze in ein universelles Kommunikationsnetz ermöglicht wird.

Ein solches Netz, das diensteintegrierende Digitalnetz (engl. Integrated Services Digital Network, ISDN), ist zur Zeit in vielen Ländern in der Einführungsphase und wird in absehbarer Zeit flächendeckend zur Verfügung stehen. Es bietet dem Benutzer in der vom CCITT standardisierten Form [24] eine für alle Dienste einheitliche Benutzer/Netzschnittstelle und beim Basisanschluß zwei 64 kbit/s-Nutzkanäle (B-Kanäle) und einen getrennten 16 kbit/s-Kanal (D-Kanal), auf dem vorwiegend Signalisier Nachrichten ausgetauscht werden. Komplexe Signalisierprotokolle erlauben neben der Inanspruchnahme der Grundfunktionen wie Verbindungsaufbau und -abbau auch die Kontrolle zusätzlicher neuer Dienstmerkmale wie

z.B. Anrufumleitung, Konferenzschaltung, Gebührenübernahme, Dienstwechsel während der Verbindung usw. Die Konzeption des ISDN basiert auf der Durchschaltvermittlung, es werden jedoch auch paketvermittelte Dienste sowohl für die B-Kanäle als auch für den D-Kanal angeboten werden. Bei Bedarf stehen im „Schmalband“-ISDN höherratige Schnittstellen bis 2 Mbit/s zur Verfügung.

Die einheitliche „Kommunikationssteckdose“ des Basisanschlusses erlaubt die Benutzung verschiedenster Teledienste, bei denen die Funktionen und Protokolle unter Einbeziehung der Kommunikationsfunktionen der Endgeräte für alle 7 Schichten des OSI-Referenzmodells [83] komplett standardisiert sind. Bisher festgelegte Teledienste sind Telephonie, Teletex, Telefax 4, Mixed Mode (Dokumente mit einer Kombination aus Text und Grafik), Bildschirmtext (Videotex) und Telex. Die Definition weiterer Teledienste für dialogorientierte, speicherorientierte und Verteil-Kommunikation wird erwartet. Neben den Telediensten sieht das Netz auch einfache Übermittlungsdienste zur Informationsübermittlung zwischen den Benutzer/Netzschnittstellen vor. Diese Übermittlungsdienste umfassen nur die Funktionalität der unteren drei Schichten des OSI-Referenzmodells und wurden für Durchschalte- und Paketvermittlung getrennt in verschiedene Kategorien eingeteilt. Durch die Liberalisierung der Vorschriften – vor allem im Bereich der Endgeräte – und die Entwicklung immer komplexerer Endsysteme wird sich auf der Basis dieser Übermittlungsdienste ein breites Spektrum verschiedener Anwendungen entwickeln.

Zur Deckung des Bedarfs an breitbandigen Verbindungen z.B. für interaktive Bewegtbildkommunikation (Videokonferenz) und sehr schnelle Datenübertragung wurde zunächst für den Zeitraum ab 1995 die Einführung eines Breitband-ISDN (B-ISDN) auf durchschaltvermittelter Basis geplant. Dieses sollte neben den im ISDN bereits vorhandenen Kanälen noch Breitbandkanäle mit 45 Mbit/s und mit 140 Mbit/s zur Verfügung stellen. In der internationalen Normung konnte sich dieses Konzept jedoch nicht durchsetzen. Zum einen herrscht eine große Unsicherheit, welche neuen Breitbanddienste sich entwickeln werden und wie groß z.B. ihre Bandbreitanforderungen und ihre Verkehrsanteile sein werden. Eine Struktur mit einer relativ groben Kanalhierarchie, die noch dazu ohne Kenntnis der genauen Anforderungen festgelegt werden müsste, erschien deshalb für ein universelles, zukunftssicheres Kommunikationsnetz zu unflexibel. Zum anderen müssten für die extrem verschiedenen Kanalraten getrennte Koppelnetzstrukturen zur Verfügung gestellt werden, da eine gemeinsame Durchschaltung sehr schwierig realisierbar ist. Dies schränkt die Flexibilität eines solchen Konzeptes weiter ein.

Statt dessen hat sich das CCITT frühzeitig auf ein B-ISDN auf der Basis des Asynchronen Transfer-Modus (ATM) festgelegt. Dieses Konzept baut auf einer paketorientierten Informationsübermittlung unter Verwendung der asynchronen Zeitmultiplextechnik auf. Es bietet eine weitestgehende Flexibilität bezüglich der Wahl der Spitzenbitraten einzelner Verbindungen und unterstützt auch zeitlich veränderliche Bandbreitanforderungen während einer Ver-

bindung, wie sie z.B. bei hochentwickelten Codierern für die Bewegtbildkommunikation auftreten. Außerdem erlaubt dieser Transfer-Modus die Vermittlung unterschiedlichster Bitraten in einheitlichen Koppelnetzen.

Ein zukünftiges integriertes Breitband-Kommunikationsnetz auf der Basis des Asynchronen Transfer-Modus, wie er in der CCITT Empfehlung I.121 [26] und in weiteren Dokumenten der CCITT Studiengruppe XVIII definiert ist, läßt sich wie folgt charakterisieren:

- Es werden universell nutzbare, genormte Benutzer/Netzschnittstellen im Bereich von 150 bzw. 600 Mbit/s zur Verfügung stehen.
- Es wird ein weiter Bereich von Audio-, Video- und Datenanwendungen unterstützt. Dazu werden Dienste definiert, die festgeschaltete und vermittelte Verbindungen im durchschaltvermittelten und im paketorientierten Modus in verschiedenen Konfigurationen (Punkt-zu-Punkt, Mehrpunkt) verwenden.
- Sämtliche Daten – einschließlich der Nutzdaten der verschiedenen Dienste, der Signalmeldungen und der Meldungen für Betrieb und Wartung des Netzes – werden in paketierter Form übertragen und vermittelt.
- Die übermittelten Blöcke haben eine einheitliche, feste Länge und bestehen aus einem Kopffeld und aus einem Informationsfeld, wobei für das Kopffeld eine Länge von 5 Oktett und für das Informationsfeld eine Länge von 48 Oktett vorgesehen ist. Die Blöcke werden als „Zellen“ bezeichnet.
- Die Informationsübermittlung erfolgt verbindungsorientiert unter Verwendung des Konzepts der virtuellen Verbindungen. Dadurch nehmen alle Zellen einer Verbindung denselben Weg durch das Netz. In den zukünftigen Normen sind auch Möglichkeiten für eine Unterstützung der verbindungslosen Kommunikation vorgesehen.
- Auf der Zellübermittlungsschicht (ATM-Schicht) wird vom Netz die korrekte Reihenfolge der Zellen innerhalb einer Verbindung garantiert. Bedingt durch das asynchrone, paketorientierte Übermittlungsverfahren können jedoch Zellverluste und variable Verzögerungen für die einzelnen Zellen einer Verbindung auftreten. Oberhalb der ATM-Schicht ist deshalb eine Anpassungsschicht vorgesehen, in der – falls nötig – durch dienstindividuelle Protokolle die Auswirkungen dieser Effekte ausgeglichen werden sollen.
- Die Signalisierung erfolgt wie im Schmalband-ISDN getrennt von der Nutzdatenübertragung in getrennten virtuellen Verbindungen. Die für das Schmalband-ISDN entwickelten Konzepte und Protokolle werden mit entsprechenden Erweiterungen und Anpassungen weiter verwendet.

In der vorliegenden Arbeit werden einige der neuartigen Probleme beschrieben und untersucht, die durch die Kombination der paketorientierten Übermittlung und des asynchronen

Multiplexverfahrens in einem so charakterisierten ATM-Netz mit den verschiedenen, teilweise sehr hohen und widersprüchlichen Anforderungen der unterschiedlichen Dienste entstehen.

## 1.2 Übersicht über die Arbeit

Im Kapitel 2 werden als Grundlage für das Verständnis der nachfolgenden Kapitel zunächst die Konzepte für die Architektur und den Betrieb von Breitbandnetzen auf der Basis des Asynchronen Transfer-Modus (ATM) vorgestellt. Nach der Erläuterung einiger grundsätzlicher Anforderungen an ein solches Netz und der möglichen Prinzipien für deren Erfüllung werden die vom CCITT bis zur Fertigstellung der Arbeit getroffenen Festlegungen für den Asynchronen Transfer-Modus und das breitbandige, diensteintegrierende Digitalnetz (B-ISDN) dargestellt. Weiterhin werden einige Aspekte der grundsätzlichen Architektur einer ATM-Vermittlungsstelle diskutiert. Bezüglich des Betriebs von ATM-Netzen wird besonderes Augenmerk auf die Funktionen zur Steuerung und Kontrolle der Verkehrsflüsse in einem solchen Netz, wie z.B. Verbindungsannahme, Wegesuche und Quellflußkontrolle, und die gegenseitigen Abhängigkeiten dieser Funktionen untereinander gelegt.

Kapitel 3 beschreibt die Fragestellungen bei der verkehrstheoretischen Modellierung von Verkehrsströmen in ATM-Netzen und zeigt Ansätze und Verfahren für deren Lösung auf. Im einzelnen werden die verschiedenen stochastischen Charakterisierungsmöglichkeiten für zeitkontinuierliche sowie zeitdiskrete Verkehrsströme, insbesondere die Verteilungen für die Zwischenankunftsabstände und den Zählprozeß, angesprochen. Dabei werden neue Algorithmen zur Berechnung der Verteilung des Zählprozesses vorgestellt und am Ende des Kapitels an Beispielen bezüglich ihrer numerischen Genauigkeit untersucht, weil diese Algorithmen die Grundlage für einige der in Kapitel 5 durchgeführten analytischen Untersuchungen darstellen. Weiterhin werden die im Zusammenhang mit Untersuchungen an ATM-Netzen verwendeten Verkehrsmodelle zusammengestellt. Neben einigen Erweiterungen der Modelle und ihrer Beschreibung wird dabei auf deren Eignung für bestimmte Problemstellungen bei ATM-Untersuchungen eingegangen.

Im Kapitel 4 werden die für Untersuchungen im Zusammenhang mit ATM-Netzen im allgemeinen und für die Untersuchungen im Kapitel 5 im besonderen geeigneten Warteschlangenmodelle und die zu ihrer Lösung verwendeten simulativen und analytischen Verfahren dargestellt. Das Hauptaugenmerk liegt dabei auf Warteverlustsystemen mit einer deterministischen Bedieneinheit, wie sie sich bei der verkehrstheoretischen Modellierung von ATM-Systemkomponenten und -mechanismen durch die einheitliche Länge der übermittelten Zellen häufig ergeben. Bei den zeitdiskreten Systemen wird das dort auftretende Phänomen der Gleichzeitigkeit von Ankunft und Bedienung systematisch behandelt. Dazu werden die Lösungsansätze entsprechend erweitert und es wird an Beispielen quantitativ der Einfluß

der verschiedenen Möglichkeiten zur Auflösung dieser Gleichzeitigkeit dargestellt. Am Ende dieses Kapitels werden am Beispiel einer Einzelquelle verschiedene Verkehrsmodelle und ihre Lösungsmethoden verglichen und bezüglich ihrer Anwendbarkeit bei den folgenden Untersuchungen bewertet.

Die so eingeführten Verkehrsquellen- und Systemmodelle werden im Kapitel 5 eingesetzt, um die bisher bekannten Mechanismen zur Implementierung der Quellflußkontrollfunktion bezüglich relevanter Kriterien zu vergleichen. Die Quellflußkontrollfunktion (engl. Policing) ist eine neuartige Netzfunktion, deren Aufgabe es ist, den von einer Quelle gesendeten Zellstrom während der gesamten Verbindungsdauer zu überwachen und – falls notwendig – zu begrenzen. Diese Funktion wird in ATM-Netzen benötigt, weil keine Flußsteuerungsmechanismen zwischen dem Benutzer und dem Netz vorgesehen sind und eine Quelle sonst durch das asynchrone Multiplexprinzip die beim Verbindungsaufbau vereinbarten Verkehrsparameter erheblich überschreiten könnte. Dadurch könnte die Dienstgüte aller Verbindungen, welche die gleichen Netzressourcen mitbenutzen, spürbar verschlechtert werden. Bei der Untersuchung wird insbesondere auf die Kontrolle der mittleren Zellrate eingegangen, und es werden verschiedene Möglichkeiten der Dimensionierung der einzelnen Verfahren diskutiert und quantitativ untersucht. Die bei diesen Untersuchungen erhaltenen Ergebnisse wurden verwendet, um zwei neue Quellflußkontrollmechanismen zu entwickeln. Diese werden im letzten Abschnitt dieses Kapitels mit den bekannten Verfahren verglichen, um die mit ihnen erreichbare Verbesserung der Leistungsfähigkeit zu dokumentieren.

Kapitel 6 enthält über die quantitativen Ergebnisse in Kapitel 5 hinausgehende, allgemeingültige Aussagen über die Möglichkeiten und Grenzen des Quellflußkontrollkonzepts sowie einige Hinweise auf zusätzliche Maßnahmen mit deren Hilfe die Quellflußkontrolle und damit die Ressourcenzuteilung in ATM-Netzen effektiver gestaltet werden kann.

## Kapitel 2

# Architektur und Betrieb von Breitbandnetzen auf der Basis von ATM

### 2.1 Anforderungen an ein integriertes Breitband-Kommunikationsnetz

Ein universelles integriertes Breitband-Kommunikationsnetz (engl. Integrated Broadband Communication Network, IBCN), das im Sprachgebrauch des CCITT als Breitband-ISDN (engl. Broadband ISDN, B-ISDN) bezeichnet wird, ist gekennzeichnet durch eine umfassende Unterstützung aller denkbaren Dienste – einschließlich der Verteildienste – und sämtlicher Verbindungskonzepte mit Hilfe von wenigen, einheitlichen Schnittstellen zum Benutzer. Dabei müssen, wie in Bild 2.1 dargestellt in zunehmendem Maße neben den direkt und fest angeschlossenen Teilnehmern der unterschiedlichsten Art auch die verschiedenen Arten von privaten Netzen auf der Basis von Nebenstellenanlagen, lokalen Netzen und privaten Hochgeschwindigkeitsnetzen (MAN) berücksichtigt werden [22, 44]. Weitere wichtige Punkte sind die Einbindung einer zunehmenden Anzahl von mobilen Teilnehmern [4, 167] und die Möglichkeit für eine weitentwickelte Satellitenkommunikation. Das IBCN soll nicht nur die grundlegenden Kommunikationsfunktionen unterstützen, sondern auch darüber hinausgehende Funktionen bereitstellen (Intelligent Network), die eine flexible und auf die individuellen Wünsche des Benutzers zugeschnittene Nutzung erlauben.

Insbesondere ist ein Netzkonzept erforderlich, das eine Integration auch im Bezug auf die übertragungs- und vermittlungstechnischen Aspekte ermöglicht, denn nur dadurch läßt sich ein solches Netz ökonomisch sinnvoll betreiben. Ebenso wichtig ist die Möglichkeit einer

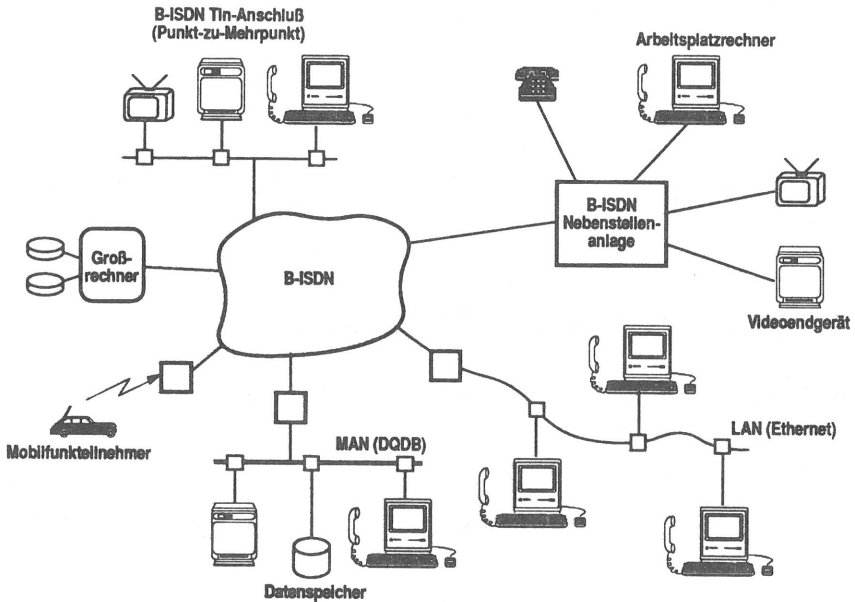


Bild 2.1: Szenario für ein B-ISDN

evolutionären Einführungsstrategie, wobei während einer Übergangsphase die bestehenden, kapitalintensiven Installationen mitverwendet werden können.

## 2.2 Prinzipien für ein integriertes Breitband-Kommunikationsnetz

Ursprünglich sollte ein geplantes IBCN konzeptionell sehr ähnlich wie das Schmalband-ISDN aufgebaut sein, d.h. es sollte auf der Durchschaltvermittlung (engl. Circuit Switching, CS) basieren. Bedingt durch die extrem unterschiedlichen Bandbreitanforderungen der verschiedenen Dienste und die damit verbundenen Probleme bei der Verwendung gemeinsamer Koppelnetze [164, 79] waren getrennte Koppelnetzstrukturen mit verschiedenen Basiskanalraten und die Zuordnung mehrerer Basiskanäle zu einer Verbindung (engl. Multi Rate Circuit Switching, MRCS) vorgesehen. Entsprechende Vermittlungsstellen mit Kanälen von maximal 140 Mbit/s wurden entwickelt und in Versuchsnetzen, wie z.B. im Bereich der Deutschen Bundespost dem Bildfernsprechversuchsnetz [10], dem Breitbandvorläufernetz (BVN) [14] und dem BERKOM-Netz [11, 50], getestet. Ende der siebziger Jahre wurden erste Konzepte für paketorientierte Hochgeschwindigkeitsnetze für den öffentlichen Bereich entwickelt,



die statistische Multiplextechniken verwendeten [63, 103, 149, 161]. Diese Konzepte wurden unter den Bezeichnungen „Fast Packet Switching“ (FPS), „Wideband Packet Technology“ (WPT) und „Asynchronous Time-Division Switching“ (ATD) bekannt gemacht. In entsprechenden Experimenten, z. B. in Frankreich (PRELUDE [43, 48, 155]) und in den Vereinigten Staaten (WPT [116], STARLITE [76], KNOCKOUT [169]) wurde die prinzipielle Implementierbarkeit dieser neuen Konzepte nachgewiesen. Gleichzeitig wurden auch Konzepte für kanalorientierte, statistische Breitbandnetze unter den Bezeichnungen „Fast Circuit Switching“ (FCS) und „Burst Switching“ publiziert [6, 7, 68].

Bei der internationalen Normung des B-ISDN, die zu Anfang der Studienperiode 84/88 des CCITT von einem kanalorientierten Konzept ausging, konnten zunächst keine entscheidenden Fortschritte erzielt werden. Die vor allem durch die Unsicherheit bezüglich der zukünftigen Dienste erforderliche Flexibilität des Netzkonzepts konnte mit diesem Ansatz nicht verwirklicht werden. Dies äußerte sich z.B. darin, daß keine Einigung über die endgültigen Bitraten der Breitbandkanäle erzielt werden konnte. Vergleiche mit den oben angeführten Konzepten (siehe [103, 114, 161]) bewirkten ein Umschwenken des CCITT – und damit der Hersteller und Netzbetreiber weltweit – auf ein paketorientiertes Konzept als Ziellösung für das B-ISDN [22, 44]. Dies wurde im CCITT Blaubuch durch die prinzipielle Definition des Asynchronen Transfer-Modus (engl. Asynchronous Transfer Mode, ATM) in der Empfehlung I.121 „Breitbandaspekte des ISDN“ [26] dokumentiert.

In den nächsten Abschnitten werden die Konzepte, die einem B-ISDN auf der Basis von ATM zugrunde liegen, und die bisher getroffenen Festlegungen erläutert, um das Verständnis der in den folgenden Kapiteln aufgegriffenen Fragestellungen zu erleichtern.

## **2.3 Zeitmultiplexverfahren und Verbindungskonzepte**

### **2.3.1 Synchrones Zeitmultiplex**

In den heutigen durchschaltvermittelnden Netzen wird das Synchrones Zeitmultiplex (engl. Synchronous Time Division Multiplexing, STD) zur simultanen Übertragung vieler getrennter Verbindungen auf einem gemeinsamen, breitbandigen Übertragungsweg verwendet. Zu diesem Zweck wird die gesamte verfügbare Übertragungskapazität des Mediums in zeitlich aufeinanderfolgende Abschnitte (Zeitschlitze) aufgeteilt. Diese Zeitschlitze werden durch geeignete Synchronisationsmaßnahmen zu periodischen Rahmen gruppiert, die jeweils die gleiche Anzahl von Zeitschlitzen enthalten. Relativ zum Rahmenanfang wird die Position eines Zeitschlitzes durch eine Zeitschlitznummer gekennzeichnet. Die Struktur eines solchen

synchronen Rahmens ist in Bild 2.2 am Beispiel des STS-1 (Synchronous Transport Signal - Level 1) Rahmens dargestellt, der die Grundeinheit für die von ANSI standardisierte SONET-Hierarchie (Synchronous Optical Network) bildet [12].

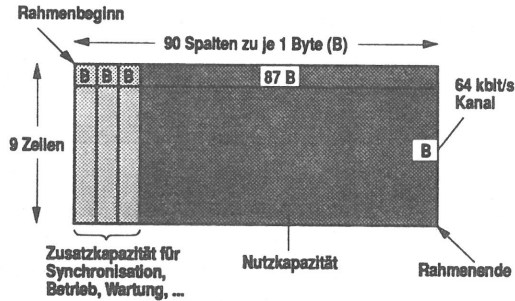


Bild 2.2: Beispiel für einen synchronen Zeitmultiplexrahmen

Der Rahmen wiederholt sich alle  $125\mu\text{s}$  und enthält 810 Bytes, was einer Gesamtbitrate von 51,84 Mbit/s entspricht. Durch die Periode von  $125\mu\text{s}$  werden die 64 kbit/s-Standardkanäle unterstützt, denen ein Zeitschlitz (Byte) pro Rahmen zugeordnet wird. Der Rahmen läßt sich mit 90 Spalten und 9 Zeilen von Zeitschlitzen darstellen, wobei die einzelnen Zeitschlitze zeilenweise von links nach rechts und von oben nach unten zeitlich hintereinander ausgesendet werden. Die beiden zuerst gesendeten Bytes dienen der Synchronisation und kennzeichnen den Rahmenanfang. Die ersten drei Spalten in dieser Darstellung können keine Nutzinformation enthalten und stellen neben Kapazität für Mechanismen zur Fehlererkennung spezielle Kanäle für Betriebs- und Wartungsaufgaben zur Verfügung.

Die CCITT Studiengruppe XVIII begann in der Studienperiode 84/88 ebenfalls mit der Normung einer neuen, synchronen Übertragungshierarchie (engl. Synchronous Digital Hierarchy, SDH) [23]. Diese orientiert sich konzeptionell an der oben beschriebenen SONET-Hierarchie und hat als Basiselement das „Synchronous Transport Module - Level 1“ (STM-1) mit einer Gesamtbitrate von 155,52 Mbit/s und einer Nutzbitrate von 149,76 Mbit/s. Dieses Signal kann z.B. aus drei STS-1-Signalen aufgebaut werden.

Die synchrone Übertragungshierarchie unterstützt direkt den Synchronen Transfer-Modus (engl. Synchronous Transfer Mode, STM), bei dem das Konzept der Durchschaltvermittlung (engl. Circuit Switching, CS) angewendet wird. Dabei wird jeder Verbindung während des Verbindungsaufbaus auf den einzelnen Übertragungsabschnitten für die gesamte Verbindungsdauer eine der maximalen Bitrate entsprechende Anzahl von bestimmten Zeitschlitzen pro Rahmen fest zugeordnet. In den zwischen den einzelnen Übertragungsabschnitten gelegenen Vermittlungsstellen können die einer Verbindung zugeordneten Zeitschlitznummern

umgesetzt werden (nicht-koinzidente Durchschaltung). Dazu werden Umsetztabelle verwendet, die beim Verbindungsaufbau geladen werden. Bei jeder Zeitlagenumsetzung wird die Informationseinheit – je nach Lage der beiden zugeordneten Zeitschlitze zueinander – um bis zu eine Rahmendauer verzögert. Das Prinzip einer solchen physikalischen (durchgeschalteten) Verbindung ist in Bild 2.3 dargestellt.

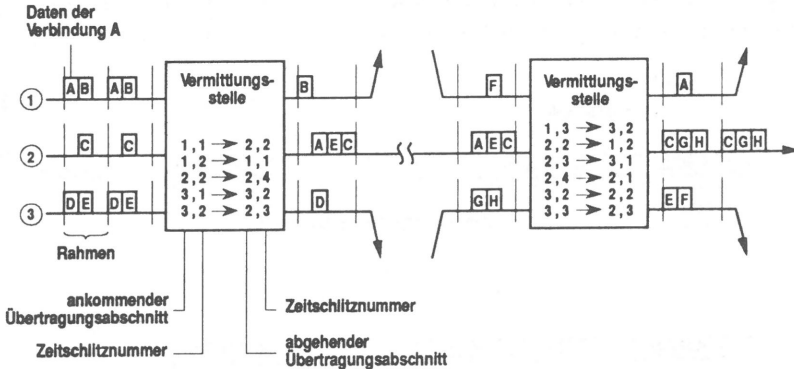


Bild 2.3: Prinzip einer physikalischen Verbindung

Die Zugehörigkeit ankommender Daten zu einer bestimmten Verbindung wird in den Vermittlungsstellen nur aufgrund der Zeitschlitznummern erkannt (engl. Position Multiplexing). Die Verzögerung in einer Vermittlungsstelle kann durch wiederholte Zeitlagenumsetzung in mehrstufigen Koppelnetzen mehrere Rahmendauern betragen, sie ist jedoch für alle Informationseinheiten einer bestimmten Verbindung gleich. Durch die feste Zeitschlitzzuordnung beeinflussen sich bestehende Verbindungen gegenseitig nicht. Informationsverluste können bei physikalischen Verbindungen nur aufgrund von Synchronisationsproblemen (Schlupf) oder Fehlfunktionen auftreten und ihr Auftreten ist unabhängig von der Art und Zahl gleichzeitig bestehender Verbindungen.

Die maximale Bitrate einer Verbindung kann durch die Kanalorientierung des Verfahrens nur ein ganzzahliges Vielfaches des Basiskanals sein, wobei es vermittlungstechnisch problematisch ist, viele Basiskanäle, die einer Verbindung zugeordnet sind, gleichzeitig durchzuschalten. Für Verbindungen mit sehr unterschiedlichen Spitzenbitraten werden deshalb i.a. getrennte Koppelnetze mit unterschiedlichen Basiskanälen vorgesehen, was die Flexibilität des Verfahrens im Hinblick auf ein alle denkbaren Dienste integrierendes Universalnetz erheblich einschränkt.

### 2.3.2 Asynchrones Zeitmultiplex

Neben der synchronen Zeitmultiplextechnik kann auch das Asynchrone Zeitmultiplex (engl. Asynchronous Time Division Multiplexing, ATD [25]) für die Mehrfachausnutzung eines Übertragungsweges verwendet werden. Dabei wird die Übertragungskapazität in Zeitschlitze aufgeteilt, die Datenblöcke konstanter Länge aufnehmen können. Diese Blöcke werden als Zellen (engl. Cells) bezeichnet und bestehen aus einem Kopffeld und dem eigentlichen Informationsfeld. Um den Beginn der einzelnen Zellen erkennen zu können, müssen geeignete Synchronisationsmaßnahmen vorgesehen werden. Beispiele hierfür sind in Bild 2.4 gezeigt.

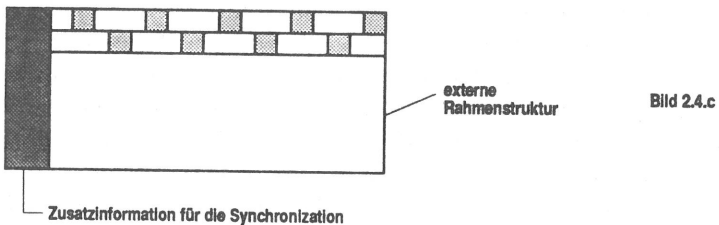
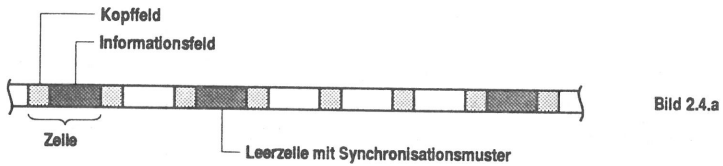


Bild 2.4: Möglichkeiten für die Zellsynchronisation bei ATD

Für die Zellsynchronisation können z.B. nicht benutzte Zellen (Leerzellen) dienen, die ein entsprechendes Synchronisationsmuster tragen (Bild 2.4.a). Ein Nachteil dieser Methode ist, daß bei einer hohen Auslastung des Übertragungsweges der Abstand zwischen den Leerzellen relativ groß und deshalb die Zeit bis zum Wiedererlangen der Synchronität unzulässig lang werden kann.

Eine andere Möglichkeit zur Zellsynchronisation auf der Basis eines reinen Zellstromes besteht durch die Mitverwendung eines im Kopffeld vorhandenen Fehlerschutzmechanismus (engl.

Header Error Control, HEC). Im Falle eines Synchronitätsverlusts wird dabei Bit für Bit überprüft, ob eine der Kopffeldlänge entsprechende Folge von Bits neben den zu schützenden Bits die Kontrollbits enthält, die nach der Schutzvorschrift zu erwarten sind (das Fehlersyndrom ist in diesem Fall Null). Falls diese Bedingung erfüllt ist, wird die Prüfung nach einer der Zelllänge entsprechenden Anzahl von Bits wiederholt. Fällt diese Überprüfung in einer bestimmten Anzahl von aufeinanderfolgenden Fällen positiv aus, dann wird eine Wiedererlangung der Zellsynchronität angenommen.

Zur Zellsynchronisation können auch periodisch in den Zellstrom eingefügte Synchronisationszellen verwendet werden (siehe Bild 2.4.b). Diese definieren dann einen Rahmen ähnlich wie beim STD, der prinzipiell für die Unterstützung eines synchronen Übermittlungsmodus mitverwendet werden könnte (hybrides System). Dafür müßte jedoch das gesamte Netz synchron betrieben werden, was beim reinen ATD nicht unbedingt nötig ist, da bei diesem Verfahren bei Bedarf beliebig Leerzellen eingefügt oder entfernt werden können, um Geschwindigkeitsunterschiede auszugleichen. Nachteilig bei diesem Zellsynchronisationsverfahren ist, daß die benötigten Synchronisationszellen die für die Nutzübertragung verfügbare Übertragungskapazität vermindern.

Falls für die Übertragung der Zellen bestehende synchrone Übertragungsstrecken verwendet werden (siehe Abschnitt 2.3.1), kann auch die dort definierte Rahmenstruktur für die Zellsynchronisation verwendet werden (Bild 2.4.c). Dadurch ergibt sich jedoch eine unerwünschte Abhängigkeit der Zellsynchronisation vom Übertragungsmedium.

Das ATD-Prinzip ist dem asynchronen Paketmultiplexverfahren ähnlich, wie es z.B. in den bestehenden Paketvermittlungsnetzen gemäß der CCITT-Empfehlung X.25 [35] angewendet wird. Beim asynchronen Paketmultiplexverfahren können die einzelnen Informationsblöcke jedoch verschieden lang sein und es können beliebige Leerzeiten zwischen den einzelnen Paketen auftreten. Deshalb müssen Anfang und Ende eines jeden Paketes individuell mit Begrenzern (engl. Flags) versehen sein, die eine entsprechende Synchronisation erlauben.

Das ATD-Prinzip unterstützt direkt den in den folgenden Abschnitten näher erläuterten Asynchronen Transfer-Modus, bei dem das Prinzip der virtuellen Verbindungen (engl. Virtual Circuit) zur Anwendung kommt (siehe Bild 2.5).

Beim Verbindungsaufbau wird dabei der Weg festgelegt, auf dem alle Zellen dieser Verbindung das Netz durchlaufen. Dadurch wird erreicht, daß sich die Zellen einer Verbindung nicht überholen können und beim Empfänger in der korrekten Reihenfolge ankommen. Gleichzeitig wird der Verbindung auf jedem Verbindungsabschnitt eine logische Kanalnummer (engl. Logical Channel Number, LCN) zugeordnet, die alle Zellen dieser Verbindung im Kopffeld tragen. In den Vermittlungsstellen wird aufgrund dieser logischen Kanalnummer die Zugehörigkeit der ankommenden Zellen zu einer bestimmten Verbindung erkannt. Die Zellen werden dann zu den entsprechenden Ausgängen der Vermittlungsstelle weitertransportiert

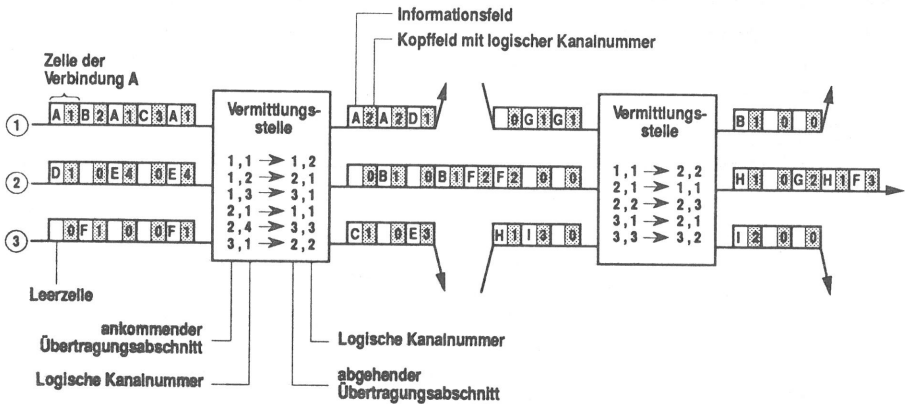


Bild 2.5: Prinzip einer virtuellen Verbindung

und erhalten die auf dem neuen Verbindungsabschnitt gültige logische Kanalnummer im Kopffeld eingetragen. Für die Auswahl des richtigen Ausgangs und für die Zuordnung der neuen logischen Kanalnummer werden Tabellen verwendet, in die beim Verbindungsaufbau die entsprechenden Daten eingetragen werden.

Im Gegensatz zur synchronen Durchschaltung brauchen die Zellen einer Verbindung nicht periodisch aufzutreten, d.h. ihr Abstand kann sich im Prinzip nach der momentan von der Verbindung benötigten Übertragungskapazität richten. Dadurch werden in natürlicher Weise unterschiedliche Spitzenbitraten unterstützt. Auch Bandbreitanforderungen, die während der Dauer einer Verbindung variieren, werden unterstützt, wobei eine virtuelle Verbindung nur dann Übertragungskapazität belegt, wenn sie wirklich benötigt wird. Die durch die asynchrone Überlagerung solcher Verbindungen entstehenden statistischen Effekte definieren die Eigenschaften einer virtuellen Verbindung. So können z.B. bei einer realistischen Dimensionierung der Netzressourcen in einer Vermittlungsstelle kurzzeitig mehr Zellen für einen bestimmten Ausgang eintreffen, als auf diesem übertragen werden können. Für diesen Fall sind in den Vermittlungsstellen in begrenztem Umfang Pufferspeicher vorgesehen, die übermäßige Zellverluste verhindern und dafür variable Verzögerungen für die verschiedenen Zellen verursachen. Diese Verzögerungen und im Extremfall auch Zellverluste durch Pufferüberlauf entstehen durch die gegenseitige Beeinflussung bestehender Verbindungen und sind abhängig von der Art und Anzahl von Verbindungen, die sich die jeweiligen Netzressourcen teilen (statistisches Multiplexen). Im Gegensatz zu der in Abschnitt 2.3.1 dargestellten physikalischen Verbindung hängt also die u.a. durch die Zellverlustwahrscheinlichkeit und den Umfang von Verzögerungen und Verzögerungsschwankungen definierte Dienstgüte (engl. Quality of Service, QOS) einer virtuellen Verbindung stark von den gleichzeitig bestehenden

Verbindungen und damit von der Auslastung des Netzes ab.

Durch die einheitliche Darstellung sämtlicher Informationen in Form von gleichartigen Zellen ermöglicht dieses Prinzip jedoch die Vermittlung von Verbindungen mit verschiedenen und evtl. veränderlichen Bitraten in einheitlichen Koppelnetzen und bietet dadurch die Flexibilität und Zukunftssicherheit, die für ein zukünftiges diensteintegrierendes Universalnetz unabdingbar ist.

## 2.4 Konzepte für das Breitband-ISDN

In der Studienperiode von 1984–1988 wurde dann in der Studiengruppe XVIII des CCITT begonnen, ATM als Ziellösung für die Konzeption eines Breitband-ISDN zu standardisieren. Als erster Schritt dazu wurde die Empfehlung I.121 „Breitbandaspekte des ISDN“ [26] in das CCITT Blaubuch aufgenommen. Diese Empfehlung soll als Richtlinie für eine weitergehende Normung in der neuen Studienperiode dienen und enthält relativ allgemeine Festlegungen für die Konzepte, Architekturmodelle, Protokolle, Dienste und Schnittstellen eines B-ISDN auf der Basis von ATM. Den folgenden Abschnitten liegen neben dieser Empfehlung auch die bisher verfügbaren Dokumente über spätere Sitzungen der Studienkommission XVIII zugrunde [36, 37, 38].

### 2.4.1 Dienste und Dienstkonzept

Ganz allgemein soll das B-ISDN ein breites Spektrum an Diensten unterstützen, worunter sowohl die vom Schmalband-ISDN her bekannten Dienste als auch spezifische Breitbanddienste fallen. Das Dienstkonzept für ein ISDN, auf dem auch die Definition und Beschreibung der Dienste im B-ISDN basiert, ist in den CCITT-Empfehlungen I.2xx [27] definiert. In der Empfehlung I.121 wird speziell auf die Breitbanddienste und deren Klassifikation eingegangen, wobei auch Beispiele für Dienste der einzelnen Dienstklassen enthalten sind.

Prinzipiell wird es weiterhin die Unterscheidung in Übermittlungsdienste (engl. Bearer Services) und Teledienste (engl. Teleservices) geben. Bei den Übermittlungsdiensten sind nur die den OSI-Schichten 1 bis 3 zuzuordnenden Funktionen festgelegt und durch entsprechende Attribute (Low Layer Attributes) beschrieben, bei den Telediensten sind zusätzlich die Funktionen der OSI-Schichten 4 bis 7 durch zusätzliche Attribute (High Layer Attributes) festgelegt.

Für die Übermittlungsdienste müssen bei einigen Attributen weitere Werte zugelassen werden. So wird es neben dem durchschaltvermittelten Transfer-Modus (engl. Circuit Mode) einen deterministischen ATM-Modus geben, bei dem während der ganzen Verbindungsdauer

die entsprechende Spitzenbitrate zugeordnet wird, sowie einen statistischen ATM-Modus, bei dem nur eine zwischen mittlerer und maximaler Bitrate liegende Kapazität reserviert wird. Auch die Übertragungsrate der Verbindung muß in der flexiblen ATM-Umgebung genauer als bisher beschrieben werden, da es keine starre Kanaleinteilung mehr gibt und die geforderte maximale Bitrate beliebige Werte annehmen kann. Zusätzlich zur Spitzenbitrate muß außerdem noch das zeitliche Verhalten der Bitrate während der Verbindung durch geeignete Angaben, z.B. den Mittelwert, charakterisiert werden. Diese Angaben werden im Netz benötigt, um z.B. zu entscheiden, ob ein Bandbreitengewinn durch statistisches Multiplexen (im statistischen ATM-Modus) erzielt werden kann und um die Parameter der benötigten Quellfluß- und Zugangskontrollmechanismen anzupassen.

Desweiteren wird das Attribut „Dienstgüte“ (engl. Quality Of Service, QOS) gegenüber dem Schmalband-ISDN eine weit größere Bedeutung erlangen, da mit diesem Attribut die Anforderungen der verschiedenen Dienste bezüglich End-zu-End-Verzögerungen, Verzögerungsschwankungen und Informationsverlusten genauer beschrieben werden können. Diese Größen sind, wie in Abschnitt 2.3.2 beschrieben, in einem ATM-Netz von Art und Anzahl gleichzeitig existierender Verbindungen bestimmt, die sich gegenseitig beeinflussen. Die entsprechenden Anforderungen der Dienste müssen durch geeignete Maßnahmen z.B. bei der Verbindungsannahme (engl. Call Acceptance Control), bei der Wegesuche (engl. Routing) oder durch die Einführung von Prioritätsmechanismen erfüllt werden.

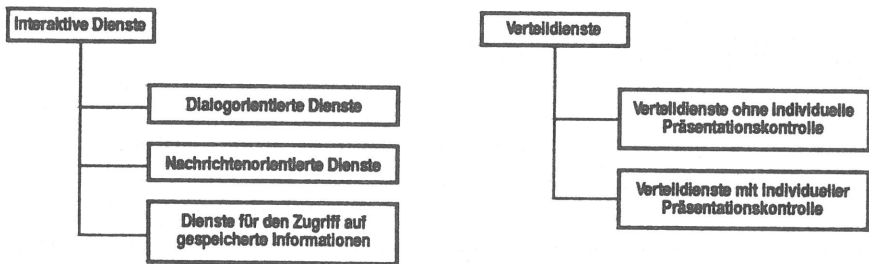


Bild 2.6: Klassifizierung der Breitbanddienste nach der I.121

Das Spektrum der zukünftigen Teledienste mit Breitbandanforderungen kann grob in zwei Kategorien eingeteilt werden, nämlich interaktive Dienste (engl. Interactive Services) und Verteildienste (engl. Distribution Services), wobei die interaktiven Dienste wie in Bild 2.6 gezeigt weiter unterteilt werden in

- *Dialogorientierte Dienste (Conversational Services)*: Bidirektionale, dialogorientierte Kommunikation mit End-zu-End Datenaustausch in Echtzeit (keine Speichervermittlung). Die Information wird von einem Benutzer erzeugt und ist für einen oder mehrere



individuelle Benutzer bestimmt. Beispiele sind Videotelephonie, Videokonferenz und Hochgeschwindigkeits-Datenübertragung.

- *Nachrichtenorientierte Dienste (Messaging Services)*: Kommunikation zwischen individuellen Benutzern nach dem Speichervermittlungsprinzip (Store-and-Forward), einschließlich Speicherungs- (engl. Mailbox), Editier-, Verarbeitungs- und Umwandlungsfunktionen (engl. Message Handling). Beispiele sind Message-Handling- und Mail-Dienste für Bewegtbilder, hochauflösende Graphik oder Audio-Information.
- *Dienste für den Zugriff auf gespeicherte Informationen (Retrieval Services)*: Zugriff auf gespeicherte, allgemein zugängliche Informationen auf individuelle Anfrage, d.h. der Benutzer bestimmt den Zeitpunkt des Beginns der Informationssequenz. Beispiele sind Datenbankdienste für Bewegtbilder und hochauflösende Graphik.

Bei den Verteildiensten gibt es eine Unterscheidung in

- *Dienste ohne individuelle Präsentationskontrolle (Broadcast Services Without User Individual Presentation Control)*: Ein kontinuierlicher Informationsfluß wird von einer zentralen Quelle an eine unbegrenzte Zahl von autorisierten Benutzern verteilt. Der Benutzer kann dabei weder den Beginn der Informationseinheiten noch deren Reihenfolge beeinflussen. Beispiele für diese Dienste sind Fernseh- und Audio-Verteildienste.
- *Dienste mit individueller Präsentationskontrolle (Distribution Services With User Individual Presentation Control)*: Hier wird die Information als Folge zyklisch wiederholter Informationseinheiten angeboten wobei der Benutzer den Anfang und die Reihenfolge der Einheiten bestimmen kann. Er kann deshalb die Präsentation immer von Anfang an verfolgen. Beispiele sind hochqualitative Videoübertragungen.

Für das B-ISDN werden auch neue Aspekte beim Dienstkonzept berücksichtigt werden müssen. So werden viele der Breitbanddienste sogenannte „Multi-Media“-Dienste sein, die verschiedene Arten von Information mit zum Teil sehr unterschiedlichen charakteristischen Eigenschaften benötigen. Beim Bildfernsprechen werden z.B. Ton- und Bildinformation benötigt, wobei die Bildinformation während des Gesprächs (engl. Call) abschaltbar sein soll. Ein anderes Beispiel ist die Fernsehverteilung mit hoher Auflösung (engl. High Definition Television, HDTV), bei der man für die Video-Komponente die Wahl zwischen normaler und hoher Auflösung und beim Ton Möglichkeiten wie Stereoton oder Zweikanalton (in verschiedenen Sprachen) haben wird. Diese Möglichkeiten soll man dynamisch je nach Bedarf kombinieren und die Kombination auch nach dem eigentlichen Verbindungsaufbau noch verändern können (engl. In-Call Modification). Zur Beschreibung solcher Dienste wurde das Konzept der „Dienstkomponenten“ (engl. Service Components) vorgeschlagen [18, 139]. Dabei werden Dienstkomponenten mit einheitlichem Informationstyp (Sprache, Text, ...) definiert, aus denen die Multi-Media-Dienste zusammengesetzt werden können. Dadurch kann die Definition eines eigenen Teledienstes für jede mögliche Kombination vermieden werden.

## 2.4.2 Prinzipielle Eigenschaften von ATM

ATM im Sinne der CCITT-Empfehlung I.121 [26] ist ein spezieller paketorientierter Übermittlungsmodus, bei dem die asynchrone Zeitmultiplextechnik (ATD, siehe Abschnitt 2.3.2) angewendet wird. Die dabei verwendeten Informationsblöcke (Zellen, engl. Cells) haben eine feste Länge und bestehen aus einem Kopffeld (engl. Header), das für den Transfer der Zellen durch das Netz benötigt wird und dessen Hauptaufgabe die Zuordnung der Zellen zu einer virtuellen Verbindung ist, sowie aus einem Informationsfeld (engl. User Information Field), das unverändert durch das Netz transportiert wird. Das Informationsfeld hat für alle Verbindungen und an allen genormten Bezugspunkten die gleiche Länge, dasselbe soll für das Kopffeld ebenfalls erreicht werden. Die Zellen werden von den aktiven Verbindungen angefordert und gemäß der vorhandenen Netzressourcen zugeteilt. Die entsprechenden Werte für das Kopffeld werden auf jedem Verbindungsabschnitt bei Bedarf zugeordnet und wieder freigegeben, wenn sie nicht mehr benötigt werden. Die korrekte Reihenfolge der Zellen einer Verbindung bleibt im Netz erhalten, es können jedoch Zellen einer Verbindung durch Fehlfunktionen oder Überlastsituationen verloren gehen.

Vom Prinzip her ist ATM ein verbindungsorientiertes Verfahren, wobei ein B-ISDN auf der Basis von ATM neben vermittelten und festgeschalteten Verbindungen für die durchschaltete und für die paketorientierte Kommunikation auch die z.B. in lokalen Netzen (engl. Local Area Network, LAN) angewendete verbindungslose Kommunikation unterstützt.

Signalisier- und Nutzinformationen werden in getrennten virtuellen Verbindungen transportiert (Außerband-Signalisierung, engl. Outband Signalling), wodurch eine flexible Signalisierung auch während der aktiven Phase einer Verbindung ermöglicht wird.

## 2.4.3 Basis-Architekturmodell

Das Basis-Architekturmodell für das B-ISDN entspricht im Prinzip dem des Schmalband-ISDN [30] mit entsprechenden Erweiterungen. Es enthält, wie in Bild 2.7 dargestellt, die lokalen Funktionseinheiten (engl. Local Functional Capabilities, LFC), die z.B. die Funktionen der Ortsvermittlungsstelle (engl. Local Exchange, LEX) umfassen. Darunter fallen u.a. die Funktionen der Benutzer/Netz-Signalisierung (engl. User/Network Signalling).

Desweiteren enthält ein B-ISDN Funktionseinheiten zur Zwischenamts-Signalisierung (engl. Inter Exchange Signalling Capabilities) und die ebenfalls vom Schmalband-ISDN her bekannten Funktionseinheiten zur Übermittlung auf der Basis von 64 kbit/s sowie die neuen Funktionseinheiten zur breitbandigen Informationsübermittlung. Unter Benutzung dieser den OSI-Schichten 1 bis 3 zuzuordnenden Funktionseinheiten (engl. Low-Layer Capabilities) wird ein Informationsaustausch zwischen Endsystemen (engl. Terminal Equipment, TE) oder

zwischen Endsystemen und ebenfalls an das Netz angeschlossenen Dienstaniern (Datenbanken usw.) ermöglicht.

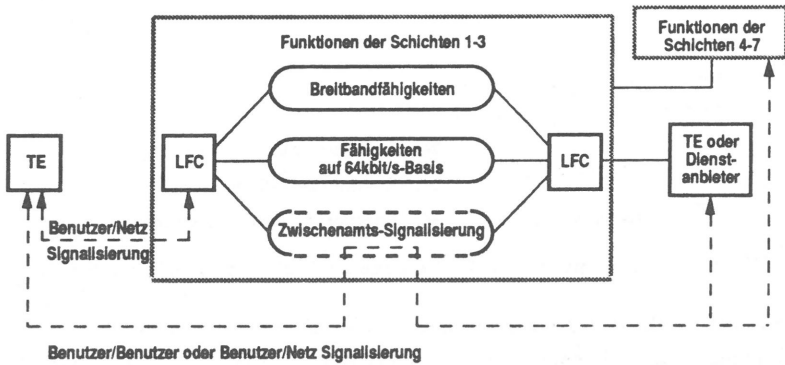


Bild 2.7: Basisarchitekturmodell für das B-ISDN

Außerdem soll das B-ISDN den OSI-Schichten 4 bis 7 zuzuordnende Instanzen (High-Layer Capabilities) enthalten, die z.B. Datenbankdienste, Möglichkeiten zur Informationsbearbeitung oder Möglichkeiten zur Protokollumsetzung bieten. Zur flexiblen Nutzung dieser Funktionen steht die Möglichkeit der Benutzer-zu-Benutzer Signallerung zur Verfügung, die transparent durch das Netz übermittle wird.

#### 2.4.4 Bezugskonfiguration

Die Festlegung einer Bezugskonfiguration (engl. Reference Configuration) mit einheitlich festgelegten Bezugspunkten für die Normung der Schnittstellen erfolgt für das B-ISDN ebenfalls analog zu den im Schmalband-ISDN verwendeten Definitionen [31] in der Empfehlung I.413 „ISDN User-Network Interface with Broadband Capabilities“ [38]. Um eine Unterscheidung zu den Definitionen des Schmalband-ISDN zu erreichen, wird bei den Definitionen für das B-ISDN an die entsprechenden Bezeichnungen jeweils der Buchstabe „B“ angefügt.

Wie in Bild 2.8 dargestellt, werden für die Bezugskonfiguration verschiedene Funktionsblöcke und Bezugspunkte definiert.

Die Benutzer/Netzschnittstelle wird an den Bezugspunkten (engl. Reference Point)  $S_B$  und  $T_B$  bezüglich ihrer logischen und physikalischen Eigenschaften international genormt. Mit B-TE (Terminal Equipment) werden die in den Endsystemen implementierten Funktionen bezeichnet, die alle Schichten des OSI-Referenzmodells umfassen können. Endsysteme, die den CCITT-Spezifikationen für die Benutzer/Netzschnittstelle des B-ISDN entsprechen und

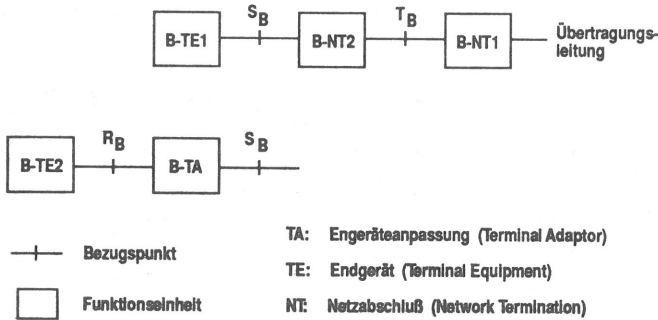


Bild 2.8: Bezugskonfiguration für die Benutzer/Netzschnittstelle des B-ISDN

deshalb direkt an das B-ISDN angeschlossen werden können, werden dabei als B-TE1 (Terminal Equipment Type 1) bezeichnet. Für Endsysteme mit anderen Schnittstellen (B-TE2) werden in einer Endgeräteanpassung (Terminal Adaptor, B-TA) entsprechende Funktionen zur Anpassung der verschiedenen am Bezugspunkt  $R_B$  vorliegenden Schnittstellen an die B-ISDN Spezifikationen vorgesehen.

Zwischen den Bezugspunkten  $S_B$  und  $T_B$  liegen die Funktionen der B-NT2 (Netzabschluß, engl. Network Termination), die z.B. die Behandlung von Protokollen der Schichten 1 bis 3, Vermittlungs-, Signalisier-, Konzentrations- und Multiplexfunktionen umfassen. Diese Funktionen können in Form von Nebenstellenanlagen (engl. Private Branch Exchange, PBX), lokalen Netzen oder einfachen Multiplexern implementiert werden [22]. Der Umfang der in einer B-NT2 enthaltenen Funktionen richtet sich nach der jeweiligen Anschlußkonfiguration und den Bedürfnissen des Benutzers. Falls diese Funktionen gar nicht benötigt werden, soll die B-NT2 zur Nullfunktion schrumpfen können oder zumindest möglichst einfach sein. Dies erfordert eine weitgehende Übereinstimmung der Spezifikationen für die Bezugspunkte  $S_B$  und  $T_B$  und stellt damit eine Anforderung für die weitere Normung dar. Wenn eine B-NT2 physikalisch vorhanden ist, wird sie dem Einflusbereich des Benutzers (engl. Customer Premises Network, CPN) zugeordnet und untersteht nicht der direkten Kontrolle des Netzbetreibers.

Die B-NT1 bildet den Netzabschluß bezüglich der OSI-Schicht 1 und bietet damit die Funktionen, die für einen ordnungsgemäßen physikalischen und elektrischen bzw. optischen Abschluß des Netzes nötig sind.

An den Bezugspunkten  $T_B$  und  $S_B$  sollen neben Punk-zu-Punkt- auch Punk-zu-Mehrpunkt-Konfigurationen mit Zugriff auf ein gemeinsames Übertragungsmedium möglich sein (siehe I.413 CCITT-Empfehlung [38]).

Am Bezugspunkt  $T_B$  beginnt der Hoheitsbereich des Netzbetreibers und alle Quellfluß- und Zugangskontrollmechanismen zum Schutz des Netzes müssen deshalb vom Benutzer aus ge-

sehen hinter diesem Bezugspunkt angesiedelt werden. Unabhängig davon können Funktionen zur Quellfußkontrolle und Ressourcenzuteilung auch in der B-NT2 vorhanden sein.

### 2.4.5 Protokollmodell

Das in der Empfehlung I.121 und in der Empfehlung I.321 „B-ISDN Protocol Reference Model and its Applications“ [38] enthaltene Protokollmodell (siehe Bild 2.9) hat prinzipiell die gleiche Struktur wie dasjenige für das Schmalband-ISDN [29]. Das Hauptmerkmal ist die Unterteilung in geschichtet strukturierte Steuerungs- und Benutzerebenen (engl. Control Plane, User Plane), die durch die Trennung von Signalisierung und Nutzdatenübertragung zustande kommt. Die Protokolle der Benutzerebene legen dabei die Übermittlung der Nutzinformationen fest, wogegen die Steuerungsebene die Protokolle für die Steuerung von Verbindungen und zusätzlichen Dienstmerkmalen enthält. Weiterhin ist eine Management-Ebene definiert, die sich über Steuerungs- und Benutzerebene erstreckt und in Funktionen für das Ebenenmanagement und in Funktionen für das Schichtenmanagement unterteilt ist. Die nicht weiter strukturierten Funktionen für das Ebenenmanagement dienen dabei zur Verwaltung gemeinsamer Daten und zur Koordination zwischen den getrennten Ebenen. Die Funktionen für das Schichtenmanagement verwalten schichtspezifische Ressourcen und Parameter und bearbeiten Nachrichten für Betriebs- und Wartungsaufgaben.

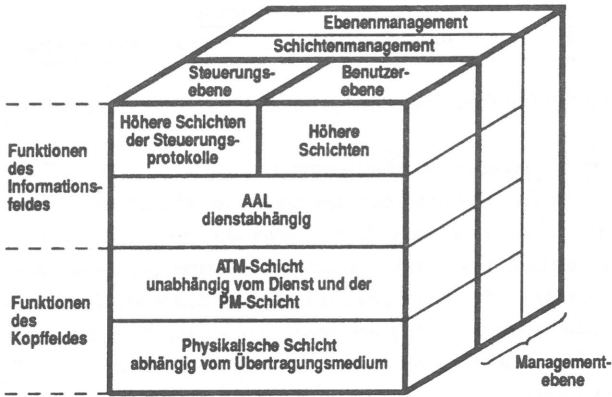


Bild 2.9: Protokollmodell für das B-ISDN auf der Basis von ATM

Die drei unteren Schichten werden von Benutzer- und Steuerungsebene gemeinsam benutzt. Die unterste Schicht, die auch als „physikalische Schicht“ (engl. Physical Layer) bezeichnet wird, ist in der Empfehlung I.432 „B-ISDN User-Network Interface – Physical Layer Specification“ [38] genauer spezifiziert. Wie in Bild 2.10 dargestellt, ist sie weiter unterteilt in eine

vom physikalischen Medium abhängige Subschicht (engl. Physical Medium Sublayer, PM) und in eine Subschicht, die die Umsetzung des Zellstroms in zur Übertragung auf dem Medium geeignete Dateneinheiten vornimmt (engl. Transmission Convergence Sublayer, TC). Die PM Subschicht legt unter anderem die physikalischen Eigenschaften der Übertragungsstrecke, den Leitungscode, die elektrisch/optische Umwandlung und die Mechanismen zur Bitsynchronisation fest.

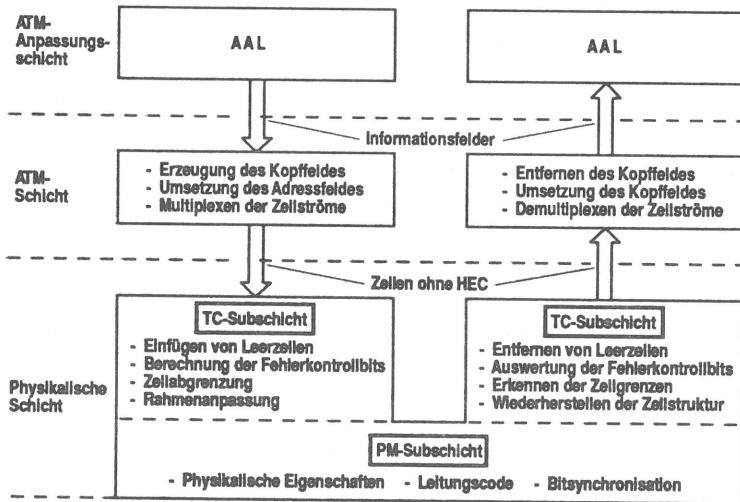


Bild 2.10: Die Funktionen der physikalischen und der ATM-Schicht

Abhängig vom jeweils verwendeten Übertragungssystem wird in der TC Subschicht – falls nötig – die Rahmenstruktur für den Übertragungsrahmen erzeugt und für eine entsprechende Strukturierung des Zellstroms gesorgt. Es sind zwei Optionen für die verwendeten Übertragungssysteme vorgesehen. Neben einem speziellen Übertragungssystem auf der Basis eines reinen ATM-Zellstroms ist dies eine Variante, in der die Zellen als Nutzinformation in einem STM-1 Signal (siehe Abschnitt 2.3.1) der neuen synchronen Übertragungshierarchie transportiert werden. Die Anpassung an diese Rahmenstruktur ist in der CCITT-Empfehlung I.432 genauer beschrieben.

Weiterhin werden in dieser Subschicht die für das Erkennen der Zellgrenzen (engl. Cell Delineation) und für den Fehlerschutz des Kopffeldes (engl. Header Error Control, HEC, siehe Abschnitt 2.3.2) nötigen Funktionen ausgeführt. Zellen mit nicht korrigierbaren Fehlern im Kopffeld werden verworfen.

Eine weitere Funktion dieser Schicht ist die Entkopplung der Rate des von der ATM-Schicht

empfangenen Zellstroms von der Kapazität des Übertragungssystems. Dazu werden Leerzellen (engl. Idle Cells) eingefügt bzw. wieder entfernt.

Oberhalb der PM- und der TC-Schicht befinden sich zwei weitere für ein ATM-Netz spezifische Protokollschichten, nämlich die ATM-Schicht (engl. ATM Layer) und die ATM-Anpassungsschicht (engl. ATM Adaptation Layer, AAL). Die ATM-Schicht ist unabhängig vom jeweils in Anspruch genommenen Dienst und bietet die Möglichkeit einer abschnittsweisen (engl. Link-by-Link) Übermittlung von ATM-Zellen. Diese Schicht umfaßt in der Hauptsache die folgenden Funktionen (siehe Bild 2.10):

- *Umsetzen des Adreßfeldes (engl. Cell VPI/VCI Translation):* Zuordnung der auf den jeweiligen Übertragungsabschnitten für die Verbindung gültigen logischen Kanalnummern in den Vermittlungsstellen.
- *Multiplexen und Demultiplexen der Zellströme (engl. Cell Multiplexing/Demultiplexing):* Zu sendende Zellen von verschiedenen Verbindungen werden unter Anwendung des statistischen Multiplex-Verfahrens zu einem Zellstrom überlagert. Dabei können Funktionen wie Zwischenpufferung und Verwerfen von Zellen nötig sein. Auf der Empfängerseite wird der überlagerte Zellstrom wieder in die einzelnen Verbindungen aufgespalten.
- *Erzeugen und Entfernen des Kopffeldes (engl. Cell Header Generation/Extraction):* Die von der AAL-Schicht empfangenen Informationsfelder der Zellen werden mit einem Kopffeld (ohne HEC) versehen. Von den zur Übergabe an die AAL-Schicht bestimmten Zellen wird das Kopffeld entfernt.
- *Generische Flußkontrolle (engl. Generic Flow Control, GFC):* Unterstützung von Mehrpunkt-Anschlußkonfigurationen mit einem gemeinsamen Übertragungsmedium.

Die oben genannten Funktionen werden durch entsprechende Felder im Kopffeld der Zellen unterstützt (siehe Abschnitt 2.4.6.2), das Informationsfeld wird auf der ATM-Schicht nicht bearbeitet.

Die AAL-Schicht schirmt den Benutzer von den spezifischen Eigenschaften der ATM-Schicht ab und unterstützt die Funktionen der höheren Protokollschichten in der Steuerungs- und in der Benutzerebene [22]. Diese Schicht ist abhängig vom verwendeten Dienst und erlaubt auch Verbindungen zwischen ATM- und Nicht-ATM-Schnittstellen (siehe CCITT I.362 „B-ISDN ATM Adaptation Layer (AAL) Functional Description“ [38]).

Für die AAL-Schicht sind innerhalb der Informationsfeldes entsprechende Felder reserviert, die je nach den Anforderungen des jeweiligen Dienstes unterschiedlich sind und deren Festlegung in der Empfehlung I.363 „ATM Adaptation Layer (AAL) Specification“ [38] erfolgt.

Der ATM-Anpassungsschicht werden auf der Sendeseite von den höheren Schichten Dateneinheiten übergeben, die je nach Anwendung auch eine variable Länge haben können. Diese werden innerhalb der AAL-Schicht in Informationsfelder von ATM-Zellen umgewandelt und

	Klasse A	Klasse B	Klasse C	Klasse D
Feste Zeitbeziehung zwischen Quelle und Ziel	besteht		besteht nicht	
Bitrate	konstant	variabel		
Verbindungsart	verbindungsorientiert			verbindungslos

Tabelle 2.1: Dienstklassen für die AAL-Schicht

an die ATM-Schicht weitergegeben. Auf der Empfangsseite werden aus den empfangenen Informationsfeldern die ursprünglichen Dateneinheiten wiedergewonnen und die Einflüsse des ATM-Netzes (Verzögerungsschwankungen, Zellverluste, Bitfehler) – falls erforderlich – ausgeglichen.

Die Funktionen der ATM-Anpassungsschicht können je nach Konfiguration in den Endsystemen, in den Netzabschlußeinheiten, in den Endgeräteanpassungen und in den Vermittlungsabschlüssen (engl. Exchange Termination, ET) implementiert sein. Zur Einbindung der Nicht-ATM-Teile des ISDN sind für die Anpassungsfunktionen sogenannte Netzwerk-anpassungen (engl. Network Adaptor, NA) vorgesehen.

Gemäß Tabelle 2.1 werden die Dienste in vier Klassen eingeteilt, wobei für jede Klasse eine bestimmte Kombination von Funktionen in der AAL-Schicht benötigt wird.

Die zur Klasse A gehörenden Dienste werden auch als CBR-Dienste (engl. Constant Bitrate Oriented) bezeichnet. Darunter versteht man solche Dienste, die einen ununterbrochenen digitalen Datenfluß benötigen wie z.B. Sprache mit 64 kbit/s PCM-Kodierung. Bei diesen Diensten müssen in abgehender Richtung die in regelmäßigen Abständen von den höheren Schichten eintreffenden Informationseinheiten aufgesammelt werden bis eine Zelle gefüllt werden kann. Auf der empfangenden Seite müssen Fehler, z.B. Zellverluste, erkannt und entsprechend behandelt werden. Außerdem muß die variable Verzögerung der einzelnen Zellen ausgeglichen und der Sendetakt wiederhergestellt werden, um ein synchrones Ausspielen der Informationseinheiten zu ermöglichen.

In die Klasse B fallen z.B. Video-, Audio- und Sprachdienste mit zeitlich veränderlicher Bitrate (engl. Variable Bitrate, VBR), für die die AAL-Schicht u.a. Mechanismen für eine Kompensation der Verzögerungsschwankungen und eine Fehlererkennung bzw. Fehlerkorrektur bereitstellen muß.

Unter die Klassen C und D fallen z.B. die bereits existierenden Datendienste. Dazu sind in der Benutzerebene verbindungslose und verbindungsorientierte Datendienste auf der Basis von Protokollen wie z.B. LLC (Logical Link Control [81]) oder X.25 [35] und in der Steuerungsebene die Signalisierprotokolle des D-Kanals auf der Basis von LAP D [32, 33] mit den



in Abschnitt 2.4.8 angesprochenen Erweiterungen zu zählen. Diese Dienste erfordern in der ATM-Anpassungsschicht Funktionen wie:

- Zerlegen der Informationsblöcke aus den höheren Schichten (variable Länge) in ATM-Zellen (engl. Segmenting)
- Zusammensetzen der Informationsblöcke aus den empfangenen ATM-Zellen (engl. Reassembly)
- Behandlung nicht vollständig gefüllter ATM-Zellen
- Kennung des Segment-Typs, d.h. Angabe ob es sich um Beginn, Fortsetzung oder Ende einer Dateneinheit der höheren Schichten oder um eine komplette Dateneinheit handelt (engl. Segment Type Identification)
- Fehlerbehandlung, z.B. Erkennung von Zellverlusten und Bitfehlern im Informationsfeld (engl. Error Recovery)
- Dienstekennung (engl. Service Indication), z.B. Angabe, ob eine verbindungsorientierte oder eine verbindungslose Kommunikation vorliegt, und
- Flußkontrolle (engl. Flow Control).

Für die Unterstützung der verbindungslosen Kommunikation werden zusätzlich zu den AAL-Funktionen der Klasse D noch weitere Funktionen in höheren Schichten benötigt. Ein Vorschlag hierzu ist der „Switched Multi-Megabit Data Service“ (SMDS [72]). Der SMDS ist ein paketvermittelter, verbindungsloser Dienst, der für Hochgeschwindigkeits-Datenübertragung über weite Strecken, z.B. zur Verbindung von lokalen Netzen oder MANs, geeignet ist. Der SMDS, der in den nächsten Jahren bereits im Bereich der MANs eingeführt werden wird, könnte in einem ATM-Netz etwa mit Hilfe von SMDS-Dienstmodulen realisiert werden, die in die Vermittlungsknoten integriert werden können [49].

Von der Funktionalität her kann die AAL-Schicht weiter unterteilt werden, und zwar in eine Konvergenz-Subschicht (Convergence Sublayer, CS) und in eine SAR-Subschicht (Segmentation and Reassembly). Der SAR-Subschicht werden dabei die für alle Dienste ähnlichen Funktionen für das Zerlegen und Zusammensetzen der verschiedenen langen Benutzer-Dateneinheiten zugeordnet, während in der CS-Schicht die zusätzlichen, mehr dienstabhängigen Funktionen zusammengefaßt werden. Die AAL-Schicht wird dabei insgesamt der OSI-Schicht 1 zugeordnet, wodurch von der OSI-Schicht 2 aufwärts die bestehenden Protokolle verwendet werden können. Die meisten dieser Protokolle führen in den höheren Schichten nochmals Funktionen wie z.B. abschnittsweise (engl. Link-By-Link) Fehlerbehandlung und Flußkontrolle aus. Dadurch sind diese Protokolle für Hochgeschwindigkeitsanwendungen nicht oder nur in gewissen Grenzen einsetzbar. Deshalb sind Dienstklassen vorstellbar, für die die ATM-Anpassungsschicht auch die Funktionen der OSI-Schichten 2, 3 und eventuell 4 umfassen wird (siehe auch [128]). Diese sogenannten „Streamline“-Protokolle führen die oben genannten Funktionen nur noch in einer Schicht und auf einer Ende-zu-Ende Basis aus, um den

gewünschten hohen Nutzdurchsatz sicherzustellen. Möglich werden solche vereinfachten Protokolle u.a. durch die kleinen Bitfehlerraten und die hohen Übertragungsgeschwindigkeiten optischer Übertragungssysteme.

Bestrebungen für weitgehende Vereinheitlichungen, wie z.B. eine Anpassung des Segment-Nutzlast-Formats an das ATM-Zellformat bei der Normung für Hochgeschwindigkeitsnetze (engl. Metropolitan Area Network, MAN) auf der Basis von DQDB (Distributed Queue Dual Bus [82]), vereinfachen die Funktionen der AAL-Schicht erheblich und ermöglichen eine Zusammenarbeit (engl. Interworking) verschiedener Netztypen mit Hilfe von relativ einfachen Netzkoppeleinheiten (engl. Gateway), wodurch sich die Akzeptanz der entsprechenden Netzkonzepte verbessern wird.

## 2.4.6 Zellformat

Die Trennung zwischen der ATM-Schicht und der ATM-Anpassungsschicht spiegelt sich in der Trennung der dem Kopffeld bzw. dem Informationsfeld zugeordneten Funktionen wieder. In der Empfehlung I.121 hat das CCITT einen Bereich von 3 bis 8 Oktett für das Kopffeld und einen Bereich von 32 bis 120 Oktett für das Informationsfeld festgelegt. Das genaue Format sollte unter Berücksichtigung der im folgenden dargestellten Kriterien erfolgen.

### 2.4.6.1 Format des Informationsfeldes

Ein sehr kurzes Informationsfeld bietet einige Vorteile, z.B.

- Kürzere Verzögerung durch das Auffüllen der Zellen bei Diensten mit niedrigen Bitraten wie z.B. Sprache. Da in diesen Fällen die Paketierung einen signifikanten Anteil an der Gesamtverzögerung ausmacht, wird durch kurze Zellen diese Verzögerung reduziert.
- Größere Reichweite ohne teure und technisch schwierige Echokompensation bei der Sprachübertragung. Störende Echos treten bei der Zusammenarbeit mit analogen zweidräftigen Systemen auf, wobei der störende Effekt mit der Übertragungsdauer der Zellen in Hin- und Rückrichtung (Round Trip Delay) zunimmt.
- Kleinerer Pufferspeicherbedarf im Netz für den Ausgleich der Verzögerungsschwankungen und zur Auflösung von Kollisionen.

Für möglichst lange Zellen sprechen dagegen

- Besseres Verhältnis von Nutz- zu Steuerinformation, da auch für kurze Zellen schon ein relativ langes Kopffeld benötigt wird. Weiterhin muß berücksichtigt werden, daß für die Funktionen der ATM-Anpassungsschicht noch Kapazität im Informationsfeld benötigt wird, wodurch der Anteil an Nutzinformation weiter reduziert wird.

- Mehr Zeit für die Bearbeitung der Steuerinformationen des Kopffeldes.
- Maximierung der Anzahl von Signalisiermeldungen, die nur eine Zelle für die Übermittlung benötigen.

Nach verschiedenen Vorschlägen für die Länge der Zellen (32 + 4, 64 + 5, 66 + 6) wurde in einer Sitzung der Studienkommission XVIII im Juni 1989 eine Einigung auf Zellen mit 48 Oktett Informationsfeld und 5 Oktett Kopffeld erzielt. Der zusätzliche Aufwand für die AAL-Schicht wird z.B. für die Unterstützung einer verbindungslosen Kommunikation 4 Oktett betragen. Dadurch beläuft sich der Overhead-Anteil in einer Zelle insgesamt auf etwa 17 %. Die Festlegung des Zellformats erfolgt in der Empfehlung I.361 „ATM Layer Specification for B-ISDN“ [38].

### 2.4.6.2 Format und Funktionen des Kopffeldes

Die Strukturierung des Kopffeldes unterscheidet sich an der Benutzer/Netzchnittstelle (engl. User/Network Interface, UNI) und an den netzinternen Schnittstellen (engl. Network Node Interface, NNI) wie in Bild 2.11 dargestellt durch ein am UNI vorhandenes Flußkontrollfeld (engl. Generic Flow Control, GFC). Dieses Feld soll Vielfachzugriffsprotokolle im Anschlußbereich, wie etwa das Medienzugriffsprotokoll von Anordnungen auf der Basis von DQDB [82], unterstützen und damit mehrpunktfähige Konfigurationen mit einem gemeinsamen Übertragungsmedium in Benutzerbereich (CPN) erlauben. Innerhalb des Netzes wird dieses Feld nicht benötigt und die entsprechenden Bits werden zur Erweiterung des eigentlichen Adreßfeldes verwendet.

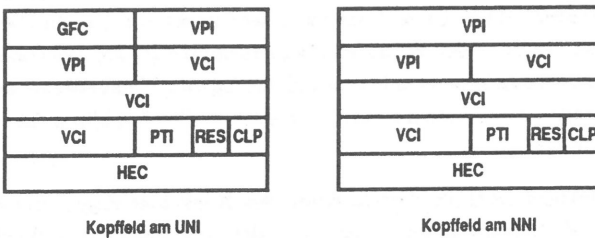


Bild 2.11: Die Struktur des Kopffeldes einer ATM-Zelle

Das Adreßfeld, auf dessen Basis die Zuordnung der Zelle zu einer virtuellen Verbindung vorgenommen wird, besteht aus einer Kennung für den virtuellen Kanal (Virtual Channel Identifier, VCI) und einer Kennung für den sogenannten „virtuellen Pfad“ (Virtual Path Identifier, VPI).

Der VCI kennzeichnet einen individuellen virtuellen Kanal für den z.B. die Erhaltung der Zellreihenfolge (engl. Cell Sequence Integrity) sichergestellt werden muß. Ein virtueller Kanal ist dabei als die logische, unidirektionale Verknüpfung zwischen den Endpunkten eines Übertragungsabschnitts definiert, die den Transfer von Zellen über diesen Abschnitt erlaubt. Die Zuordnung eines VCI erfolgt beim Verbindungsaufbau durch die für die Verbindungssteuerung zuständigen Instanzen. Am UNI und am NNI stehen für den VCI jeweils 16 Bit zur Verfügung.

Mit Hilfe des VPI können virtuelle Kanäle logisch zusammengefaßt und die zugehörigen Zellen als einheitlicher Zellstrom behandelt werden, der einem gemeinsamen virtuellen Pfad folgt [3, 38, 141]. Für einen virtuellen Pfad wird eine einheitliche Dienstgüte angegeben, die sich an den Anforderungen des anspruchsvollsten beteiligten Dienstes orientiert. Die Gruppen von Kanälen müssen in als „Virtual Path Terminator“ bezeichneten Einheiten zusammengesetzt und wieder in einzelne Kanäle aufgespalten werden. Anwendungen dieses zweistufigen Konzepts sind sowohl für den Benutzer als auch für den Netzbetreiber vorhanden, so z.B.

- Aufbau logisch getrennter privater Netze, wobei im öffentlichen Netz nur der VPI ausgewertet wird und der VCI transparent durchgeschaltet wird.
- Logische Trennung von Diensten mit verschiedenen Dienstgüteanforderungen unter Benutzung gemeinsamer Übertragungseinrichtungen.
- Vereinfachung und Beschleunigung der VerbindungsaufbauprozEDUREN, da die Wegesuche entfällt und nur Anfangs- und Endpunkte der virtuellen Pfade beteiligt sind.
- Umleitung der Zellströme auf alternative Wege bei Ausfällen im Netz auf der Basis der virtuellen Pfade (engl. Protection Switching).

Durch die Anwendung des VPI-Konzeptes können desweiteren reine Durchgangsvermittlungstellen zu sogenannten „ATM Cross Connects“ vereinfacht werden, die die Vermittlungsfunktionen nur auf der Basis der virtuellen Pfade ausführen [4, 38]. Virtuelle Pfade für die Nutzung durch Teilnehmer können entweder auf Antrag vom Netzbetreiber oder direkt über entsprechende Signalisierprozeduren eingerichtet werden. Für den VPI stehen am UNI 8 und am NNI 12 Bit zur Verfügung.

Die in der Empfehlung I.121 als Kernfunktion des Kopffeldes eingestufte Kennzeichnung von Leerzellen wird ebenfalls unter Verwendung des Adressfeldes verwirklicht, indem eine bestimmte, einheitlich festgelegte VPI/VCI-Kombination für die Leerzellen reserviert wird (VPI = 0, VCI = 0). Weitere reservierte VPI/VCI-Kombinationen sind z.B. für die Meta-Signalisierung (siehe Abschnitt 2.4.8) und für die Kennzeichnung von Zellen für Schicht-1-Wartungszwecke vorgesehen.

Für die Unterscheidung der in den Zellen enthaltenen Art von Nutzinformation steht ein „Payload Type Identifier“ (PTI) mit 2 Bit zur Verfügung. Zunächst war hierbei nur eine

Unterscheidung zwischen Benutzerdaten und internen Daten zur Netzsteuerung vorgesehen. Inzwischen werden jedoch auch andere Anwendungen diskutiert, wie z.B.

- Information über Überlastsituationen
- Markierung von Zellen, welche die der Verbindung zugeordnete Bandbreite überschritten haben (engl. Violation Tag)
- Kennzeichnung von Zellen für Betriebs- und Wartungsaufgaben.

Desweiteren enthält das Kopffeld ein Prioritätsbit (engl. Cell Loss Priority, CLP), mit dem angezeigt werden kann, welche Zellen bei Überlast im Netz bevorzugt verworfen werden können (CLP = 1). Dieses Bit kann vom Benutzer gesetzt und bei Bedarf vom Netz modifiziert werden.

Zusätzlich enthält das Kopffeld noch ein Bit, dessen Funktion noch nicht festgelegt ist und das für zukünftige Erweiterungen reserviert ist.

Das gesamte Kopffeld ist durch einen Fehlersicherungsmechanismus (engl. Header Error Control, HEC) geschützt, für dessen Prüfbits ein Oktett reserviert ist. Das dabei verwendete Generatorpolynom

$$x^8 + x^2 + x + 1$$

erlaubt die Korrektur eines einzelnen Bitfehlers und gleichzeitig das Erkennen von mehrfachen Bitfehlern. Dadurch werden die Zellverluste durch Fälschungen im Kopffeld sowie die Wahrscheinlichkeit für die Zuordnung der Zellen zu der falschen Verbindung (engl. Cell Insertion) reduziert. Das HEC-Feld wird, wie in Abschnitt 2.3.2 beschrieben, zur Erkennung der Zellgrenzen mitbenutzt (siehe auch CCITT Empfehlung I.432 [38, 165]).

## 2.4.7 Benutzer/Netzschnittstelle

In der Empfehlung I.121 [26] sind zwei Benutzer/Netzschnittstellen (UNI) mit verschiedenen Bitraten vorgesehen, von denen die eine im Bereich von 150 Mbit/s und die andere im Bereich von 600 Mbit/s liegen soll. Diese Schnittstellen sollen reine ATM-Schnittstellen sein und neben den Breitbanddiensten auch Schmalbanddienste unterstützen.

Die Gesamtbitrate am Bezugspunkt  $T_B$  wird 155,520 Mbit/s betragen, wobei 149,760 Mbit/s für die Nutzübertragung zur Verfügung stehen, was genau den Spezifikationen für das STM-1-Signal entspricht (siehe Abschnitt 2.3.1). Die Benutzer/Netzschnittstelle soll symmetrisch sein und an den Bezugspunkten  $T_B$  und  $S_B$  den Einsatz von Punkt-zu-Mehrpunkt-Konfigurationen erlauben. Die 600 Mbit/s-Schnittstelle soll eine Gesamtbitrate von 622,080 Mbit/s haben.

Die bei einem kanalorientierten Konzept grundlegend wichtige Festlegung der Kanalhierarchie für die Breitbandkanäle reduziert sich in der ATM-Umgebung auf die Angabe der maximalen unterstützten Bitrate pro Verbindung. Bei dieser noch nicht getroffenen Festlegung muß neben der Kapazität der Benutzer/Netz-Schnittstelle noch der zusätzliche Aufwand für Steuerung, Signalisierung, Betrieb und Wartung berücksichtigt werden.

## 2.4.8 Signalisierung

Im B-ISDN wird die Signalisierinformation in eigenen virtuellen Verbindungen (engl. Signalling Virtual Channel, SVC) getrennt von der Nutzinformation übermittelt, wobei ein Benutzer mehrere getrennte SVCs gleichzeitig unterhalten können soll. Für die Signalisierverbindungen steht eine gewisse Menge von logischen Kanalnummern (engl. Signalling VCI, SVCI) zur Verfügung. Für mehrpunktfähige Schnittstellenkonfigurationen müssen neben den Punkt-zu-Punkt-Signalisierverbindungen auch Möglichkeiten zur „Broadcast“-Signalisierung, z.B. für den Aufbau von Verbindungen auf der gerufenen Seite, vorgesehen werden. Im Prinzip sollen zur Benutzer/Netz-Signalisierung weiterhin die in der I.441 [33] und in der I.451 [34] festgelegten Protokolle für die Schichten 2 und 3 benutzt werden. Durch die Verwendung des Konzepts der virtuellen Verbindungen und durch die neuen Anforderungen der Breitbanddienste sind Erweiterungen nötig, wie sie an Beispielen im folgenden beschrieben werden.

Falls an einer Schnittstelle keine dauernd bestehenden virtuellen Signalisierverbindungen vorhanden sind, wird für den Aufbau, die Unterhaltung und den Abbau dieser Verbindungen eine sogenannte „Meta-Signalisierung“ benötigt [38, 126, 147]. Zu diesem Zweck wird an der Benutzer/Netzschnittstelle eine bestimmte VPI/VCI-Kombination reserviert, die überall gleich ist. Die bei der Meta-Signalisierung verwendeten Prozeduren zur Steuerung von SVCs können in ähnlicher Weise definiert werden wie diejenigen für die Vergabe der Schicht-2-Adresse (Terminal Endpoint Identifier, TEI) im D-Kanal-Protokoll [33], da dort ähnliche Anforderungen gestellt werden. Im einzelnen werden benötigt [147]:

- Eine Prozedur zum Aufbau der Signalisierverbindungen, in der z.B. die SVCI's zugeteilt werden
- Eine Prozedur mit der überprüft werden kann, ob eine bestehende Signalisierverbindung noch benötigt wird
- Eine Prozedur zum Abbau von Signalisierverbindungen, in der z.B. die SVCI's wieder freigegeben werden.

Für die Signalisierung – vor allem in der Schicht 3 – werden durch die erweiterten Dienstkonzepte einige neue Funktionen notwendig [18, 113]. Für Multi-Media-Gespräche muß eine Beschreibung der einzelnen Dienstelemente (siehe 2.4.1) und ihrer Zuordnung zu einem Gespräch ermöglicht werden. Da für die einzelnen Dienstelemente eventuell getrennte, aber

miteinander synchronisierte Netzwerkverbindungen nötig sind und gewisse Dienstkomponenten nicht notwendigerweise während der gesamten Gesprächsdauer bestehen, müssen Meldungen und Prozeduren eingeführt werden, die den Auf- und Abbau korrelierter Netzwerkverbindungen innerhalb eines Gesprächs erlauben. Diese Möglichkeiten müssen auch für Gespräche mit Punkt-zu-Mehrpunkt- und Mehrpunkt-zu-Mehrpunkt-Konfigurationen unterstützt werden (siehe auch CCITT I.311 „B-ISDN General Network Aspects“ [38]).

Auf die Zwischenamtssignalisierung können durch zukünftige Entwicklungen, wie etwa die zunehmende Mobilität der Teilnehmer, ebenfalls neue Aufgaben zukommen. Die Mobilität der Teilnehmer ist dadurch gekennzeichnet, daß die Rufnummern individuellen Teilnehmern und nicht mehr ortsfesten Endgeräten zugeordnet werden. Solche „persönlichen“ Nummern können dann in irgendeiner Form mitgenommen und von beliebigen Endgeräten aus benutzt werden. Um solche Konzepte verwirklichen zu können, werden verteilte Datenbanken – ähnlich wie in Mobilfunknetzen [167] – zur Speicherung der teilnehmerbezogenen Daten benötigt. Die charakteristischen Daten (Berechtigungen, ...) des Teilnehmers werden in der „Heimatvermittlungsstelle“ gespeichert und bei Bedarf in die Vermittlungsstelle im jeweiligen Aufenthaltsbereich übertragen. Umgekehrt müssen z.B. Daten über den jeweiligen Aufenthaltsort oder Gebühreninformationen über die Zwischenamtssignalisierung zur Heimatvermittlungsstelle übermittelt werden. Diese intensive Kommunikation zwischen den Datenbanken kann, wenn man von einer großen Zahl von mobilen Teilnehmern ausgeht, eine erhebliche Last für das Signalisiersystem bedeuten [4].

## 2.5 Prinzipielle Architektur einer ATM-Vermittlungsstelle

Bedingt durch die hohen Durchsatz-, Geschwindigkeits- und Verzögerungsanforderungen (Verzögerung pro Vermittlungsstelle im Bereich von  $\leq 1$  ms) muß für den Aufbau einer ATM-Vermittlungsstelle eine verteilte Struktur mit der Möglichkeit zu konsequenter Parallelverarbeitung gewählt werden. Darüberhinaus müssen möglichst viele Funktionen mit anwendungsspezifischen integrierten Spezialbausteinen implementiert werden, da die Ausführung entsprechender Programme in vielen Fällen zu lange dauern würde. Die prinzipiellen Funktionseinheiten einer ATM-Vermittlungsstelle sind in Bild 2.12 dargestellt.

Die Leitungsmodul schließen die Übertragungsabschnitte in kommender und in gehender Richtung ab und führen die abschnittsbezogenen Protokollfunktionen (PM- und TC-Schicht, ATM-Schicht, siehe Abschnitt 2.4.5) aus. Dazu gehört neben der Synchronisation und der Fehlerkontrolle für das Kopffeld auch die Umsetzung der logischen Kanalnummern (VCI/VPI) auf die Werte, die der jeweiligen Verbindung auf dem abgehenden Übertragungsabschnitt

zugeordnet sind. Dafür werden lokale Zuordnungstabellen verwendet, die beim Verbindungsaufbau geladen werden.

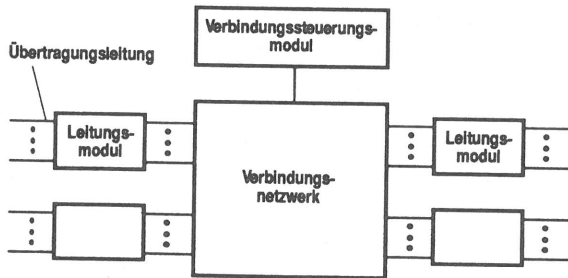


Bild 2.12: Prinzipielle Funktionseinheiten einer ATM-Vermittlungsstelle

Wenn das Verbindungsnetzwerk im sogenannten „Self Routing“ oder „Digit Controlled Routing“-Mode betrieben wird, dann wird in den Leitungsmodulen die lokale Steuerinformation für das Auffinden des richtigen Verbindungsnetzausgangs hinzugefügt bzw. in abgehender Richtung wieder entfernt. Dabei werden ebenfalls entsprechende Zuordnungstabellen verwendet, so daß auch diese Funktion mit Hilfe von schneller Spezialhardware realisiert werden kann. Weitere Funktionen der Leitungsmodule umfassen – je nach Architektur des Vermittlungsknotens –

- Die Abbildung der ATM-Zellen auf das intern verwendete Datenformat, wobei Seriell-Parallel-Wandlungen und weitere Manipulationen (z.B. Zerlegen der Zellen in kleinere Einheiten) vorgenommen werden können
- Die Anpassung an die knoteninterne Taktfrequenz, die zur Verminderung von Blockiereffekten erhöht sein kann
- Die Zwischenspeicherung von Zellen bei Blockierungen, um Zellverluste zu vermindern
- Das Wiederherstellen der korrekten Zellreihenfolge innerhalb einer Verbindung, wenn mehrere Wege zwischen bestimmten Ein- und Ausgängen möglich sind und die Betriebsweise der Verbindungsnetzwerks ein gegenseitiges Überholen von Zellen erlaubt.

Im Gegensatz zu den Leitungsmodulen, die auf der Zeitebene der ATM-Zellen arbeiten (Dauer einer ATM-Zelle:  $2,8\mu\text{s}$  bei  $150\text{ Mbit/s}$ ,  $0,7\mu\text{s}$  bei  $600\text{ Mbit/s}$ ), bearbeitet das Verbindungssteuerungsmodul nur Vorgänge auf der Verbindungsebene. Da die Funktionen der Verbindungssteuerung überdies wesentlich komplexer und vielseitiger sind, kommen hier Softwarelösungen zur Anwendung. Die Hauptaufgabe der Verbindungssteuerung besteht darin, beim Verbindungsaufbau geeignete abgehende Übertragungsabschnitte mit ausreichend freier Kapazität für die neue Verbindung und entsprechende Wege durch das Verbindungsnetzwerk



zu finden (engl. Routing). Für erfolgreich aufgebaute Verbindungen müssen dann die Verbindungszustandsspeicher sowie die Tabellen in den Leitungsmodulen modifiziert werden. Bei Änderungen innerhalb der Verbindungen und beim Verbindungsabbau müssen die Tabellen jeweils aktualisiert werden. Neben den direkt verbindungsbezogenen Aufgaben können diesem Funktionsblock auch Betriebs- und Wartungsaufgaben zugeordnet werden.

Das Verbindungsnetzwerk (engl. Interconnection Network, Switch Fabric) selbst weist in der Regel eine verteilte Steuerung auf, um durch die simultane Vermittlung vieler Zellen die Erfüllung der gestellten Durchsatzanforderungen zu erlauben. Die Verbindungsnetzwerke sind meist mehrstufig, da die Größe der eigentlichen Basismodule (Koppelemente, engl. Switching Element) durch technologische Randbedingungen auf etwa 32 Ein- und Ausgänge beschränkt ist. Bei den Koppelementen kann man prinzipiell unterscheiden, ob sie Pufferspeicher zur Reduzierung von Zellverlusten im Kollisionsfall enthalten oder nicht. Da die Kollisionen nicht nur durch innere Blockierungen sondern auch durch prinzipbedingte, kurzzeitige Überlastsituationen auf den abgehenden Übertragungsabschnitten (siehe 2.3.2) verursacht werden, müssen bei ungepufferten Verbindungsnetzwerken Pufferspeicher an den Eingängen oder an den Ausgängen des Verbindungsnetzwerks vorgesehen werden, auch wenn das Verbindungsnetzwerk intern blockierungsfrei ist.

Für die Realisierung der Koppelemente gibt es ein breites Spektrum an Möglichkeiten, die z.B. in [109, 135] dargestellt und bezüglich Leistungsfähigkeit und anderer Kriterien verglichen werden. Unter den möglichen Strukturen sind u.a. solche auf Matrix-, Bus- oder Ringbasis sowie solche mit einem zentralen Speicher, die von den in der synchronen Zeitmultiplextechnik verwendeten Zeitstufen abgeleitet sind. Die Leistungsfähigkeit gepufferter Koppelemente hängt neben ihrer Größe vor allem von der Platzierung der Pufferspeicher ab. Untersuchungen [74, 89, 134] zeigen, daß eine - physikalische oder logische - Zuordnung der Speicher zu den Ausgängen des Koppelements günstig ist, da dann alle internen Blockiereffekte vermieden werden können und nur noch die unvermeidbaren Blockierungen durch die Abnehmerleitungen auftreten. Ebenfalls vorteilhaft ist es, wenn zusätzlich die Puffer von möglichst vielen Ein- und Ausgängen gemeinsam benutzt werden können, weil dann aufgrund statistischer Effekte eine Reduzierung des Gesamtspeicherbedarfs erzielt werden kann.

Die aus den Koppelementen aufgebauten mehrstufigen Verbindungsnetzwerke kann man von der Topologie her in solche mit nur einem Weg zwischen einem Eingangs-/Ausgangs-Paar (engl. Single Path Networks) und in solche mit mehreren möglichen Wegen (engl. Multiple Path Networks) unterteilen [112, 135].

Die von der Architektur für Multiprozessorsysteme her bekannte, sehr allgemeine Klasse von Single Path oder „Banyan“ Netzwerken [62] kann weiter unterteilt werden [78, 152], wobei die wichtigste Unterklasse die der „Delta“-Netzwerke [125] ist, der die meisten bekannten Implementierungen angehören. Banyan-Strukturen, die in einigen Prototyp-Implementierungen für paketorientierte Hochgeschwindigkeits-Vermittlungsstellen verwendet wurden [5],

haben den Nachteil, daß sie bereits bei relativ geringer Last starke interne Blockierungen aufweisen. Diese Blockierungen werden hauptsächlich durch die Anzahl und Plazierung der Pufferspeicher, durch die Größe der einzelnen Kopelemente sowie durch die Anzahl der Stufen im Verbindungsnetzwerk beeinflusst [78, 152]. Zur Vermeidung übermäßiger Zellverluste muß die Auslastung der internen Zwischenleitungen in Banyan-Strukturen weit unterhalb der theoretischen Durchsatzgrenze liegen. Die dafür erforderliche Aufweitung kann z.B. durch eine interne Geschwindigkeitserhöhung, mehrfache Zwischenleitungsführung oder Lastteilung zwischen mehreren parallelen Verbindungsebenen [99] erreicht werden. Durch das Vorhandensein eines eindeutigen Weges reduziert sich die Wegesuche auf die Bestimmung der Adresse des gewünschten Ausgangs.

Die Strukturen mit mehrfacher Wegemöglichkeit weisen weniger oder gar keine internen Blockierungen und i.a. eine höhere Fehlertoleranz auf [2], dafür werden jedoch wesentlich komplexere Wegesuchalgorithmen benötigt [9]. Die einfachste Art eine mehrfache Wegemöglichkeit zu schaffen, ist die Erweiterung einer Banyan-Struktur um zusätzliche Stufen mit gleicher Funktionalität, wodurch z.B. die sogenannten „Benes“-Netzwerke [13] entstehen. Die zusätzlich vorhandenen Stufen können jedoch auch andere Funktionen erfüllen, z.B. eine gleichmäßige Verteilung des Verkehrs bewirken [5, 21] oder das Sortieren der ankommenden Zellen anhand der lokalen Zieladressen [5, 76]. Im letzteren Fall kann die nachfolgende Banyan-Struktur blockierungsfrei betrieben werden, sofern nicht zwei Zellen für den gleichen Ausgang bestimmt sind. Neben diesen erweiterten Banyan-Strukturen eignen sich auch die aus der klassischen Durchschaltvermittlung bekannten Koppelnetzstrukturen, wie z.B. Umkehrgruppierungen oder Clos-Netzwerke [41] für den Einsatz in paketorientierten Breitbandvermittlungstellen.

Neben der Topologie werden die Verbindungsnetzwerke durch ihre Betriebsweise charakterisiert. In einer ATM-Umgebung mit Zellen konstanter Länge bietet sich ein synchroner Betrieb an, bei dem eine zellsynchrone, getaktete Weitergabe von Stufe zu Stufe erfolgt. Bei der Kopplung der einzelnen Stufen besteht die Möglichkeit, einen Rückstaumechanismus (engl. Backpressure Mechanism) zu implementieren, der die Steuerung eines Kopelements informiert, wenn die nächste Stufe keine Zelle aufnehmen kann [160]. Dadurch werden Zellverluste innerhalb des eigentlichen Verbindungsnetzwerks verhindert und an dessen Eingang verlagert. Dort können Zellverluste – wenn nötig – durch zusätzliche Pufferspeicher abgefangen werden [134].

Bezüglich der Steuerinformation für den Transport durch das Verbindungsnetzwerk existiert neben dem Self-Routing-Konzept, bei dem die komplette lokale Steuerinformation bereits vor dem Verbindungsnetzwerk bestimmt und hinzugefügt wird, noch die Möglichkeit des sogenannten „VCI-Routing“. Bei dieser Art von Durchschaltung wird das eigentliche Adressfeld der Zelle in jedem Kopelement ausgewertet und eventuell umgesetzt.

Die ATM-Vermittlungsknoten müssen in der Lage sein, Verteildienste auf vermittelter

Basis zu unterstützen. Dabei werden im Netz dynamisch Verteilbäume aufgebaut, damit auf jedem Übertragungsabschnitt nur eine Kopie der jeweiligen Zellen übertragen werden muß [162]. Für dieses Konzept müssen die Vermittlungsknoten und eventuell auch die anderen Netzkomponenten, wie z.B. Multiplexer, in der Lage sein, ankommende Zellen von Verteilverbindungen zu kopieren und die einzelnen Kopien in verschiedene Richtungen weitzuschicken. Im Verbindungsnetzwerk kann die Realisierung so aussehen, daß zusätzliche, dedizierte Stufen für die Kopierfunktion eingesetzt werden (Copy Network [163]). Bei bestimmten Koppellement-Architekturen, wie z.B. bei solchen mit Zentralspeicher, läßt sich die Kopierfunktion auch durch Modifikationen in der Koppellementsteuerung ohne weiteren zusätzlichen Aufwand realisieren.

## **2.6 Betrieb von ATM-Breitbandnetzen**

Von den vielfältigen Aspekten des Betriebs von ATM-Netzen sollen in diesem Abschnitt nur diejenigen dargestellt werden, die einen unmittelbaren Bezug zu den im weiteren Verlauf der Arbeit dargestellten Mechanismen haben. Angaben über die nicht dargestellten Betriebs- und Wartungsaspekte sind z.B. in der Empfehlung I.610 „B-ISDN OAM Principles“ zu finden.

### **2.6.1 Verbindungsannahme und Wegesuche**

Der Hauptvorteil von ATM-Netzen liegt in der Flexibilität, die durch die paketorientierte Übermittlung in Kombination mit dem asynchronen Multiplexverfahren erreicht wird. Dadurch ergibt sich aber auch, wie schon erwähnt, eine gegenseitige Beeinflussung der bestehenden Verbindungen, die z.B. durch geeignete Maßnahmen bei der Verbindungsannahme kontrolliert werden muß.

Neben der Zuteilung der maximal benötigten Bandbreite für die gesamte Verbindungsdauer (engl. Peak Rate Allocation), wie sie auch bei durchschaltvermittelten Netzen angewendet wird, gibt es in einem ATM-Netz bei Verbindungen mit schwankendem Bandbreitebedarf auch die Möglichkeit, der Verbindung nur eine „effektive Bandbreite“ zuzuordnen. Diese effektive Bandbreite liegt zwischen der mittleren und der maximalen benötigten Bandbreite und muß unter Berücksichtigung der Verkehrscharakteristik der Verbindung bestimmt werden [26, 53, 59]. Durch diese Art der Bandbreitezuteilung kann – abhängig von Art und Mischungsverhältnis der Verbindungen – ein Multiplexgewinn und damit eine bessere Auslastung der Netzressourcen erzielt werden. Dabei muß jedoch darauf geachtet werden, daß die hierbei bewußt in Kauf genommene Verschlechterung der Dienstgüte in einem tolerierbaren Bereich bleibt.

Deshalb kommt der Funktion, die über die Annahme einer neuen Verbindung entscheidet (engl. Call Acceptance Control, CAC), eine große Bedeutung zu. Sie muß die Anzahl gleichzeitig bestehender Verbindungen so begrenzen, daß die Dienstgüteanforderungen aller dieser Verbindungen eingehalten werden können. Dazu muß in der CAC-Funktion der aktuelle Lastzustand der von der neuen Verbindung betroffenen Netzressourcen (Übertragungsabschnitte, Vermittlungsknoten, etc.) ermittelt werden. Dies kann z.B. durch laufende Messungen geschehen, wobei die hohen Geschwindigkeiten der Vorgänge und die zu erwartende große Dynamik bei der Änderung der Lastzustände eine sinnvolle Implementierung sehr schwierig macht.

Eine andere Möglichkeit besteht darin, aufgrund der während des Verbindungsaufbaus ausgehandelten Parameter der Verkehrsquelle (siehe 2.4.1) mit statistischen Methoden den Lastzustand und damit die Dienstgüteparameter der verschiedenen Verbindungen näherungsweise zu ermitteln. Ein solches Vorgehen erfordert im einzelnen:

- die ausreichend genaue Charakterisierung der verschiedenen, dienstspezifischen Verkehrsquellen mit wenigen Parametern,
- die Charakterisierung des Überlagerungsverkehrs, der von Typ und Mischung der Einzelverkehre bestimmt wird,
- die realistische Modellierung der Netzressourcen und
- die Berechnung der verbindungsindividuellen Dienstgüteparameter.

Eine Randbedingung für die CAC-Funktion ist, daß die Entscheidung in Echtzeit getroffen werden muß. Deshalb muß der Entscheidungsalgorithmus möglichst einfach sein oder aber die Werte müssen vorher berechnet und in mehrdimensionalen Tabellen abgelegt werden. Da im Prinzip pro Verkehrstyp eine Dimension benötigt wird, wird dieses Verfahren schnell sehr aufwendig.

Neben der Kontrolle der Dienstgüte innerhalb einer Verbindung fällt der CAC-Funktion auch die Aufgabe zu, die Abweiswahrscheinlichkeiten auf Verbindungsebene zwischen den verschiedenen Diensten auszugleichen, da sonst hochratige Verbindungen im Verhältnis sehr hohe Blockierungen aufweisen [53, 79]. Im Prinzip können hierzu Mechanismen eingesetzt werden, wie sie von der Mehrkanaldurchschaltung in kanalorientierten Netzen her bekannt sind [90, 79].

Verschiedene Konzepte und Methoden für die Bandbreitenzuteilung in ATM-Netzen sind z.B. in [53, 57, 58, 59, 80, 87, 123, 132, 142] ausführlich beschrieben.

Während des Verbindungsaufbaus wird auch der Weg festgelegt, den alle Zellen einer bestimmten Verbindung durch das Netz nehmen (Wegesuche, engl. Routing), wobei aus den zur Verfügung stehenden, geeigneten Wegen nach bestimmten Kriterien der optimale ausgewählt werden muß. Neben den bekannten Lastkriterien [53] können in ATM-Netzen auch

andere Aspekte eine Rolle spielen. So kann es z.B. ein Ziel des Wegesuchalgorithmus sein, möglichst viele Verbindungen mit gleichartiger Verkehrscharakteristik auf eine bestimmte Übertragungsleitung zu leiten, weil dadurch der erzielbare Multiplexgewinn maximiert werden kann (siehe z.B. [123]).

## 2.6.2 Quellflußkontrolle während der Verbindung (Policing)

Bedingt durch das asynchrone Multiplexprinzip könnte eine Verkehrsquelle – abweichend von den beim Verbindungsaufbau vereinbarten Werten – mit einer Rate senden, die prinzipiell nur durch die Kapazität der Benutzer/Netzschnittstelle begrenzt ist und damit die Dienstgüte aller Verbindungen drastisch verschlechtern. Dies ist bei einem ATM-Netz – im Gegensatz zu herkömmlichen Paketnetzen auf der Basis von X.25 [35] – möglich, weil keine Flußkontrollprotokolle, z.B. Fenstermechanismen (engl. Window Flow Control), zwischen dem Netz und dem Benutzer vorgesehen sind, mit deren Hilfe die Senderate der Quelle auf einen vom Netz bestimmten Wert begrenzt werden kann.

Gründe für ein von den getroffenen Vereinbarungen abweichendes Verhalten können neben Fehlfunktionen auch Böswilligkeit oder – weit wahrscheinlicher – mögliche ökonomische Vorteile sein. Die Einhaltung der vereinbarten Werte muß deshalb während der gesamten Verbindungsdauer durch die sogenannte „Quellflußkontroll“-Funktion (engl. Policing oder Usage Parameter Control) überwacht werden, wie sie in der Empfehlung CCITT I.311 „B-ISDN General Network Aspects“ [38] definiert wird. Die Quellflußkontrollfunktion muß in Echtzeit ausgeführt werden. Damit das gesamte Netz geschützt wird, muß sie so nahe wie möglich am Rand des Netzes angesiedelt werden und kann nicht zentralisiert werden, um eine effektive Auslastung der entsprechenden Einheiten zu erreichen. Als Konsequenz daraus muß der Algorithmus für die Quellflußkontrollfunktion relativ einfach und mit wenig Aufwand in Spezialhardware zu implementieren sein, weil die Funktion an jeder Benutzer/Netzschnittstelle mehrfach vorhanden sein muß. Die Quellflußkontrollfunktion kann auf der Basis einzelner virtueller Kanäle, virtueller Pfade oder insgesamt für den auf einem Übertragungsabschnitt transportierten Verkehr ausgeführt werden. Weil sie das Netz gegen ein Fehlverhalten der Benutzer schützen soll, muß sich die Quellflußkontrollfunktion unter der ausschließlichen Kontrolle des Netzbetreibers befinden. Als mögliche Stelle für die Implementierung der Quellflußkontrollfunktion kommen – je nach Anschlußkonfiguration – Netzelemente wie z.B. Multiplexer, Konzentratoren, Ortsvermittlungsstellen oder auch Durchgangsvermittlungsstellen in Frage. Quellflußkontrollfunktionen können auch im Hoheitsbereich des Teilnehmers, z.B. in ATM-Nebenstellenanlagen, implementiert sein, der Netzbetreiber darf sich darauf jedoch nicht verlassen. Eine Funktion mit den gleichen Funktionen muß auch an den Übergangsstellen zwischen den Subnetzen verschiedener Betreiber vorgesehen werden, da kein Betreiber den Schutz seines Netzes anderen überlassen kann.

Die für die Charakterisierung der Verkehrsquellen beim Verbindungsaufbau vorgesehenen – und deshalb von der Quellflußkontrollfunktion zu überwachenden – Parameter sind (siehe CCITT-Empfehlung I.311 [28])

- die mittlere Bitrate der Verbindung,
- die maximale Bitrate der Verbindung und
- die Dauer der Phasen mit hoher Aktivität bei Diensten mit zeitlich variabler Bitrate.

Die erste Hauptaufgabe der Quellflußkontrollfunktion liegt darin, Abweichungen von den vereinbarten Werten zu erkennen. Die Schwierigkeit liegt dabei darin, die normalen statistischen Schwankungen im Quellverhalten von wirklichen Parameterverletzungen zu unterscheiden. Im letzteren Fall muß die Quellflußkontrollfunktion Maßnahmen ergreifen, die dazu geeignet sind, die anderen Verbindungen vor Zellverlusten oder übermäßigen Verzögerungen durch Überlastung von Puffern im Netz zu schützen. Möglichkeiten für solche Maßnahmen sind unter anderem:

- *Das Verwerfen von Zellen, bei denen eine Abweichung erkannt wurde.* Durch die statistische Natur der Verkehrsquellen und die eventuell unzureichende Beschreibung mit einigen wenigen Parametern kann es vorkommen, daß auch Zellen von solchen Quellen verworfen werden, die die vereinbarten Parameter einhalten. Die so entstehenden Zellverluste müssen vom Netz verantwortet werden und deshalb auf Werte beschränkt werden, die um Größenordnungen unter den entsprechenden Dienstgüteanforderungen des jeweiligen Dienstes liegen. Dieses Verfahren wird in Kapitel 5 näher untersucht.
- *Das Markieren von Zellen, bei denen eine Abweichung erkannt wurde.* Diese markierten Zellen werden dann im Netz bevorzugt verworfen, wenn an irgendeiner Stelle eine kritische Lastsituation entsteht [55]. Dies setzt voraus, daß im Netz Prioritätsmechanismen implementiert sind, die zwischen markierten und nichtmarkierten Zellen unterscheiden können. Ein weiterer Nachteil ist, daß Zellen, die irgendwo im Netz verworfen werden, bis dorthin eine „Blindlast“ darstellen. Bei geringer Netzauslastung, z.B. nachts, kann fast ohne Verluste ein erheblich höherer Durchsatz als vereinbart erzielt werden. Die Verbindungsgebühren können deshalb nur bedingt auf der Basis der vereinbarten Parameter erhoben werden und müssen das tatsächlich gesendete Datenvolumen mit berücksichtigen.
- *Das Verzögern von Zellen, bei denen Abweichungen erkannt wurden.* Bei diesem Verfahren [93] wird zwar die Verkehrscharakteristik von büschelförmigen Verkehren geglättet, die End-zu-End-Verzögerung steigt jedoch an, was vor allem bei Echtzeit-Diensten problematisch sein kann. Außerdem bedeuten die für diese Methode erforderlichen Pufferspeicher einen erheblichen Kostenaufwand.
- *Das Informieren der Verbindungssteuerung.* Diese kann dann z.B. die Verbindung abrechnen oder erhöhte Verbindungsgebühren in Rechnung stellen (engl. Punitive

Charging). Hierbei sind die Reaktionszeiten jedoch so lang, daß ein Schutz der aktuell bestehenden Verbindungen nicht gewährleistet ist.

Eine weitere Aufgabe der Quellflußkontrollfunktion ist die Kontrolle, ob die in den jeweiligen Zellen verwendeten Adreßfelder gültig sind.

### 2.6.3 Dienstgüteklassen und Prioritäten

Die in einem B-ISDN zu erwartenden Dienste stellen sehr unterschiedliche Anforderungen an die Dienstgüte des Netzes. Paketierte Sprache z.B. ist relativ robust gegenüber Zellverlusten, wogegen sie durch Verzögerungsschwankungen (engl. Delay Jitter) zwischen den einzelnen Zellen, die nicht ausgeglichen werden, relativ schnell unverständlich wird. Auch relativ große Werte für die absolute Verzögerung müssen vermieden werden, da sich diese – ähnlich wie bei heutigen Satellitenverbindungen – negativ auf die Konversation auswirken. Datendienste hingegen sind empfindlicher gegen Informationsverluste und toleranter gegen Verzögerungen und Verzögerungsschwankungen. Bestimmte Kodieralgorithmen für Bewegtbildkommunikation erfordern wiederum sowohl geringe Verzögerungen als auch eine sehr kleine Zellverlustwahrscheinlichkeit. Bei einer Gleichbehandlung aller Zellen im Netz müßte die Dimensionierung der Netzressourcen auf die Anforderungen des empfindlichsten Dienstes (mit nennenswertem Verkehrsaufkommen) ausgerichtet werden, was eine sehr geringe Auslastung der Netzkomponenten zur Folge hätte. Um den Vorteil der Flexibilität konsequent ausnutzen zu können, wird allgemein die Notwendigkeit für mehrere Dienstklassen mit verschiedenen Anforderungen an die Übermittlungsqualität (engl. Network Performance) anerkannt [37], die grob eingeteilt werden könnten in

- Dienste mit konstanter Bitrate (engl. Constant Bit Rate Service, CBR), bei denen die Verfügbarkeit der Bandbreite garantiert wird und
- Dienste mit variabler Bitrate (engl. Variable Bit Rate Service, VBR), die
  - empfindlich gegen Verluste und Verzögerungen sind
  - empfindlich gegen Verluste / unempfindlich gegen Verzögerungen sind
  - unempfindlich gegen Verluste / empfindlich gegen Verzögerungen sind
  - unempfindlich gegen Verluste und Verzögerungen sind.

Bei Bedarf kann eine Verfeinerung der obigen Klassen in verschiedene Empfindlichkeitsstufen vorgenommen werden.

Für die Realisierung der verschiedenen Klassen innerhalb des Netzes kann z.B. eine Trennung der für die jeweiligen Klassen benutzten Wege erfolgen (Route Separation). Diese Wege können dann je nach den Dienstgüteanforderungen mehr oder weniger ausgelastet werden.

In den Netzkomponenten selbst ergeben sich durch diese Maßnahme keine Änderungen, da sie allein von der Verbindungssteuerung realisiert werden kann. Dieses Verfahren arbeitet nur dann zufriedenstellend, wenn die Verkehrsanteile der verschiedenen Klassen bekannt sind und die gesamten vorhandenen Netzressourcen entsprechend aufgeteilt werden [97].

Eine weitere Möglichkeit der Realisierung liegt in der Einführung von verschiedenen Prioritäten auf Verbindungs- oder Zellebene [69, 97, 123, 150]. Bei einer Prioritätsvergabe auf Verbindungsebene werden die Prioritäten je nach Dienst an alle Zellen der Verbindung vergeben, bei der Prioritätsvergabe auf Zellebene können die Zellen einer Verbindung verschiedene Prioritäten erhalten. Dies ist z.B. bei Videocodecs sinnvoll, die die Signale in den Frequenzbereich transformieren und vor der Übertragung in mehrere Spektralanteile zerlegen, die dann getrennt übertragen werden (Subband Coding). Da die verschiedenen Spektralanteile für die subjektive Bildqualität unterschiedlich wichtig sind, bietet sich hier die Vergabe verschiedener Prioritäten innerhalb einer Verbindung an. Diese Art der Prioritätsvergabe wird durch die Einführung des CLP-Bits im Kopffeld (siehe Abschnitt 2.4.6.2) unterstützt.

Die Zellverlustwahrscheinlichkeit kann durch sogenannte „räumliche Prioritäten“ (engl. Spatial Priorities) beeinflusst werden. Bei diesen Verfahren sind die Pufferspeicher im Netz entweder nur zu einem Teil für die Zellen mit niedriger Priorität zugänglich (Partial Buffer Sharing [97]) oder die Zellen mit der höheren Priorität sind in der Lage, niederprioritäre Zellen aus den Pufferspeichern zu verdrängen (Push-Out Mechanisms [69, 150]). Bei den zuletzt genannten Mechanismen muß darauf geachtet werden, daß bei der Verdrängung die Reihenfolge der Zellen innerhalb einer Verbindung erhalten bleibt.

Zur Beeinflussung der Verzögerungen und Verzögerungsschwankungen können im Prinzip die zeitlichen Prioritäten verwendet werden, wie sie schon seit langem bekannt sind. Da die durch die Pufferung im Netz entstehenden Verzögerungen in der Regel gegenüber den durch die Paketierung/Depaketierung bzw. durch die Signallaufzeiten verursachten Verzögerungen ziemlich klein sind, ist die Auswirkung dieser Verfahren relativ gering.

Denkbar sind für die oben definierten Dienstgüteklassen auch Kombinationen von räumlichen und zeitlichen Prioritäten.

Durch die verschiedenen Prioritätsmechanismen wird bei der Steuerungslogik der Pufferspeicher ein zusätzlicher Aufwand nötig, und es muß geprüft werden, ob die höhere Auslastung der Netzressourcen einen entsprechenden Vorteil bietet.

## **2.6.4 Zusammenhang zwischen den Verkehrssteuerungsmechanismen**

Die oben erläuterten Komponenten der Verkehrssteuerung sowie die Überlastabwehr-Mechanismen im Netz und die Vergebühung hängen sehr stark voneinander ab. Wie in Bild 2.13



dargestellt, bildet die reale Verkehrsquelle die Basis für alle diese Operationen. Für diese Quelle muß vom Netz eine bestimmte Übermittlungsqualität garantiert werden, damit unter Berücksichtigung der Funktionalität der höheren Protokollschichten dem Benutzer die erforderliche Dienstgüte bereitgestellt werden kann.

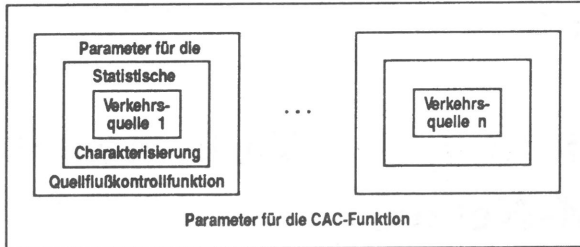


Bild 2.13: Zusammenhang zwischen den Verkehrssteuerungsmechanismen

Die erste Aufgabe ist dann die Charakterisierung des Quellverhaltens durch einen Satz von wenigen, relevanten Parametern. Diese Parameter und ihre jeweiligen Werte werden dem Netz beim Verbindungsaufbau mitgeteilt und bilden die Grundlage für die Auswahl und Dimensionierung der Quellflußkontrollfunktion. Bedingt durch die statistische Natur der Verkehrsquellen und ihre eventuell unzureichende Beschreibung durch die angegebenen Parameter ist es i.a. nicht möglich, alle Parameter exakt zu überwachen. Daher müssen die Parameter der Quellflußkontrollfunktion in der Regel überdimensioniert werden, um der Quelle einen gewissen Spielraum in ihrem Verhalten zu erlauben, ohne daß unangemessen viele Zellen verworfen oder markiert werden.

Auf der Basis der Parameter der Quellflußkontrollfunktion kann ermittelt werden, welche Charakteristik der Verkehr hat, der schlimmstenfalls von der Quelle in das Netz gelangen kann. Die Beschreibung dieses „Worst Case“-Verkehrs kann dann von der CAC-Funktion benutzt werden, um über die Annahme der neuen Verbindung zu entscheiden. Dabei kann auch eine Verhandlung über die Parameter erfolgen, falls eine Verbindung mit den ursprünglichen Parametern nicht akzeptiert werden kann.

Die Wegesuch- und Überlastabwehrmechanismen im Netz müssen auf die eingesetzten Quellfluß- und Zugangskontrollfunktionen abgestimmt sein und nötigenfalls geeignete Prioritätsmechanismen für verschieden markierte Zellen vorsehen. Die Gebührenstruktur, die sich im Prinzip auf die Parameter der Quelle oder auf die der Quellflußkontrollfunktion abstützen kann, erlaubt z.B. eine Bevorzugung von gleichförmigen Verkehren, die effektiv gemultiplext werden können und dadurch weniger Netzkapazität belegen als solche mit gleichem Mittelwert aber mit büschelförmigem Charakter. Durch ein solches Konzept mit aufeinander abgestimmten Funktionen kann eine wirtschaftliche Ausnutzung der Netzressourcen bei gleichzeitiger Einhaltung der Dienstgütereanforderungen der Benutzer erreicht werden.

## Kapitel 3

# Modellierung von Verkehrsströmen in ATM-Netzen

### 3.1 Modellierungsaspekte

Stochastische Modelle, wie sie im folgenden beschrieben werden, werden als Hilfsmittel zur Leistungsbewertung und Dimensionierung von Kommunikationssystemen und Komponenten eingesetzt. Dabei wird das zeitliche Verhalten der verwendeten zufallsabhängigen Prozesse – wie z.B. das Verhalten von Verkehrsquellen – mit statistischen Methoden charakterisiert. Durch die statistische Behandlung werden die in der Realität vorhandenen Abhängigkeiten und Kausalzusammenhänge zwischen bestimmten Ereignissen ganz oder teilweise vernachlässigt. In Verbindung mit entsprechenden Modellen für die Betriebsmittel und Steuerungsabläufe können mit Hilfe von mathematischen Analyse- und Simulationsmethoden Aussagen über das Verhalten und die Leistungsfähigkeit der betrachteten Systeme gemacht werden.

Wie bereits im vorigen Kapitel erwähnt, kommt der geeigneten Charakterisierung und Modellierung der Verkehrsquellen und der Verkehrsströme aufgrund des angewendeten statistischen Multiplexverfahrens eine besondere Bedeutung bei der Leistungsbewertung zu. Im Gegensatz zu den durchschaltvermittelten Netzen, bei denen eine relativ einfache Charakterisierung der Verkehrsquellen bezüglich Dauer und Häufigkeit der jeweiligen Verbindungen ausreicht, hat in einem paketorientierten ATM-Netz das Verhalten der Verkehrsströme innerhalb einer Verbindung einen entscheidenden Einfluß auf die Beanspruchung der Netzressourcen. Das zeitliche Verhalten innerhalb einer Verbindung ist dienstspezifisch und kann, z.B. bei Videocodecs mit variabler Ausgangsbitrate, relativ komplexen Mechanismen gehorchen.

In Bild 3.1 sind die prinzipiellen Einflüsse bei der Definition von Modellen für Verkehrsströme in ATM-Netzen dargestellt. Ausgehend von einem bestimmten Kommunikationsdienst kann

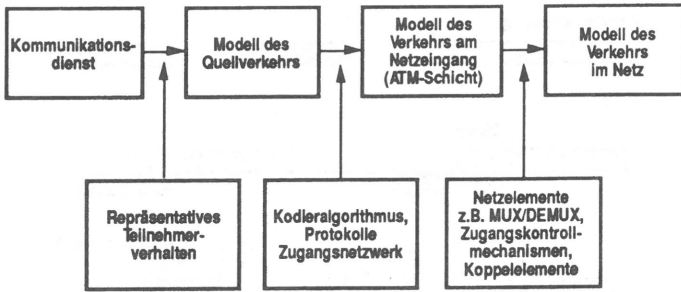


Bild 3.1: Einflüsse auf Verkehrsströme in ATM-Netzen

zunächst das Verhalten direkt an der Quelle des Verkehrs untersucht und statistisch ausgewertet werden. Schwierigkeiten bereiten hier besonders die Quellen mit hohen Datenraten, da bei diesen aufgrund des Datenumfangs nur relativ kurze, nicht repräsentative Echtzeitausschnitte ausgewertet werden können. Bei noch nicht implementierten Diensten ist man hier auf Vermutungen und Extrapolationen von ähnlichen, existierenden Diensten beschränkt. Ein Modell des wirklichen Quellverkehrs kann in der Regel nicht direkt zur Bewertung der Netzkomponenten verwendet werden, da der Verkehr auf dem Weg zu der jeweiligen Komponente durch verschiedene Einflüsse stark beeinflusst werden kann. Die ursprüngliche Charakteristik der Daten wird z.B. bei Video- und Sprachdiensten durch eventuell verwendete redundanzmindernde Codierverfahren sehr stark verändert. Bei Datendiensten haben die durchlaufenen höheren Protokollschichten mit ihren Formatier- und eventuell Flußsteuermechanismen ebenfalls einen starken Einfluß. Desweiteren können die Eigenschaften des Verkehrs im Bereich des Netzzugangs weiter verändert werden, wenn man etwa an Teilnehmerinstallationen in Form von Nebenstellenanlagen oder Hochgeschwindigkeits-LANs mit Mehrfachzugriffsprotokollen denkt. Der so geformte Verkehr erreicht den Netzeingang und damit die eigentlich interessierende ATM-Schicht und kann z.B. zur Untersuchung von verbindungsindividuellen Quellflußkontrollverfahren verwendet werden. Für die Beurteilung von Komponenten im Netz oder von End-zu-End-Verzögerungen müssen noch die Einflüsse der zuvor durchlaufenen Komponenten (statistische Multiplexer, Koppelnetze, usw.) mit berücksichtigt werden. In den meisten Fällen können die Modelle für die Verkehrsströme dabei jedoch dadurch wieder vereinfacht werden, daß der Einfluß der individuellen Quellcharakteristik durch die Überlagerung einer sehr großen Zahl von Verbindungen vernachlässigbar wird.

Bei der Modellierung der Verkehre können natürlich nicht alle oben beschriebenen Einflüsse umfassend berücksichtigt werden, je nach Anwendung des Modells muß jedoch entschieden werden, welche Effekte mit beachtet werden müssen.

Ein sehr allgemeines, hierarchisches Modell für eine ATM-Verkehrsquelle ist in Bild 3.2 dargestellt (ähnliche Modelle finden sich z.B. in [57, 80, 142]). In der obersten Hierarchiestufe

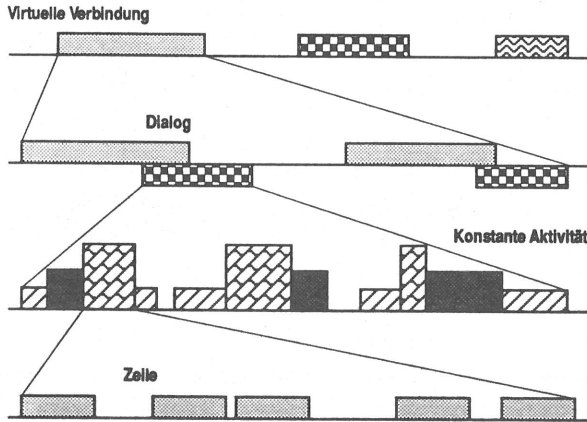


Bild 3.2: Hierarchische Modellierung von ATM-Verkehrsströmen

ist die Ebene der virtuellen Verbindungen dargestellt. Diese Ebene ist von der Durchschaltvermittlung her bekannt und kann durch Angaben über die Dauer der Verbindungen und deren Abstand charakterisiert werden. Die typischen Dauern liegen bei dieser Ebene im Bereich von Sekunden bis zu Stunden. Die Werte für die Abstände der Verbindungswünsche hängen dabei nicht nur vom jeweiligen Dienst ab, sondern sie sind zeitlichen Schwankungen unterworfen, die sich z.B. am Tagesrhythmus orientieren. Solche Effekte werden bei der Modellierung z.B. durch die Definition von Hauptverkehrsstunden mitberücksichtigt [122].

Für das Verhalten einer Quelle während einer Verbindung ergeben sich verschiedene Ebenen mit unterschiedlicher Charakteristik. Zunächst existiert bei den meisten Diensten eine Ebene, die durch die Abhängigkeiten zwischen den an der Kommunikation beteiligten Partnern gekennzeichnet ist. Diese „Dialogebene“, für die typische Zeiten im Sekunden- bis Minutenbereich liegen, ergibt sich z.B. bei der Sprachübertragung durch die Dialogstruktur während eines Gesprächs und bei der interaktiven Datenkommunikation durch das Wechselspiel von Anfragen und Antworten. Bei der interaktiven Videokommunikation hat diese Ebene keine charakteristische Ausprägung, dafür ergibt sich aber durch die Abfolge von einzelnen Szenen eine Struktur, deren Zeitdauern im gleichen Bereich liegen.

Darunter liegt eine Ebene, die oft als „Burst“-Ebene bezeichnet wird. Bei Sprachübertragung mit Pausenunterdrückung ergeben sich auf dieser Ebene während des Sprechens Aktivitätsphasen (engl. Bursts), die durch die Pausen zwischen Sätzen, Worten und Silben unterbrochen werden. Bei Datendiensten ist ebenfalls eine ausgeprägte Ein/Aus-Charakteristik zu beobachten, deren Phasendauern von der übertragenen Datenmenge bestimmt werden. Bei Bewegtbildkommunikation mit variabler Ausgangsbitrate der Codecs erkennt man auf dieser

Ebene die Struktur der einzelnen Bilder, wobei die momentane Aktivität in einem weiten Bereich variieren kann.

Die unterste Ebene dieser hierarchischen Modellierung ist die Ebene der ATM-Zellen, bei der die typischen Zeiten vom Mikrosekunden-Bereich für die Zelldauer bis hin zum Millisekunden-Bereich für die Zeiten zwischen zwei Zellen bei langsameren Quellen reichen.

Eine exakte Modellierung auf der Zellebene müßte neben den leicht am Zellstrom meßbaren Größen, wie z.B. Zellabstand und Anzahl von Zellen in einem Intervall, auch die inhärenten Abhängigkeiten zwischen den Zellen sowie die schwer erfassbaren Einflüsse der anderen Ebenen mit berücksichtigen, was im allgemeinen nur teilweise möglich ist.

Bei der Modellierung von Verkehrsströmen gibt es von der Methodik her zwei Ansätze. Zum einen kann man versuchen, mit Hilfe der gemessenen statistischen Größen des Verkehrs (Momente, Korrelationsfunktionen, ...) geeignete, analytisch behandelbare stochastische Prozesse anzupassen und so einen Ersatzverkehr zu erhalten, der sich bezüglich seiner Auswirkungen auf die Netzressourcen gleich verhält wie der Originalverkehr. Ein Beispiel für dieses Vorgehen wird im Abschnitt 4.3 gezeigt, wo sporadische Quellen mittels eines Markoff-modulierten Poisson-Prozesses (siehe Abschnitt 3.3.3) approximiert werden.

Bei dem anderen Ansatz, der vor allem für simulative Untersuchungen geeignet ist, wird versucht, das reale Quellverhalten mit seinen spezifischen Eigenheiten möglichst genau nachzubilden. Dazu wird ein Zustandsmodell für das Quellverhalten erstellt, und es werden den Zuständen bestimmte - eventuell aus Messungen gewonnene - Dauern oder Wahrscheinlichkeiten zugeordnet. In den einzelnen Zuständen erzeugen diese Generatormodelle dann verschiedene, dem Verhalten der Originalquelle angepaßte Ereignisfolgen. Durch dieses Vorgehen lassen sich insbesondere direkte Abhängigkeiten innerhalb des Verkehrsstroms explizit darstellen.

## **3.2 Charakterisierung von Verkehrsströmen**

### **3.2.1 Zwischenankunftsabstände**

Die am meisten verwendeten statistischen Kenngrößen für Verkehrsströme sind die Verteilungen der Zwischenankunftsabstände bzw. deren Momente. Die entsprechenden Größen für Ströme mit kontinuierlichen bzw. diskreten Zwischenankunftsabständen sind in Tabelle 3.1 zusammengestellt.  $A$  bezeichnet dabei jeweils die Zufallsvariable für die Zwischenankunftsabstände. Im kontinuierlichen Fall bezeichnet die Variable „ $t$ “ die Zeit. Im diskreten Fall können die Ankünfte nur zu bestimmten Zeitpunkten erfolgen, welche in einem der Basiseinheiten (z.B. Länge eines Zeitschlitzes) entsprechenden Abstand voneinander liegen. Die

KENNGRÖSSE	KONTINUIERLICH	DISKRET
Verteilungsfunktion	$A(t) = P\{A \leq t\}$ $= \int_0^t a(t)dt$	$A(k) = P\{A \leq k\}$ $= \sum_{i=0}^k a_i$
Verteilungsdichtefunktion	$a(t) = \frac{dA(t)}{dt}$	$a(k) = \sum_{i=0}^k a_i \delta(k - i)$
Verteilung		$a_k = P\{A = k\}$ $= \frac{1}{i!} \left. \frac{d^k a_{EF}(z)}{dz^k} \right _{z=0}$
Erwartungswert $E[g(A)]$	$\int_0^\infty g(t)a(t)dt$	$\sum_{k=0}^\infty g(k)a_k$
$\ell$ -tes gewöhnliches Moment  $m_\ell = E[A^\ell]$	$\int_0^\infty t^\ell a(t)dt =$ $(-1)^\ell \left. \frac{d^\ell a_{LST}(s)}{ds^\ell} \right _{s=0}$	$\sum_{k=0}^\infty k^\ell a_k$
$\ell$ -tes zentrales Moment  $\mu_\ell = E[(A - m_1)^\ell]$	$\int_0^\infty (t - m_1)^\ell a(t)dt$	$\sum_{k=0}^\infty (k - m_1)^\ell a_k$
Mittelwert $E[A] = m_1$	$- \left. \frac{da_{LST}(s)}{ds} \right _{s=0}$	$\left. \frac{da_{EF}(z)}{dz} \right _{z=1} = G'(1)$
Varianz $VAR[A] = \mu_2$		$G''(1) + G'(1) - [G'(1)]^2$
Standardabweichung	$\sigma_a = \sqrt{VAR[A]}$	
Variationskoeffizient	$c_a = \frac{\sigma_a}{m_1}$	
Transformationen	<p>Laplace-Stieltjes</p> $a_{LST}(s) = \int_0^\infty e^{-st} dA(t)$ $= \int_0^\infty e^{-st} a(t)dt$	<p>Erzeugende Funktion</p> $a_{EF}(z) = \sum_{k=0}^\infty a_k z^k$

Tabelle 3.1: Verteilungen und Momente am Beispiel der Zwischenankunftsabstände

Variable „ $k$ “ ist eine ganzzahlige, auf die Basiszeiteinheit normierte Größe. Mit  $\delta(k-i)$  wird der Dirac-Impuls an der Stelle  $i$  bezeichnet.

Durch die Beschreibung der Verkehrsströme mit Hilfe der Zwischenankunftsverteilung werden die Abhängigkeiten zwischen aufeinanderfolgenden Ankunftsabständen nicht erfasst. Bei Analysen auf dieser Basis wird dementsprechend der reale Ankunftsprozeß durch einen Erneuerungsprozeß (rekurrenter Prozeß) mit gleicher Verteilung angenähert, der dadurch gekennzeichnet ist, daß die aufeinanderfolgenden Ereignisabstände unabhängig voneinander und gleich verteilt sind.

### 3.2.2 Zählprozeß

Eine andere Methode zur Charakterisierung von stationären Verkehrsströmen verwendet den sogenannten Zählprozeß  $N_t$ , der die Anzahl von Ankunftsereignissen in einem Intervall der Länge  $t$  angibt. Diese Methode hat Vorteile, wenn mehrere unabhängige Verkehrsströme überlagert werden sollen, weil sich in diesem Fall die Zufallsvariable für den Zählprozeß der Überlagerung aus der (leicht berechenbaren) Summe der entsprechenden Zufallsvariablen für die Einzelströme ergibt. Diese Art der Charakterisierung berücksichtigt auch Abhängigkeiten innerhalb des Ankunftsstromes.

#### 3.2.2.1 Verteilung

Für kontinuierliche, stationäre Erneuerungsprozesse läßt sich die Verteilung des Zählprozesses im Prinzip wie folgt berechnen. Bild 3.3 zeigt die Situation für einen beliebig gewählten Anfangspunkt, der nicht mit einem Ankunftsereignis zusammenfällt. Für die Berechnung

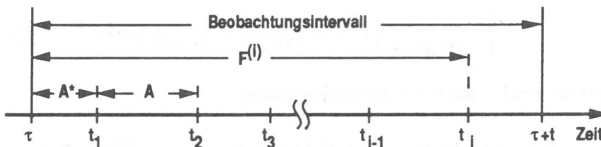


Bild 3.3: Der Zählprozeß für Erneuerungsprozesse

der Verteilung des Zählprozesses wird zunächst eine Zufallsvariable (ZV)  $F^{(i)}$  für die Zeit bis zum  $i$ -ten Ereignis benötigt. Dabei ist zu beachten, daß die Zeit bis zum ersten Ereignis i. a. eine andere Verteilung aufweist als alle übrigen (modifizierter Erneuerungsprozeß). Diese

Zeit entspricht der Vorwärtsrekurrenzzeit  $A^*$  des Ankunftsabstandes. Die Verteilung der Vorwärtsrekurrenzzeit berechnet sich für kontinuierliche Ankunftsabstände als

$$a^*(t) = \frac{1}{E[A]} (1 - A(t)) = \frac{1}{E[A]} \left( 1 - \int_0^t a(\tau) d\tau \right) \quad (3.1)$$

Da für den Erneuerungsprozeß die Ankunftsabstände unabhängig voneinander sind, ergibt sich für die Zeit  $F^{(i)}$

$$F^{(i)} = A^* + \underbrace{A + \dots + A}_{(i-1) \text{ Terme}} \quad (3.2)$$

Für die Verteilungsdichtefunktionen gilt entsprechend

$$f^{(i)}(t) = a^*(t) \star \underbrace{a(t) \star \dots \star a(t)}_{(i-1) \text{ Terme}} \quad (3.3)$$

wobei „ $\star$ “ die kontinuierliche Faltungsoperation bezeichnet. Die Wahrscheinlichkeit  $n_i(t)$  für  $i$  Ankünfte in einem Intervall der Länge  $t$  wird durch

$$\begin{aligned} n_i(t) &= P\{N_t = i\} = P\{F^{(i)} \leq t < F^{(i+1)}\} \\ &= P\{F^{(i)} \leq t\} - P\{F^{(i+1)} \leq t\} \\ &= F^{(i)}(t) - F^{(i+1)}(t) \end{aligned} \quad (3.4)$$

beschrieben. Für  $n_i(0)$  gilt

$$n_i(0) = P\{A^* > t\} = 1 - A^*(t) \quad (3.5)$$

Die Laplace-Transformierte (LT)  $n_{LT,i}(s)$  ist gegeben durch

$$n_{LT,i}(s) = \begin{cases} \frac{1}{s} \left( 1 - \frac{1}{E[A]s} (1 - a_{LST}(s)) \right), & i = 0 \\ \frac{1}{E[A]s^2} (1 - a_{LST}(s))^2 [a_{LST}(s)]^{i-1}, & i \geq 1 \end{cases} \quad (3.6)$$

Durch eine Division erhält man die rekursive Form

$$n_{LT,i}(s) = n_{LT,i-1}(s) a_{LST}(s) \quad \text{für } i \geq 2 \quad (3.7)$$

Im Zeitbereich läßt sich die Verteilung des Zählprozesses also mit Hilfe der Beziehungen

$$n_i(t) = \begin{cases} 1 - A^*(t), & i = 0 \\ E[A] (a^*(t) \star a^*(t)), & i = 1 \\ n_{i-1}(t) \star a(t), & i \geq 2 \end{cases} \quad (3.8)$$



angeben. Für den Erwartungswert  $E[N_t]$  gilt

$$\begin{aligned} E[N_t] &= \sum_{i=1}^{\infty} i n_i(t) \\ &= F^{(1)}(t) + 2F^{(2)}(t) + 3F^{(3)}(t) + \dots \\ &\quad - 1F^{(2)}(t) - 2F^{(3)}(t) - \dots \\ &= \sum_{i=1}^{\infty} F^{(i)}(t) \end{aligned} \quad (3.9)$$

und transformiert

$$\mathcal{LT} \{E[N_t]\} = \frac{1}{E[A]s^2} (1 - a_{LST}(s)) \sum_{i=1}^{\infty} [a_{LST}(s)]^{i-1} = \frac{1}{E[A]s^2} \quad (3.10)$$

Damit ergibt sich  $E[N_t]$  zu  $t/E[A]$ , was auch durch direkte Überlegung bestätigt werden kann.

Für diskrete Erneuerungsprozesse kann die Verteilung des Zählprozesses analog berechnet werden. Wenn man den Anfang des betrachteten Intervalls unmittelbar vor einen möglichen Ereigniszeitpunkt legt, dann ist die Verteilung der Vorwärtsrekurrenzeit  $a_k^*$  durch

$$a_k^* = \frac{1}{E[A]} (1 - A(k)) = \frac{1}{E[A]} \left( 1 - \sum_{i=0}^k a_i \right) \quad (3.11)$$

gegeben. Diese Beziehung läßt sich auch rekursiv darstellen als (siehe z.B. [133, 157])

$$a_k^* = \begin{cases} \frac{1 - a_0}{E[A]}, & k = 0 \\ a_{(k-1)}^* - \frac{a_k}{E[A]}, & k > 0 \end{cases} \quad (3.12)$$

Die Gleichung für die Verteilung der Dauer bis zum  $i$ -ten Ankunftsereignis lautet dann

$$f_k^{(i)} = a_k^* \underbrace{\star a_k \star \dots \star a_k}_{(i-1) \text{ Terme}}, \quad (3.13)$$

wobei „ $\star$ “ die diskrete Faltungsoperation

$$d_k = b_k \star c_k = \sum_{i=-\infty}^{\infty} a_{k-i} b_i \quad (3.14)$$

bezeichnet. Die Verteilung des Zählprozesses  $n_i(k)$  von beliebigen zeitdiskreten Erneuerungsprozessen ergibt sich dann zu

$$n_i(k) = P\{F^{(i)} < k\} - P\{F^{(i+1)} < k\} = \sum_{l=0}^{k-1} f_l^{(i)} - \sum_{l=0}^{k-1} f_l^{(i+1)} \quad (3.15)$$

Nach Anwendung der Z-Transformation

$$a_{ZT}(z) = \sum_{k=0}^{\infty} a_k z^{-k} \quad (3.16)$$

erhält man unter Verwendung der Beziehungen

$$a_{ZT}^*(z) = \frac{1}{E[A]} \frac{1 - a_{ZT}(z)}{1 - z^{-1}} \quad \text{und} \quad (3.17)$$

$$ZT \left\{ \sum_{i=0}^{k-1} a_i \right\} = a_{ZT}(z) \frac{z^{-1}}{1 - z^{-1}} \quad (3.18)$$

für die Transformierte der Verteilung des Zählprozesses

$$n_{ZT,i}(z) = \begin{cases} \frac{1}{1 - z^{-1}} \left( 1 - \frac{z^{-1}}{E[A](1 - z^{-1})} (1 - a_{ZT}(z)) \right), & i = 0 \\ \frac{z^{-1}}{E[A](1 - z^{-1})^2} (1 - a_{ZT}(z))^2 [a_{ZT}(z)]^{i-1}, & i \geq 1 \end{cases} \quad (3.19)$$

Nach der Rücktransformation findet man die der Gleichung 3.8 für kontinuierliche Verteilungen entsprechende Beziehung

$$n_i(k) = \begin{cases} 1 - A^*(k-1), & i = 0 \\ E[A] (a_{(k-1)}^* \star a_{(k-1)}^*), & i = 1 \\ n_{i-1}(k) \star a_{(k-1)}, & i \geq 2 \end{cases} \quad (3.20)$$

die für Intervallängen größer null gültig ist (in einem Intervall der Länge null kann keine Ankunft erfolgen).

Für die Z-Transformierte des Mittelwertes  $E[N_k]$  für den Zählprozeß ergibt sich ähnlich wie im kontinuierlichen Fall

$$ZT \{E[N_k]\} = \frac{z^{-1}}{E[A](1 - z^{-1})^2} \quad (3.21)$$

Nach der Rücktransformation erhält man

$$E[N_k] = \frac{k}{E[A]} \quad (3.22)$$

Die für die numerische Berechnung der Verteilung des Zählprozesses nach Gleichung 3.20 benötigte Rechenzeit kann durch den Einsatz der „schnellen Fourier-Transformation“ (engl. Fast Fourier Transform, FFT [73]) für Zwischenankunftsverteilungen mit vielen Elementen

erheblich reduziert werden. Statt die Faltungen im Zeitbereich auszuführen, erfolgt eine Transformation in den Frequenzbereich mit Hilfe der FFT, wo die der Faltung entsprechende Multiplikation ausgeführt wird. Anschließend erfolgt die Rücktransformation in den Zeitbereich. Durch Speicherung der Transformatierten von  $n_{i-1}(k)$  und  $a_{(k-1)}$  erfordert die Berechnung eines Elementes der Verteilung nur eine komplexe Multiplikation der zwei Vektoren und eine Rücktransformation. Der Algorithmus kann auch zur approximativen Bestimmung des Zählprozesses für kontinuierliche Verteilungen angewendet werden. Der bei der erforderlichen Diskretisierung der Verteilungen entstehende Fehler kann so gering gehalten werden, daß er das Ergebnis nicht signifikant beeinflusst [157]. Beispiele für die Anwendung des beschriebenen Algorithmus finden sich in Abschnitt 3.4.

### 3.2.2.2 Index of Dispersion for Counts

Der sogenannte „Index of Dispersion for Counts“ (IDC) ist eine aus dem Zählprozeß abgeleitete Kenngröße. Der IDC ist als Verhältnis von Varianz zu Mittelwert des Zählprozesses definiert [45].

$$IDC(t) = \frac{VAR[N_t]}{E[N_t]} \quad (3.23)$$

Der Grenzwert des IDC für eine gegen unendlich strebende Intervalllänge ist bei Erneuerungsprozessen gleich dem Quadrat des Variationskoeffizienten der Zwischenankunftsabstände.

$$\lim_{t \rightarrow \infty} IDC(t) = c_a^2 \quad (3.24)$$

Für die Überlagerung von  $n$  unabhängigen Strömen ergibt sich wegen der Additivität von Varianz und Mittelwert für den IDC des Summenstromes ( $IDC^s(t)$ )

$$IDC^s(t) = \frac{\sum_{i=1}^n VAR[N_{t,i}]}{\sum_{i=1}^n E[N_{t,i}]} = \sum_{i=1}^n IDC_i(t) \frac{1}{E[A_i] \sum_{\ell=1}^n \frac{1}{E[A_\ell]}} \quad (3.25)$$

Für den Sonderfall identischer Ströme gilt daher (siehe [70])

$$IDC^s(t) = IDC(t) \quad (3.26)$$

Für den Grenzwert gilt mit 3.24

$$\lim_{t \rightarrow \infty} IDC^s(t) = \sum_{i=1}^n c_{a,i}^2 \frac{1}{E[A_i] \sum_{\ell=1}^n \frac{1}{E[A_\ell]}} \quad (3.27)$$

Für zeitdiskrete Prozesse gelten die entsprechenden Beziehungen. Der IDC gibt Aufschluß über das Korrelationsverhalten des Verkehrsstromes und damit über Abweichungen vom Poisson-Prozeß ( $IDC(t) = 1$  unabhängig von  $t$ , siehe Abschnitt 3.3.1.1.1), die z.B. anhand der Momente der Zwischenankunftsabstände nicht immer sichtbar sind. Beispiele für die Berechnung des IDC werden in Abschnitt 3.4 gezeigt.

### 3.2.3 Autokovarianz

Bedingt z.B. durch redundanzmindernde Codierverfahren weisen die Quellverkehrsströme mit variabler Bitrate in ATM-Netzen oft sehr ausgeprägte Korrelationseigenschaften auf, die einen signifikanten Einfluß auf die Belegung der Netzressourcen haben. Eine in diesem Zusammenhang oft verwendete Kenngröße ist deshalb die Autokovarianzfunktion  $C(t)$  der momentanen Bitrate  $\lambda(t)$

$$C(t) = E[\lambda(\tau)\lambda(\tau + t)] - m_1^2 \quad (3.28)$$

die am realen Verkehrstrom relativ einfach meßbar ist. Problematisch ist hier jedoch die Anpassung eines stochastischen Modells bezüglich der Autokovarianzfunktion. Ansätze für eine solche Anpassung sind z.B. in [66, 77, 110] zu finden.

## 3.3 Verwendete Modelle für Verkehrsströme

In diesem Abschnitt sollen die für die Untersuchungen verwendeten Quellmodelle für Verkehrsströme kurz eingeführt und ihre spezifischen Eigenschaften dargestellt werden. Desweiteren soll kurz auf ihre Anwendbarkeit auf verkehrstheoretische Probleme in ATM-Netzen eingegangen werden.

### 3.3.1 Erneuerungsprozesse

Erneuerungsprozesse im Sinne dieses Abschnitts sind stationäre Punktprozesse, die dadurch gekennzeichnet sind, daß ihre Zwischenankunftsabstände unabhängig voneinander sind und der gleichen Verteilung gehorchen (engl. Independent and Identically Distributed, IID). Diese Prozesse sind durch die Verteilung ihrer Zwischenankunftsprozesse vollständig beschreibbar.

### 3.3.1.1 Zeitkontinuierliche Erneuerungsprozesse

#### 3.3.1.1.1 Negativ-exponentielle Verteilung (M)

Diese Verteilung besitzt als einzige kontinuierliche Verteilung die Eigenschaft der „Gedächtnisfreiheit“, die auch als Markoff-Eigenschaft bezeichnet wird. Die Gedächtnisfreiheit besagt, daß von einem beliebigen Zeitpunkt  $t$  an die Verteilung für die Zeit bis zum nächsten Ereignis (Vorwärtsrekurrenzzeit) unabhängig von der Zeitspanne zwischen dem letzten Ereignis und  $t$  ist. Es gilt also

$$a^*(t) = a(t) \quad (3.29)$$

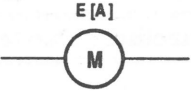
$A(t)$	$a(t)$	$a_{LST}(s)$	$E[A]$	$VAR[A]$	$c_a$
$1 - e^{-\lambda t}$	$\lambda e^{-\lambda t}$	$\frac{\lambda}{s + \lambda}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	1
Symbolische Darstellung					
					

Tabelle 3.2: Charakteristische Größen der negativ-exponentiellen Verteilung

Aus Gleichung 3.6 läßt sich leicht berechnen, daß sich für die Anzahl von Ereignissen in einem beliebigen Intervall der Länge  $t$  (Zählprozeß) die Poisson-Verteilung (siehe 3.3.1.2.5)

$$n_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t} \quad (3.30)$$

ergibt. Ein Ankunftsprozeß mit negativ-exponentiell verteilten Zwischenankunftsabständen wird deshalb auch als Poisson-Prozeß bezeichnet. Aus den Eigenschaften der Poisson-Verteilung (siehe 3.3.1.2.5) ergibt sich für den IDC unabhängig von  $t$

$$IDC(t) = \frac{VAR[N_t]}{E[N_t]} = \frac{\lambda t}{\lambda t} = 1 \quad (3.31)$$

Eine oft benutzte Eigenschaft des Poisson-Prozesses ist, daß die Überlagerung unabhängiger Poisson-Prozesse wieder einen Poisson-Prozeß mit der Summenankunftsrate der Einzelströme

ergibt. Auch eine Überlagerung von vielen Verkehrsströmen mit anderen Zwischenankunftsverteilungen nähert sich unter bestimmten Bedingungen [75] einem Poisson-Prozeß an. Aus diesem Grund wird der Poisson-Prozeß oft als Näherungsprozeß für eine solche Überlagerung verwendet.

Eine bei der Systemanalyse oft verwendete Eigenschaft des Poisson-Prozesses ist die sogenannte „PASTA“-Eigenschaft (Poisson Arrivals See Time Averages [168]), die besagt, daß gemäß einem Poisson-Prozeß in einem System ankommende Anforderungen die im zeitlichen Mittel vorliegenden Zustandswahrscheinlichkeiten „sehen“.

### 3.3.1.1.2 Deterministische Verteilung (D)

Bei dieser Verteilung erfolgen die Ereignisse deterministisch im Abstand  $t_0$ . Sie wird zur Beschreibung von Verkehrsströmen mit konstanter Bitrate (CBR) verwendet. Ein weiterer wichtiger Anwendungsfall ist die Modellierung von Übertragungszeiten für Zellen auf einem Verbindungsabschnitt.


$A(t)$	$a(t)$	$a_{LST}(s)$	$E[A]$	$VAR[A]$	$c_a$
$s(t - t_0)$	$\delta(t - t_0)$	$e^{-st_0}$	$t_0$	0	0
Symbolische Darstellung					
					

Tabelle 3.3: Charakteristische Größen der deterministischen Verteilung

### 3.3.1.1.3 Erlang-k-Verteilung ( $E_k$ )

Die Erlang-k-Verteilung ergibt sich aus der Summe von  $k$  negativ-exponentiell verteilten Zufallsvariablen mit Mittelwert  $1/k\lambda$ . Diese Verteilung wird verwendet, wenn ein Variationskoeffizient kleiner als 1 benötigt wird. Es sind nur bestimmte Werte für  $c_a$  möglich, die sich aus der Phasenanzahl ergeben.

Die Verteilungsdichtefunktion der Vorwärtsrekurrenzzeit ergibt sich im Zeitbereich zu

$$a^*(t) = \lambda \sum_{i=0}^{k-1} \frac{(k\lambda t)^i}{i!} e^{-k\lambda t} \quad (3.32)$$

$A(t)$	$a(t)$	$a_{LST}(s)$	$E[A]$	$VAR[A]$	$c_a$
$1 - \sum_{i=0}^{k-1} \frac{(k\lambda t)^i}{i!} e^{-k\lambda t}$	$(k\lambda) \frac{(k\lambda t)^{k-1}}{(k-1)!} e^{-k\lambda t}$	$\left(\frac{k\lambda}{k\lambda + s}\right)^k$	$\frac{1}{\lambda}$	$\frac{1}{k\lambda^2}$	$\frac{1}{\sqrt{k}}$
Symbolische Darstellung					

Tabelle 3.4: Charakteristische Größen der Erlang-k-Verteilung

und im Laplace-Bereich zu

$$a_{LST}^*(s) = \frac{1}{k} \sum_{i=0}^{k-1} \left[ \frac{k\lambda}{k\lambda + s} \right]^{i+1} \quad (3.33)$$

Damit erhält man für die Verteilung des Zählprozesses

$$n_i(t) = \begin{cases} \frac{1}{k} \sum_{j=0}^{k-1} \sum_{\ell=0}^j \frac{(k\lambda t)^\ell}{\ell!} e^{-k\lambda t}, & i = 0 \\ \frac{1}{k} \sum_{j=0}^{k-1} \sum_{\ell=j+k(i-1)+1}^{j+ki} \frac{(k\lambda t)^\ell}{\ell!} e^{-k\lambda t}, & i \geq 1 \end{cases} \quad (3.34)$$

### 3.3.1.1.4 Hyperexponentielle Verteilung 2. Ordnung ( $H_2$ )

Diese Verteilung entsteht durch eine Auswahl zwischen zwei Zufallsvariablen  $T_1$  und  $T_2$  mit Mittelwert  $1/\lambda_1$  bzw.  $1/\lambda_2$ , wobei  $T_1$  mit der Wahrscheinlichkeit  $p$  ausgewählt wird. Diese Verteilung kann verwendet werden, wenn ein Variationskoeffizient größer als 1 benötigt wird. Der dritte freie Parameter wird meist durch die Symmetriebedingung

$$\frac{p}{\lambda_1} = \frac{(1-p)}{\lambda_2} \quad (3.35)$$

festgelegt.

$A(t)$	$a(t)$	$a_{LST}(s)$	$E[A]$	$VAR[A]$
$p(1 - e^{-\lambda_1 t}) +$ $(1-p)(1 - e^{-\lambda_2 t})$	$p\lambda_1 e^{-\lambda_1 t} +$ $(1-p)\lambda_2 e^{-\lambda_2 t}$	$p \frac{\lambda_1}{\lambda_1 + s} +$ $(1-p) \frac{\lambda_2}{\lambda_2 + s}$	$\frac{p}{\lambda_1} +$ $\frac{(1-p)}{\lambda_2}$	$\frac{2p-p^2}{\lambda_1^2} - \frac{p(1-p)}{\lambda_1\lambda_2} +$ $\frac{2(1-p) - (1-p)^2}{\lambda_2^2}$
$c_a$	$p$		$\lambda_1$	$\lambda_2$
$> 1$	$\frac{1}{2} \left( 1 + \sqrt{\frac{c_a^2 - 1}{c_a^2 + 1}} \right)$		$\frac{2p}{E[A]}$	$\frac{2(1-p)}{E[A]}$
<b>Symbolische Darstellung</b>				

Tabelle 3.5: Charakteristische Größen der hyperexponentiellen Verteilung

### 3.3.1.2 Zeitdiskrete Erneuerungsprozesse

Zeitdiskrete Erneuerungsprozesse sind solche, bei denen die Ereignisse nur zu bestimmten Zeitpunkten erfolgen können, die durch ganzzahlige Vielfache einer Basiszeiteinheit gegeben sind. Die Variable „k“ sowie die mit der Dimension „Zeit“ behafteten Kenngrößen sind auf diese Basiszeiteinheit normiert. Neben der in der Tabelle 3.1 definierten Erzeugenden Funktion wird zur Charakterisierung dieser Prozesse oft die Z-Transformation der Verteilung für die Zwischenankunftsabstände

$$a_{ZT}(z) = \sum_{k=0}^{\infty} a_k z^{-k} \tag{3.36}$$

verwendet, die im folgenden in den entsprechenden Tabellen angegeben ist. Die Erzeugende Funktion kann prinzipiell durch Ersetzen von  $z$  durch  $z^{-1}$  aus der Z-Transformierten gewonnen werden.



### 3.3.1.2.1 Bernoulli-Verteilung (BER)

Die Bernoulli-Verteilung wird bei der Analyse zeitdiskreter Systeme sehr oft für die Beschreibung der Ankunfts-wahrscheinlichkeit einer Zelle in einem Zeitschlitz verwendet. Unter der Annahme, daß diese Wahrscheinlichkeit konstant gleich  $(1 - q)$  ist und daß die Wahrscheinlichkeiten in den einzelnen Zeitschlitzen unabhängig voneinander sind, ergibt sich für den Zwischenankunftsabstand der Zellen eine verschobene geometrische Verteilung mit einem minimalen Zwischenankunftsabstand von einer Zeiteinheit.

$a_k$	$a_{zT}(z)$	$E[A]$	$VAR[A]$	$c_a$
$a_0 = q$ $a_1 = 1 - q$ $0 \leq q \leq 1$	$q + (1 - q)z^{-1}$	$(1 - q)$	$(1 - q)q$	$\sqrt{\frac{q}{1 - q}}$

Tabelle 3.6: Charakteristische Größen der Bernoulli-Verteilung

### 3.3.1.2.2 Vershobene Geometrische Verteilung (GEO<sub>m</sub>)

Die verschobene geometrische Verteilung (GEO<sub>m</sub>) entsteht aus der normalen geometrischen Verteilung (GEO<sub>0</sub>) durch Verschieben der Verteilungselemente um  $m$  Zeiteinheiten in Richtung der positiven Zeitachse. Dabei werden die ersten  $m$  Elemente der verschobenen Verteilung identisch null und es ergibt sich ein minimaler Zwischenankunftsabstand von  $m$  Zeiteinheiten. Die beiden Sonderfälle GEO<sub>0</sub> und GEO<sub>1</sub> sind diejenigen diskreten Verteilungen, welche die Eigenschaft der Gedächtnisfreiheit aufweisen. Beobachtet man den Prozeß unmittelbar vor einem diskreten Zeitpunkt (an dem ein Ereignis möglich ist), so weist die GEO<sub>0</sub>-Verteilung, bei einer Betrachtung unmittelbar nach einem diskreten Zeitpunkt die GEO<sub>1</sub>-Verteilung diese Eigenschaft auf [157]. In Analogie zur PASTA-Eigenschaft gilt für diese Verteilungen die GASTA-Eigenschaft (Geometric Arrivals See Time Averages) [67].

$a_k$	$a_{zT}(z)$	$E[A]$	$VAR[A]$	$c_a$
$(1 - q)q^{k-m}$ $0 \leq q \leq 1$ $m \geq 0, k \geq m$	$\frac{(1 - q)z^{-m}}{1 - qz^{-1}}$	$\frac{q}{1 - q} + m$	$\frac{q}{(1 - q)^2}$	$\frac{\sqrt{q}}{q + m(1 - q)}$

Tabelle 3.7: Charakteristische Größen der verschobenen geometrischen Verteilung

Die Verteilung des Zählprozesses für die GEO<sub>0</sub>-Verteilung berechnet sich wie folgt:

Da der Ankunftsabstand 0 mit der Wahrscheinlichkeit  $(1 - q)$  auftritt, ist die Anzahl von gleichzeitig eintreffenden Ankünften  $b_i$  geometrisch verteilt mit

$$b_i = (1 - q)^i q \quad (3.37)$$

Die Z-Transformierte für die Anzahl von Ankünften in einem Intervall der Länge  $k$  ergibt sich also zu

$$n_{ZT,i}(z) = \frac{q^k}{(1 - (1 - q)z^{-1})^k} \quad (3.38)$$

Nach der Rücktransformation erhält man unter Verwendung der Beziehung

$$ZT^{-1} \left\{ \frac{1}{(1 - \alpha z^{-1})^{n+1}} \right\} = \binom{i+n}{n} \alpha^i \quad (3.39)$$

die Negativ-Binomial-Verteilung (siehe 3.3.1.2.4) <sup>1</sup>

$$n_i(k) = \binom{i+k-1}{k-1} (1-q)^i q^k \quad k \geq 1 \quad \text{und} \quad i \geq 0 \quad (3.40)$$

Bei der  $GEO_1$ -Verteilung ergibt sich

$$b(i) = \begin{cases} q, & i = 0 \\ 1 - q, & i = 1 \end{cases} \quad (3.41)$$

und

$$n_{ZT,i} = (q + (1 - q)z^{-1})^k \quad (3.42)$$

was der Z-Transformierten der Binomial-Verteilung (siehe 3.3.1.2.3) entspricht. Für die Verteilung des Zählprozesses der  $GEO_1$ -Verteilung gilt also

$$n_i(k) = \binom{k}{i} (1 - q)^i q^{k-i} \quad (3.43)$$

Dieses Ergebnis erhält man auch bei einer direkten Auswertung von Gleichung 3.20. Der IDC ist für die  $GEO_0$ - und für die  $GEO_1$ -Verteilung unabhängig von der Länge des Betrachtungsintervalls, wie es im kontinuierlichen Fall bei der negativ-exponentiellen Verteilung der Fall ist. Als Werte ergeben sich aus den charakteristischen Größen der Negativ-Binomial- bzw. der Binomial-Verteilung  $1/q$  für die  $GEO_0$ -Verteilung und  $q$  für die  $GEO_1$ -Verteilung. Diese Werte entsprechen jeweils dem Quadrat des Variationskoeffizienten der Zwischenankunftsabstände, womit die Gleichung 3.24 für den Grenzwert für eine gegen unendlich strebende Intervallbreite erfüllt ist.

---

<sup>1</sup>Es gilt  $\binom{n}{m} = \binom{n}{n-m}$

### 3.3.1.2.3 Binomial-Verteilung (BIN)

Die Binomialverteilung ist eine endliche, diskrete Verteilung, bei der der Abstand zwischen zwei Ankünften maximal  $n$  Zeiteinheiten betragen kann. Die Binomial-Verteilung wird oft für die Beschreibung der Anzahl von Ankünften pro Zeitschlitz bei der Überlagerung unabhängiger GEO<sub>1</sub>-Ströme verwendet. Dabei entspricht  $n$  der Anzahl der überlagerten Ströme und  $k$  der Anzahl der Zellen.

$a_k$	$a_{zT}(z)$	$E[A]$	$VAR[A]$	$c_a$
$\binom{n}{k} (1-q)^k q^{n-k}$ $0 \leq q \leq 1$ $0 \leq k \leq n$	$(q + (1-q)z^{-1})^n$	$n(1-q)$	$nq(1-q)$	$\sqrt{\frac{q}{n(1-q)}}$

Tabelle 3.8: Charakteristische Größen der Binomial-Verteilung

### 3.3.1.2.4 Negativ-Binomial-Verteilung (NEGBIN)

Die Negativ-Binomial-Verteilung ist eine unendliche diskrete Verteilung. Mit diesem Verteilungstyp können beliebige Variationskoeffizienten in einem weiten Bereich von annähernd null bis sehr viel größer als eins realisiert werden, solange die Bedingung  $E[A]c_a^2 > 1$  erfüllt ist.

$a_k$	$a_{zT}(z)$	$E[A]$	$VAR[A]$	$c_a$
$\binom{y+k-1}{k} (1-q)^y q^k$ $0 \leq q \leq 1$ $y \text{ reell}$	$\left(\frac{1-q}{1-qz^{-1}}\right)^y$	$\frac{yq}{1-q}$	$\frac{yq}{(1-q)^2}$	$\frac{1}{\sqrt{yq}}$
$q$		$y$		
$1 - \frac{1}{E[A]c_a^2}$		$\frac{E[A]}{E[A]c_a^2 - 1}$		

Tabelle 3.9: Charakteristische Größen der Negativ-Binomial-Verteilung

Aufgrund dieser Eigenschaft ist die Negativ-Binomial-Verteilung sehr gut für Parameterstudien bei zeitdiskreten Systemen geeignet. Sie kann auch mit relativ guter Genauigkeit als Näherung für die entsprechenden kontinuierlichen Verteilungstypen verwendet werden, wobei die Abweichung mit Zunahme des Variationskoeffizienten größer wird.

### 3.3.1.2.5 Poisson-Verteilung (POIS)

Die Poisson-Verteilung ergibt sich bei der Berechnung der Verteilung des Zählprozesses für negativ-exponentiell verteilte Zwischenankunftsabstände. Für diese Verteilung ist der Mittelwert immer gleich der Varianz.

$a_k$	$a_{zT}(z)$	$E[A]$	$VAR[A]$	$c_a$
$\frac{y^k}{k!} e^{-y}$ $k \geq 0, y \text{ reell}$	$e^{y(z^{-1}-1)}$	$y$	$y$	$\frac{1}{\sqrt{y}}$

Tabelle 3.10: Charakteristische Größen der Poisson-Verteilung

### 3.3.2 Modell für eine sporadische Quelle

Dieses Quellenmodell wird in ATM-Untersuchungen sehr oft verwendet, weil es sowohl den büschelförmigen Charakter der ATM-Verkehre in der Burst- bzw. Dialogebene (siehe Abschnitt 3.1) als auch die periodische, durch den Paketiervorgang entstehende Struktur in der Zellebene modelliert. Der entsprechende Ankunftsprozeß für dieses Modell ist in Bild 3.4 beispielhaft dargestellt. Während der Aktivitätsphasen auf Burst- bzw. Dialog-Ebene werden

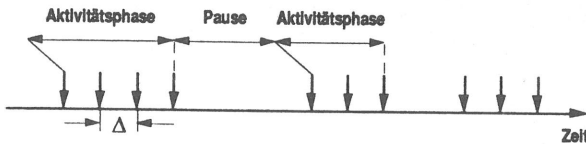


Bild 3.4: Ankunftsprozeß einer sporadischen Quelle

Zellen mit einem konstanten Ankunftsabstand  $\Delta$  erzeugt, wobei eine Aktivitätsphase mindestens eine Zelle umfaßt. Die erste Zellankunft erfolgt um  $\Delta$  verschoben nach dem Beginn

der Aktivitätsphase, da die entsprechende Zelle erst paketierrt werden muß bevor sie gesendet werden kann. In der darauffolgenden Pausenphase werden keine Zellen erzeugt.

$A(t)$		$a(t)$	
$[(1 - \alpha\Delta) + \alpha\Delta(1 - e^{-\beta(t-\Delta)})]s(t - \Delta)$		$(1 - \alpha\Delta)\delta(t - \Delta) + \alpha\Delta\beta e^{-\beta(t-\Delta)}s(t - \Delta)$	
$a_{LST}(s)$	$E[A] = 1/\lambda$	$VAR[A]$	$c_a$
$\left( (1 - \alpha\Delta) + \frac{\alpha\Delta\beta}{s + \beta} \right) e^{-s\Delta}$	$\Delta + \frac{\alpha\Delta}{\beta}$	$\frac{1 - (1 - \alpha\Delta)^2}{\beta^2}$	$\sqrt{\frac{1 - (1 - \alpha\Delta)^2}{(\alpha\Delta + \beta\Delta)^2}}$
Symbolische Darstellung			

Tabelle 3.11: Charakteristische Größen der zweiphasigen sporadischen Quelle

In der Regel wird angenommen, daß die Anzahl der Zellen pro Aktivitätsphase geometrisch ( $GEO_1$ ) und die Pausendauern negativ-exponentiell verteilt sind <sup>2</sup>. Mit dieser Annahme wird der entsprechende Ankunftsprozeß ein kontinuierlicher Erneuerungsprozeß, der durch drei Parameter, z.B. die mittlere Anzahl von Zellen pro Aktivitätsphase  $E[X]$ , die mittlere Pausendauer  $E[S]$  und den Zellabstand in der Aktivitätsphase  $\Delta$ , vollständig beschrieben werden kann. Nach Einführung der mittleren Ankunftsrate  $\lambda$  und der Raten  $\alpha$  für das Verlassen des aktiven Zustands und  $\beta$  für das Verlassen des Pausenzustands mit

$$\lambda = \frac{1}{E[A]}, \quad \alpha = \frac{1}{E[X] \cdot \Delta}, \quad \beta = \frac{1}{E[S]} \quad (3.44)$$

ergeben sich für die Kenngrößen des Ankunftsprozesses die in der Tabelle angegebenen Werte.

Durch entsprechende Umrechnung lassen sich auch andere Parametersätze für die Beschreibung des Modells verwenden, z.B.

<sup>2</sup>Je nach Lösungsmethode des verwendeten Warteschlangenmodells werden manchmal auch die Dauern der Aktivitätsphasen als negativ-exponentiell verteilt oder die Pausendauern als geometrisch verteilt angenommen

- Mittlere Zellrate, Burstiness-Faktor <sup>3</sup>, Variationskoeffizient oder
- Mittlere Zellrate, maximale Zellrate, mittlere Anzahl von Zellen pro Aktivitätsphase,

wobei die einzelnen Parameter jeweils unabhängig voneinander verändert werden können. Die obigen Parameter werden z.B. in der CCITT-Empfehlung I.311 [28] als relevante Parameter zur Beschreibung des statistischen Verhaltens einer Verbindung empfohlen.

In der Literatur wurde dieses Modell mit entsprechenden, teilweise aus Messungen gewonnenen, Parameterwerten als realistisches Modell für Dienste wie paketierte Sprache mit Pausenunterdrückung [46, 70, 148], interaktive Standbildübertragung [59, 61] und interaktive Datenübertragung verwendet.

Die Verteilung des Zählprozesses für den Ankunftsprozeß der sporadischen Quelle läßt sich wie folgt explizit angeben. Nach Gleichung 3.6 ergibt sich für die Laplace-Transformation der Verteilung des Zählprozesses  $n_{LT,i}(s)$

$$n_{LT,i}(s) = \begin{cases} \frac{1}{s} - \frac{\lambda}{s^2} + \frac{\lambda}{s^2} a_{LST}(s), & i = 0 \\ \frac{\lambda}{s^2} [a_{LST}(s)]^{i-1} - 2[a_{LST}(s)]^i + [a_{LST}(s)]^{i+1}, & i \geq 1 \end{cases} \quad (3.45)$$

Unter Verwendung von Gleichung 3.10 ergibt sich nach der Rücktransformation

$$n_i(t) = \begin{cases} 1 - \lambda t + \lambda \left[ t - \Delta - \frac{\alpha \Delta}{\beta} (1 - e^{-\beta(t-\Delta)}) \right] s(t - \Delta), & i = 0 \\ \lambda \left[ \mathcal{L}T^{-1} \left\{ \frac{[a_{LST}(s)]^{i-1}}{s^2} \right\} - 2\mathcal{L}T^{-1} \left\{ \frac{[a_{LST}(s)]^i}{s^2} \right\} \right. \\ \left. + \mathcal{L}T^{-1} \left\{ \frac{[a_{LST}(s)]^{i+1}}{s^2} \right\} \right], & i \geq 1 \end{cases} \quad (3.46)$$

Durch Anwendung des Binomischen Satzes auf  $[a_{LST}(s)]^n$ , Rücktransformation und anschließende Integration läßt sich zeigen, daß gilt:

$$\mathcal{L}T^{-1} \left\{ \frac{[a_{LST}(s)]^n}{s^2} \right\} = \left[ \sum_{k=0}^n \binom{n}{k} (1 - \alpha \Delta)^{n-k} (\alpha \Delta)^k \times \right. \\ \left. \left( t - n\Delta - \frac{1}{\beta} \sum_{i=1}^k \left( 1 - e^{-\beta(t-n\Delta)} \left( \sum_{j=0}^{i-1} \frac{(\beta(t-n\Delta))^j}{j!} \right) \right) \right) \right] s(t - n\Delta) \quad (3.47)$$

womit  $n_i(t)$  vollständig bestimmt ist. Aus dieser Verteilung – die für eine endliche Intervallbreite  $t$  nur endlich viele Elemente besitzt – lassen sich alle Momente exakt berechnen.

<sup>3</sup>Der Burstiness-Faktor ist definiert als  $\frac{\text{Maximale Zellrate}}{\text{Mittlere Zellrate}}$

### 3.3.3 Markoff-modulierter Poisson-Prozeß (MMPP)

Der Markoff-modulierte Poisson-Prozeß (engl. Markov Modulated Poisson Process, MMPP) ist ein doppelt-stochastischer Poisson-Prozeß mit  $n$  Zuständen, wobei der Zustandsprozeß durch eine zeitkontinuierliche Markoff-Kette bestimmt ist und die Ankunftsraten vom jeweiligen Zustand abhängen. Ein Beispiel für einen zweiphasigen MMPP ist in Abbildung 3.5 dargestellt. Der MMPP ist kein Erneuerungsprozeß. Dieser Prozeß ist ein Spezialfall des all-

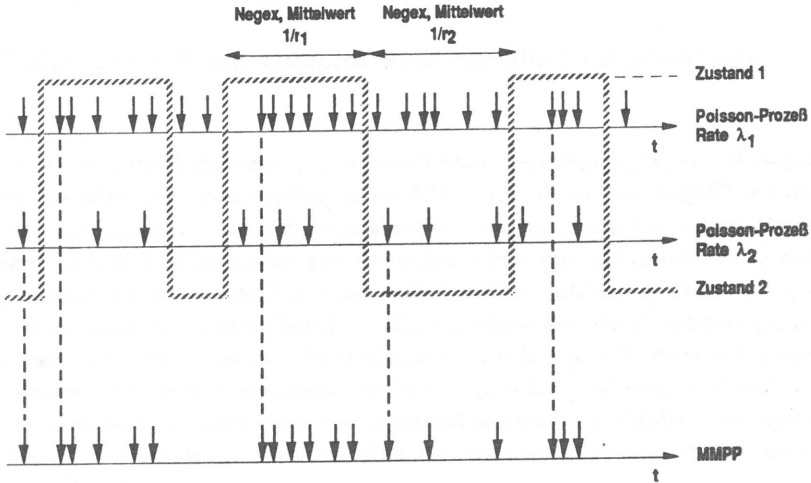


Bild 3.5: Beispiel für einen MMPP mit 2 Phasen

gemeinen geschalteten Poisson-Prozesses (engl. Switched Poisson Process, SPP) und enthält den unterbrochenen Poisson-Prozeß (engl. Interrupted Poisson Process, IPP) als Sonderfall [156]. Der zweiphasige MMPP ist durch vier Parameter gekennzeichnet, nämlich die mittleren Verweildauern in den beiden Zuständen,  $r_1^{-1}$  und  $r_2^{-1}$ , sowie die Ankunftsraten in den Zuständen,  $\lambda_1$  und  $\lambda_2$ .

Für die mittlere Ankunftsrate  $\lambda$  und den Variationskoeffizienten der Zwischenankunftsabstände  $c_a$  gilt (siehe z.B. [70])

$$\lambda = \frac{\lambda_1 r_2 + \lambda_2 r_1}{r_1 + r_2}, \quad c_a = \sqrt{1 + \frac{2(\lambda_1 - \lambda_2)^2 r_1 r_2}{(r_1 + r_2)^2 (\lambda_1 r_2 + \lambda_2 r_1)}} \quad (3.48)$$

In [70] ist auch die Ableitung des IDC und anderer Größen aus der Erzeugenden Funktion des Zählprozesses dargestellt.

Der MMPP wird in der Literatur mehrfach (z.B. in [70, 138]) als Approximation für den Überlagerungsprozeß von ATM-Verkehrsströmen verwendet, weil er auf der einen Seite die Korrelationseigenschaften dieses Überlagerungsprozesses nachbildet, und auf der anderen Seite die Lösung der entsprechenden Warteschlangenmodelle mit analytischen Methoden (siehe Abschnitt 4.2.1.3) erlaubt. Der MMPP weist jedoch auf der Zellebene keine Periodizität auf, wie sie z.B. bei der Überlagerung von sporadischen Quellen zu beobachten ist. Diese Periodizität hat einen signifikanten Einfluß auf das Verlustverhalten der Verkehrsmodelle [131], weshalb das MMPP-Modell nur bedingt einsetzbar ist.

### 3.3.4 Allgemein modulierter deterministischer Prozeß (GMDP)

Der allgemein modulierte deterministische Prozeß (engl. Generally Modulated Deterministic Process, GMDP) [102] ist wie der MMPP ein doppelt-stochastischer Nichterneuerungs-Prozeß. Er basiert auf einem endlichen Automatenmodell mit – im allgemeinen Fall –  $m$  Zuständen. Der Zustandsprozeß wird durch eine Übergangsmatrix  $P = (p_{ij})$  beschrieben, wobei  $p_{ij}$  die Wahrscheinlichkeit für einen Übergang vom Zustand  $i$  in den Zustand  $j$  angibt. In den einzelnen Zuständen werden die Zellen in deterministischen, zustandsspezifischen Abständen  $d_i$  erzeugt. Die Anzahl der in einem Zustand erzeugten Zellen kann einer allgemeinen Verteilung gehorchen, und es sind auch Pausenzustände ohne Ankünfte erlaubt. Ein Beispiel für einen GMDP mit 2 aktiven Zuständen und einem Pausenzustand ist in Bild 3.6 dargestellt. Die Berechnung der Momente für die Zwischenankunftsabstände ist im allgemei-

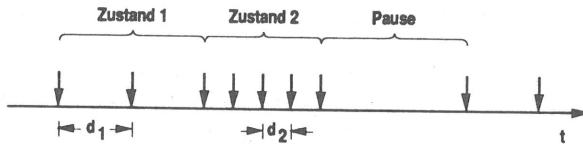


Bild 3.6: Beispiel für einen GMDP mit 3 Phasen

nen relativ aufwendig und erfolgt durch wiederholtes Ableiten der Erzeugenden Funktion, die in [154] hergeleitet und explizit angegeben ist. Falls die Anzahl von Zellen in allen Zuständen geometrischen Verteilungen gehorcht, kann auch das Korrelationsverhalten des Prozesses analytisch untersucht werden [17].

Das GMDP-Modell bildet das Verhalten eines korrelierten ATM-Zellstromes mit einer über der Zeit variablen Zellrate nach, wie er z.B. bei Videocodecs mit variabler Ausgangsbitrate beobachtet werden kann. Ein Problem dieses Quellmodells liegt – wie bei allen auf allgemeinen Zustandsautomaten basierenden Modellen – in der mit der Anzahl der Zustände



sehr schnell anwachsenden Zahl von Parametern und deren Belegung mit sinnvollen Werten. Ansätze für eine Anpassung der GMDP-Parameter an Meßwerte für einen Videocodec sind z.B. in [42] zu finden.

### 3.3.5 Autoregressives Modell für paketierte Videoübertragung

Um eine effektive digitale Übermittlung von Videosignalen zu ermöglichen, wird versucht, die benötigte Datenrate durch Verringerung der redundanten und irrelevanten Signalanteile zu reduzieren. Dabei kommen Methoden wie die Differenz-Pulsmodulation (DPCM) und die Transformationscodierung zur Anwendung. Bei der DPCM werden statistische Abhängigkeiten zwischen geometrisch (innerhalb eines Bildes, engl. intraframe) oder zeitlich (zwischen aufeinanderfolgenden Bildern, engl. interframe) benachbarten Bildpunkten ausgenutzt. Bei der Transformationscodierung werden die PCM-codierten Signale z.B. mittels der diskreten Cosinus-Transformation (DCT) in den Spektralbereich transformiert, wo etwa durch Weglassen von subjektiv weniger wichtigen (hohen) Spektralanteilen sowie Codierung mit variabler Codewortlänge eine Datenreduktion erreicht werden kann. Eine Übersicht über die verschiedenen Methoden wird z.B. in [115, 117] gegeben. Von Natur aus haben die so entstehenden Digitalsignale eine zeitlich variable Bitrate, durch Pufferung im Codec und – falls nötig – zeitweise Verminderung der Codierungsqualität können aber auch Signale mit einer reduzierten, konstanten Bitrate erzeugt werden. Durch die Kombination von Datenreduktion und fester Ausgangsbitrate ergeben sich Schwankungen in der Bildqualität, die sich besonders bei schnell bewegten Szenen oder bei Szenenwechseln bemerkbar machen. Durch den Einsatz von Videocodecs mit variabler Ausgangsbitrate (engl. Variable Bitrate Video Codec, VBR Video Codec), wie er in zukünftigen ATM-Netzen möglich sein wird, können solche Qualitätsschwankungen weitgehend vermieden werden, da bei Bedarf kurzfristig eine sehr hohe Übertragungskapazität in Anspruch genommen werden kann [110].

Die von solchen Quellen erzeugten Zellströme weisen – im Gegensatz zu der Ein/Aus-Charakteristik von z.B. Sprachquellen – zeitlich variable Bitraten ohne wirkliche Pausen sowie ausgeprägte Korrelationen in verschiedenen Zeitebenen auf. Zum einen sind Korrelationen zwischen den einzelnen Bildern (engl. Frames) im Sub-Sekundenbereich zu beobachten, zum anderen zeigen sich auch Korrelationen im Bereich von einigen Sekunden, die durch die Dauer einzelner Szenen bestimmt sind.

Um dieses Verhalten nachzubilden, haben Maglaris et al. ein autoregressives Markoff-Modell [110] vorgeschlagen. Die verwendeten Zahlenwerte für die Modellparameter wurden aus Meßwerten gewonnen. Die Messung umfaßte eine Dauer von etwa 10 Sekunden oder 300 Bildern, wobei eine typische Bildschirmtelefonszene mit wenig Bewegung zugrunde gelegt wurde. Der verwendete Codec war ein interframe DPCM-Codec mit bedingter Auffrischung (engl. Conditional Replenishment). Der Codec erzeugte 30 Bilder pro Sekunde, wobei in jedem Bild

250000 Bildpunkte (engl. Pixel) zu codieren waren. Bei einem gemessenen Mittelwert von 0.52 Bit/Pixel ergab sich eine mittlere Bitrate von 3,9 Mbit/s. Die maximale und die minimale Bitrate (gemittelt über ein Bild) ergaben sich in der Messung zu 10,575 Mbit/s bzw. 600 kbit/s.

Das Modell beschreibt die Änderung der momentanen Bitrate von Bild zu Bild, wobei das Verhalten innerhalb eines Bildes nicht betrachtet wird, weil das Meßintervall gleich der Dauer eines Bildes gewählt wurde. Die Bitrate im  $n$ -ten Bild ergibt sich aus derjenigen im  $(n-1)$ -ten als

$$\lambda(n) = a\lambda(n-1) + bw(n) \quad (3.49)$$

Die Parameter  $a$  und  $b$  sind hierbei konstante Gewichtungsfaktoren und  $w(n)$  ist eine Gaussverteilte Zufallsvariable mit Mittelwert  $\eta$  und Varianz 1, die die Änderung der Bitrate bewirkt. Durch Anpassung der mittleren Bitrate und eine exponentielle Näherung für die Autokovarianz (siehe Abschnitt 3.2.3) ergeben sich für die Parameter

$$a = 0,8781, \quad b = 0,1108 \text{ Bit/Pixel}, \quad \eta = 0.572 \quad (3.50)$$

Für Untersuchungen von Quellflußkontrollverfahren ist dieses Modell von der zeitlichen Auflösung her zu grob, da für ein Bild bis zu etwa 920 Zellen erzeugt werden. Deshalb wurden für diese Untersuchungen wie in Bild 3.7 dargestellt zwei extreme Fälle für das Verhalten des Zellstromes innerhalb eines Bildes definiert:

- Die Zellen werden der Bitrate des jeweiligen Bildes entsprechend in deterministischen Abständen erzeugt. Dies ist der gleichmäßigste Zellstrom, der mit diesem Modell erzeugt werden kann. Aus der Simulation ergibt sich für diese Variante ein Variationskoeffizient der Zwischenankunftsabstände von etwa 2.
- Die Zellen werden äquidistant, der maximalen Bitrate des Codecs entsprechend, am Anfang eines Bildes gesendet. Daraus ergibt sich ein büschelförmiger Zellstrom mit einer Aktivitäts- und einer Pausenphase pro Bild. Diese Variante des Modells liefert einen Variationskoeffizienten von ca. 10,6.

Es kann erwartet werden, daß das tatsächliche Kurzzeitverhalten des Codecs zwischen beiden Extremen liegt, wobei die erste, für das Netz günstigere Variante z.B. durch entsprechende Pufferung im Codec realisiert werden könnte. Das oben beschriebene Modell ist nur für die simulative Behandlung geeignet, deshalb wurde in [110] ein zweites Modell vorgeschlagen, welches sich näherungsweise analytisch behandeln läßt. Bei diesem kann sich die momentane Bitrate nur in diskreten Stufen ändern, wobei die Bitratenänderung durch einen eindimensionalen Geburts- und Sterbeprozess festgelegt ist. In der Literatur wurden in der letzten Zeit weitere, zum Teil sehr ähnliche Modelle und Erweiterungen vorgeschlagen

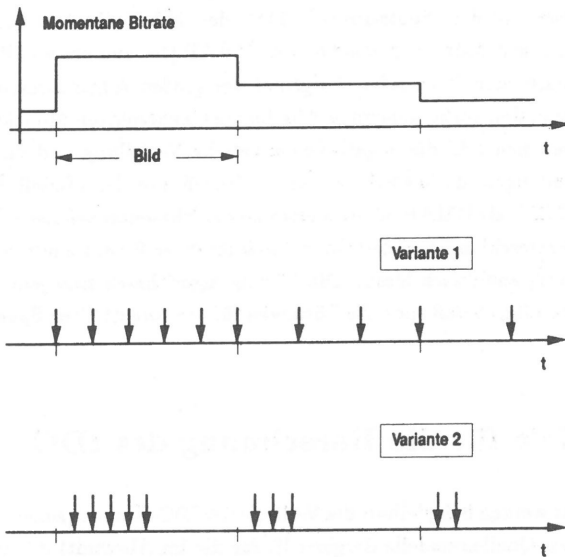


Bild 3.7: Varianten für das Kurzzeitverhalten eines Videocodex

[52, 77, 121, 124, 143, 144], die zum Teil auf eigenen Messungen und zum Teil auf den Daten in [110] basieren.

Allgemein ist zu sagen, daß die Eigenschaften des Zellstromes sowohl stark vom Codieralgorithmus als auch vom Bildinhalt abhängig sind. Die anderen Experimente, wie z.B. in [77], haben deshalb zum Teil signifikant andere Meßergebnisse erbracht, die sich in anderen Modellparametern niederschlagen. Es ist also nicht möglich, ein allgemeingültiges Modell für die Videokommunikation anzugeben, das verwendete autoregressive Modell zeigt jedoch zumindest einige typische Eigenschaften eines solchen Verkehrs. Die autoregressive Struktur des Prozesses sowie die Tatsache, daß im Mittel pro gezogener Zufallszahl etwa 340 Zellen erzeugt werden, machen simulative Untersuchungen mit diesem Modell sehr aufwendig.

### 3.3.6 Andere Modelle für ATM-Verkehrsströme

Zusätzlich zu den oben beschriebenen, allgemein verwendeten Modellen wurden in der Literatur noch weitere Modelle für ATM-Verkehrsströme vorgeschlagen. Darunter befinden sich z.B. das Packet Train-Modell [85] für Datenverkehr sowie das Filter-Modell [66], bei dem durch Filterung mit Hilfe eines Transversalfilters aus weißem Rauschen ein Ankunftsprozeß mit vorgebarbarer Korrelationsstruktur erzeugt wird. Diese Modelle sind nur für simulative

Untersuchungen anwendbar. Daneben gibt es eine Gruppe von Modellen, die sehr allgemeine Markoffsche Prozesse wie den Neuts-Prozeß [118], den MAP (General Markovian Arrival Process) [107], dessen zeitdiskretes Analogon, den DMAP [16] und ein von Ramaswami [130] vorgeschlagenes Phasenmodell umfaßt. Aufgrund der großen Allgemeinheit dieser Modelle lassen sich viele der in den vorhergehenden Abschnitten eingeführten Modelle als Sonderfälle darstellen. So lassen sich z.B. die negativ-exponentielle Verteilung und die aus ihr zusammengesetzten Verteilungen als MAP bzw. Neuts-Prozeß und das Modell der sporadischen Quelle sowie der GMDP als DMAP oder als allgemeiner Phasenprozeß nach Ramaswami darstellen. Einfache Warteschlangenmodelle lassen sich für diese Prozesse mit den Methoden der Matrixgeometrie [119] analytisch lösen. Die Lösungsalgorithmen sind jedoch sehr komplex und rechenzeitaufwendig, so daß spezielle Lösungen für die vorgestellten Spezialfälle meistens vorteilhafter sind.

### 3.4 Beispiele für die Berechnung des IDC

In diesem Abschnitt werden beispielhaft die Verläufe der IDC-Kurven (siehe Abschnitt 3.2.2.2) für einige derjenigen Quellenmodelle dargestellt, für die im Abschnitt 3.3 explizite Formeln für die Verteilung des Zählprozesses angegeben wurden. Durch Vergleiche mit den Ergebnissen der direkten Berechnung wird die numerische Genauigkeit des in Abschnitt 3.2.2.1 vorgestellten zeitdiskreten Algorithmus abgeschätzt.

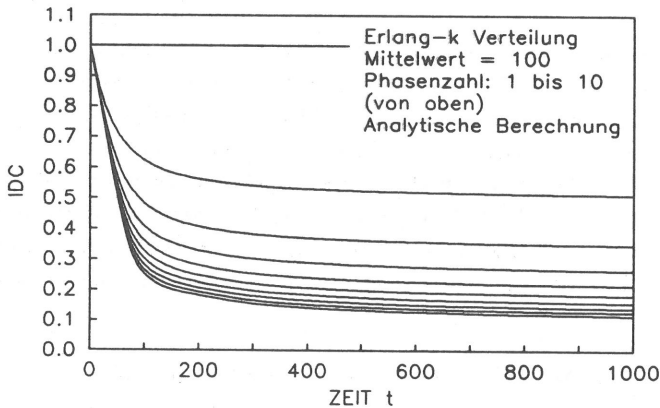


Bild 3.8: Verlauf des IDC für Erlang-k-Verteilungen mit verschiedenen Phasenzahlen

In Bild 3.8 ist der Verlauf der IDC-Kurven für Erlang-k-Verteilungen mit verschiedenen Phasenzahlen dargestellt. Für eine Phasenzahl von 1 ergibt sich die negativ-exponentielle Verteilung und entsprechend ein IDC von konstant 1. Bei mehreren Phasen zeigen die Kurven

eine monoton fallende Tendenz, wobei der Grenzwert nach Gleichung 3.24 jeweils durch  $1/k$  gegeben ist.

Bild 3.9 zeigt den Einfluß des Variationskoeffizienten der Ankunftsabstände auf den Verlauf des IDC am Beispiel der Negativ-Binomial-Verteilung. Die Kurven für dieses Bild wurden mit Hilfe des zeitdiskreten Algorithmus berechnet. Wie erwartet nähern sich die Kurven asymptotisch dem Quadrat des jeweiligen Variationskoeffizienten. Die Konvergenzgeschwindigkeit nimmt mit steigendem Variationskoeffizienten ab.

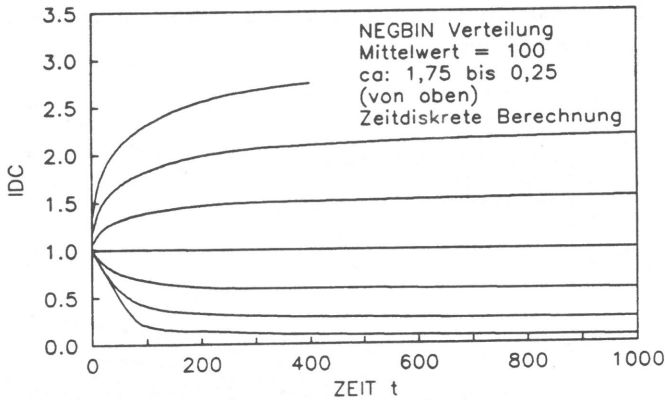


Bild 3.9: Verlauf des IDC in Abhängigkeit vom Variationskoeffizienten

Der Verlauf für den IDC der zweiphasigen sporadischen Quelle wird in Bild 3.10 dargestellt. Als Basisparameter wurden die in [70] angegebenen Werte verwendet, wie sie sich für eine paketierte Sprachübertragung mit Pausenunterdrückung bei Verwendung der adaptiven Puls-Code-Modulation (ADPCM) ergeben. Dabei wird alle  $125\mu\text{s}$  eine Abtastprobe von 4 Bit Länge erzeugt, wodurch sich bei der in [70] zugrunde gelegten Informationsfeldlänge von 64 Byte in der Aktivitätsphase ein Zellabstand von 16ms ergibt. Mit einer mittleren Anzahl von 22 Zellen pro Aktivitätsphase und einer mittleren Pausendauer von 650ms ergibt sich ein mittlerer Zellabstand  $E[A]$  von etwa  $45,54\text{ms}$ , ein Variationskoeffizient von etwa 4,2538 und ein Burstiness-Faktor von etwa 2,8466. Bei der in Bild 3.10 dargestellten Variation des minimalen Zellabstandes wurde der mittlere Ankunftsabstand konstant gehalten. Um den Variationskoeffizienten des Ankunftsabstandes  $c_a$  ebenfalls konstant zu halten, wurden die mittlere Anzahl von Zellen pro Aktivitätsphase  $E[X]$  und die mittlere Pausendauer  $E[S]$

entsprechend der Vorschrift

$$E[X] = \frac{1}{2} \left( \left( \frac{E[A]c_a}{E[A] - \Delta} \right)^2 + 1 \right), \quad E[S] = (E[A] - \Delta) E[X] \quad (3.51)$$

gewählt. Dadurch nähern sich alle Kurven – allerdings mit verschiedener Konvergenzgeschwindigkeit – dem gleichen Grenzwert 18,1 an. Der Minimalabstand zwischen den Zellen äußert sich in dem periodischen Verhalten, das am Anfang der Kurven besonders ausgeprägt ist und im weiteren Verlauf durch den Einfluß der negativ-exponentiellen Pausenphasen verwischt wird. Der Abfall der Kurven im ersten, dem jeweiligen minimalen Zellabstand entsprechenden Teil der einzelnen Kurven kommt durch den zufälligen Zeitpunkt für den Beginn des Beobachtungsintervalls zustande.

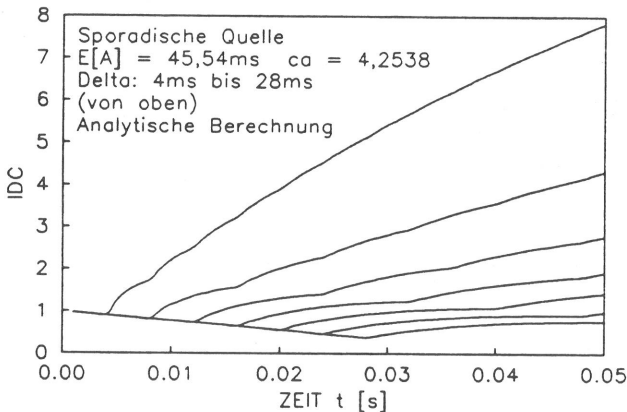


Bild 3.10: Einfluß des minimalen Zellabstands auf den IDC der sporadischen Quelle

Die folgenden Tabellen zeigen einige Vergleiche, die es erlauben, die Genauigkeit der verschiedenen Algorithmen zur Berechnung der Verteilung des Zählprozesses – und damit des IDC – zu beurteilen. In Tabelle 3.12 sind die mit den entsprechenden Programmen berechneten Werte des IDC der negativ-exponentiellen Verteilung dargestellt, der nach Abschnitt 3.3.1.1.1 konstant 1 sein muß. Die Spalte, deren Werte mit dem in Abschnitt 3.3.1.1.3 vorgestellten Algorithmus zur Berechnung des Zählprozesses für die Erlang-k-Verteilung berechnet wurden, zeigt eine sehr gute Übereinstimmung mit Abweichungen kleiner  $10^{-8}$  über den ganzen Bereich. Die Abweichungen kommen dadurch zustande, daß die Auswertung abgebrochen wird, wenn die Summe der Elemente der Zählprozeßverteilung um weniger als  $10^{-10}$  von eins abweicht. Die einzelnen Verteilungselemente sind unabhängig von den Parametern bis auf mindestens 13 signifikante Stellen genau.

Die nächste Spalte zeigt die Abweichungen, die man erhält, wenn man den durch Gleichung 3.20 definierten zeitdiskreten Algorithmus anwendet. Als Ankunftsprozeß dient dabei

$E[A]$	$t$	Direkt	IDC $I(t)$		
			Erlang-1, analytisch	M diskretisiert, zeitdiskret	NEGBIN, $c_a = 1$ , zeitdiskret
1	1	1	$-8,38 \cdot 10^{-9}$	$+2,55 \cdot 10^{-1}$	$+1,70 \cdot 10^{-1}$
	10		$-2,73 \cdot 10^{-9}$		$+1,82 \cdot 10^{-2}$
	100		$+3,50 \cdot 10^{-9}$		$+1,82 \cdot 10^{-3}$
10	1	1	$-7,40 \cdot 10^{-9}$	$+2,50 \cdot 10^{-3}$	$+6,78 \cdot 10^{-2}$
	10		$-8,38 \cdot 10^{-9}$		$+2,73 \cdot 10^{-2}$
	100		$-2,73 \cdot 10^{-9}$		$+3,17 \cdot 10^{-3}$
100	1	1	$-1,65 \cdot 10^{-9}$	$+2,35 \cdot 10^{-5}$	$+9,26 \cdot 10^{-3}$
	10		$-7,40 \cdot 10^{-9}$	$+2,46 \cdot 10^{-5}$	$+7,52 \cdot 10^{-3}$
	100		$-8,38 \cdot 10^{-9}$	$+2,40 \cdot 10^{-5}$	$+2,85 \cdot 10^{-3}$

Tabelle 3.12: Genauigkeit der Algorithmen zur Berechnung des IDC am Beispiel der negativ-exponentiellen Verteilung

eine nach dem in [86] beschriebenen Verfahren diskretisierte negativ-exponentielle Verteilung. Es ist zu erkennen, daß die Übereinstimmung mit dem erwarteten Wert von eins vor allem von der Diskretisierungsgenauigkeit abhängt, die mit zunehmendem Mittelwert besser wird. Die relative Abweichung der einzelnen Verteilungselemente liegt nahezu unabhängig vom absoluten Wert der Elemente (bei einem Mittelwert von 100) bei etwa  $10^{-4}$  und ist somit bei geeigneter Diskretisierung ausreichend klein. Bei Verwendung der NEGBIN-Verteilung mit Variationskoeffizient 1 rührt die Abweichung primär von der Approximation durch einen anderen Verteilungstyp her und der gewählte Mittelwert spielt nur eine untergeordnete Rolle.

Tabelle 3.13 zeigt den Vergleich des zeitdiskreten Algorithmus für die  $GEO_1$ -Verteilung mit den nach Abschnitt 3.3.1.2.2 zu erwartenden exakten Werten. Hier tritt kein Diskretisierungsfehler auf und die Größe der Abweichung wird vor allem dadurch bestimmt, daß der Algorithmus abgebrochen wird, wenn die Summe der Elemente um weniger als  $10^{-10}$  von eins abweicht. Bei der Erzeugung des Verteilungsvektors sowie nach jeder Faltung werden außerdem alle Elemente eliminiert, die kleiner als  $10^{-10}$  sind. Die einzelnen Elemente der Verteilung des Zählprozesses zeigen beim Vergleich mit denen der entsprechenden Binomialverteilung keine signifikanten Abweichungen.

Die Anwendbarkeit des zeitdiskreten Algorithmus für die Untersuchung der zweiphasigen sporadischen Quelle wird in Tabelle 3.14 gezeigt. Die Genauigkeit hängt hier vor allem vom Fehler bei der Diskretisierung der negativ-exponentiellen Pausendauer ab, die ebenfalls nach dem Verfahren von Jans [86] vorgenommen wird. Das Optimum für die Diskretisierung liegt bei 2ms pro Zeiteinheit. Bei einer größeren Diskretisierung mit einer Diskretisierungseinheit von 4ms bzw. 16ms ergeben sich geringfügig schlechtere Werte.

$E[A]$	$t$	IDC $I(t)$	
		Direkt	Zeitdiskret
2	1	0,50	$-6,28 \cdot 10^{-10}$
	10		$-4,95 \cdot 10^{-9}$
	100		$-9,33 \cdot 10^{-8}$
10	1	0,90	$-4,32 \cdot 10^{-10}$
	10		$-1,03 \cdot 10^{-8}$
	100		$-3,28 \cdot 10^{-7}$
100	1	0,99	$-3,05 \cdot 10^{-10}$
	10		$-6,08 \cdot 10^{-8}$
	100		$-6,59 \cdot 10^{-7}$

Tabelle 3.13: Genauigkeit der Algorithmen zur Berechnung des IDC am Beispiel der  $GEO_1$ -Verteilung

$t [ms]$	Diskretisierungseinheit [ms]	IDC $I(t)$	
		Analytisch	Zeitdiskret
16	1	$6,48702 \cdot 10^{-1}$	$-9,54 \cdot 10^{-8}$
	2		$-1,16 \cdot 10^{-7}$
	4		$-3,72 \cdot 10^{-7}$
	16		$-5,75 \cdot 10^{-6}$
96	1	$3,40317 \cdot 10^0$	$-5,75 \cdot 10^{-5}$
	2		$-6,14 \cdot 10^{-7}$
	4		$-1,50 \cdot 10^{-6}$
	16		$+2,17 \cdot 10^{-5}$
496	1	$1,08445 \cdot 10^1$	$+2,94 \cdot 10^{-4}$
	2		$-4,98 \cdot 10^{-7}$
	4		$+4,38 \cdot 10^{-6}$
	16		$+8,58 \cdot 10^{-5}$
992	1	$1,40776 \cdot 10^1$	$-1,46 \cdot 10^{-2}$
	2		$+3,56 \cdot 10^{-5}$
	4		$-5,32 \cdot 10^{-5}$
	16		$-3,37 \cdot 10^{-4}$

Tabelle 3.14: Genauigkeit der Algorithmen zur Berechnung des IDC am Beispiel der Verteilung der sporadischen Quelle



Eine noch feinere Diskretisierung mit 1ms ergibt eine Verschlechterung der Werte. Diese kommt vor allem dadurch zustande, daß nach jeder Faltung der Ergebnisvektor aus Speicherplatzgründen auf eine maximale Länge ( $2^{15}$  Elemente) abgeschnitten und neu normiert wird. Bei einer sehr feinen Diskretisierung und entsprechend vielen Faltungen macht sich dieser Effekt signifikant bemerkbar.

Die Genauigkeit der Verteilungselemente ist nahezu unabhängig von deren Absolutwert, auch wenn dieser sehr klein wird. Deshalb kann der zeitdiskrete Algorithmus mit guter Genauigkeit für die Analysen in Kapitel 5 verwendet werden.

# Kapitel 4

## Systemmodelle und Analysemethoden

In diesem Kapitel werden die Systemmodelle und Analyseverfahren beschrieben, die für Untersuchungen an ATM-Netzen und ihren Komponenten geeignet sind. Insbesondere wird ausführlicher auf die in Kapitel 5 verwendeten Verfahren eingegangen.

### 4.1 Simulation

#### 4.1.1 Prinzip der zeitreuen Simulation

Die Rechnersimulation von Kommunikationssystemen und -mechanismen wird in der Regel dann eingesetzt, wenn die Komplexität der Modelle und Vorgänge eine rein analytische Behandlung des Problems nicht erlaubt.

Bei der in dieser Arbeit verwendeten zeitreuen Simulationstechnik (engl. Event-by-Event Simulation) [101] wird dazu ein Programm auf einem Digitalrechner erstellt, das sowohl die Strukturkomponenten des Systems als auch den angebotenen Verkehr und das aufgrund der systeminternen Mechanismen erzeugte dynamische Verhalten hinreichend genau nachbildet. Zeitspannen für die zeitbehafteten Vorgänge (z.B. Bedienzeiten oder Zwischenankunftsabstände) werden dabei entsprechend ihrer statistischen Charakteristiken mit Hilfe von Pseudozufallszahlen erzeugt und zeitlich geordnet in einem Ereigniskalender gespeichert. Die im Ereigniskalender enthaltenen Ereignisse werden dann einzeln nacheinander abgearbeitet, wobei eine Variable den jeweils aktuellen Zeitpunkt auf der simulierten Zeitachse (Systemzeit) angibt. Bei der Abarbeitung eines Ereignisses werden alle damit verbundenen Änderungen des Systemzustandes in geeigneter Weise gespeichert. Zu den gespeicherten Änderungen

zählen sowohl diejenigen in den Zuständen der Betriebsmittel (Warteschlangen, Bedieneinheiten) als auch solche, die sich in den Zuständen der Protokollmechanismen ergeben, welche den internen Datenfluß steuern.

Weiterhin werden die relevanten Meßdaten erfaßt und zur späteren statistischen Auswertung gespeichert. Für diesen Zweck werden Zeitmarken verwendet, die z.B. die Erzeugungszeit einer Nachricht oder den Zeitpunkt für den Eintrag einer Nachricht in eine Warteschlange kennzeichnen. Aus der Differenz zwischen der Zeitmarke und der aktuellen Systemzeit können dann z.B. Warte- oder Durchlaufzeiten bestimmt werden.

Folgeereignisse, die gemäß der internen Abarbeitungsregeln initiiert werden, werden ebenfalls zur späteren Bearbeitung in den Ereigniskalender aufgenommen. Insbesondere wird bei der Abarbeitung eines externen Ankunftsereignisses das nächste Ereignis für den betreffenden Generator eingeplant, um den Fortgang der Simulation zu gewährleisten.

Der Rechenzeitbedarf der Simulation wird vorwiegend von der Anzahl der zu bearbeitenden Ereignisse und der dafür benötigten Zeit bestimmt, weil die Zeit zwischen den Ereignissen (in der sich der Systemzustand nicht ändert) übersprungen wird.

Die Simulation ist vom Prinzip her ein experimentelles Verfahren und die Genauigkeit der Ergebnisse hängt deshalb von der Anzahl von Stichproben für die einzelnen Messungen ab. Um die Aussagesicherheit des durch Simulation erhaltenen Schätzwertes beurteilen zu können, wird ein Simulationslauf normalerweise in Teiltests unterteilt, aus deren Einzelergebnissen neben den Endergebnissen auch sogenannte „Vertrauensintervalle“ berechnet werden können. Diese geben für eine vorgegebene Aussagesicherheit den Bereich um den Schätzwert an, in dem der wirkliche Wert liegt.

Bei der stationären Simulation ist zusätzlich zu den gemessenen Teiltests eine „Warmlaufphase“ nötig, um das simulierte System in den eingeschwungenen Zustand zu bringen.

#### **4.1.2 Simulationsbibliothek**

Im Rahmen der Arbeit wurde unter Verwendung einiger bereits vorhandener Programm-Module eine Programmbibliothek in der Programmiersprache PASCAL erstellt, mit deren Hilfe sich die verschiedenen, für die Untersuchungen benötigten Simulationsprogramme leicht konfigurieren lassen. Eine Übersicht über die verwendeten Module und ihre Verknüpfung untereinander ist in Bild 4.1 dargestellt. Ein Modul besteht dabei aus einem Definitionsteil, der die Vereinbarungen von Datenstrukturen und Prozeduren bzw. Funktionen enthält sowie aus einem Implementierungsteil, der Variablendefinitionen und die Anweisungsteile der Prozeduren und Funktionen enthält. Die Module sind nach funktionalen Gesichtspunkten untergliedert und stellen Einheiten dar, deren Implementierungsteile nach Änderungen unabhängig

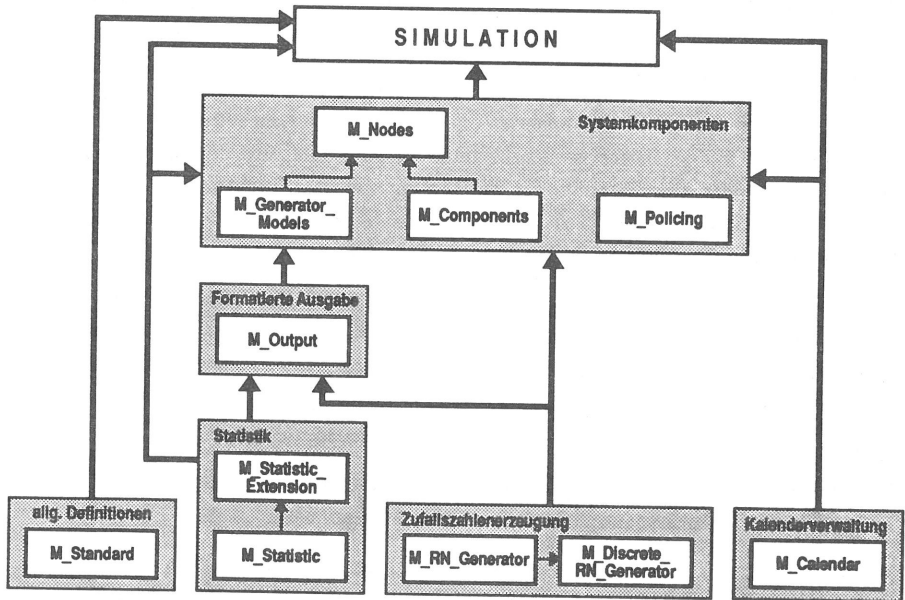


Bild 4.1: Die Module der Simulationsbibliothek

voneinander übersetzt werden können, solange sich die Schnittstellen zu den anderen Modulen nicht ändern. Die Pfeile zeigen an, welche Module die Definitionen von anderen Modulen übernehmen, wodurch sich eine hierarchische Struktur ergibt. Nach jeder Änderung des Definitionsteils eines Moduls müssen alle Module, die diese Definitionen übernehmen, neu übersetzt werden. Die Module in den unteren Ebenen werden in jeder Simulation gebraucht und sind allgemein verwendbar, während die Module für die Definition der Systemkomponenten Teile enthalten, die zu einem gewissen Grad spezifisch für bestimmte Simulationsmodelle sind.

Die allgemein verwendbaren Module sind

- *M\_Calendar*<sup>1</sup>: Kalendermodul, enthält Prozeduren für den Aufbau und die Verwaltung des Ereigniskalenders. Der Kalender ist in Form eines binären Baumes realisiert, um einen effektiven Zugriff auf die einzelnen Elemente zu erlauben. Im Rahmen der Arbeit wurde der Kalender so erweitert, daß bei Zeitgleichheit mehrerer Ereignisse – wie sie bei zeitdiskreten Systemen oft vorkommt (siehe Abschnitt 4.2.2.1) – eine Kontrolle der Ein- und Austragsreihenfolge durch zusätzliche Prioritäten möglich ist.

<sup>1</sup>Diese Module waren bereits vorhanden und wurden lediglich geringfügig angepaßt bzw. erweitert

- *M\_RN\_Generator*<sup>1</sup>: Pseudozufallszahlen-Generator für kontinuierliche Zufallsvariablen, enthält Routinen für die Erzeugung der Zufallszahlen. Basis für die Erzeugung der Pseudozufallszahlen sind gleichverteilte Zufallszahlen im Bereich  $[0, 1)$ , die durch den Aufruf eines vom Betriebssystem bereitgestellten Zufallszahlengenerators erzeugt werden. Wahlweise kann auch ein eigener Zufallszahlengenerator verwendet werden, der mit einer Kombination von zwei auf der Basis der multiplikativen Kongruenzmethode arbeitenden Generatoren implementiert ist und der eine sehr große Periodenlänge aufweist [56].
- *M\_Discrete\_RN\_Generator*: Pseudozufallszahlen-Generator für diskrete Zufallsvariablen, unterstützt die entsprechenden Funktionen für diskrete Verteilungen. Dieses Modul benutzt den Zufallszahlengenerator des kontinuierlichen Moduls mit, kann aber auch so konfiguriert werden, daß es ohne das kontinuierliche Modul funktioniert.
- *M\_Statistic*<sup>1</sup>: Statistikmodul, enthält Datenstrukturen und Routinen zur Erfassung und statistischen Auswertung von Meßwerten für
  - relative Häufigkeiten (Wahrscheinlichkeiten),
  - Zeitspannen,
  - Zählerstände und
  - zeitgemittelte Größen.
- *M\_Stat\_Extension*: Erweiterung des Statistikmoduls um die Messung von Häufigkeitsverteilungen und um die Messung des IDC.
- *M\_Standard*: Definition allgemein gebräuchlicher Variablen für die Simulationssteuerung wie z.B. Systemzeit, Zahl der Teiltests usw.

#### Die modellspezifischen Module umfassen

- *M\_Output*: Formatierte Ausgabe von Eingabeparametern und Meßergebnissen, enthält Routinen für die einheitliche Ausgabe in tabellarischer Form.
- *M\_Components*: Definition der Systemkomponenten wie Wartespeicher und Bedieneinheiten und Routinen für deren Verwaltung. Die Routinen enthalten die Meßwerterfassung für alle charakteristischen Größen, wobei neben Mittelwertmessungen auch Verteilungsmessungen implementiert sind. Da die Definition von verschiedenen Verkehrsklassen vorgesehen ist, werden auch sämtliche Messungen klassenindividuell durchgeführt. Der Umfang der Meßwerterfassung wird klassenindividuell durch Steuervariablen kontrolliert, so daß die gewünschten Messungen über die Eingabedaten selektiert werden können.
- *M\_Generator\_Models*: Definition der in Abschnitt 3.3 vorgestellten Verkehrsquellenmodelle und deren Verwaltung. Dieses Modul macht sowohl die kontinuierlichen und

diskreten Standardverteilungen als auch die aus ihnen zusammengesetzten, komplexeren Quellenmodelle über einheitliche Prozeduren zugänglich.

- *M\_Policing*: Enthält Routinen für die in Kapitel 5 dargestellten Quellflußkontrollmechanismen, wobei die Meßwerterfassung für alle interessierenden Größen enthalten ist.

Um die Erstellung des eigentlichen Simulationsprogramms weiter zu erleichtern, wurde noch ein weiteres Modul (*M\_Nodes*) implementiert, das einen Netzknoten bestehend aus einer Bedieneinheit und einer Warteschlange mit Verbindungspunkten am Ein- und Ausgang des Knotens definiert. Diese Verbindungspunkte werden verwendet, um mit ebenfalls im Modul vorhandenen Prozeduren die einzelnen Knoten zu einfachen Netzen verbinden zu können (siehe Bild 4.2).

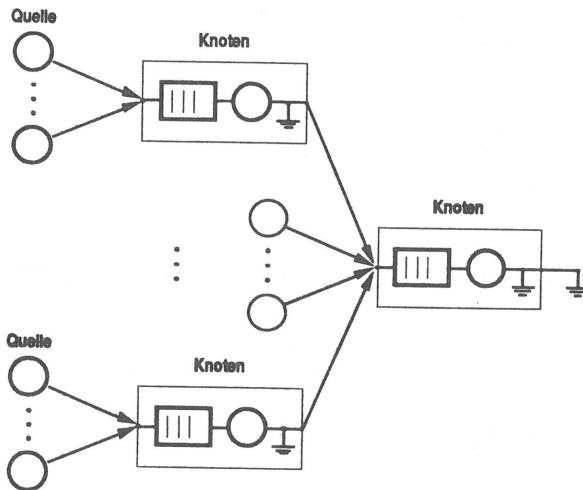


Bild 4.2: Beispiel für ein Simulationsmodell

Durch die Festlegung der Konfiguration im Eingabedatensatz wird eine weitestgehende Flexibilität des Simulationsprogramms gewährleistet.

Die Verbindungspunkte stellen gleichzeitig Meßpunkte dar, an denen auf Wunsch detaillierte, klassenindividuelle Messungen für die ankommenden und abgehenden Verkehrsströme vorgenommen werden können. Die in den Knoten dargestellten Verkehrsabzweigungen können wahlweise zufällig mit festen Wahrscheinlichkeiten oder auf der Basis von virtuellen Verbindungen erfolgen. Diese Funktion wurde bei nicht in der Arbeit dargestellten Untersuchungen von ATM-Koppelnetzen und von End-zu-End-Verzögerungen in ATM-Netzen benötigt.

Zusätzlich zu den in Bild 4.2 dargestellten Konfigurationsmöglichkeiten können die einzelnen Quellen individuell durch die verschiedenen Quellflußkontrollmechanismen überwacht werden.

Neben den bereits angesprochenen Routinen enthalten alle Module jeweils spezifische Ein- und Ausgaberoutinen sowie Routinen zur Initialisierung der entsprechenden Datenstrukturen. Weiterhin sind Routinen zum definierten Zurücksetzen der Meßwerterfassung bzw. der sonstigen Variablen z.B. nach der Warmlaufphase oder für die instationäre Simulation enthalten.

Bei Verwendung der angesprochenen Module reduziert sich der Aufwand für das Erstellen eines Simulationsprogramms auf die Organisation des Simulationsablaufs. Der aufwendigste Teil besteht dabei in der Definition der Aktionen für die Bearbeitung der verschiedenen Kalenderereignisse (Ankünfte, Bedienendeereignisse und Ereignisse für die Simulationssteuerung).

Um die Simulation möglichst effektiv zu gestalten, wurden bei der Implementierung der Module folgende Gesichtspunkte beachtet:

- Verwendung dynamischer Datenstrukturen (Zeigerstrukturen), um den Speicherplatzbedarf des Simulationsprogrammes zu minimieren. Solche Datenstrukturen werden nur dann erzeugt, wenn sie wirklich benötigt werden. Dadurch wird die Zahl der relativ aufwendigen Seitenaustausch-Operationen (engl. Paging) reduziert, die wegen des begrenzten Hauptspeichers bei der Verwaltung des virtuellen Speichers vorgenommen werden müssen.
- Anlegen von Stapeln (engl. Stacks) für Kalenderereignisse und Zellen, von denen die Elemente bei Bedarf entnommen und auf die sie nach dem Gebrauch wieder zurückgelegt werden. Dadurch wird die abwechselnde Generierung und Rückgabe der Elemente vermieden, die durch die benötigten Betriebssystem-Aufrufe relativ zeitaufwendig ist. Neue Elemente werden nur dann erzeugt, wenn auf einem Stapel keine Elemente mehr vorhanden sind.
- Weitergabe der Zellen innerhalb des simulierten Systems durch Änderung der entsprechenden Zeiger statt durch Umspeichern der gesamten Datenstruktur.

### 4.1.3 Anwendbarkeit simulativer Methoden

Beim Einsatz simulativer Methoden für Untersuchungen in ATM-Netzen ergeben sich einige prinzipielle Beschränkungen. Während bei der Simulation von durchschaltvermittelnden Systemen, wo die Ereignisse (Verbindungsanforderung, Verbindungsende) relativ große zeitliche Abstände haben, in der Regel ein Zeitraffereffekt erzielt werden kann, ergibt sich bei der Untersuchung von paketvermittelnden Hochgeschwindigkeitsnetzen durch die explizite

Betrachtung einzelner Zellen bzw. Pakete eine Zeitdehnung. Diese Zeitdehnung kann trotz der Rechenleistung heutiger Digitalrechner so extrem sein, daß mit vertretbarem Rechenaufwand nur wenige Sekunden Echtzeit simuliert werden können. Im Gegensatz zu dem Geschehen auf der Zell- bzw. Burstebene (siehe Abschnitt 3.1), das hauptsächlich durch die Eigenschaften des Systems bestimmt wird, richten sich die Ereignisse auf der Dialog- und Verbindungsebene weiterhin nach dem Benutzerverhalten und haben dadurch einen um mehrere Größenordnungen anderen Zeitmaßstab. Mit heutigen Mitteln ist es deshalb wegen dieser Dynamikunterschiede nicht möglich, die Ereignisse auf Verbindungs- und Zellebene mit einem einheitlichen Simulationsprogramm zu untersuchen und es werden statt dessen unterschiedliche Programme verwendet. Für die Untersuchungen auf der Zellebene wird dabei die Verbindungsebene als statisch („eingefroren“) betrachtet, was bei der kurzen simulierten Echtzeitspanne meistens eine zulässige Näherungsannahme ist. Bei den Untersuchungen auf Verbindungsebene wird meist versucht, die typischen Eigenschaften der Zellebene durch einfache Systemkomponenten, z.B. Bedieneinheiten mit entsprechender Bedienzeitcharakteristik, mit zu berücksichtigen (Aggregation).

Eine weitere Einschränkung ergibt sich durch die Bereiche, in denen die geforderten Ergebnisse liegen. Die zu messenden Dienstgüteparameter sind neben Mittelwerten und Verteilungen von Verzögerungen und Durchlaufzeiten vor allem Verlustwahrscheinlichkeiten. Für die Dimensionierung der Netzfunktionen und Komponenten sind hierbei Werte im Bereich von  $10^{-10}$  und sogar darunter im Gespräch. Da bei einem experimentellen Verfahren wie der Simulation eine bestimmte Anzahl von ausgezeichneten Ereignissen, z.B. Zellverlusten, erfaßt werden muß, um eine genügende Aussagesicherheit zu gewährleisten, liegen die für die Simulation solcher seltener Ereignisse insgesamt benötigten Ereigniszahlen um Größenordnungen höher als die mit heutigen Mitteln erreichbaren. Mit vertretbarem Aufwand können in der Regel bei den untersuchten Modellen nur Verlustwahrscheinlichkeiten bis  $10^{-5}$  einigermaßen verläßlich ermittelt werden. In solchen Fällen wird die Simulation oft mit unrealistischen Parametern durchgeführt, um zumindest das prinzipielle Verhalten der Komponenten und Mechanismen zu bestimmen oder um analytische Lösungen und Näherungsverfahren zu validieren. In vielen Fällen weisen die Kurven bei abnehmenden Verlustwahrscheinlichkeiten ab einem gewissen Punkt ein exponentiell abfallendes Verhalten auf, was eine relativ gute Genauigkeit bei der Extrapolation der Ergebnisse erlaubt.

Die Grenzen für die Anwendbarkeit simulativer Analysetechniken können durch den Einsatz immer leistungsfähigerer Rechner und durch parallele Simulation, z.B. mit Hilfe von Transputer-Systemen oder durch andere Methoden, wie etwa die Perturbations-Analyse [151], verschoben werden, ohne daß dabei jedoch das prinzipielle Problem gelöst wird. In letzter Zeit wurden auch spezielle Methoden für die Simulation seltener Ereignisse vorgeschlagen (z.B. Reverse Time Simulation), die aber noch nicht so weit entwickelt sind, daß sie allgemein anwendbar sind.



## 4.2 Analytische Modelle und Methoden

Bedingt durch die Beschränkungen der simulativen Methoden muß versucht werden, die Leistungsuntersuchungen mit exakten oder approximativen analytischen Methoden durchzuführen. Dabei ist es – eine genügende numerische Stabilität der Algorithmen vorausgesetzt – keine Schwierigkeit, die geforderten kleinen Verlustwahrscheinlichkeiten zu berechnen.

Ein Problem bei der analytischen Behandlung liegt darin, daß die Systeme im allgemeinen sehr viel gröber modelliert werden müssen als bei der Simulation, um sie den Analysetechniken zugänglich zu machen. So müssen z.B. oft die beteiligten Prozesse (Ankunfts- bzw. Bedienprozesse) als gedächtnisfreie oder zumindest als Erneuerungsprozesse modelliert werden. Auch die Protokolle und sonstigen Mechanismen können oft nur unvollständig nachgebildet werden. Bei Durchlaufzeit- und Koppelnetzuntersuchungen wird z.B. meistens die Verzweigung des Verkehrs als zufälliger Vorgang mit einer dem jeweiligen Verkehrsanteil entsprechenden Wahrscheinlichkeit modelliert, während sie in der Realität auf der Basis von virtuellen Verbindungen erfolgt. Durch diese Vereinfachung können sich signifikante Unterschiede in den Ergebnissen ergeben (siehe z.B. [135]).

Weiterhin sind auch bei solchen vereinfachten Modellen oft noch weitere Näherungen, z.B. Unabhängigkeitsannahmen, nötig, um eine analytische Behandlung zu ermöglichen. Es muß also, etwa mittels Simulation des detaillierten Modells in entsprechenden Parameterbereichen, der Einfluß der Näherungen bei der Modellierung und bei der Analyse abgeschätzt werden, um die Zuverlässigkeit der getroffenen Aussagen überprüfen zu können.

Für die in Kapitel 5 untersuchten, von Natur aus relativ einfachen Mechanismen eignen sich analytische Methoden sehr gut, weil die Modellierung und Analyse in den meisten Fällen exakt sein kann und sich im Vergleich zur Simulation ein entscheidender Rechenzeitgewinn erzielen läßt.

Bei der Untersuchung von ATM-Netzen sind Systeme mit einer deterministischen Bedienzeit ( $D$ ) von besonderer Bedeutung. Das rührt daher, daß die betrachteten Dateneinheiten, nämlich die Zellen, alle eine einheitliche Länge haben, so daß ihre Übertragung immer gleich lang dauert. Bei der Modellierung ergeben sich deshalb für Multiplexer und andere Netzkomponenten entsprechende Systemmodelle mit deterministischen Bedieneinheiten.

Für die Untersuchungen im nächsten Kapitel ist als Dienstgüteparameter vor allem die Zellverlustwahrscheinlichkeit interessant, weswegen im folgenden hauptsächlich auf die Warteverlustsysteme eingegangen wird, die aus einer Bedieneinheit und einem endlichen Wartepeicher bestehen. Die Zellverlustwahrscheinlichkeit entspricht dabei der Wahrscheinlichkeit, daß zum Ankunftszeitpunkt einer Anforderung das System, d.h. die Wartepplätze und die Bedieneinheit, voll belegt sind. Auf die Berechnung der anderen Systemcharakteristiken, wie z.B.

- Wartewahrscheinlichkeit,
- Mittelwert und Verteilung der Wartezeit,
- mittlere Warteschlangenlänge und
- Systemauslastung

wird aus Platzgründen nicht ausführlich eingegangen, da sie für die Untersuchungen im Kapitel 5 nicht benötigt werden. Diese Größen können wie z.B. in [100] dargestellt aus den Zustandswahrscheinlichkeiten oder zum Teil durch Mittelwertbetrachtungen berechnet werden. Die Wartezeiten ebenso wie die Durchlaufzeiten sind bei den Systemen mit deterministischer Bedienzeit und Abfertigung in Ankunftsreihenfolge (engl. First In First Out, FIFO) direkt über die konstante Bedienzeit mit den Zustandswahrscheinlichkeiten zum Ankunftszeitpunkt verknüpft.

Vom Prinzip her sind die betrachteten ATM-Netze zeitdiskrete Systeme, wobei die Zeiteinheit durch die Übertragungs- bzw. Bearbeitungszeit für eine Zelle gegeben ist. In vielen Fällen, z.B. bei der Untersuchung von statistischen Multiplexern, sind die Verhältnisse jedoch so, daß die kontinuierlichen Systeme eine relativ gute Näherung darstellen und oft verwendet werden, weil für sie eine ganze Reihe wohlbekannter Analysetechniken existieren.

Die im folgenden verwendete Kennzeichnung der Systeme basiert auf der Notation von Kendall [92] für einstufige Warteschlangenmodelle. Das Grundmodell wird bei dieser Notation durch den Typ von Ankunfts- und Bedienprozeß, die Anzahl der Bedieneinheiten und die Anzahl der Warteplätze charakterisiert. Ein Modell mit einem allgemeinem Erneuerungsprozeß (GI) als Ankunftsprozeß, allgemein verteilten Bedienzeiten (G),  $n$  Bedieneinheiten und  $s$  Warteplätzen wird in der Kurznotation als System vom Typ

$$GI/G/n-s$$

bezeichnet. Zur genaueren Bestimmung der beteiligten Prozesse werden die in Abschnitt 3.3 eingeführten Kurzbezeichnungen verwendet. Bei reinen Verlustsystemen wird die Anzahl der Warteplätze explizit mit 0 angegeben, bei reinen Wartesystemen mit einem unendlichen Wartespeicher wird die Anzahl der Warteplätze weggelassen. Falls nicht explizit eine andere Abarbeitungsstrategie für die Warteschlange angegeben ist, wird angenommen, daß die Anforderungen gemäß ihrer Ankunftsreihenfolge bedient werden (FIFO).

## 4.2.1 Systeme mit zeitkontinuierlichem Ankunftsprozeß

### 4.2.1.1 M/D/1-s

Dieses System wird oftmals als Modell für Komponenten innerhalb des Netzes verwendet, da sich bei diesen der Ankunftsprozeß aus einer Überlagerung sehr vieler Einzelquellen er-

gibt und angenommen wird, daß sich der Überlagerungsverkehr deshalb wie ein Poisson-Prozeß verhält. Bei einer genügenden Zahl mehr oder weniger gleichartiger Quellen ist diese Annahme gerechtfertigt, wobei die benötigte Zahl von Quellen sowohl von deren Charakteristika wie von der Auslastung der modellierten Systemkomponenten abhängt. Falls die Zahl überlagerter Quellen zu klein wird oder einige höherratige Quellen einen dominierenden Anteil zum Gesamtverkehr beitragen, ist diese Annahme nicht mehr gerechtfertigt und man erhält Ergebnisse, die um Größenordnungen von den wirklichen abweichen können. Dies gilt insbesondere bei Untersuchungen an Einzelquellen (siehe Abschnitt 4.3).

Durch die Gedächtnisfreiheit des Poisson-Ankunftsprozesses können für dieses System mit Hilfe einer zu den Bedienendezeitpunkten eingebetteten Markoff-Kette alle interessierenden Größen berechnet werden, da die PASTA-Eigenschaft (siehe Abschnitt 3.3.1.1.1) gilt. Die Berechnung der Zustandswahrscheinlichkeiten kann dabei für allgemeine Bedienzeitverteilungen durch Abschneiden und Umnormieren aus den bekannten Zustandswahrscheinlichkeiten für das reine Wartesystem (M/G/1) [65, 91] oder direkt durch die Anwendung der Methode der Supplementaritätsvariablen [108] erfolgen. Bei der Berechnung über das reine Wartesystem ergibt sich der Nachteil, daß eine Lösung nur für Verkehrsangebote kleiner eins existiert, weil sonst das reine Wartesystem instationär wird.

Eine weitere, sehr einfache Methode ist für den Sonderfall einer deterministischen Bedienzeit in [84] beschrieben. Die  $s$  ersten Gleichungen für die zeitgemittelten Zustandswahrscheinlichkeiten  $p_i$  sind die gleichen wie für die Zustandswahrscheinlichkeiten zum Bedienende, weil sich diese Wahrscheinlichkeiten nur durch einen konstanten Faktor unterscheiden. Setzt man das Gleichungssystem für den stationären Fall an und läßt die letzte, linear abhängige Gleichung außer acht, dann erhält man die  $s$  Gleichungen

$$\begin{aligned} p_0 &= p_0 n_0 + p_1 n_0 \\ p_1 &= p_0 n_1 + p_1 n_1 + p_2 n_0 \\ &\vdots \\ p_{s-1} &= p_0 n_{s-1} + p_1 n_{s-1} + p_2 n_{s-2} + \cdots + p_s n_0 \end{aligned} \tag{4.1}$$

wobei  $n_i$  die Wahrscheinlichkeit bezeichnet, daß innerhalb einer Bedienzeit  $i$  Ankünfte erfolgen. Diese Wahrscheinlichkeiten sind nach Abschnitt 3.3.1.1.1 durch eine Poisson-Verteilung mit Parameter  $\lambda D$  gegeben, wobei  $\lambda$  die Ankunftsrate und  $D$  die deterministische Bedienzeit repräsentiert. Die fehlenden Gleichungen für die Berechnung der  $s + 2$  zeitgemittelten Zustandswahrscheinlichkeiten erhält man aus der Normierungsbedingung

$$\sum_{i=0}^{s+1} p_i = 1 \tag{4.2}$$

und aus dem Erhaltungssatz, der besagt, daß der angebotene Verkehr  $A_{OL} = \lambda D$  gleich der Summe aus dem akzeptierten und dem abgewiesenen Verkehr ist. Diese Werte ergeben sich wegen der PASTA-Eigenschaft aus den Zustandswahrscheinlichkeiten zu

$$A_{OL} = 1 - p_0 + A_{OL} \cdot p_{s+1} \quad (4.3)$$

Das resultierende lineare Gleichungssystem kann entweder mit Standardmethoden direkt ausgewertet oder nach dem in [84] dargestellten Algorithmus rekursiv aufgelöst werden. Aus den zeitgemittelten Zustandswahrscheinlichkeiten ergeben sich alle charakteristischen Größen auf einfache Weise, insbesondere gilt wie oben erwähnt für die Verlustwahrscheinlichkeit  $p_v$

$$p_v = p_{s+1} \quad (4.4)$$

#### 4.2.1.2 M/D/n-0

Dieses System ist ein reines Verlustsystem ohne Wartespeicher. Die Zustandswahrscheinlichkeiten sind identisch zu denen des M/M/n-0-Systems, weil das System unempfindlich gegenüber der Bedienzeitverteilung ist [136]. Die Herleitung der Zustandswahrscheinlichkeiten  $p_i$  kann über einen einfachen Geburts- und Sterbeprozess erfolgen (siehe z.B. [100]), wobei sich Erlang's Formel für die Zustandswahrscheinlichkeiten

$$p_i = \frac{\frac{A_{OL}^i}{i!}}{\sum_{x=0}^n \frac{A_{OL}^x}{x!}} \quad (4.5)$$

ergibt, die für allgemeine Bedienzeiten gültig ist. Die Verlustwahrscheinlichkeit  $p_v$  ist wegen der PASTA-Eigenschaft durch die Wahrscheinlichkeit gegeben, daß alle Bedieneinheiten belegt sind.

$$p_v = p_n = \frac{\frac{A_{OL}^n}{n!}}{\sum_{x=0}^n \frac{A_{OL}^x}{x!}} \quad (4.6)$$

#### 4.2.1.3 Systeme mit allgemeinen Markoffschen Ankunftsprozessen

Auf der Basis eines allgemeinen Markoffschen Phasenprozesses lassen sich sehr allgemeine Ankunftsprozesse wie der MAP (General Markovian Arrival Process [107]) oder der noch

allgemeinere N-Prozeß (Neuts-Prozeß [118]) definieren, die im allgemeinen keine Erneuerungseigenschaft besitzen.

Systeme mit diesen Ankunftsprozessen und einer Bedieneinheit lassen sich für allgemeine Bedienzeitverteilungen mit den Methoden der Matrixgeometrie [15, 19, 106, 107, 118, 119, 120, 129] analytisch behandeln. Obwohl diese Ankunftsprozesse insgesamt keine Erneuerungseigenschaft besitzen, ist die Entwicklung ab einem bestimmten Zeitpunkt nur von der Phase abhängig, in der sich der Prozeß zu diesem Zeitpunkt befindet. Unter Einbeziehung des Phasenzustandes läßt sich deshalb – ähnlich wie beim M/G/1-System [65, 95, 120] – zu den Bedienenden eine Markoff-Kette einbetten. Aus dieser lassen sich dann im Prinzip die Zustandswahrscheinlichkeiten zu den Bedienendezeitpunkten berechnen. Um diese Berechnung effektiv vornehmen zu können, werden die entstehenden Gleichungen in Matrix-Form dargestellt und mit Hilfe von Matrixoperationen, wie z.B. Inversion, Multiplikation, Ableitung, Exponentiation, Faltung, Laplace-Transformation, Eigenwertbildung usw., unter Anwendung der entsprechenden Theoreme umgeformt. Mit Hilfe des „Key Renewal Theorem“ [40] werden aus den Zustandswahrscheinlichkeiten zum Bedienende die zeitgemittelten Zustandswahrscheinlichkeiten erzeugt. Durch die Elimination der Abhängigkeit vom Phasenzustand lassen sich dann die Zustandswahrscheinlichkeiten zu den Ankunftszeitpunkten und die Kenngrößen berechnen, wie z.B. die mittlere Warteschlangenlänge, die mittlere Wartezeit und – beim endlichen System – die Verlustwahrscheinlichkeit. Für die Lösung des endlichen Systems kann neben der direkten Methode [15], bei der die Größe der zu verarbeitenden Matrizen – und damit der Berechnungsaufwand – direkt von der Anzahl der Wartespeicher abhängt, auch ein Näherungsverfahren verwendet werden, bei dem die Zustandswahrscheinlichkeiten des endlichen Systems aus denen des reinen Wartesystems abgeleitet werden [19]. Dieses Verfahren ist für kleine Verlustwahrscheinlichkeiten ausreichend genau und hat den Vorteil, daß der Berechnungsaufwand nur von der Phasenzahl des Ankunftsprozesses abhängt. Deshalb erlaubt es die Berechnung beliebiger Wartespeichergößen, solange das Verkehrsangebot an das System kleiner ist.

Diese Lösungsverfahren wurden am Institut implementiert (siehe [19, 39]) und werden teilweise in dieser Arbeit für Untersuchungen an dem MMPP/D/1-s-System verwendet. Wegen der Komplexität der Verfahren und Ableitungen wird hier auf eine detaillierte Darstellung verzichtet und auf die entsprechenden Literaturstellen verwiesen.

#### 4.2.1.4 UAS-Modell

Ein in der Literatur oft verwendetes Modell ist das sogenannte „UAS“ Modell (Uniform Arrival and Service) [8, 46, 159], das manchmal auch als „Fluid Flow“ Modell bezeichnet wird. Bei diesem Modell wird sowohl die Gliederung des Ankunftsstromes als auch die des

abgehenden Verkehrsstromes in endliche Einheiten (Zellen) vernachlässigt, so daß sich für die Analyse ein kontinuierlicher Zu- bzw. Abfluß ergibt. Unter dieser Näherungsannahme kann das System mit deterministischer Bedienzeit für Ankunftsprozesse gelöst werden, wie sie von einer oder mehreren sporadischen Quellen oder – näherungsweise – von Videoquellen mit variabler Ausgangsbitrate erzeugt werden [110]. Die Analyse beruht auf der Lösung eines Systems von Differentialgleichungen und ergibt bezüglich der Verlustwahrscheinlichkeit eine relativ gute Näherung. Für das System mit einer sporadischen Quelle als Ankunftsprozeß<sup>1</sup> ergibt sich mit

$$k_1 = \frac{\alpha \Delta D}{D - \Delta}, \quad k_2 = \beta D \quad (4.7)$$

für die Verlustwahrscheinlichkeit im UAS-Modell [61, 159]

$$p_v = \frac{D - \Delta}{D} \cdot \frac{k_1 - k_2}{k_1 e^{[(k_1 - k_2)(s+1)]} - k_2} \quad (4.8)$$

wobei  $D$  die deterministische Bedienzeit bezeichnet und  $\alpha, \beta$  sowie  $\Delta$  durch die Parameter der sporadischen Quelle gegeben sind (siehe Abschnitt 3.3.2).

## 4.2.2 Systeme mit zeitdiskretem Ankunftsprozeß

### 4.2.2.1 Gleichzeitigkeit von Ankunft und Bediende

Bei Systemen mit zeitdiskretem Ankunftsprozeß tritt durch die Limitierung der möglichen Ereigniszeitpunkte auf ganzzahlige Vielfache der Basiszeiteinheit systematisch das Phänomen auf, daß gleichzeitig Ankunfts- und Bedienereignisse erfolgen. Dieses Phänomen ist auch bei wirklichen Implementierungen zu beobachten, z.B. beim Versuch von gleichzeitigen Lese- und Schreibzugriffen auf einen Speicher in einem taktgesteuerten System. Eine Möglichkeit zur Auflösung dieser Gleichzeitigkeit ist, daß logisch zunächst das Ankunftsereignis abgearbeitet wird und dann erst das Bedienendeereignis. Diese Strategie wird im folgenden als AF-Strategie (Arrival First) bezeichnet. Der umgekehrte Fall, in dem zunächst das Bediende abgearbeitet wird, wird entsprechend als DF-Strategie (Departure First) bezeichnet. Der Verlauf des Systemzustandes für die beiden Strategien ist in Bild 4.3 beispielhaft dargestellt. Die Ereigniszeitpunkte werden dabei symbolisch als Intervalle gezeichnet, um die im Modell gleichzeitig ablaufenden Ereignisse in ihrer logischen Reihenfolge darstellen zu können. Die Wahl der Strategie beeinflusst die charakteristischen Größen des Systems. Die Verlustwahrscheinlichkeit zum Beispiel ist im AF-Fall größer, weil eine Anforderung, die zum Ankunftszeitpunkt einer neuen Anforderung ihre Bedienung beendet, bei dieser Strategie noch einen Wartespeicher belegt.

---

<sup>1</sup>Die Dauer der Aktivitäts- und Pausenphasen werden jeweils als negativ-exponentiell verteilt angenommen

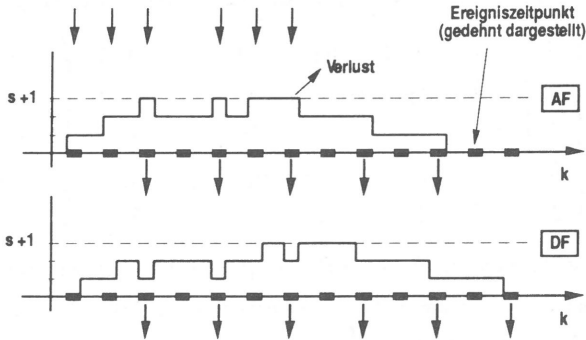


Bild 4.3: Der Verlauf des Systemzustands für die Strategien AF und DF

#### 4.2.2.2 $GEO_1/D/1-s$

Bei Untersuchungen an ATM-Systemkomponenten wird oft argumentiert, daß durch die Überlagerung vieler unabhängiger Verkehrsströme in einem zeitdiskreten System die Wahrscheinlichkeiten für die Ankünfte von Zellen in jedem Zeitschlitz gleich sind. Deshalb wird zur Modellierung oft das  $GEO_1/D/1-s$ -System verwendet. Für die Zulässigkeit dieser Argumentation gelten die bereits im Abschnitt 4.2.1.1 für den Poisson-Ankunftsprozeß genannten Bedingungen.

Der Unterschied zwischen dem kontinuierlichen  $M/D/1-s$ -System und dem zeitdiskreten  $GEO_1/D/1-s$ -System ist bei einer Bedienzeit  $D$ , die viel größer als 1 ist, nicht mehr signifikant (siehe Abschnitt 4.3) und beide Systeme können für solche Fälle als gleichwertig betrachtet werden.

Prinzipiell können bei der Lösung der zeitdiskreten Systeme dieselben Argumente verwendet werden wie bei den zeitkontinuierlichen, wodurch sich für das  $GEO_1/D/1-s$ -System der gleiche Ansatz ergibt wie für das oben beschriebene  $M/D/1-s$ -System. Für die Zustandswahrscheinlichkeiten  $\pi_i$  zum Bedienende ergibt sich ein Gleichungssystem analog zu Gleichung 4.1, das sich (unter Einbeziehung der letzten, redundanten Gleichung) in Matrixschreibweise als

$$\Pi = Q\Pi \quad (4.9)$$

darstellen läßt, wobei  $\Pi$  der Spaltenvektor mit den Zustandswahrscheinlichkeiten  $\pi_i$  ist und  $Q$  die Matrix, welche die Übergänge zwischen den einzelnen Zuständen beschreibt. Läßt man die  $(s+1)$ -te, redundante Gleichung für  $\pi_s$  weg, ersetzt sie durch die Normierungsbedingung

und formt die Gleichungen um, dann ergibt sich das lineare Gleichungssystem

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} n_0 - 1 & n_0 & 0 & \cdots & 0 \\ n_1 & n_1 - 1 & n_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_{s-1} & n_{s-1} & n_{s-2} & \cdots & n_0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \pi_0 \\ \pi_1 \\ \vdots \\ \pi_{s-1} \\ \pi_s \end{pmatrix} \quad (4.10)$$

Die Wahrscheinlichkeiten  $n_i$  für die AF-Strategie ergeben sich aus dem Zählprozess der  $GEO_1$ -Verteilung und gehorchen für eine Bedienzeit von  $D$  Zeiteinheiten der Binomial-Verteilung (siehe Abschnitt 3.3.1.2.2)

$$n_i = \binom{D}{i} (1-q)^i q^{D-i}, \quad 0 \leq i \leq D \quad (4.11)$$

Der Unterschied für die Strategie DF ist aus Bild 4.4 zu ersehen (siehe auch [71]). Bei einem Übergang, der von einem leeren System aus beginnt, enthält die Bedienphase, die durch die nächste Ankunft ausgelöst wird, bei der DF-Strategie einen möglichen Ereigniszeitpunkt weniger als bei der AF-Strategie, nämlich  $D - 1$ .

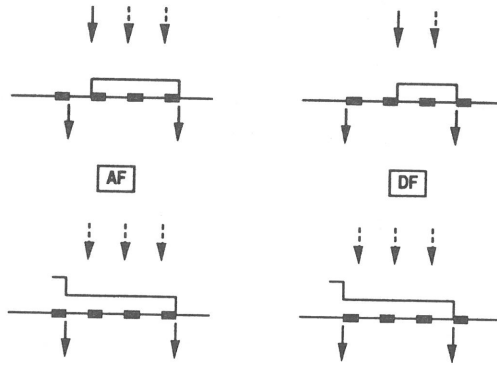


Bild 4.4: Unterschied bei der Berechnung für die Strategien AF und DF

Während einer Belegphase des Systems tritt dieser Unterschied nicht auf. Für diese Strategie muß also lediglich die erste Spalte der Matrix  $Q$ , die die Übergänge vom leeren System aus repräsentiert, durch die Werte

$$n'_i = \binom{D-1}{i} (1-q)^i q^{(D-1)-i}, \quad 0 \leq i \leq D-1 \quad (4.12)$$



ersetzt werden. Die Auflösung des linearen Gleichungssystems kann wiederum mit Hilfe einer Standardmethode oder rekursiv vorgenommen werden. Aus den so berechneten Zustandswahrscheinlichkeiten lassen sich unter Verwendung der GASTA-Eigenschaft (siehe Abschnitt 3.3.1.2.2) für den Fall AF die Zustandswahrscheinlichkeiten zu den Ankunftszeitpunkten  $p_i$  berechnen. Nach [64] gelten die Beziehungen

$$p_i = \begin{cases} \frac{\pi_i}{A_{OL} + c \cdot \pi_0}, & 0 \leq i \leq s \\ 1 - \frac{1}{A_{OL} + c \cdot \pi_0}, & i = s + 1 \end{cases} \quad (4.13)$$

$A_{OL}$  bezeichnet dabei den angebotenen Verkehr

$$A_{OL} = \frac{D}{E[A]} = D \cdot (1 - q) \quad (4.14)$$

und  $c$  bezeichnet einen konstanten Faktor, der für die AF-Strategie gleich 1 und für die DF-Strategie gleich  $q$  ist. Die Verlustwahrscheinlichkeit ist für beide Strategien durch die Wahrscheinlichkeit  $p_{s+1}$  gegeben.

Eine andere Lösungsmethode für das  $GEO_1/D/1-s$ -System mit AF-Strategie ist in [105] angegeben.

#### 4.2.2.3 $GI^{[X]}/D/1-s$

Zur Lösung von Wartesystemen mit allgemeinem Ankunfts- und Bedienprozeß kann im kontinuierlichen Zeitbereich die Lindleysche Integralgleichung [104] verwendet werden, die eine Beziehung zwischen den Beträgen der im System befindlichen Restarbeit zu aufeinanderfolgenden Ankunftszeitpunkten angibt. Da bei einer FIFO-Abfertigungsstrategie die bei der Ankunft angetroffene Restarbeit genau der Wartezeit entspricht, können die charakteristischen Größen aus der Restarbeitsverteilung berechnet werden.

Das zeitdiskrete Analogon zur Lindleyschen Integralgleichung, das z.B. in [157] ausführlich hergeleitet wird, erlaubt eine numerische Lösung des zeitdiskreten Systems mit allgemeinem Ankunfts- und Bedienprozeß und unbegrenztem Wartespeicher. Die Auswertung der Gleichung kann dabei nach [1] entweder iterativ im Zeitbereich [140] oder direkt im Frequenzbereich [133] erfolgen. Bei beiden Algorithmen werden Methoden aus der Signalverarbeitung eingesetzt, wie z.B. die Schnelle Fourier Transformation (FFT) oder das Konzept des komplexen Cepstrums. Eine ausführliche Darstellung der Methodik ist in [157] enthalten.

Der iterative Lösungsansatz kann bei entsprechender Modifikation auch für die Lösung des Warteverlustsystems angewendet werden [146]. Das Problem, das sich bei der Anwendung

stellt ist, daß aufgrund des Ansatzes die Begrenzung des Wartespeichers durch ein Restarbeitskriterium gegeben ist, anstatt wie sonst üblich durch die Anzahl der Anforderungen im System. Deshalb ist nur die Berechnung des Systems mit deterministischen Bedienzeiten möglich, weil für dieses System ein direkter Zusammenhang zwischen der Restarbeit und der Anzahl von Anforderungen im System besteht. Die iterative Lösung für das System mit allgemeinen Gruppenankünften ( $GI^{[X]}$ ), einer Bedienzeit  $D = 1$  und verschiedenen Abweisstrategien für Gruppen, die nicht mehr vollständig in den Wartespeicher aufgenommen werden können, ist in [158] dargestellt. Dabei wird die Strategie DF unterstellt.

Im folgenden wird die für die weiteren Untersuchungen benötigte Erweiterung auf Bedienzeiten  $D > 1$  unter Einbeziehung der Strategien AF und DF erläutert und ein direktes Lösungsverfahren angegeben. Bei diesem Verfahren werden die Zustandsgleichungen für die Ankunftszeitpunkte direkt aufgestellt, um das entstehende Gleichungssystem mit Standardmethoden lösen zu können. Diese Methode ist für kleine bis mittlere Systeme wesentlich effektiver als der iterative Algorithmus, da dieser – besonders bei einem Verkehrsangebot nahe bei eins – sehr langsam konvergiert und sich dadurch Rechenzeiten ergeben, die im Stundenbereich liegen können (auf einem VAX-Rechner 3600).

Anhand von Abbildung 4.5 soll die Vorgehensweise bei der Analyse am Beispiel eines Systems mit DF-Abfertigung erklärt werden. Es wird angenommen, daß von einer ankommenden Gruppe so viele Zellen angenommen werden, wie im Speicher Platz haben und nur der Rest verworfen wird. Diese Strategie wird in [158] als „Blocking Policy 1“ (BP1) bezeichnet.

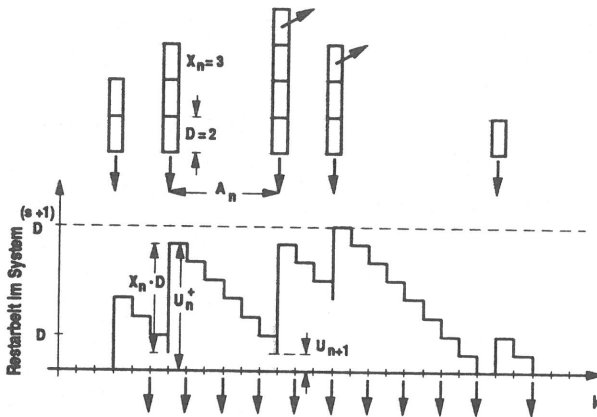


Bild 4.5: Verlauf der Restarbeit bei dem System mit DF und BP1

Zur Beschreibung werden folgende Zufallsvariablen benötigt:

$A_n$  Zufallsvariable für den Ankunftsabstand zwischen der  $n$ -ten und der  $(n+1)$ -ten Gruppe.

$X_n$  Zufallsvariable für die Anzahl der Zellen in der  $n$ -ten Gruppe

$U_n$  Zufallsvariable für die im System befindliche Restarbeit unmittelbar vor der Ankunft der  $n$ -ten Gruppe. Da es sich um ein System mit DF-Strategie handelt, ist die in der letzten Zeiteinheit vor der Ankunft abgearbeitete Restarbeit bereits abgezogen.

$U_n^+$  Zufallsvariable für die Restarbeit im System unmittelbar nach der Ankunft der  $n$ -ten Gruppe.

Im Beispiel wurde eine Bedienzeit  $D = 2$  und eine Anzahl von Warteplätzen  $s = 3$  gewählt. Die maximal im System (Warteschlange und Bedieneinheit) befindliche Restarbeit ergibt sich deshalb zu  $U_{max} = (s + 1)D = 8$ . Die Restarbeit im System unmittelbar nach der Ankunft der  $n$ -ten Gruppe ergibt sich aus der Summe von  $U_n$  und der Bedienzeit der angenommenen Zellen

$$U_n^+ = \begin{cases} U_n + D \cdot X_n, & X_n \leq I(U_n) \\ U_n + D \cdot I(U_n), & X_n > I(U_n) \end{cases} \quad (4.15)$$

$I(U_n)$  gibt die Anzahl der freien Warteplätze in Abhängigkeit von  $U_n$  an und berechnet sich zu

$$I(U_n) = \left( s + 1 - \left\lceil \frac{U_n}{D} \right\rceil \right) \quad (4.16)$$

wobei  $\lceil R \rceil$  die kleinste ganze Zahl größer oder gleich  $R$  bezeichnet. Während der Zeit bis zur Ankunft der  $(n + 1)$ -ten Gruppe werden  $A_n$  Einheiten der Restarbeit abgearbeitet, wenn das System nicht vorher leer wird. Es gilt also

$$U_{n+1} = \begin{cases} U_n^+ - A_n, & U_n^+ > A_n \\ 0, & \text{sonst} \end{cases} \quad (4.17)$$

Durch Einsetzen erhält man schließlich

$$U_{n+1} = \begin{cases} U_n + D \cdot X_n - A_n, & X_n \leq I(U_n) \text{ und } U_n^+ > A_n \\ U_n + D \cdot I(U_n) - A_n, & X_n > I(U_n) \text{ und } U_n^+ > A_n \\ 0, & \text{sonst} \end{cases} \quad (4.18)$$

Setzt man den stationären Fall voraus, wo die einzelnen Größen nicht mehr von der Ordnungsnummer der Ankunft abhängen ( $U_{n+1} = U_n = U$ , etc.), und geht auf die Wahrchein-

lichkeiten über, dann ergibt sich für  $k > 0$

$$\begin{aligned}
 u_k &= P\{U = k\} \\
 &= P\{U + D \cdot X - A = k, X \leq I(U), U^+ > A\} \\
 &\quad \times \\
 &\quad P\{U + D \cdot I(U) - A = k, X > I(U), U^+ > A\} \\
 &= P\{\text{Fall I}\} \times P\{\text{Fall II}\}
 \end{aligned} \tag{4.19}$$

weil *Fall I* und *Fall II* disjunkt sind. Für  $P\{\text{Fall I}\}$  ergibt sich nach dem Gesetz der totalen Wahrscheinlichkeit

$$P\{\text{Fall I}\} = \sum_{j=0}^{D(s+1)} P\{\text{Fall I}|U = j\}u_j \tag{4.20}$$

Die bedingte Wahrscheinlichkeit  $P\{\text{Fall I}|U = j\}$  ist gegeben durch

$$\begin{aligned}
 P\{\text{Fall I}|U = j\} &= \sum_{i=0}^{\infty} P\{D \cdot X - A = k - j, i \leq I(U = j)|X = i\}x_i \\
 &= \sum_{i=Max(0, \lceil \frac{k-j}{D} \rceil)}^{I(U=j)} a_{Di-k+j} \cdot x_i
 \end{aligned} \tag{4.21}$$

Analog dazu ergibt sich für  $P\{\text{Fall II}|U = j\}$

$$\begin{aligned}
 P\{\text{Fall II}|U = j\} &= \sum_{i=I(U=j)+1}^{\infty} P\{D \cdot I(U = j) - A = k - j, i > I(U = j)|X = i\}x_i \\
 &= a_{DI(U=j)+j-k} \sum_{i=I(U=j)+1}^{\infty} x_i
 \end{aligned} \tag{4.22}$$

wobei alle  $a_i$  für  $i < 0$  identisch null sind. Durch Einsetzen in Gleichung 4.19 ergibt sich ein Gleichungssystem mit  $D(s+1)$  Gleichungen für die Elemente der Restarbeitsverteilung  $u_k$  mit  $k > 0$ .

$$u_k = \sum_{j=0}^{D(s+1)} \left( \sum_{i=Max(0, \lceil \frac{k-j}{D} \rceil)}^{I(U=j)} a_{Di-k+j} \cdot x_i + a_{DI(U=j)+j-k} \sum_{i=I(U=j)+1}^{\infty} x_i \right) \cdot u_j \tag{4.23}$$

Die Gleichung für  $k = 0$  ist linear abhängig und wird deshalb – weil sie die aufwendigste ist – durch die Normierungsbedingung ersetzt. Die Aufstellung des Gleichungssystems kann bei der Programmierung dadurch effektiver gemacht werden, daß die Summe für die bedingte

Wahrscheinlichkeit im Fall I bzw. im Fall II gespeichert und modifiziert wird, anstatt sie für jedes Element neu zu berechnen (für größere Systeme wird ein signifikanter Anteil der Rechenzeit für den Aufbau des Gleichungssystems verbraucht).

Die Zellverlustwahrscheinlichkeit  $p_v$  berechnet sich aus dem Verhältnis der verlorenen Zellen pro Gruppenankunft zur mittleren Zahl von ankommenden Zellen pro Gruppe  $E[X]$ . Es ergibt sich also

$$p_v = \frac{1}{E[X]} \left( \sum_{k=0}^{D(s+1)} u_k \cdot \left( \sum_{j=I(U=k)+1}^{\infty} (j - I(U=k)) \cdot x_j \right) \right) \quad (4.24)$$

Bei der alternativen Abweisstrategie BP2 wird die gesamte Gruppe verworfen, wenn sie nicht komplett in den Wartespeicher übernommen werden kann. Dadurch ändert sich Gleichung 4.15 zu

$$U_n^+ = \begin{cases} U_n + D \cdot X_n, & X_n \leq I(U_n) \\ U_n, & X_n > I(U_n) \end{cases} \quad (4.25)$$

Daraus ergibt sich für die Restarbeit  $u_k$ ,  $k > 0$  wieder eine Unterscheidung in zwei Fälle, wobei der erste identisch mit dem oben besprochenen Fall I ist. Für den zweiten Fall ergibt sich

$$\begin{aligned} P\{\text{Fall II} | U = j\} &= \sum_{i=I(U=j)+1}^{\infty} P\{A = j - k, i > I(U = j) | X = i\} x_i \\ &= a_{j-k} \sum_{i=I(U=j)+1}^{\infty} x_i \end{aligned} \quad (4.26)$$

und daraus

$$u_k = \sum_{j=0}^{D(s+1)} \left( \sum_{i=Max(0, \lceil \frac{k-i}{D} \rceil)}^{I(U=j)} a_{Di-k+j} \cdot x_i + a_{j-k} \sum_{i=I(U=j)+1}^{\infty} x_i \right) \cdot u_j \quad (4.27)$$

Da in diesem Fall bei einer Abweisung die ganze Gruppe verloren geht, berechnet sich die Zellverlustwahrscheinlichkeit zu

$$p_v = \frac{1}{E[X]} \left( \sum_{k=0}^{D(s+1)} u_k \cdot \left( \sum_{j=I(U=k)+1}^{\infty} j \cdot x_j \right) \right) \quad (4.28)$$

Der Einfluß der AF-Strategie für die Auflösung der Gleichzeitigkeit ist aus Bild 4.6 zu ersehen. Die Übergangswahrscheinlichkeiten zwischen den Zuständen sind bei AF und DF gleich (Bild 4.6 a). Eine Ausnahme bildet der Fall, daß eine ankommende Anforderung ein leeres

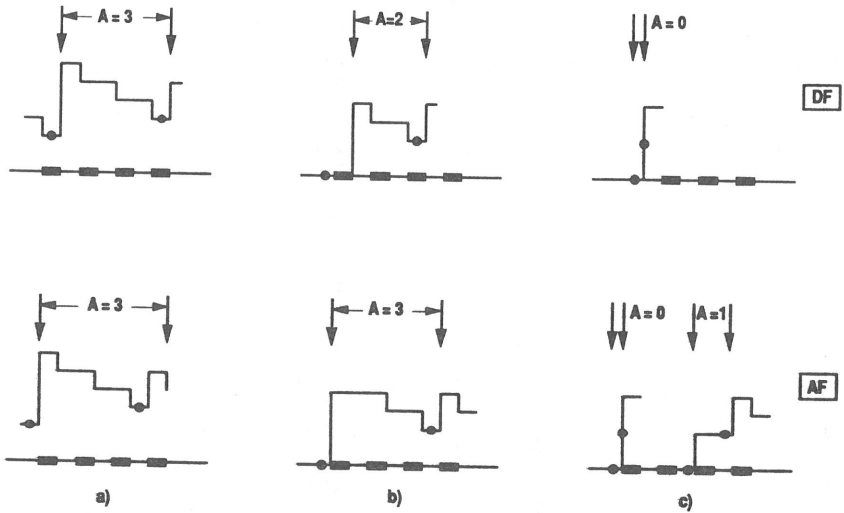


Bild 4.6: Verlauf der Restarbeit bei Systemen mit AF- und DF-Strategie

System vorfindet und der Übergang eine Dekrementierung der Restarbeit um mindestens eine Zeiteinheit erfordert (Bild 4.6 b). In diesem Fall wird zum Übergang in den gleichen Zustand wie bei DF ein um eine Zeiteinheit größerer Ankunftsabstand benötigt, weil zum Eintreffzeitpunkt keine Dekrementierung der Restarbeit erfolgt und zu den folgenden Zeitpunkten die Dekrementierung – logisch gesehen – erst nach dem möglichen Ankunftszeitpunkt stattfindet. Dies gilt im Prinzip auch für den Fall, daß beim Antreffen des leeren Systems keine Dekrementierung der Restarbeit für den Übergang erforderlich ist, weil bei der Gruppenankunft genau  $k$  Zeiteinheiten neue Arbeit in das System übernommen werden. Allerdings liefert dann auch ein Ankunftsabstand von null Zeiteinheiten den gleichen Übergang (Bild 4.6 c). Die Gleichung 4.23 für die BP1-Strategie muß deshalb dahingehend modifiziert werden, daß die Terme für  $j = 0$  durch

$$u_0 \cdot \left( \sum_{i=Max(0, \lceil \frac{k}{D} \rceil)}^{s+1} \left[ a_{Di-k+1} \cdot x_i + \begin{cases} a_0 \cdot x_{\frac{k}{D}}, & \text{für } Di=k \\ 0, & \text{sonst} \end{cases} \right] + \right. \\
 \left. \sum_{i=s+2}^{\infty} x_i \cdot \left[ a_{D(s+1)-k+1} + \begin{cases} a_0, & \text{für } D(s+1)=k \\ 0, & \text{sonst} \end{cases} \right] \right) \quad (4.29)$$

ersetzt werden, wobei die Terme in den geschweiften Klammern den ersten Fall in Abbildung 4.6 c repräsentieren, bei dem aus dem leeren System der Zustand  $k$  direkt erreicht wird und der nächste Ankunftsabstand 0 Zeiteinheiten beträgt. Entsprechend ergibt sich für die BP2-Abweisstrategie bei AF in Gleichung 4.27 eine Änderung in

$$u_0 \cdot \left( \sum_{i=\text{Max}(0, \lceil \frac{k}{D} \rceil)}^{s+1} \left[ a_{D_i-k+1} \cdot x_i + \left\{ \begin{array}{ll} a_0 \cdot x_{\frac{k}{D}}, & \text{für } D_i=k \\ 0, & \text{sonst} \end{array} \right\} \right] \right) \quad (4.30)$$

Der zweite Term in Gleichung 4.27 ist für  $j = 0$  und  $k > 0$  immer gleich null, weil die entsprechenden Elemente der Zwischenankunftsverteilung identisch null sind.

Die Berechnung des Systems mit AF-Strategie ist auch bei der in [158] beschriebenen, iterativen Methode anwendbar. Dazu muß nach jedem Iterationsschritt die berechnete Restarbeit unmittelbar vor dem Ankunftszeitpunkt entsprechend der Vorschrift

$$\begin{aligned} u'_0 &= u_0 \cdot a_0 \\ u'_1 &= u_1 + u_0 \cdot (1 - a_0) \end{aligned} \quad (4.31)$$

angepaßt werden, bevor die neu hinzugekommene Arbeit addiert werden kann (siehe [166]).

Das zeitdiskrete GI/D/1-s-System ergibt sich als Sonderfall aus dem oben beschriebenen System, wenn man eine deterministische Gruppengröße von eins wählt. Für diesen Fall sind die Abweisstrategien BP1 und BP2 identisch, weil jedesmal nur eine Zelle ankommt. Für den in Kapitel 5 verwendeten Fall mit Gruppengröße eins und DF-Strategie vereinfacht sich Gleichung 4.23 zu

$$u_k = \sum_{j=0}^{sD} u_j \cdot a_{D-k+j} + \sum_{j=sD+1}^{(s+1)D} u_j \cdot a_{j-k}, \quad 0 < k \leq (s+1)D \quad (4.32)$$

Die Verlustwahrscheinlichkeit ist für diesen Fall durch

$$p_v = \sum_{i=sD+1}^{(s+1)D} u_i \quad (4.33)$$

gegeben.

Die Anzahl der Gleichungen wird in allen Fällen durch die Anzahl der Elemente der Restarbeitsverteilung bestimmt und beträgt deshalb  $D(s+1) + 1$ . Bis zu einer – durch die Größe des virtuellen Speicherbereichs begrenzten – Gleichungsanzahl von etwa 850 kann das Gleichungssystem mit Hilfe einer Prozedur aus der Standard-Programmbibliothek IMSL aufgelöst werden. Diese Prozedur verwendet einen in [111] beschriebenen, für Systeme mit virtueller

Speicherverwaltung optimierten Gauss-Eliminations-Algorithmus. Dieser Algorithmus unterteilt die Matrix in kleinere Blöcke, auf die separat ein Eliminationsverfahren angewendet wird. Dadurch wird die Lokalität des Programmes vergrößert und es werden weniger Seitenaustauschoperationen benötigt. Der Algorithmus benötigt für die maximale Systemgröße eine Rechenzeit im Minutenbereich (VAX-Rechner 3600). Um die Berechnung größerer Systeme zu ermöglichen, kann z.B. das oben erwähnte Verfahren so implementiert werden, daß die nicht benötigten Blöcke explizit in einer Datei zwischengespeichert werden und deshalb nicht im virtuellen Speicher verwaltet werden müssen. Dies geht natürlich auf Kosten der Rechenzeit, weil Dateizugriffe über das Betriebssystem wesentlich zeitaufwendiger sind als Seitenwechseloperationen im virtuellen Speicher.

Eine andere Möglichkeit ist der Einsatz iterativer Lösungsmethoden [170] wie z.B. des Gauss-Seidel-Verfahrens, bei dem es genügt, eine Zeile der Matrix im Speicher zu halten. Dadurch lassen sich sehr große Systeme berechnen, wobei die Systemgröße hier durch die Rechenzeit begrenzt ist. Der Schnittpunkt, bei dem die iterativen Verfahren effektiver werden als die direkten, liegt nach den Erfahrungen aus der Anwendung der verschiedenen Algorithmen im Bereich von etwa 2000 Gleichungen.

Die Lösung des zeitdiskreten GI/D/1-s-Systems ist auch für die zweiphasige sporadische Quelle als Ankunftsprozeß anwendbar. Dazu muß lediglich die negativ-exponentielle Verteilung der Pausendauer entsprechend diskretisiert werden. Wie die Beispiele für die Berechnung des IDC in Abschnitt 3.4 zeigen, bei denen die gleiche Approximation verwendet wird, ist dies ohne signifikanten Einfluß auf die Ergebnisse möglich.

#### 4.2.2.4 DMAP/D/1-s

Der DMAP [16, 153] ist das zeitdiskrete Analogon zum zeitkontinuierlichen MAP (siehe Abschnitt 4.2.1.3) und ein Warte- oder Warteverlustsystem mit diesem Ankunftsprozeß sowie allgemeiner Bedienzeitverteilung kann im Prinzip mit dem gleichen Ansatz und den gleichen Methoden wie das kontinuierliche System gelöst werden. Eine ausführliche Beschreibung der Methodik und der Ergebnisse ist in [19] enthalten.

Da der GMDP (siehe Abschnitt 3.3.4) als DMAP darstellbar ist, kann mit den Methoden der Matrixgeometrie auch das GMDP/D/1-System analytisch behandelt werden. Ein Problem dabei ist, daß schon bei der Darstellung eines GMDP mit wenigen Zuständen die Phasenzahl des DMAP sehr groß werden kann und sich dadurch mit den heutigen Algorithmen Rechenzeiten im Bereich von mehreren Stunden ergeben können (VAX-Rechner 3600).



### 4.2.3 Andere Modelle

Neben den oben beschriebenen Systemen und Lösungsmethoden wurden in der Literatur für ATM-Untersuchungen, z.B. an statistischen Multiplexern, noch einige weitere vorgeschlagen. Von besonderem Interesse sind hierbei Modelle für die Überlagerung sporadischer Quellen [20, 46, 171]. Bei diesen Modellen wird neben der Anzahl der Zellen im System auch der Zustand der Quellen (Aktivität/Pause) mit einbezogen. Ein anderer Ansatz wird in [148] vorgeschlagen, bei dem der Überlagerungsverkehr durch zwei Momente charakterisiert wird. Ein Vergleich verschiedener Systemmodelle für ATM-Untersuchungen ist in [98] enthalten.

Weitere Modelle behandeln die Überlagerung periodischer Verkehrsströme, wie sie von stromförmigen Diensten (CBR, siehe Abschnitt 2.4.5) erzeugt werden [54, 88]. Als Modell ergibt sich hierbei das  $\Sigma D/D/1$ -System mit endlichem bzw. unendlichem Wartespeicher. Der ungünstigste Fall, bei dem alle Quellen gleichzeitig, d.h. ohne Phasenverschiebung senden, ist relativ leicht anzugeben, die Wahrscheinlichkeit für das Auftreten dieses Falles ist jedoch für unabhängige Verkehrsquellen in den meisten Fällen sehr gering. Durch die Berücksichtigung der verschiedenen Möglichkeiten für die Phasenbeziehungen der Quellen ergeben sich sehr komplizierte Lösungsalgorithmen, besonders wenn man Quellen mit verschiedenen Periodendauern oder sogar sporadische Quellen mit einer festen Pausendauer und einer festen Anzahl von Zellen pro Aktivitätsphase zuläßt<sup>2</sup>. Auch bei der Simulation dieser Modelle ergeben sich durch den periodischen Charakter der Verkehre Probleme, weil sich dabei ebenfalls eine unerwünschte feste Phasenbeziehung einstellt, die durch geeignete Maßnahmen in der Simulationssteuerung verhindert werden muß.

## 4.3 Beispiele für die Berechnung der Verlustwahrscheinlichkeit

Zunächst soll der Einfluß der Bedienzeit und der Auflösungsstrategie für die Gleichzeitigkeit bei zeitdiskreten Systemen anhand des  $GEO_1/D/1$ -s-Systems aufgezeigt werden.

Für eine zunehmende Bedienzeit steigt die Anzahl der möglichen Ankunftszeitpunkte pro Bedienzeit an – was einer feineren Diskretisierung entspricht – und das Verhalten nähert sich dem des entsprechenden kontinuierlichen Systems an. Bild 4.7 zeigt den relativen Unterschied zwischen den Verlustwahrscheinlichkeiten der zeitdiskreten Systeme mit AF bzw. DF und dem zeitkontinuierlichen  $M/D/1$ -s-System bei einem Angebot von  $A_{OL} = 0,9$ . Erwartungsgemäß nimmt die Differenz mit steigender Bedienzeit ab. Bei Quellen, für die der Abstand

---

<sup>2</sup>Für einige dieser Modellvarianten sind noch keine analytischen Lösungen bekannt

zwischen den Zellen signifikant größer ist als die Übertragungs- oder Bedienzeit, können deshalb bei den Untersuchungen ohne weiteres kontinuierliche Modelle verwendet werden, falls dies Vorteile bringt.

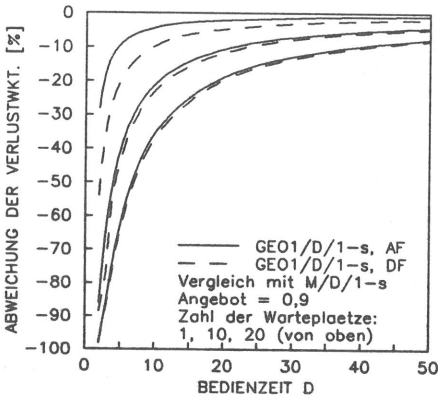


Bild 4.7: Die Verlustwahrscheinlichkeit in Abhängigkeit von der Bedienzeit

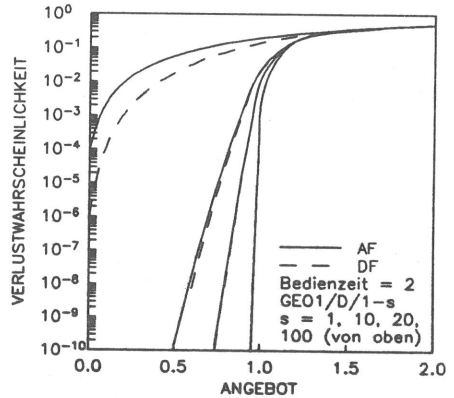


Bild 4.8: Die Verlustwahrscheinlichkeit in Abhängigkeit vom Angebot

Das System, bei dem die Ankünfte vor den Bedieneereignissen abgearbeitet werden (AF), liefert – wie erwartet – eine höhere Verlustwahrscheinlichkeit als das System mit DF. Der Unterschied ist jedoch nur für Systeme mit sehr wenigen Wartepätzen signifikant.

In Bild 4.8 ist die Abhängigkeit der Verlustwahrscheinlichkeit vom angebotenen Verkehr und von der Wartespeichergröße dargestellt. Es ist zu erkennen, daß der Unterschied zwischen den Strategien AF und DF für alle Angebote wiederum nur bei kleiner Wartespeichergröße entscheidend ist.

Die numerische Stabilität der direkten Lösungen für das M/D/1-s und das GEO<sub>1</sub>/D/1-s-System ist bei Verwendung doppelt genauer Arithmetik bis zu Verlustwahrscheinlichkeiten kleiner 10<sup>-20</sup> gewährleistet. Eine Validierung des zeitdiskreten Algorithmus mit Hilfe der direkten Lösung für geometrische Ankunftsprozesse ergab bis ca. 10<sup>-15</sup> keine signifikanten Abweichungen, wodurch dieser Algorithmus ohne Einschränkungen für die Untersuchungen im nächsten Kapitel eingesetzt werden kann.

Als nächstes Beispiel soll das Verhalten eines Systems mit deterministischer Bedieneinheit dargestellt werden, wenn verschiedene Modelle für Einzelquellen und die entsprechenden Lösungsverfahren angewendet werden. Das System mit einer zweiphasigen sporadischen Quelle und den in [70] angegebenen Parametern für paketierte Sprache (siehe Abschnitt 3.4) wird dabei mittels der direkten Analyse des GI/D/1-s-Systems sowie mit Hilfe des UAS-Modells untersucht. Zum Vergleich wird eine Quelle mit negativ-exponentieller Ankunfts-

verteilung und gleichem Mittelwert bzw. eine Quelle mit einer negativ-binomialen Ankunftsverteilung verwendet, bei der die ersten zwei Momente an diejenigen der sporadischen Quelle angepaßt wurden. Außerdem wird eine zweiphasige MMPP-Quelle verwendet, wie sie ebenfalls für die Untersuchung von Quellflußkontrollverfahren vorgeschlagen wurde. Die vier Parameter des MMPP wurden dabei nach dem in [70] vorgeschlagenen Algorithmus an die Kenngrößen der sporadischen Quelle angepaßt. Die dabei verwendeten Kenngrößen der sporadischen Quelle sind:

- die mittlere Zellankunftsrate  $\lambda$ ,
- der IDC für ein Intervall der Länge  $t_1$ ,
- der Grenzwert des IDC für eine gegen unendlich strebende Intervalldauer (dieser Grenzwert ist identisch mit dem Quadrat des Variationskoeffizienten der Ankunftsabstände  $c_a^2$ , siehe Abschnitt 3.2.2.2) und
- das dritte zentrale Moment der Verteilung für die Anzahl von Ankünften in einem Intervall der Länge  $t_2$ .

Für eine sinnvolle Wahl von  $t_1$  und  $t_2$  wurde in [70] keine Vorschrift angegeben. Die Abbildungen 4.9 und 4.10 zeigen deshalb den Einfluß der Wahl des Zeitpunktes  $t_2$  für einen festen Wert  $t_1=1000\text{ms}$ .

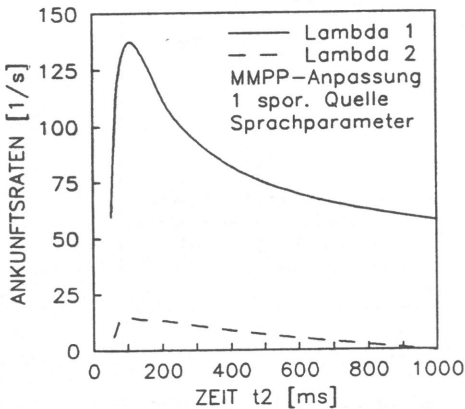


Bild 4.9: MMPP-Parameter  $\lambda_1$  und  $\lambda_2$  in Abhängigkeit von der Zeit  $t_2$

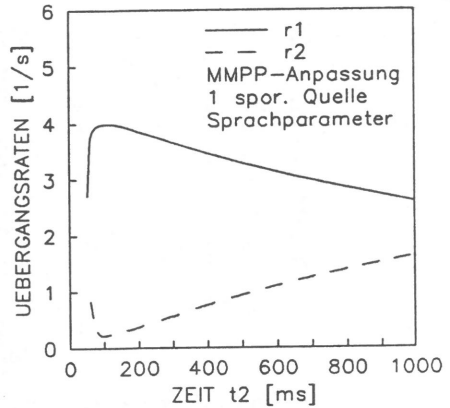


Bild 4.10: MMPP-Parameter  $r_1$  und  $r_2$  in Abhängigkeit von der Zeit  $t_2$

Zur Verdeutlichung des Einflusses der Wahl von  $t_2$  auf die Verlustwahrscheinlichkeit wurden die Kurven für  $t_2=100\text{ms}$ , wo die Parameter Extremwerte aufweisen, sowie für  $t_2=900\text{ms}$  verwendet, wo sich näherungsweise ein unterbrochener Poisson-Prozeß ergibt. Der mittlere Ankunftsabstand für diesen Parametersatz beträgt im aktiven Zustand  $16,7\text{ms}$ , die mittlere Verweildauer im aktiven Zustand beträgt  $368,26\text{ms}$  und diejenige im (nahezu) passiven

Zustand beträgt im Mittel 662,38ms. Die entsprechenden Werte der sporadischen Quelle ergeben sich zu 16ms, 352ms und 650ms. Die Werte ähneln sich sehr stark, wobei jedoch beim MMPP eine höhere Verlustwahrscheinlichkeit zu erwarten ist, weil die Ankünfte im aktiven Zustand nicht so gleichmäßig erfolgen wie bei der sporadischen Quelle.

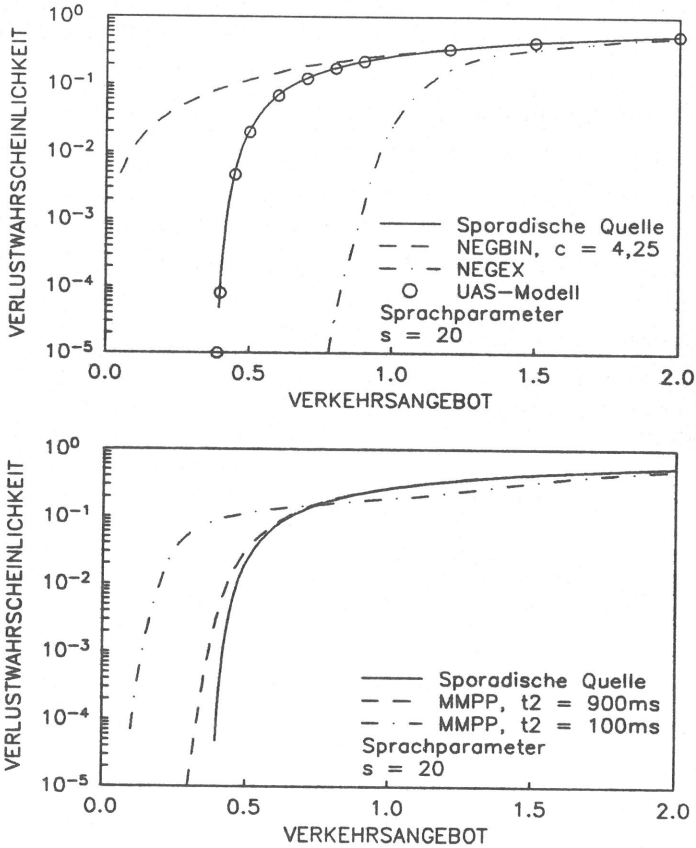


Bild 4.11: Verlustwahrscheinlichkeit für das GI/D/1-20-System

Bild 4.11 zeigt die Verlustwahrscheinlichkeit für ein System mit 20 Warteplätzen für die oben angeführten Quellenmodelle. Das UAS-Modell ergibt eine sehr gute Übereinstimmung mit der direkten Lösung für die (diskretisierte) sporadische Quelle. Beide Modelle ergeben für eine Bedienzeit, die sich an den minimalen Zellabstand annähert ( $A_{OL} = 0,351$ ), eine gegen null strebende Verlustwahrscheinlichkeit. Die anderen Modelle weisen in diesem Bereich durch die bei ihnen vorkommenden kleineren Ankunftsabstände noch eine von null verschiedene Verlustwahrscheinlichkeit auf. Die Quelle mit negativ-exponentiellen Zwischen-

ankunftsabständen ergibt im ganzen Bereich viel zu optimistische Ergebnisse und ist deshalb als Approximation nicht geeignet. Die Quelle mit negativ-binomialen Zwischenankunftsabständen überschätzt die Verlustwahrscheinlichkeit bei kleinen bis mittleren Verkehrsangeboten drastisch. Dies ist dadurch bedingt, daß die Wahrscheinlichkeitsmasse bei dieser Verteilung gleichmäßig verteilt ist, während sie bei der sporadischen Quelle auf den Bereich des minimalen Zellabstands konzentriert ist.

Die MMPP-Verlustkurve für  $t_2=900\text{ms}$  zeigt ein ähnliches Verhalten wie diejenige für die sporadische Quelle, während die Kurve für  $t_2=100\text{ms}$  einen signifikant anderen Verlauf aufweist. Ähnliche Effekte treten auch bei der Verwendung des MMPP als Approximation für die Überlagerung sporadischer Quellen [70] auf. Dadurch ist dieses Verfahren ohne eine Vorschrift für eine günstige Wahl von  $t_1$  und  $t_2$  nur bedingt einsetzbar.

Als letztes Beispiel soll der Einfluß einer zusätzlichen Phase bei einer sporadischen Quelle untersucht werden. Mit dieser Phase soll ein Verhalten besser nachgebildet werden, wie es bei Videocodern mit veränderlicher Ausgangsbitrate zu beobachten ist. Dazu wird ein dreiphasiger GMDP mit den in Bild 4.12 angegebenen Parametern verwendet<sup>3</sup>, für den der Verlauf der Verlustwahrscheinlichkeit bei Systemen mit verschiedenen großen Wartespeichern in Bild 4.13 dargestellt ist.

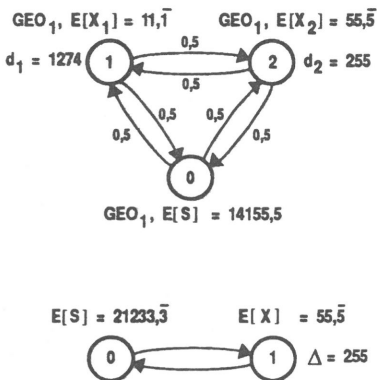


Bild 4.12: Parameter der Quellen

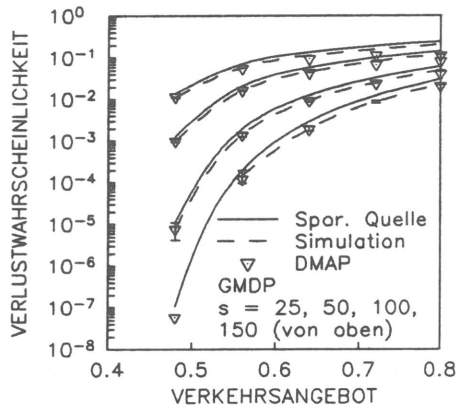


Bild 4.13: Einfluß einer zusätzlichen Phase

Die Ergebnisse wurden mit der in [19] dargestellten Approximation für die Verlustwahrscheinlichkeit des DMAP/D/1-s-Systems berechnet. Da diese nur für kleine Verlustwahrscheinlichkeiten genau ist, wurden die Werte zusätzlich simuliert. Es ist deutlich erkennbar, daß die Abweichung der Approximation mit der Verlustwahrscheinlichkeit zunimmt.

<sup>3</sup>Ein GMDP mit diesen Parametern wurde in RACE R1022 als Standardquellmodell für die Untersuchungen an Koppelnetzen verwendet. Es ergibt sich mit diesen Parametern  $E[A] = 637.2$  und  $c_a = 3.925$  [154].

Zum Vergleich wurden die Kurven für die ebenfalls in Bild 4.12 dargestellte zweiphasige sporadische Quelle aufgetragen, die bezüglich der mittleren und der maximalen Zellrate an den GMDP angepaßt wurde. Der dritte Parameter der sporadischen Quelle wurde so gewählt, daß die mittlere Anzahl der Zellen pro Aktivitätsphase derjenigen im hochratigen Zustand des GMDP entspricht. Die Kurven zeigen, daß der zweiphasige Prozeß, der durch das direkte Umschalten zwischen null und der hohen Zellrate einen höheren Variationskoeffizienten der Verteilung der Zwischenankunftsabstände ( $c_a = 6,29$ ) aufweist, in allen Fällen eine höhere Verlustwahrscheinlichkeit ergibt. Die Abweichungen sind jedoch – zumindest bei einer GMDP-Quelle mit diesen Parametern – so gering, daß die im nächsten Abschnitt gemachten Aussagen auch für den GMDP gültig sind und keine eigenen Untersuchungen gemacht wurden.

# Kapitel 5

## Leistungsuntersuchung von Quellflußkontrollmechanismen

In diesem Kapitel werden die in den vorhergehenden Kapiteln eingeführten Verkehrsquellen- und Systemmodelle verwendet, um die Leistungsfähigkeit von Quellflußkontrollmechanismen (siehe Abschnitt 2.6.2) zu untersuchen.

### 5.1 Anforderungen und Voraussetzungen

Aus den in Abschnitt 2.6.2 dargestellten Aufgaben einer Quellflußkontrollfunktion auf Zellebene ergeben sich einige grundsätzliche Anforderungen an einen Mechanismus, der diese Funktion implementieren soll. Zu Beginn der Untersuchungen wurden deshalb folgende Kriterien für den Vergleich der Verfahren formuliert<sup>1</sup>:

- *Selektivität des Mechanismus.* Ein Quellflußkontrollmechanismus muß eine gute Selektivität aufweisen. Das bedeutet, daß er für eine „gutartige“ Verkehrsquelle, die sich an die beim Verbindungsaufbau ausgehandelten Parameter (Verkehrsvertrag) hält, transparent sein und deren statistische Schwankungen nicht als Vertragsverletzungen werten sollte. Auf der anderen Seite sollen möglichst alle wirklichen Verletzungen der Vertragsparameter erkannt und durch entsprechende Aktionen (siehe Abschnitt 2.6.2) unterbunden werden, um das Netz und damit die gleichzeitig existierenden Verbindungen effektiv zu schützen. Auf der Basis dieser Anforderung läßt sich ein idealer Verlauf der Detektionskurve für Verletzungen des Verkehrsvertrags angeben, der in den folgenden Abschnitten für die Bewertung der verschiedenen Verfahren verwendet wird. Bei der Bewertung der Selektivität muß beachtet werden, daß alle Einflüsse

---

<sup>1</sup>Ähnliche Kriterien finden sich inzwischen auch in der CCITT Empfehlung I.311 [28]

(Verluste, Verzögerungen), welche die Quellflußkontrollfunktion als Netzfunktion auf gutartige Quellen ausübt, bei der Bestimmung der Netzgüte mit einbezogen werden müssen. Die Wahrscheinlichkeit für ungerechtfertigte Zellverluste durch Fehlentscheidungen des Quellflußkontrollmechanismus muß deshalb in dem Bereich liegen, der für andere Netzkomponenten, wie z.B. Multiplexer oder Vermittlungsknoten, zulässig ist. Dieser Bereich liegt – nach heutiger Ansicht – unter  $10^{-10}$ , was eine sehr harte Anforderung an die Transparenz der Quellflußkontrollfunktion darstellt.

- *Im ungünstigsten Fall durchgelassener Verkehr (engl. Worst Case).* Unter Berücksichtigung der Parameter des Verkehrsvertrags und der spezifischen Eigenschaften des Quellflußkontrollverfahrens läßt sich ein (deterministischer) Verkehr definieren, der den gebotenen Spielraum optimal ausnutzt. Gesichtspunkte bei der Definition dieses Verkehrs sind z.B. die mittlere Anzahl von Zellen, die pro Zeiteinheit in das Netz geschickt werden können und die Eignung des resultierenden Verkehrs für ein effektives statistisches Multiplexen. Der Unterschied zwischen der gutartigen statistischen Verkehrsquelle und dieser extremen Quelle ist ein weiteres Kriterium für die Güte des Verfahrens.
- *Dynamisches Verhalten.* Die Puffer im Netz sind in der Regel so dimensioniert, daß sie keine Überlastsituationen im Zeitrahmen der Burst-Ebene oder darüber ausgleichen können, sondern nur solche auf der Zellebene. Deshalb muß die dynamische Reaktionszeit des Mechanismus für die Quellflußkontrolle kurz sein, um die Dauer der durch Verletzungen des Verkehrsvertrags verursachten Überlastsituationen möglichst gut zu begrenzen.
- *Implementierungsaufwand.* Wie bereits in Abschnitt 2.6.2 erläutert, bestehen hohe Anforderungen bezüglich der Echtzeittauglichkeit, die nur durch eine Implementierung der Mechanismen mit schnellen integrierten Spezialbausteinen erfüllt werden können. Weiterhin erfordert die Anzahl der pro Teilnehmer zur Verfügung zu stellenden Quellflußkontrollfunktionen eine kosteneffektive Implementierbarkeit. Deshalb sollte ein Quellflußkontrollmechanismus möglichst einfach sein und nur solche Operationen verwenden, die schnell und mit wenig Aufwand von entsprechenden logischen Schaltwerken ausgeführt werden können.
- *Flexibilität.* Die verschiedenen im Verkehrsvertrag enthaltenen Parameter (siehe Abschnitt 2.6.2) können nur eingeschränkt mit einem einzelnen Quellflußkontrollmechanismus überwacht werden, weil dabei verschiedene Zeitbereiche relevant sind und entsprechend eine unterschiedliche Dimensionierung der Mechanismen nötig ist. Deshalb ist die Überwachung durch mehrere Mechanismen pro Verbindung vorteilhaft. Ein Quellflußkontrollmechanismus sollte deshalb durch entsprechende Parametrisierung für die Überwachung der verschiedenen Parameter angepaßt werden können. Weiterhin sollte er flexibel genug sein, um eine optimale Anpassung an die Charakteristika und



Anforderungen der verschiedenen Dienste zu ermöglichen. Dies erlaubt die Definition eines einheitlichen Quellflußkontroll-Bausteins, was vom ökonomischen Standpunkt aus sehr vorteilhaft ist.

Für die weiteren Untersuchungen wurden einige Annahmen getroffen, die im folgenden kurz dargestellt werden. Zunächst wird angenommen, daß die Verkehrsquellen die Zellen mit ihrer natürlichen – durch den Dienst und die Codierung bestimmten – Charakteristik ohne Rückwirkung vom Netz (Flußkontrolle) senden. Als Quellmodell wird vorwiegend die zweiphasige sporadische Quelle (siehe Abschnitt 3.3.2) mit verschiedenen Parametersätzen verwendet, weil sie trotz ihrer Einfachheit einige der für ATM-Verkehre typischen Eigenschaften aufweist und eine gezielte Variation der relevanten Parameter erlaubt. Die anderen Quellenmodelle werden verwendet, um die Sensitivität der Ergebnisse bezüglich der Quellcharakteristik zu untersuchen.

Eine weitere Annahme ist, daß die Eigenschaften des Quellverkehrs auf dem Weg der Zellen bis zum Quellflußkontrollmechanismus nicht verändert werden. Eine solche Veränderung könnte in der Realität z.B. durch eine als lokales Netz oder ATM-Nebenstellenanlage ausgebildete Netzabschlußeinheit (B-NT2, siehe Abschnitt 2.4.4) hervorgerufen werden, weil sich in solchen Einheiten durch die Zwischenspeicherung Verzögerungsschwankungen ergeben können. Solche Effekte wurden bewußt nicht mit einbezogen<sup>2</sup>, um den direkten Einfluß der Quellcharakteristik aufzuzeigen.

Die Auswirkungen der verschiedenen Mechanismen und Dimensionierungsansätze werden anhand der Überwachung der mittleren Zellrate untersucht. Dabei wird implizit davon ausgegangen, daß ein separater Überwachungsmechanismus vorhanden ist, der die maximale Zellrate kontrolliert. Bezüglich der mittleren Zellrate wird vorausgesetzt, daß diese exakt bekannt ist und sich über der Zeit nicht verändert.

Bei der Untersuchung wird davon ausgegangen, daß die Quellflußkontrollfunktion auf der Basis einzelner virtueller Verbindungen ausgeführt wird, weil dadurch die direkteste Kontrolle der einzelnen Quellen erreicht werden kann.

Als Aktion des Quellflußkontrollmechanismus beim Erkennen einer Parameterverletzung wird das Verwerfen der Zellen vorausgesetzt. Diese Reaktion wird in der CCITT Empfehlung I.311 [28] als die bevorzugte angegeben und eignet sich für die Untersuchung der verschiedenen Mechanismen besonders gut, weil die Ergebnisse nur von der Quellcharakteristik und vom verwendeten Mechanismus und nicht von anderen Größen, wie z.B. der Netzauslastung, abhängen.

---

<sup>2</sup>Eine Untersuchung dieser Effekte findet sich z.B. in [137]

## 5.2 Beschreibung und Modellierung der Mechanismen

In den nachfolgenden Vergleich werden neben dem sogenannten „Leaky Bucket“-Mechanismus [55, 60, 61, 96, 145, 161] noch weitere, weniger bekannte Verfahren einbezogen, die z.B. in [51, 47] beschrieben wurden.

### 5.2.1 Leaky Bucket

Der ursprünglich von Turner [161] vorgeschlagene „Leaky Bucket“-Mechanismus (LB) besteht im Prinzip aus einem Zähler mit endlichem Wertebereich, der bei jeder Zellankunft um eins inkrementiert und – falls sein Wert größer null ist – in festen Zeitabständen wieder um eins dekrementiert wird. Das erste Zeitintervall für die Dekrementierung wird von der ersten, beim Zählerstand null ankommenden, Zelle gestartet. Hat der Zähler zum Ankunftszeitpunkt einer Zelle seinen Höchststand bereits erreicht, dann wird eine Verletzung der Verkehrsparameter angenommen und die entsprechende Reaktion (siehe Abschnitt 2.6.2) veranlaßt. Das Verfahren hat zwei Parameter, nämlich die Länge des Dekrementierintervalls  $D$  und den Höchststand des Zählers  $N$ . Der Kehrwert des Dekrementierintervalls bestimmt die Zellrate  $\lambda_p$ , die kontrolliert werden soll ( $D = 1/\lambda_p$ ). Die Zählergrenze bestimmt die Toleranz gegenüber einem kurzfristigen Überschreiten dieser Zellrate.

Es wurden auch Varianten des LB-Mechanismus vorgeschlagen, bei denen eine Zellankunft eine Inkrementierung des Zählers um mehr als eins bewirkt oder bei denen eine Dekrementierung um mehr als eins am Ende des Dekrementierintervalls erfolgt. Diese Variationen ändern nichts am prinzipiellen Verhalten des Verfahrens und werden im folgenden nicht berücksichtigt.

Das LB-Verfahren ist sehr einfach zu implementieren. Man kommt dabei mit einem Zähler für den aktuellen Systemzustand, einem Zähler für die Messung des Dekrementierintervalls sowie zwei Variablen für die Zählergrenze bzw. die Dauer des Dekrementierintervalls aus. Der Aufwand für

- die Auswertung der Adresse im Kopffeld (VCI/VPI, siehe Abschnitt 2.4.6.2),
- die Entscheidung, ob eine Parameterverletzung vorliegt und
- die Implementierung der entsprechenden Reaktion auf eine Verletzung

wird hierbei nicht berücksichtigt, weil dieser für alle Verfahren identisch ist. Der Bereich für die Zählergrenze und für das Dekrementierintervall ist nur durch den Wertebereich der Zähler begrenzt und kann deshalb mit geringem Aufwand sehr groß gemacht werden.

Bezüglich der Wahrscheinlichkeit für das Erkennen einer Parameterverletzung kann der LB-Mechanismus als  $G/D/1-s$  Warteverlustsystem modelliert werden. Die deterministische Bedienzeit  $D$  entspricht dabei dem Dekrementierintervall, die Zählergrenze  $N$  wird durch die Anzahl der Wartepplätze berücksichtigt. Da sich bei dem  $G/D/1-s$  Modell noch eine Anforderung in der Bedieneinheit befinden kann, gilt für die Anzahl der Wartepplätze

$$s = N - 1 \quad (5.1)$$

Damit können alle im Kapitel 4 dargestellten Warteverlustsysteme für die Untersuchung des LB-Verfahrens verwendet werden.

Es soll hier betont werden, daß das Modell nur für die Verlustwahrscheinlichkeit gültig ist, da im realen System die Zellen nicht zwischengespeichert sondern sofort weitergegeben werden<sup>3</sup>. Insbesondere entspricht der Ausgangsprozeß des LB-Mechanismus nicht dem des  $G/D/1-s$  Systems. Der Ausgangsprozeß des LB-Verfahrens entspricht vielmehr dem Ankunftsprozeß der angenommenen Anforderungen. Dieser ergibt sich aus dem normalen Ankunftsprozeß durch Ausblenden der verworfenen Zellen.

### 5.2.2 Jumping Window

Der „Jumping Window“-Mechanismus (JW) basiert – wie auch die nachfolgend beschriebenen Mechanismen – auf einem Zeitfenster der Dauer  $T$ , in dem die Anzahl der akzeptierten Ankünfte auf ein Maximum von  $N$  begrenzt wird. Beim JW-Mechanismus sind die Zeitfenster auf der Zeitachse so angeordnet, daß sie ohne Überlappung oder Zwischenraum direkt hintereinander liegen. Der Startzeitpunkt des ersten Fensters ist dabei unabhängig von den Zellankünften. Während eines Zeitfensters wird die Anzahl der Ankünfte erfaßt. Kommen innerhalb eines Zeitfensters mehr als  $N$  Zellen an, dann wird eine Verletzung der Vertragsparameter angenommen und die überzähligen Zellen werden markiert oder verworfen. Beim Beginn eines neuen Zeitfensters wird der Zähler für die Ankünfte wieder auf null gesetzt. Eine Zelle beeinflusst deshalb den Zählerstand – je nach Ankunftszeitpunkt relativ zum Intervallanfang – für eine Zeitspanne, die zwischen null und der Intervalldauer  $T$  liegt. Das Verhältnis der Anzahl pro Intervall höchstens akzeptierter Zellen  $N$  zur Fensterbreite  $T$  bestimmt die Zellrate, die von dem Verfahren kontrolliert wird ( $\lambda_p = N/T$ ). Eine Erhöhung beider Werte im gleichen Verhältnis ändert an der Rate nichts, erhöht aber die Toleranz gegenüber einer kurzfristigen Überschreitung dieser Zellrate.

Der JW-Mechanismus ist vom Implementierungsaufwand her mit dem LB-Mechanismus vergleichbar. Es werden auch Zähler für die Messung des Intervalls  $T$  und für die Erfassung der

<sup>3</sup>Lediglich in [145] wird die Implementierung eines wirklichen Warteverlustsystems zur Formung des Verkehrs vorgeschlagen und als Leaky Bucket bezeichnet

Zahl von Ankünften sowie Variablen für die Speicherung der zwei Parameter benötigt. Der realisierbare Parameterbereich wird hier ebenfalls nur vom Wertebereich der verwendeten Zähler beschränkt.

Die Wahrscheinlichkeit für eine Verletzung des Verkehrsvertrages kann für Ankunftsprozesse mit Erneuerungseigenschaft direkt über die Verteilung des Zählprozesses  $n_i(t)$  berechnet werden, die die Wahrscheinlichkeit dafür angibt, daß in einem Intervall der Länge  $t$  genau  $i$  Ankünfte erfolgen (siehe Abschnitt 3.2.2.1). Da von  $N + i$  in einem Intervall ankommenden Zellen genau  $i$  verworfen werden, ergibt sich für die Verlustwahrscheinlichkeit beim JW-Verfahren

$$p_v = \frac{\sum_{i=1}^{\infty} i n_{N+i}(T)}{\sum_{i=1}^{\infty} i n_i(T)} \quad (5.2)$$

Die Summe im Nenner für die mittlere Anzahl von Ankünften im Intervall der Länge  $T$  ergibt nach Gleichung 3.10  $\lambda T$ , wobei  $\lambda$  die mittlere Ankunftsrate des Ankunftsprozesses repräsentiert. Die Summe im Zähler läßt sich wie folgt umformen

$$\begin{aligned} \sum_{i=1}^{\infty} i n_{N+i}(T) &= \sum_{i=N+1}^{\infty} (i - N) n_i(T) \\ &= \sum_{i=0}^{\infty} i n_i(T) - \sum_{i=0}^N i n_i(T) - N \left( \sum_{i=0}^{\infty} n_i(T) - \sum_{i=0}^N n_i(T) \right) \\ &= \lambda T - N - \sum_{i=0}^{N-1} (i - N) n_i(T) \end{aligned} \quad (5.3)$$

Diese Form hat den Vorteil, daß nur die ersten  $N$  Elemente der Verteilung des Zählprozesses benötigt werden. Insgesamt ergibt sich für die Verlustwahrscheinlichkeit also

$$p_v = \frac{\lambda T - N + \sum_{i=0}^{N-1} (N - i) n_i(T)}{\lambda T} \quad (5.4)$$

Diese Gleichung gilt sowohl für kontinuierliche als auch für zeitdiskrete Ankunftsverteilungen. Für beliebige zeitdiskrete oder diskretisierte Ankunftsverteilungen kann deshalb die Verlustwahrscheinlichkeit mit dem durch Gleichung 3.20 definierten Algorithmus bestimmt werden. Für die Ankunftsverteilungen M, Erlang-k, GEO<sub>0</sub>, GEO<sub>1</sub> und insbesondere für

die zweiphasige sporadische Quelle kann die Verlustwahrscheinlichkeit direkt mit Hilfe der in Abschnitt 3.3 explizit angegebenen Zählprozessverteilungen berechnet werden. Für die Berechnung bei der sporadischen Quelle ist es sinnvoll, die Summe im Zähler wie folgt umzuwandeln:

$$\begin{aligned} \sum_{i=1}^{\infty} i n_{N+i}(T) &= F^{(N+1)}(T) + 2F^{(N+2)}(T) + 3F^{(N+3)}(T) + \dots \\ &\quad - 1F^{(N+2)}(T) - 2F^{(N+3)}(T) - \dots \quad (5.5) \\ &= \sum_{i=N+1}^{\infty} F^{(i)}(T) \end{aligned}$$

$F^{(i)}(T)$  bezeichnet dabei die Verteilungsfunktion für die Zeit bis zur  $i$ -ten Ankunft (siehe Abschnitt 3.2.2.1). Nach der Transformation in den Laplace-Bereich ergibt sich analog zu Gleichung 3.10

$$\begin{aligned} \mathcal{LT} \left\{ \sum_{i=1}^{\infty} i n_{N+i}(T) \right\} &= \frac{\lambda}{s^2} (1 - a_{LST}(s)) \sum_{i=N+1}^{\infty} [a_{LST}(s)]^{i-1} \\ &= \frac{\lambda}{s^2} (1 - a_{LST}(s)) \left( \sum_{i=0}^{\infty} [a_{LST}(s)]^i - \sum_{i=0}^{N-1} [a_{LST}(s)]^i \right) \\ &= \frac{\lambda}{s^2} (1 - a_{LST}(s)) \left( \frac{1}{1 - a_{LST}(s)} - \frac{1 - [a_{LST}(s)]^N}{1 - a_{LST}(s)} \right) \\ &= \frac{\lambda}{s^2} [a_{LST}(s)]^N \quad (5.6) \end{aligned}$$

Für die Verlustwahrscheinlichkeit ergibt sich damit

$$p_v = \frac{\mathcal{LT}^{-1} \left\{ \frac{[a_{LST}(s)]^N}{s^2} \right\}_{t=T}}{T} \quad (5.7)$$

Die Vorschrift für die Rücktransformation des Zählers ist dabei durch Gleichung 3.47 gegeben.

### 5.2.3 Triggered Jumping Window

Wie bereits erwähnt ist beim JW-Mechanismus das Zeitfenster nicht mit dem Zellstrom synchronisiert. Um die dadurch auftretende Mehrdeutigkeit des Verfahrens zu vermeiden,

wurde der „Triggered Jumping Window“-Mechanismus vorgeschlagen, bei dem die Zeitfenster nicht direkt aufeinanderfolgen. Vielmehr wird nach dem Ablauf eines Zeitfensters das nächste erst gestartet, wenn die nächste Zelle ankommt. Die erste Zelle, die nach dem Ablauf eines Zeitfensters ankommt, beeinflusst also den Zählerstand für genau eine Intervalldauer, während alle folgenden – wie beim JW-Mechanismus – nur für eine Zeitdauer berücksichtigt werden, die zwischen null und der Intervalldauer liegt. Das Verfahren hat die gleichen zwei Parameter wie das JW-Verfahren. Durch die zwischen den Intervallen auftretenden Pausen ist jedoch das Verhältnis  $N/T$  nicht gleich der kontrollierten Zellrate, sondern bildet nur eine obere Grenze für diese.

Die Modifikation im Vergleich zum JW-Mechanismus ändert am Implementierungsaufwand nichts Entscheidendes, er ist ähnlich wie der für die bereits beschriebenen Verfahren.

Die Verlustwahrscheinlichkeit für das TJW-Verfahren kann ähnlich wie diejenige für das JW-Verfahren berechnet werden. Der einzige Unterschied ist der, daß man bei der Berechnung der Verteilung für die Anzahl von Ankünften pro Zeitfenster davon ausgehen muß, daß zum Startzeitpunkt des Intervalls immer eine Ankunft erfolgt, die den Zählerstand sofort um eins erhöht. Dadurch erhält man einen gewöhnlicher Erneuerungsprozeß, bei dem sich die Verteilungsdichte für die Zeit bis zum  $i$ -ten Ereignis aus Gleichung 3.3 dadurch ergibt, daß die Zeit bis zum ersten Ereignis nicht durch die Vorwärtsrekurrenzzeit gegeben sondern immer identisch null ist.

$$f^{(i)}(t) = \underbrace{a(t) \star \dots \star a(t)}_{(i-1) \text{ Terme}} \quad (5.8)$$

Analog wie in Abschnitt 3.2.2.1 ergibt sich für die Laplace-Transformierte  $n_{LT,i}(s)$  der Verteilung für die Anzahl der Ankünfte in einem Intervall der Länge  $t$

$$n_{LT,i}(s) = \begin{cases} 0, & i = 0 \\ \frac{1}{s} (1 - a_{LST}(s)) [a_{LST}(s)]^{i-1}, & i \geq 1 \end{cases} \quad (5.9)$$

Nach der Rücktransformation ergibt sich für  $n_i(t)$  die rekursive Form

$$n_i(t) = \begin{cases} 0, & i = 0 \\ 1 - A(t), & i = 1 \\ n_{i-1}(t) \star a(t), & i \geq 2 \end{cases} \quad (5.10)$$

Weil sich die mittlere Anzahl von Ankünften im allgemeinen Fall nicht einfach durch die Parameter der Ankunftsverteilung ausdrücken läßt, muß die Verlustwahrscheinlichkeit für

den TJW-Mechanismus nach der Gleichung

$$p_v = \frac{\sum_{i=1}^{\infty} i n_{N+i}(T)}{\sum_{i=1}^{\infty} i n_i(T)} \quad (5.11)$$

berechnet werden, wobei die Summen für Quellen mit einer endlichen maximalen Zellrate nur endlich viele Elemente umfassen. Zur Berechnung der Verlustwahrscheinlichkeit für die sporadische Quelle verwendet man sinnvollerweise die Beziehung

$$n_i(t) = \mathcal{L}T^{-1} \left\{ \frac{1}{s} a_{LST}(s)^{i-1} \right\}_{t=T} - \mathcal{L}T^{-1} \left\{ \frac{1}{s} a_{LST}(s)^i \right\}_{t=T} \quad n \geq 1 \quad (5.12)$$

Die Rücktransformationen sind dabei durch die Korrespondenzgleichung

$$\mathcal{L}T^{-1} \left\{ \frac{1}{s} a_{LST}(s)^n \right\} = \left[ \sum_{k=0}^n \binom{n}{k} (1 - \alpha\Delta)^{n-k} (\alpha\Delta)^k \left( 1 - e^{-\beta(t-n\Delta)} \left( \sum_{i=0}^{k-1} \frac{(\beta(t-n\Delta))^i}{i!} \right) \right) \right] \sigma(t - n\Delta) \quad (5.13)$$

gegeben. Ähnlich wie für das JW-Verfahren kann mit Hilfe der Gleichungen 5.10 und 5.11 ein numerischer Algorithmus für beliebige zeitdiskrete oder diskretisierte Ankunftsprozesse implementiert werden.

Die Verlustwahrscheinlichkeit ist – bei gleicher Dimensionierung und gleichem Ankunftsprozeß – für das TJW-Verfahren immer größer als diejenige für das JW-Verfahren. Dies ist dadurch bedingt, daß beim TJW-Verfahren die Zeiten ohne Zellankunft nach dem Ende eines Intervalls „abgeschnitten“ werden, weil das neue Intervall erst durch die erste ankommende Zelle gestartet wird. Beim JW-Mechanismus läuft das neue Intervall bereits und die Zellen werden deshalb in diesem Fall weniger lang beim Zählerstand berücksichtigt.

### 5.2.4 Moving Window

Bei dem „Moving Window“-Mechanismus (MW) werden ebenfalls Zeitfenster verwendet, wobei jedoch jede akzeptierte Zelle – im Gegensatz zu den bisher beschriebenen Fenstermechanismen – genau eine Fensterbreite lang beim Zählerstand berücksichtigt wird. Dadurch wird in jedem beliebigen Zeitintervall der Länge  $T$  die Anzahl akzeptierter Zellen auf maximal  $N$  begrenzt. Um dies zu erreichen, wird die Ankunftszeit jeder akzeptierten Zelle gespeichert und gleichzeitig der Zähler um eins erhöht. Nach genau einer Fensterbreite wird der Zähler

wieder um eins erniedrigt. Diesen Mechanismus kann man deshalb als Zeitfenster interpretieren, das sich gleichmäßig entlang der Zeitachse bewegt. Wie beim JW-Mechanismus ist die überwachte Zellrate durch den Quotienten aus der Zählergrenze und der Fensterbreite gegeben ( $\lambda_p = N/T$ ).

Die Implementierung dieses Verfahrens ist auch für relativ niedrige Zählergrenzen schon sehr viel aufwendiger als diejenige für die bisher vorgestellten Verfahren, weil in irgendeiner Form die Ankunftszeiten von  $N$  Zellen gespeichert werden müssen. Durch diese direkte Abhängigkeit der Implementierungskomplexität von der Zählergrenze ist auch die Flexibilität bei der Definition der Parameterbereiche stärker eingeschränkt als bei den anderen Mechanismen.

Der MW-Mechanismus kann als reines Verlustsystem mit mehreren deterministischen Bedieneinheiten ( $G/D/n-0$ ) modelliert werden. Die deterministischen Bedienzeiten sind dabei durch die Fensterbreite  $T$  bestimmt und die Anzahl der Bedieneinheiten entspricht der Zählergrenze  $N$ . Für einen Poisson-Ankunftsstrom kann die Verlustwahrscheinlichkeit des MW-Mechanismus unter Anwendung der Erlang'schen Verlustformel (siehe Abschnitt 4.2.1.2) berechnet werden. Lösungen für das  $G/D/n-0$  System mit einem zustandsabhängigen Poisson-Ankunftsprozess können durch Aufstellen der Zustandsgleichungen ebenfalls angegeben werden. Für die sporadische Quelle wird im folgenden die Untersuchung mittels Simulation vorgenommen.

Generell läßt sich sagen, daß die Verlustwahrscheinlichkeit für das MW-Verfahren höher ist als diejenige für das JW- und das TJW-Verfahren, wenn man die gleiche Dimensionierung und den gleichen Ankunftsprozeß voraussetzt. Das kommt daher, daß das MW-Verfahren die Anzahl der Ankünfte in allen möglichen Intervallen der Dauer  $T$  auf genau  $N$  begrenzt, während bei den anderen Verfahren Intervalle der gleichen Länge auftreten, in denen mehr als  $N$  Zellen akzeptiert werden.

### 5.2.5 Exponentially Weighted Moving Average

Der „Exponentially Weighted Moving Average“-Mechanismus (EWMA) benutzt feste, unmittelbar aufeinanderfolgende Zeitfenster der Dauer  $T$  wie der JW-Mechanismus. Im Unterschied zu diesem und den anderen bisher besprochenen Verfahren ist der andere Parameter des Verfahrens, nämlich die Zählergrenze, nicht fest sondern hängt von der Vorgeschichte ab. Dazu wird für die Zählergrenze im  $i$ -ten Intervall ( $N_i$ ) eine Funktion gewählt, in die neben der mittleren Zählergrenze  $N$  eine exponentiell gewichtete Summe der Anzahl von Ankünften in den vorhergehenden Intervallen ( $X_i$ ) eingeht.  $N_i$  berechnet sich nach der Vorschrift

$$N_i = \frac{N - \gamma S_{i-1}}{1 - \gamma} \quad 0 \leq \gamma < 1 \quad (5.14)$$



wobei für  $S_{i-1}$  gilt

$$S_{i-1} = (1 - \gamma)X_{i-1} + \gamma S_{i-2} \quad (5.15)$$

Diese Berechnungsvorschrift kann auch in der Form

$$N_i = \frac{N - (1 - \gamma)(\gamma X_{i-1} + \dots + \gamma^{i-1} X_1) - \gamma^{i+1} S_0}{1 - \gamma} \quad (5.16)$$

dargestellt werden, bei der man die exponentiell gewichtete Berücksichtigung der vergangenen Intervalle explizit sieht.  $S_0$  ist dabei der Anfangswert, mit dem das dynamische Verhalten des Verfahrens in der Einschwingphase beeinflusst werden kann. Ein weiterer Parameter des Verfahrens, der Dämpfungsfaktor  $\gamma$ , bestimmt die Toleranz des Verfahrens gegenüber einer kurzzeitigen Überschreitung der kontrollierten Zellrate. Für  $\gamma = 0$  ergibt sich ein fester Wert für  $N_i$  und das Verfahren ist identisch mit dem JW-Mechanismus. Ein  $\gamma$  größer null läßt einen büschelförmigeren Ankunftsverkehr zu. Der Quotient  $N/T$  gibt die kontrollierte Zellrate  $\lambda_p$  an, die sich ergibt, wenn  $i$  gegen unendlich strebt. Bild 5.1 zeigt den Einfluß des Faktors  $\gamma$  auf die Wahrscheinlichkeit der Verletzung des Quellflußkontroll-Kriteriums für eine sporadische Quelle mit den in [70] angegebenen Parametern für paketierte Sprache ( $E[X] = 22$ ,  $E[S] = 650ms$ ,  $\Delta = 16ms$ ). Es wird angenommen, daß die kontrollierte Zellrate gleich der mittleren Zellrate der Verbindung ist. Der Faktor  $\gamma$  beeinflusst bei sonst gleichen Bedingungen das Verhalten so, daß die Verletzungswahrscheinlichkeit mit zunehmendem  $\gamma$  abnimmt. Das bedeutet, daß das Verfahren damit toleranter auf kurzfristige Schwankungen der momentanen Zellrate reagiert, wie sie bei büschelförmigen Verkehrsströmen auftreten.

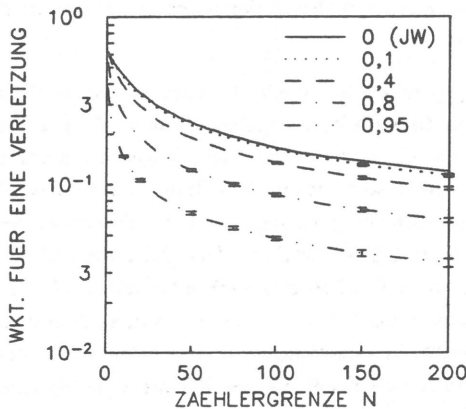


Bild 5.1: Einfluß des Faktors  $\gamma$  auf die Verletzungswahrscheinlichkeit

Für bestimmte Werte von  $\gamma$  (Zweierpotenzen, siehe [51]) kann die Berechnung der Zählergrenze, die zu Beginn jedes neuen Zeitfensters vorgenommen werden muß, sehr effektiv gemacht werden. Trotzdem ist das Verfahren in der Implementierung etwas aufwendiger als

z.B. das LB-Verfahren. Der Unterschied ist jedoch nicht so groß, daß er die Anwendbarkeit des EWMA-Verfahrens einschränkt.

### 5.3 Dimensionierung der Mechanismen

Wie in den vorigen Abschnitten erläutert, kann der Quotient aus Zählergrenze  $N$  und Fensterbreite  $T$  zur Festlegung der kontrollierten Zellrate  $\lambda_p$ , der vier Fenstermechanismen verwendet werden. Beim LB-Verfahren ist die kontrollierte Zellrate durch den Kehrwert des Dekrementierintervalls  $D$  gegeben. Bei einer Wahl von

$$T = \frac{N}{\lambda_p} \quad \text{und} \quad D = \frac{1}{\lambda_p} \quad (5.17)$$

sind deshalb alle Verfahren auf die gleiche kontrollierte Zellrate eingestellt. Die kontrollierte Zellrate ist bei den folgenden Untersuchungen nicht immer gleich der mittleren Zellrate des Ankunftsprozesses ( $\lambda$ ). Deshalb wird ein Faktor  $C$  eingeführt, der das Verhältnis dieser beiden Zellraten angibt

$$C = \frac{\lambda_p}{\lambda} \quad (5.18)$$

Für  $C = 1$  sind die Mechanismen also so dimensioniert, daß exakt die mittlere Zellrate der Quelle überwacht wird.

Statistische Verkehrsquellen werden wegen der kurzzeitigen, statistischen Fluktuationen in ihrer momentanen Zellrate das Überwachungskriterium mit einer bestimmten Wahrscheinlichkeit verletzen, obwohl über einen (unendlich) langen Zeitraum das Mittelwertkriterium eingehalten wird. Für eine feste kontrollierte Zellrate  $\lambda_p$ , kann diese Wahrscheinlichkeit beim LB-Verfahren durch eine Erhöhung der Zählergrenze reduziert werden, um die Transparenzforderung (siehe Abschnitt 5.1) zu erfüllen. Das gleiche gilt für die Fenstermechanismen, wenn dabei die Fensterbreite ebenfalls entsprechend erweitert wird, sodaß das Verhältnis von  $N$  und  $T$  konstant bleibt. Beim EWMA-Verfahren kann zu diesem Zweck auch der Faktor  $\gamma$  erhöht werden. Der Nachteil dieser Vorgehensweise besteht darin, daß sowohl eine Erhöhung der Zählergrenze als auch eine Erhöhung des Faktors  $\gamma$  die dynamische Reaktionszeit der Mechanismen verlängert.

Eine andere Möglichkeit, die Verletzungswahrscheinlichkeit zu vermindern, ist durch eine Überdimensionierung des Quellflußkontrollmechanismus gegeben, wobei ein Faktor  $C > 1$  verwendet wird. Dadurch wird jedoch die Fähigkeit der Verfahren zur Erkennung wirklicher, länger andauernder Überlast reduziert und die Verfahren verlieren an Selektivität.

Bei einer Wahl von  $N = 1$  und  $D = T$  sind der LB-Mechanismus, der MW-Mechanismus und der TJW-Mechanismus identisch und können im Prinzip mit dieser Dimensionierung zur Überwachung der maximalen Zellrate verwendet werden. Der JW- und der EWMA-Mechanismus sind mit dieser Dimensionierung nicht für eine Überwachung der maximalen Zellrate geeignet, weil der Startzeitpunkt der Fenster nicht durch eine Zellankunft bestimmt wird. Wenn der minimale Zellabstand durch Verzögerungsschwankungen auf dem Weg der Zellen bis zur Quellfußkontrollfunktion unterschritten werden kann, dann müssen ebenfalls die oben genannten Möglichkeiten zur Verringerung der Verletzungswahrscheinlichkeit angewendet werden (siehe z.B. [137]), wobei dann alle Mechanismen prinzipiell einsetzbar sind.

Bild 5.2 zeigt das prinzipielle Verhalten der Mechanismen für eine Zählergrenze von  $N = 2$  und eine mittlere Ankunftsrate von  $\lambda = 1/2$  Zellen pro Zeiteinheit.

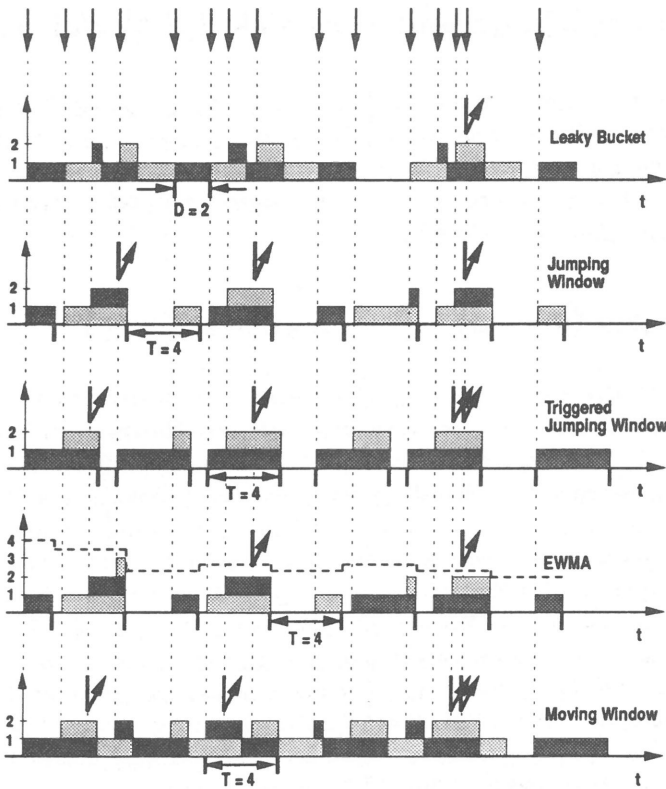


Bild 5.2: Prinzipielles Verhalten der Quellfußkontrollverfahren

Unter Zugrundelegung eines Faktors  $C = 1$  ergibt sich daraus eine Fensterbreite von  $T = 4$  Zeiteinheiten für die Fenstermechanismen und ein Dekrementierintervall von  $D = 2$  Zeiteinheiten für das LB-Verfahren. Beim EWMA-Mechanismus wurde  $\gamma = 0,5$  gesetzt und die gestrichelte Linie gibt jeweils die aktuelle Zählergrenze an. Es ist zu erkennen, daß beim MW-Verfahren jede Zelle genau 4 Zeiteinheiten lang beim Zählerstand berücksichtigt wird und deshalb innerhalb dieser Zeitspanne nie mehr als 2 Zellen akzeptiert werden. Für alle anderen Verfahren lassen sich Intervalle der Länge  $T = 4$  finden, in denen mehr als zwei Zellen angenommen werden. Dadurch weisen diese Verfahren allgemein eine kleinere Verlustwahrscheinlichkeit auf als das MW-Verfahren. Der Unterschied in der Lage der Zeitfenster zwischen dem JW-Mechanismus und dem TJW-Mechanismus ist ebenfalls aus Bild 5.2 ersichtlich.

## 5.4 Untersuchung und Vergleich der Verfahren

Bei den folgenden Untersuchungen wird angenommen, daß die maximale Zellrate durch einen separaten Quellflußkontrollmechanismus mit entsprechender Dimensionierung überwacht wird. Folglich wird bei einer Variation der mittleren Zellrate – sofern möglich – die maximale Zellrate konstant gehalten, d.h. es wird angenommen, daß die Quelle die vereinbarte maximale Zellrate nicht überschreitet.

### 5.4.1 Exakte Überwachung der mittleren Zellrate

Zunächst soll für eine sporadische Quelle mit den Standardparametern für paketierte Sprache ( $E[X] = 22$ ,  $E[S] = 650ms$ ,  $\Delta = 16ms$ ) versucht werden, den Mittelwert der Quelle exakt zu überwachen ( $C = 1$ ,  $\lambda = \lambda_p$ ). Da die Quellen über einen längeren Zeitraum das Mittelwertkriterium respektieren und sich deshalb an den vereinbarten Verkehrsvertrag halten, müssen die Mechanismen so dimensioniert werden, daß trotz der statistischen Schwankungen der Zellrate die Transparenzforderung erfüllt ist. Bild 5.3 zeigt den Einfluß der Zählergrenze auf die Verlustwahrscheinlichkeit der verschiedenen Verfahren. Für den Parameter  $\gamma$  des EWMA-Verfahrens wurde ein Wert von 0,8 angenommen. Es ist aus den Kurven klar ersichtlich, daß bei allen Verfahren unrealistisch hohe Zählergrenzen nötig wären, um in einen Bereich kleiner  $10^{-9}$  für die Verlustwahrscheinlichkeit zu kommen. Selbst bei einer Reduzierung dieser Forderung auf eine Verlustwahrscheinlichkeit im Bereich von  $10^{-5}$  würde z.B. der LB-Mechanismus, der den günstigsten Verlauf der Verlustwahrscheinlichkeit aufweist, eine Zählergrenze von über 100000 benötigen. Das würde aber bedeuten, daß eine Quelle vor dem ersten Verlust weit über 154000 Zellen ohne Pause mit der maximalen Zellrate senden könnte, was einer Dauer von fast zwei Stunden entspricht.

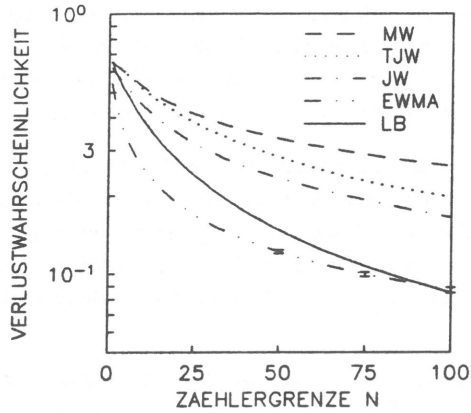


Bild 5.3: Einfluß der Zählergrenze auf die Verlustwahrscheinlichkeit

### 5.4.2 Überdimensionierung der Verfahren

Als nächstes soll der Einfluß einer Überdimensionierung der Verfahren ( $C > 1$ ,  $\lambda < \lambda_p$ ) für eine Quelle mit denselben Parametern wie im vorigen Abschnitt untersucht werden. Der in Bild 5.4 verwendete Faktor  $C = 2,277$  wurde dabei so gewählt, daß für den LB-Mechanismus bei einer Zählergrenze von 100 eine Verlustwahrscheinlichkeit im Bereich von  $10^{-9}$  erreicht wird.

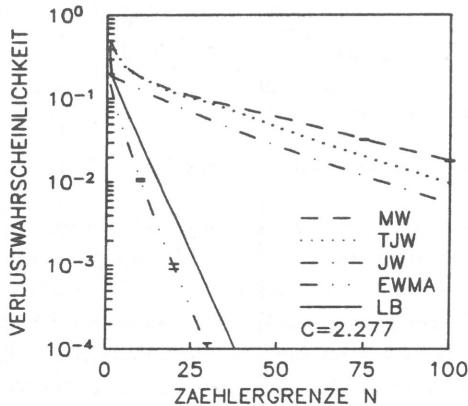


Bild 5.4: Einfluß der Zählergrenze bei einer Überdimensionierung

Es muß hier erwähnt werden, daß bei dieser Dimensionierung die überwachte Zellrate von 50 Zellen pro Sekunde schon relativ nahe bei der maximalen Zellrate der Quelle von 62,5

Zellen pro Sekunde liegt. Der LB- und der EWMA-Mechanismus zeigen eine ähnliche Reaktion auf die Überdimensionierung, während für die anderen Fensterverfahren nur bei einer kontrollierten Zellrate, die praktisch der maximalen Zellrate entspricht, mit vergleichbaren Zählergrenzen eine entsprechende Verlustwahrscheinlichkeit erzielt werden kann.

Für die Kontrolle einer näher am Mittelwert liegenden Zellrate mit  $C = 1,1$  wäre für den LB-Mechanismus eine Zählergrenze von etwa 1800 nötig, um die gleiche Verlustwahrscheinlichkeit zu erreichen. Eine solche Dimensionierung würde es der Quelle erlauben, ein Bündel von 2933 Zellen (entsprechend einer Dauer von ca. 47 Sekunden) ohne Unterbrechung mit der maximalen Zellrate zu senden bevor die erste Zelle verloren geht, wenn zu Beginn der Zählerstand 0 war.

Der Einfluß des Überdimensionierungsfaktors  $C$  auf die Fähigkeit der Mechanismen, eine wirkliche, länger andauernde Überlast zu erkennen, ist in den Bildern 5.5 und 5.6 dargestellt.

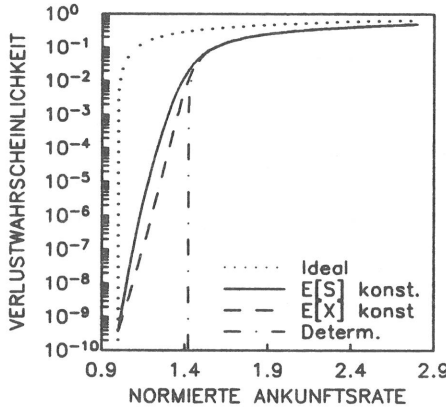


Bild 5.5: Einfluß der Überlast beim LB

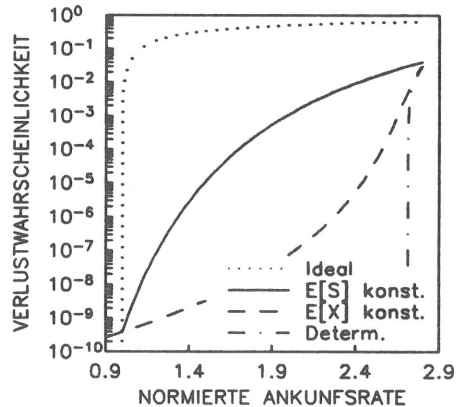


Bild 5.6: Einfluß der Überlast beim JW

Um den JW-Mechanismus bei einer Zählergrenze von 100 auf eine Verlustwahrscheinlichkeit im Bereich von  $10^{-10}$  dimensionieren zu können – ohne  $\lambda_P$  gleich der maximalen Zellrate der Quelle setzen zu müssen – wurden die Parameter der Quelle so verändert, daß sich ein weniger büschelförmiger Verkehr mit gleicher mittlerer und maximaler Zellrate ergibt. Dazu wurde  $E[X] = 5$  und  $E[S] = 147,72ms$  gewählt, was einen Variationskoeffizienten von  $c_a = 1,95$  im Vergleich zu  $c_a = 4,25$  bei den Sprachparametern ergibt. Die Verlustwahrscheinlichkeit ist über der wirklichen mittleren Zellrate der Quelle aufgetragen, die auf die beim Verbindungsaufbau angegebene mittlere Zellrate normiert ist.

Die gepunktete Kurve in den Bildern gibt den idealen Verlauf der Verlustwahrscheinlichkeit an. Dieser Verlauf ergibt sich aus der Selektivitätsanforderung:

- Bis zur im Verkehrsvertrag vereinbarten mittleren Zellrate soll der Mechanismus für die Quelle transparent sein, d.h. bis zu einer normierten Zellrate von 1 gilt  $p_v = 0$ .
- Im Überlastbereich, wo der Verkehrsvertrag verletzt wird, sollen alle überschüssigen Zellen verworfen werden. Die Verlustwahrscheinlichkeit in diesem Bereich berechnet sich deshalb zu

$$p_v = \frac{\lambda - \lambda_p}{\lambda} \quad (5.19)$$

Der JW- und der LB-Mechanismus wurden für die nominale, beim Verbindungsaufbau vereinbarte Zellankunftsrate auf die gleiche Verlustwahrscheinlichkeit ( $p_v \approx 3 \cdot 10^{-10}$ ) dimensioniert. Bei gleicher Zählergrenze von  $N = 100$  ist hierzu beim LB-Verfahren ein Faktor  $C = 1,42$  nötig, während für das JW-Verfahren ein Faktor  $C = 2,71$  benötigt wird. Für die Erhöhung der mittleren Zellrate wurden zwei Varianten verwendet. Bei der ersten wurde die mittlere Pausendauer konstant gehalten und die mittlere Anzahl von Zellen pro Aktivitätsphase erhöht, bei der zweiten wurde diese Anzahl konstant gehalten und dafür die mittlere Pausendauer verkürzt. Wie erwartet nähert sich der Verlauf der Verlustwahrscheinlichkeit beim LB-Verfahren mit  $C = 1,42$  für beide Varianten schneller an die Idealkurve an als derjenige für den JW-Mechanismus mit  $C = 2,71$ , was eine bessere Sensitivität bezüglich der Überlast anzeigt.

Eine Verkehrsquelle, die den von den Verfahren gebotenen Spielraum voll ausnutzt, z.B. eine deterministische Quelle, kann mit der vollen Zellrate  $\lambda_p$  senden, ohne dabei Verluste zu erleiden. Das bedeutet, daß eine solche Quelle die vereinbarte mittlere Zellrate bis zu einem Faktor  $C$  (2,71 beim JW-Verfahren!) überschreiten kann, ohne daß dies als Verletzung erkannt wird.

Will man den JW-Mechanismus so dimensionieren, daß er bei gleicher Verlustwahrscheinlichkeit für gutartige Quellen ähnlich sensitiv auf Überlast reagiert wie der LB-Mechanismus, kann man einen ähnlichen Faktor  $C$  wählen. Dann benötigt man allerdings eine Zählergrenze im Bereich von 750 und ein entsprechend langes Intervall  $T$ . Bei einer solchen Dimensionierung wäre deshalb das dynamische Verhalten des JW-Verfahrens wesentlich schlechter als das des LB-Verfahrens (siehe unten).

Das MW-Verfahren hat bei gleicher Dimensionierung immer eine höhere Verlustwahrscheinlichkeit als das JW-Verfahren und weist deshalb ein noch ungünstigeres Verhalten auf. Außerdem wäre die Implementierung dieses Verfahrens für die oben genannten Zählergrenzen schon relativ aufwendig (siehe Abschnitt 5.2.4).

Bild 5.7 zeigt den entsprechenden Vergleich zwischen dem LB-Mechanismus ( $C = 1,42$ ) und dem EWMA-Mechanismus. Für den EWMA-Mechanismus wurden zwei Dimensionierungsvarianten untersucht, die eine mit  $\gamma = 0,8$  und  $C = 1,42$ , die andere mit  $\gamma = 0,91$  und  $C = 1,27$ . Beide Varianten ergeben bei der verwendeten Zählergrenze von  $N = 45$  für die

nominale Ankunftsrate die gleiche Verlustwahrscheinlichkeit wie das LB-Verfahren. Bedingt durch die Beschränkungen der simulativen Methoden (siehe Abschnitt 4.1.3), die für die Untersuchung des EWMA-Verfahrens angewendet wurden, mußte für diesen Vergleich eine nominale Verlustwahrscheinlichkeit im Bereich von  $10^{-5}$  verwendet werden. Die Erhöhung der mittleren Zellrate erfolgte durch Verlängerung der aktiven Phase bei gleicher Pausendauer.

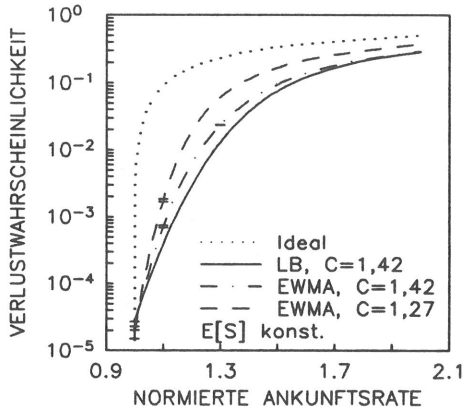


Bild 5.7: Einfluß der Überlast beim LB und beim EWMA

Für diese Dimensionierung zeigt das EWMA-Verfahren mit Faktor  $\gamma = 0,8$  bei geringer Überlast ein etwas günstigeres Verhalten bezüglich der Überlasterkennung, das jedoch mit einer längeren dynamischen Reaktionszeit (siehe nächsten Abschnitt) erkauft werden muß. Das Gleiche gilt im Prinzip auch für den EWMA-Mechanismus mit  $\gamma = 0,91$ .

### 5.4.3 Dynamisches Verhalten der Verfahren

Neben dem stationären Verhalten der Verfahren ist das dynamische Verhalten von besonderer Bedeutung für die Bewertung der Mechanismen. Bild 5.8 zeigt deshalb das prinzipielle Einschwingverhalten der verschiedenen Verfahren im Vergleich. Die Verfahren wurden dabei so dimensioniert, daß sich für eine sporadische Quelle mit Sprachparametern und einen Faktor  $C = 2,277$  eine stationäre Verlustwahrscheinlichkeit von  $10^{-2}$  bei allen Verfahren einstellt<sup>4</sup> (vergleiche auch Bild 5.4). Bei dieser Dimensionierung ergeben sich folgende Zählergrenzen:

- EWMA ( $\gamma = 0,8$ ):  $N = 10$ ,
- LB:  $N = 15$ ,

<sup>4</sup>Dieser Wert ist unrealistisch hoch, mußte aber wegen der erforderlichen simulativen Analyse gewählt werden



- JW:  $N = 80$ ,
- TJW:  $N = 100$ ,
- MW:  $N = 124$ .

Es wird angenommen, daß die Quelle mit dem 1,5-fachen der vereinbarten mittleren Zellrate sendet ( $E[X]$  wurde zu 45,25 gewählt).

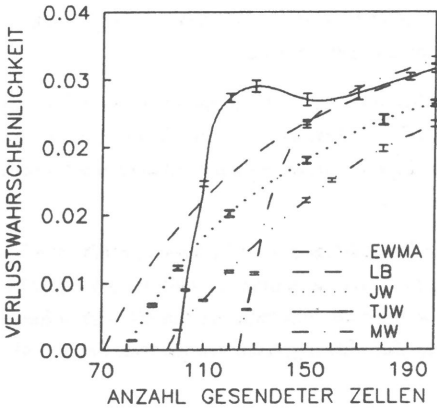


Bild 5.8: Dynamisches Verhalten der verschiedenen Verfahren

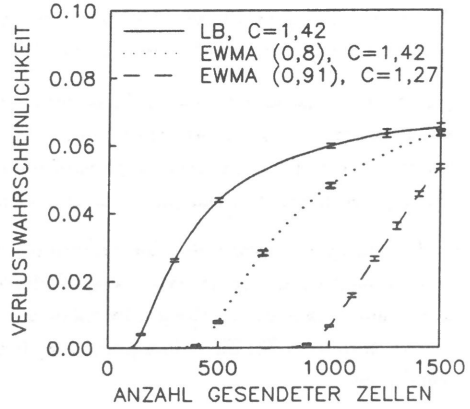


Bild 5.9: Dynamisches Verhalten des LB- und des EWMA-Verfahrens

Die Verlustwahrscheinlichkeit für das MW- und das TJW-Verfahren zeigen einen steilen Anstieg, wobei die Ordnungsnummer der Ankunft, bei der dieser Anstieg beginnt, durch die jeweilige Zählergrenze gegeben ist. Der steile Anstieg kommt dadurch zustande, daß bei beiden Verfahren der Startzeitpunkt des Fensters durch die Zellankünfte bestimmt wird. Dadurch ist die Anzahl akzeptierter Zellen in einem Zeitintervall der Länge  $T$  nach der ersten Zellankunft strikt auf  $N$  limitiert. Der Bereich abnehmender Verlustwahrscheinlichkeit beim TJW-Mechanismus wird durch den sprunghaften Wechsel in der Lage des Zeitfensters bedingt.

Die drei anderen Verfahren akzeptieren mehr als  $N$  Zellen in den ersten  $T$  Zeiteinheiten nach der ersten Zellankunft. Beim LB-Verfahren kommt dies dadurch zustande, daß bis zur  $N$ -ten Ankunft der Zähler bereits mehrfach dekrementiert worden ist (siehe Abschnitt 5.4.5). Beim JW-Verfahren ist dies dadurch bedingt, daß die Lage des Zeitfensters nicht durch die erste Zellankunft bestimmt wird. Das EWMA-Verfahren wiederum akzeptiert für  $S_0 = 0$  im ersten Intervall maximal  $N_i = N/(1 - \gamma)$  Zellen<sup>5</sup>, das sind in diesem Fall 50. Da die Quelle pro Intervall  $T$  maximal 12 Zellen senden kann, sind einige Fensterdauern erforderlich, bis sich die

<sup>5</sup>Durch eine Wahl von  $S_0 > 0$  kann das Verhalten zu Beginn einer Verbindung gezielt beeinflusst werden

Zählergrenze so verringert hat, daß entsprechend viele Zellen verworfen werden. Durch diese Eigenschaften weisen diese Verfahren einen flacheren Anstieg der Verlustwahrscheinlichkeit in der transienten Phase auf und die Ordnungsnummer der Ankünfte, bei denen der Anstieg beginnt, ist signifikant größer als die Zählergrenze  $N$ . Dadurch, daß diese Verfahren jedoch im Vergleich zu den beiden ersten Verfahren eine wesentlich niedrigere Zählergrenze benötigen, weisen sie ein durchaus vergleichbares – wenn nicht besseres – dynamisches Verhalten auf. Darüberhinaus nimmt der Unterschied zwischen den benötigten Zählerständen bei einer Dimensionierung auf kleinere Verlustwahrscheinlichkeiten weiter zu (siehe Bild 5.4), wodurch weitere Vorteile für das LB- und das JW-Verfahren zu erwarten sind.

Für die dynamische Reaktion auf Lastwechsel während einer Verbindung läßt sich im Prinzip ein ähnliches Verhalten erwarten, wobei beim EWMA-Verfahren – im Gegensatz zu den anderen Verfahren – die ganze Vorgeschichte mit eingeht und deshalb der oben angenommene Anfangszustand nicht mehr ganz erreicht wird.

Bild 5.9 zeigt die entsprechenden Kurven für den Vergleich des LB- mit dem EWMA-Mechanismus für das Beispiel aus dem vorigen Abschnitt. Es ist zu erkennen, daß der EWMA-Mechanismus wegen des relativ großen Faktors  $\gamma$  von 0,8 bzw. 0,91 und der damit verbundenen hohen Startwerte für die Zählergrenze eine längere Einschwingphase benötigt als der LB-Mechanismus.

#### 5.4.4 Einfluß der Quellcharakteristik

Am Beispiel des LB-Mechanismus soll hier die Sensitivität der Ergebnisse bezüglich der Annahmen für die Verkehrsquellen untersucht werden. In Bild 5.10 ist die Abhängigkeit der Verlustwahrscheinlichkeit vom Burstiness-Faktor  $B^6$  und vom Variationskoeffizienten dargestellt. Als Quelle wurde eine sporadische Quelle verwendet, bei der der Mittelwert konstant gehalten wurde. Der Überdimensionierungsfaktor  $C$  wurde zu 1,25 gesetzt und die Zählergrenze betrug  $N = 51$ . Es ist deutlich erkennbar, daß besonders der Variationskoeffizient der Zwischenankunftsabstände einen entscheidenden Einfluß auf die Verlustwahrscheinlichkeit hat. Selbst wenn er nur im Bereich von 1 bis 5 variiert wird, ergeben sich bei der Verlustwahrscheinlichkeit Unterschiede von fast 10 Dekaden.

Bild 5.11 zeigt einen Vergleich für Quellen mit realistischen, bestimmten Diensten entsprechenden Parametern. Dazu werden neben einer Poisson-Quelle und einer sporadischen Quelle mit Sprachparametern folgende Quellen verwendet:

- Eine sporadische Quelle, bei der die Parameter nach [61] so gewählt wurden, daß eine interaktive Standbildübertragung mit einer Spitzenbitrate von 2 Mbit/s ( $\Delta = 256\mu s$ )

---

<sup>6</sup>Der Burstiness-Faktor  $B$  ist definiert als  $\frac{\text{Maximale Zellrate}}{\text{Mittlere Zellrate}}$

und einer mittleren Bitrate von 86,96 kbit/s modelliert wird. Die so dimensionierte Quelle weist sehr viel längere Aktivitätsphasen auf ( $E[X] = 1953,13$ ), die durch die Datenmenge für einen Bildschirminhalt bestimmt sind. Auf der anderen Seite sind auch die Pausenintervalle wesentlich länger ( $E[S] = 11s$ ), wodurch sich ein sehr hoher Variationskoeffizient von  $c_a = 59,77$  ergibt. Ein ähnliches Verhalten ist auch bei interaktiven Datendiensten zu erwarten.

- Eine Videoquelle, wie sie in Abschnitt 3.3.5 dargestellt wird. Dabei werden die dort beschriebenen, extremen Varianten für das Verhalten innerhalb eines Bildes verwendet. Als Variante I wird diejenige bezeichnet, bei der die Zellen möglichst gleichmäßig über die ganze Dauer des Bildes verteilt werden, Variante II bezeichnet den Fall, bei dem alle Zellen mit der maximalen Zellrate des Codecs zu Beginn des Rahmens übertragen werden.

Bei der Berechnung der entsprechenden Zellraten wird wie bei den Sprachparametern eine Informationsfeldlänge von 64 Byte angenommen. Diese Informationsfeldlänge stimmt zwar nicht mit der von CCITT beschlossenen (48 Byte, siehe Abschnitt 2.4.6.1) überein, für die hier vorgestellten Untersuchungen spielt diese Abweichung jedoch keine Rolle.

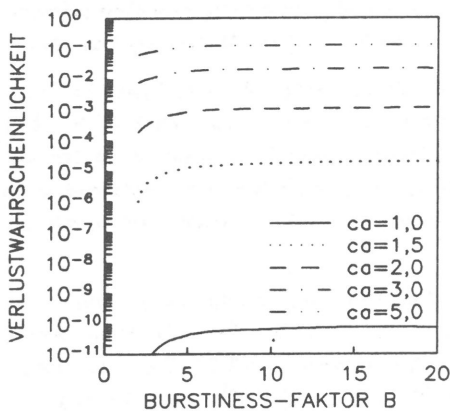


Bild 5.10: Einfluß der Quellparameter der sporadischen Quelle

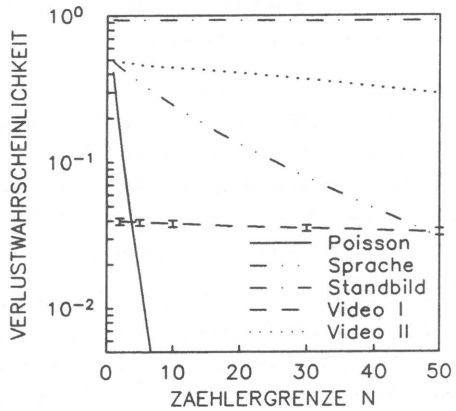


Bild 5.11: Verhalten des LB-Verfahrens für realistische Quellen

Die Poisson-Quelle liefert – wie erwartet – sehr optimistische Ergebnisse, während sich für den Fall der Standbildübertragung eine im Vergleich zur Sprachquelle deutlich höhere Verlustwahrscheinlichkeit ergibt. Für eine Überwachung in der Nähe der mittleren Zellrate ( $C = 1,1$ ) wäre für eine solche Quelle eine Zählergrenze von ungefähr 400000 nötig, um eine Verlustwahrscheinlichkeit im Bereich von  $10^{-10}$  zu erreichen. Eine solche Dimensionierung ist offensichtlich unrealistisch und es bleibt deshalb nur die Möglichkeit bei einer solchen Quelle

die kontrollierte Zellrate in die Nähe der maximalen Zellrate zu legen. Ähnliches gilt auch für die beiden Videoquellen, bei denen die Zählergrenzen um Größenordnungen höher angesetzt werden müßten, um tolerierbare Verlustwahrscheinlichkeiten zu erhalten.

#### 5.4.5 Im ungünstigsten Fall zugelassener Verkehr

Der im ungünstigsten Fall von den Verfahren ins Netz gelassene Verkehr bestimmt entscheidend die Güte eines Quellflußkontrollverfahrens. Dieser Verkehr beeinflusst auch sehr stark die Dimensionierung der CAC-Funktion, die über die Annahme neuer Verbindungen entscheidet (siehe Abschnitt 2.6.4).

Ein erstes Kriterium bei der Definition eines solchen Verkehrs ist, daß in jedem Fall die kontrollierte Zellrate  $\lambda_p$  voll ausgeschöpft wird. Dieses Kriterium wird z.B. von einer deterministischen Quelle erfüllt. Ein deterministischer Verkehr ist jedoch sehr gleichmäßig und läßt sich deshalb sehr effektiv Multiplexen. Ein weiteres Kriterium ist deshalb die Büschelförmigkeit des Verkehrs, die den beim statistischen Multiplexen erzielbaren Bandbreitengewinn signifikant reduzieren kann. Als „Worst Case“-Verkehr wird deshalb derjenige Verkehr definiert, der bei voller Ausnutzung der kontrollierten Zellrate die höchste Büschelförmigkeit aufweist. Der Verkehr wird dementsprechend durch ein periodisches Ein/Aus-Muster gegeben sein.

Beim MW-Verfahren ist dieses Muster durch eine alternierende Folge von Aktivitätsphasen mit genau  $N$  Zellen und von Pausenphasen mit einer konstanten Dauer von  $T - (N - 1)\Delta$  Zeiteinheiten gegeben und hat damit eine Periodendauer von  $T$  Zeiteinheiten.  $\Delta$  bezeichnet dabei die maximale Zellrate der Verbindung, die - bei Fehlen einer entsprechenden zusätzlichen Überwachung - im Extremfall gleich der maximalen Nutzübertragungsrate der Benutzer/Netz-Schnittstelle sein kann.

Wenn sich beim JW-Mechanismus die Quelle auf das Zeitfenster synchronisiert - was relativ einfach bewerkstelligt werden kann - ist es möglich, auf den beiden Seiten der Fenstergrenze zwei Aktivitätsphasen mit je  $N$  Zellen zu einer Aktivitätsphase mit  $2N$  Zellen zu vereinen. Danach wird dann eine Pause der Dauer  $2T - (2N - 1)\Delta$  Zeiteinheiten benötigt und es entsteht ein Muster mit Periodendauer  $2T$ . Das gleiche Muster charakterisiert auch den ungünstigsten Verkehr im Falle eines EWMA-Mechanismus im stationären Fall. Allerdings wird hier, bedingt durch die variable Zählergrenze, für eine bestimmte Zeit ein noch ungünstigerer Verkehr zugelassen.

Beim TJW-Mechanismus kann sich die Quelle durch das Senden einer Zelle, die den Start des Zeitfensters hervorruft, auf dieses synchronisieren. Ist dies geschehen, dann ergibt sich ein ähnliches Muster wie beim JW-Mechanismus. Der einzige Unterschied ist, daß jeweils eine Zelle benötigt wird, um das jeweils nächste Fenster anzustoßen. Das gesamte,  $2T$  Zeiteinheiten lange Muster ergibt sich also aus einer Folge von einer einzelnen Zelle, einer Pause

der Dauer  $T - (N - 2)\Delta$ , einer Aktivitätsphase mit  $2N - 1$  Zellen und einer weiteren Pause von  $T - (N - 1)\Delta$  Zeiteinheiten.

Beim LB-Verfahren kann der ungünstigste Verkehr dadurch erzeugt werden, daß man abwechselnd den Zähler in möglichst kurzer Zeit auf seinen Maximalstand bringt und dann wartet, bis er wieder null erreicht hat. Die maximale Anzahl von Zellen pro Aktivitätsphase  $X_{max}$  hängt beim LB-Verfahren von der maximalen Ankunftsrate der Quelle ab und ist gleich der kleinsten ganzen Zahl, die die Ungleichung

$$X_{max} > \frac{N - 1}{1 - \Delta\lambda_p}, \quad \text{für } \Delta < \frac{1}{\lambda_p} \quad (5.20)$$

erfüllt. Die erforderliche Pause hat eine Länge von  $X_{max}(1/\lambda_p - \Delta) + \Delta$  und das Muster damit eine Periodendauer von  $X_{max}/\lambda_p$ . Für den Vergleich in den Bildern 5.5 und 5.6 mit einer Zählergrenze von  $N = 100$  würde sich für den LB-Mechanismus eine Aktivitätsphase von 198 Zellen ergeben, die fast identisch ist mit derjenigen des JW-Verfahrens (200 Zellen). Die nach einer solchen Aktivitätsphase erforderliche Pause hätte beim LB-Verfahren eine Dauer von 3184ms, beim JW-Mechanismus jedoch nur von 172ms, was deutlich schlechter ist. Für den Vergleich von LB- und EWMA-Verfahren in Bild 5.7 ergibt sich bei einer Zählergrenze von  $N = 45$  eine Aktivitätsphase mit 89 Zellen für den LB-Mechanismus gegenüber 90 Zellen beim EWMA-Verfahren. Die Pausendauern mit 1440ms für den LB-Mechanismus und 1456ms für den EWMA-Mechanismus haben in diesem Fall ähnliche Werte. Wie bereits erwähnt, gestattet der EWMA-Mechanismus aber für einen bestimmten Zeitraum ein noch ungünstigeres Verhalten.

Aus den obigen Untersuchungen ist klar ersichtlich, daß das LB-Verfahren und das EWMA-Verfahren deutliche Vorteile gegenüber den anderen drei Verfahren aufweisen. Diese reagieren wesentlich unflexibler auf die statistischen Schwankungen in der momentanen Zellrate und erfordern deshalb eine für das Netz weitaus ungünstigere Dimensionierung. Beim MW-Mechanismus ist darüberhinaus die Implementierung für realistische Zählergrenzen so aufwendig, daß dieses Verfahren kaum anwendbar ist.

Das LB- und das EWMA-Verfahren haben eine vergleichbare Leistungsfähigkeit, wobei das LB-Verfahren leichte Vorteile beim dynamischen Verhalten aufweist, wogegen das EWMA-Verfahren – abhängig von der Dimensionierung – Vorteile bei der Erkennung von langandauernder Überlast hat. Der Mehraufwand bei der Implementierung des EWMA-Verfahrens ist so gering, daß er bei der Bewertung keine Rolle spielt.

## 5.5 Vorschläge für verbesserte Mechanismen

Die Untersuchungen im vorigen Abschnitt zeigen, daß eine Erfüllung aller in Abschnitt 5.1 aufgeführten Anforderungen mit den bekannten Mechanismen nicht möglich ist. Das prinzipielle Problem dabei ist, daß es nicht möglich ist, auf der Basis einer relativ kleinen Stichprobe (z.B. Stichprobe der Dauer  $T$  bei den Fensterverfahren) den Mittelwert, der von der Quelle erst über längere Zeiträume einigermaßen exakt eingehalten wird, mit der erforderlichen Genauigkeit vorherzusagen. Dieses Problem kann auch mit komplizierteren Verfahren nicht umgangen werden. Das Verhalten der Quellflußkontrollfunktion kann jedoch dadurch verbessert werden, daß gezielt die Toleranz der Verfahren gegenüber den statistischen Schwankungen erhöht wird, wobei spezifische Eigenschaften der Quelltypen mit berücksichtigt werden können. In diesem Abschnitt werden deshalb zwei Verfahren vorgestellt und untersucht, die durch zusätzliche Parameter eine größere Flexibilität aufweisen und trotzdem noch einfach zu implementieren sind. Die Verfahren sind auf die Eigenschaften der sporadischen Quelle abgestimmt, wie sie sich in vielen Anwendungsfällen ergeben wird.

### 5.5.1 Adaptive Counter

Die Zellverluste sind hauptsächlich dadurch bedingt, daß die Verfahren nicht an die Büschelförmigkeit des Quellverkehrs angepaßt sind. Deshalb wird hier das „Adaptive Counter“-Verfahren (AC) vorgeschlagen, das sich an den Zustand der Quelle (Ein/Aus) anpaßt. Das Verfahren funktioniert folgendermaßen:

- Es werden Zeitintervalle der Dauer  $T$  definiert, über die hinweg die Einhaltung einer mittleren Zellrate  $\lambda_p$  kontrolliert wird. Die Zeitintervalle schließen – wie beim JW-Mechanismus – direkt aneinander an, wodurch auch die Einhaltung dieses Mittelwertes über die ganze Verbindungsdauer (bis auf das letzte Teilintervall) überwacht wird.
- Die Anzahl angekommener und akzeptierter Zellen wird mit Hilfe eines Zählers registriert. Für den Zählerstand wird ein Maximum  $N$  definiert und Zellen, die den Zähler auf seinem maximalen Stand antreffen, werden verworfen.
- Während sich die Quelle in einer Aktivitätsphase befindet, wird der Zähler – sofern sein Wert größer null ist – jeweils nach einer Zeit  $D_A$  dekrementiert, die sich zu

$$D_A = \frac{F_A}{\lambda_p}, \quad F_A \leq 1 \quad (5.21)$$

ergibt und damit kürzer ist als das mittlere Dekrementierintervall  $D = 1/\lambda_p$  ist.  $F_B$  ist dabei ein frei wählbarer Faktor kleiner (oder gleich) eins, dessen kleinster, sinnvoller Wert durch den minimalen Zellabstand bestimmt wird.

- Während sich die Quelle in einer Pausenphase befindet, wird der Zähler jeweils nach einer Zeit  $D_P$  dekrementiert, die sich entsprechend zu

$$D_P = \frac{F_P}{\lambda_p}, \quad F_P \geq 1 \quad (5.22)$$

ergibt und länger ist als das mittlere Dekrementierintervall  $D$ .

- Um die mittlere Zellrate von  $\lambda_p$  einhalten zu können, wird der Zähler in jedem Intervall  $T$  maximal  $X_A$  mal nach  $D_A$  Zeiteinheiten und maximal  $X_P$  mal nach  $D_P$  Zeiteinheiten dekrementiert.  $X_A$  und  $X_P$  ergeben sich dabei aus der Bedingung

$$T\lambda_p = X_A + X_P = X_A F_A + X_P F_P \quad (5.23)$$

Daraus berechnet sich  $X_P$  zu

$$X_P = \left\lfloor \frac{1 - F_A}{F_P - F_A} (T\lambda_p) \right\rfloor \quad (5.24)$$

wobei  $\lfloor Q \rfloor$  die größte ganze Zahl kleiner gleich  $Q$  bezeichnet. Es wird vorausgesetzt, daß die Intervallbreite ein ganzzahliges Vielfaches des mittleren Dekrementierintervalls ist. Für  $X_A$  ergibt sich daraus

$$X_A = (T\lambda_p) - X_P \quad (5.25)$$

Ist einer der beiden Maximalwerte innerhalb eines Intervalls ausgeschöpft, dann wird unabhängig vom Zustand der Quelle mit der Rate dekrementiert, deren Maximalwert noch nicht ausgeschöpft ist. Durch die Bedingung, daß die Maximalwerte ganzzahlig sein müssen, wird am Ende des Intervalls  $T$  ein Ausgleichsintervall der Dauer  $T_B$  benötigt. Diese Dauer berechnet sich zu

$$T_B = T - \frac{X_A F_A + X_P F_P}{\lambda_p} \quad (5.26)$$

und kann bei geeigneter Wahl der Faktoren  $F_A$  und  $F_P$  auf null reduziert werden, wodurch die Implementierung des Verfahrens vereinfacht wird.

- Der Zustand der Quelle wird durch Beobachtung der Zeitspanne ermittelt, die zwischen der neuen Zellankunft und der vorhergehenden verstrichen ist. Ist diese kleiner oder gleich dem Dekrementierintervall in der aktiven Phase, dann wird angenommen, daß sich die Quelle in einer Aktivitätsphase befindet<sup>7</sup>.

<sup>7</sup>Durch die Einführung mehrerer Stufen für die Dekrementierintervalle und entsprechender Schranken könnte das Verfahren für Quellen mit variabler Zellrate erweitert werden

Für den Sonderfall  $F_A = F_P = 1$ , bei dem das Dekrementierintervall unabhängig vom Zustand der Quelle immer gleich dem mittleren Dekrementierintervall  $D$  ist, ergibt sich das normale LB-Verfahren. Das AC-Verfahren erfordert im Vergleich zum LB-Verfahren zwar einen erhöhten Implementierungsaufwand, der Unterschied ist jedoch nicht so gravierend daß er den Einsatz des Verfahrens unmöglich macht.

Bild 5.12 zeigt den Einfluß der Wahl von  $X_P$  auf die Verlustwahrscheinlichkeit für verschiedene Intervallauern  $T$ , wobei auf der Abszisse jeweils der Anteil der längeren Dekrementierintervalle an der Gesamtzahl während eines Intervalls ( $X = X_A + X_P$ ) angegeben ist. Als Quellmodell wurde die sporadische Quelle mit den auch im vorigen Abschnitt verwendeten Parametern für paketierte Sprache angenommen. Das Verfahren wurde mit  $C = 1$  auf eine exakte Überwachung des Mittelwerts dimensioniert. Für den Faktor  $F_A$  wurde der minimale Wert gewählt, der sich aus  $\lambda_p \cdot \Delta$  ergibt.  $F_P$  wurde jeweils so gewählt, daß sich die entsprechenden Anteile für die langsamere Dekrementierung ergaben, ohne daß ein Ausgleichsintervall nötig war. Als Zählergrenze wurde  $N = 50$  verwendet. Zum Vergleich wurde der Wert für die Verlustwahrscheinlichkeit des LB-Verfahrens eingezeichnet, der über den gesamten Bereich konstant ist.

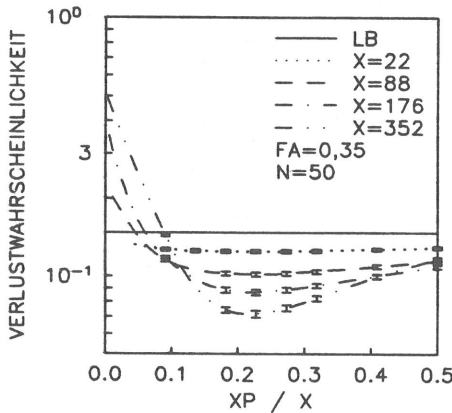


Bild 5.12: Einfluß des Faktors  $X_P$  auf die Verlustwahrscheinlichkeit für das AC-Verfahren

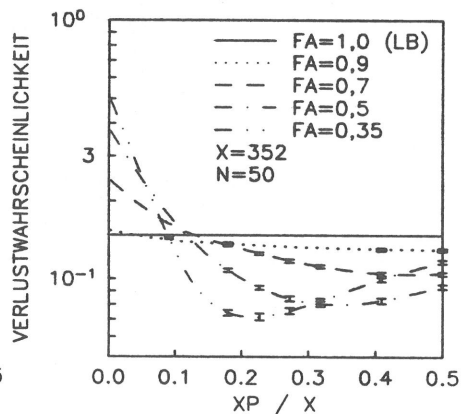


Bild 5.13: Einfluß des Faktors  $F_A$  auf die Verlustwahrscheinlichkeit für das AC-Verfahren

Wie erwartet kann mit den AC-Verfahren eine Verlustwahrscheinlichkeit erreicht werden, die kleiner ist als diejenige des LB-Verfahrens. Die erreichbare Differenz nimmt mit einer Verlängerung der Intervalldauer  $T$  ( $T = X/\lambda_p$ ) zu, weil der Mittelwertausgleich über einen längeren Zeitraum erfolgen kann. Gleichzeitig wird der Bereich für eine günstige Wahl des Verhältnisses von  $X_A$  zu  $X_P$  kleiner und es zeichnet sich deutlich ein Optimum ab. Wird



der Anteil der längeren Dekrementierintervalle zu klein gewählt, kann sich sogar ein schlechteres Verhalten als beim LB-Verfahren ergeben. Dies ist dadurch bedingt, daß bei einer solchen Dimensionierung diese Intervalle sehr lang sein müssen und während dieser Intervalle entsprechend viele Zellen verworfen werden.

Bild 5.13 zeigt den Einfluß des Faktors  $F_A$  auf die Verlustwahrscheinlichkeit für ein Intervall der Länge  $T = 16,03s$  ( $X = 352$ ) und eine Zählergrenze von  $N = 50$  bei gleichen Quellparametern. Es ist zu erkennen, daß sich die größte Verbesserung im Vergleich zum LB-Verfahren mit einer schnellstmöglichen Dekrementierung während der aktiven Phase der Quelle erreicht werden kann.

### 5.5.2 Leaky Bucket mit variabler Zählergrenze

Eine andere Möglichkeit, das Verhalten des LB-Verfahrens zu verbessern, besteht darin, daß die Zählergrenze variabel gemacht wird. Dies kann z.B. in der Form geschehen, daß Zeitfenster wie beim JW-Verfahren definiert werden, in denen die Anzahl akzeptierter Zellen erfaßt wird. Am Ende eines solchen Fensters wird dann nach der in Abschnitt 5.2.5 für das EWMA-Verfahren beschriebenen Methode die Zählergrenze für das nächste Intervall bestimmt. Die Intervallbreite wurde in den folgenden Untersuchungen so gewählt, daß sie dem mittleren Ankunftsabstand der Quelle multipliziert mit der mittleren Zählergrenze entspricht<sup>8</sup>. Ansonsten arbeitet der so definierte LB-VL-Mechanismus (Leaky Bucket - Variable Limit) gleich wie der normale LB-Mechanismus.

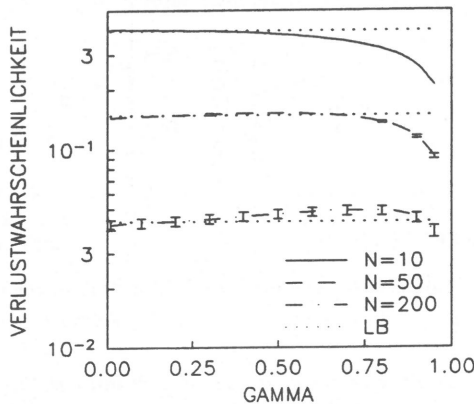


Bild 5.14: Einfluß des Faktors  $\gamma$  beim LB-VL-Verfahren

<sup>8</sup>Es wäre auch möglich, ähnlich wie bei der Überdimensionierung des EWMA-Verfahrens hier einen zusätzlichen Proportionalitätsfaktor einzuführen

Die Auswirkung der variablen Zählergrenze ist in Bild 5.14 dargestellt. Es wurden wieder die Parameter für die paketierte Sprachübertragung und ein Faktor  $C = 1$  verwendet. Variiert wurde der Faktor  $\gamma$  für die Berechnung der aktuellen Zählergrenze, wobei verschiedene mittlere Zählergrenzen angenommen wurden. Es zeigt sich, daß sich für große Werte von  $\gamma$  bei allen mittleren Zählergrenzen ein günstigeres Verhalten als bei dem zum Vergleich mit angegebenen LB-Verfahren ergibt. Für zunehmende mittlere Zählergrenzen nimmt der Bereich für eine günstige Wahl von  $\gamma$  ab und es ergibt sich ein Bereich von  $\gamma$ , in dem das Verfahren sogar mehr Verluste verursacht als das LB-Verfahren.

Der zusätzliche Implementierungsaufwand hält sich auch bei diesem Verfahren in vertretbaren Grenzen.

### 5.5.3 Vergleich mit dem LB-Verfahren

Zur Beurteilung der vorgeschlagenen Verfahren werden sie in diesem Abschnitt mit dem LB-Verfahren verglichen. Bild 5.15 zeigt die Abhängigkeit der Verlustwahrscheinlichkeit von der Zählergrenze  $N$  für eine Dimensionierung mit  $C = 1$ , d.h. für eine exakte Überwachung der mittleren Zellrate.

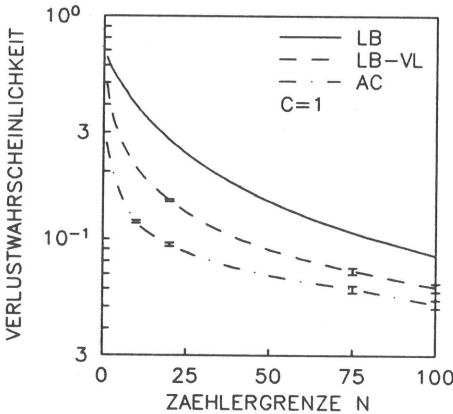


Bild 5.15: Einfluß der Zählergrenze bei einem Faktor  $C = 1$

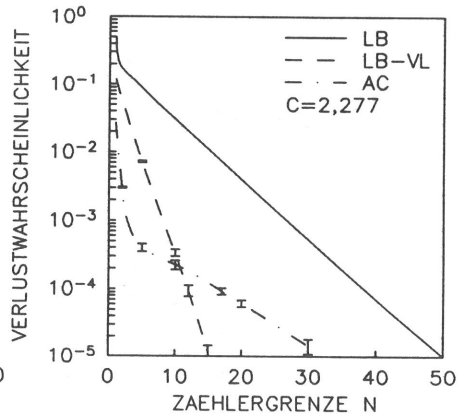


Bild 5.16: Einfluß der Zählergrenze bei einem Faktor  $C = 2,277$

Als Verkehrsquelle wurde wieder das sporadische Modell mit den Parametern für paketierte Sprache verwendet. Der AC-Mechanismus wurde bei einer Intervalldauer  $T = 16,03s$  ( $X = 352$ ) auf die günstigste Verlustwahrscheinlichkeit dimensioniert ( $F_A = 0,35$ ), wobei sich ein Maximalwert  $X_P = 80$  für die Anzahl langsamer Dekrementierungen pro Intervall ergab. Für den LB-VL-Mechanismus wurde ein Faktor  $\gamma = 0,95$  gewählt. Beide Verfahren zeigen ein

günstigeres Verlustverhalten als das normale LB-Verfahren, was einer besseren Transparenz für gutartige Quellen entspricht. Das gleiche gilt bei einer Überdimensionierung der Verfahren mit einem Faktor  $C = 2,277$  (siehe Bild 5.16). Der Unterschied zum LB-Verfahren ist hier größer, wobei der AC-Mechanismus besonders bei kleinen Zählergrenzen ein sehr günstiges Verhalten aufweist.

Die Bilder 5.17 und 5.18 zeigen den Verlauf der Verlustwahrscheinlichkeit, wenn die mittlere Zellrate durch eine Verlängerung der Aktivitätsphasen bzw. durch eine Verkürzung der Pausenphasen erhöht wird. Alle Verfahren wurden dabei für  $C = 2,277$  auf eine Verlustwahrscheinlichkeit von ca.  $8 \cdot 10^{-4}$  bei Einhaltung der mittleren Zellrate dimensioniert. Dabei ergab sich für das LB-VL-Verfahren eine Zählergrenze von  $N = 12$ , beim AC-Verfahren eine Zählergrenze von  $N = 17$  und beim LB-Verfahren eine Zählergrenze von  $N = 39$  (vergleiche Bild 5.16). Bei der Erhöhung der Anzahl von Zellen pro Aktivitätsphase (Bild 5.17) zeigt nur das LB-VL-Verfahren ein günstigeres Verhalten als der LB-Mechanismus, während bei einer Verkürzung der Pausenphasen beide Verfahren deutliche Vorteile aufweisen.

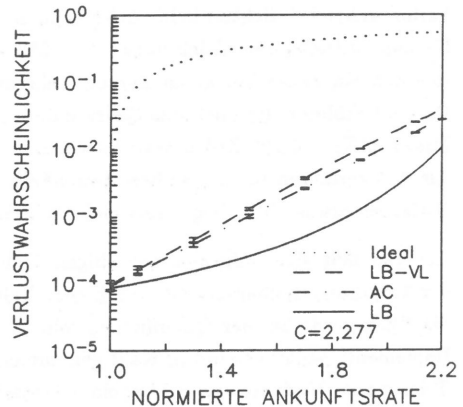
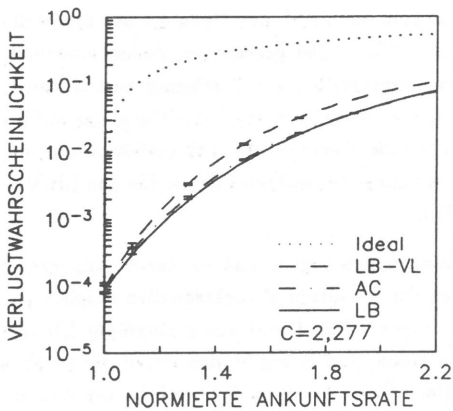


Bild 5.17: Verlustwahrscheinlichkeit bei einer Verlängerung der Aktivitätsphase

Bild 5.18: Verlustwahrscheinlichkeit bei einer Verkürzung der Pausenphase

Das dynamische Verhalten der Verfahren ist, wie in Bild 5.19 dargestellt, nicht entscheidend schlechter als das des LB-Verfahrens. Bei diesem können sich nach Gleichung 5.20 bereits ab der Zelle mit der Ordnungsnummer 190 Verluste ergeben.

Beim LB-VL-Verfahren treten, bedingt durch die hohe Zählergrenze von  $N = 240$  zu Beginn der Verbindung, erst ab der dreihundertsten gesendeten Zelle merkbare Verluste auf. Beim AC-Verfahren beginnen die Verluste ebenfalls erst in diesem Bereich, was durch die 272 möglichen Dekrementierungen mit der maximalen Zellrate ( $F_A = 0,35$ ) und das dadurch bedingte verzögerte Ansteigen des Zählerstandes erklärt werden kann. Die Bereiche fallender

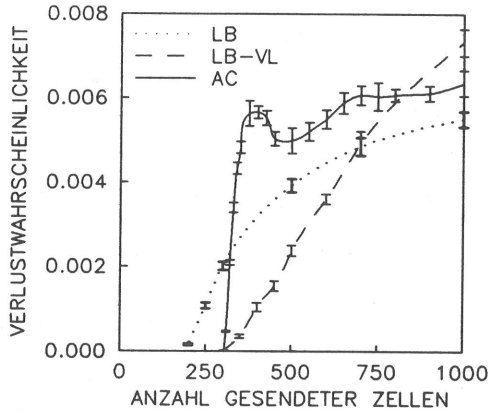


Bild 5.19: Dynamisches Verhalten der Mechanismen

Verlustwahrscheinlichkeit beim AC-Mechanismus kommen durch den Neustart des Intervalls für den Mittelwertausgleich nach  $X = 352$  bzw.  $2X = 704$  gesendeten Zellen zustande, wodurch ein neuer Vorrat an kurzen Dekrementierintervallen zur Verfügung gestellt wird. Da die Erhöhung der mittleren Zellrate durch eine Verlängerung der Aktivitätsphase auf im Mittel  $E[X] = 45,25$  Zellen bewirkt wurde, streben die Werte für das LB-Verfahren und für das AC-Verfahren auf den selben stationären Grenzwert zu, während dieser für das LB-VL-Verfahren etwas höher liegt (vergleiche Bild 5.17).

Generell läßt sich aufgrund der obigen Untersuchungen sagen, daß es durch Anpassung der Verfahren an charakteristische Eigenschaften der jeweiligen Verkehrsquellen möglich ist, die Effektivität bei der Quellflußkontrolle zu steigern. Dabei muß ein geringfügig höherer Implementierungsaufwand in Kauf genommen werden, und es ergibt sich durch die erhöhte Toleranz der Verfahren gegenüber einer kurzzeitigen Überschreitung der vereinbarten Zellrate ein etwas schlechteres dynamisches Verhalten.

## Kapitel 6

# Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde das Konzept eines diensteintegrierenden Breitbandnetzes (B-ISDN) auf der Basis des asynchronen Transfer-Modus (ATM) dargestellt, und es wurden die verschiedenen Funktionen für die Kontrolle und Steuerung der auftretenden Verkehrsflüsse angesprochen. Diese Funktionen umfassen sowohl diejenigen, die beim Verbindungsaufbau auszuführen sind, wie z.B. Verbindungsannahmекontrolle und Wegesuche, als auch Funktionen wie die Quellflußkontrolle, die während der gesamten Verbindungsdauer aktiv sein müssen. Während die Fragestellungen im Zusammenhang mit den zuerst genannten Funktionen im Prinzip aus den klassischen durchschalte- bzw. paketvermittelnden Netzen bekannt sind, ist die Quellflußkontrolle eine neuartige Netzfunktion, die durch das asynchrone Multiplexprinzip bei gleichzeitigem Fehlen einer Flußkontrolle zwischen Netz und Benutzer nötig wird.

Da die entsprechenden Netze und Netzkomponenten zur Zeit noch nicht für notwendige Messungen zur Verfügung stehen, ist die verkehrstheoretische Modellierung und Analyse von besonderer Bedeutung für die Beurteilung der Leistungsfähigkeit der Systemkomponenten und Mechanismen und für deren Dimensionierung. Dabei spielt die realistische Modellierung der Verkehrsströme, wie sie von den verschiedenartigen Diensten erzeugt werden, eine entscheidende Rolle. Deshalb wurden die verschiedenen Aspekte bei der Modellierung solcher Verkehrsströme erläutert, und es wurden die bei der Untersuchung von ATM-Netzen eingesetzten Quellmodelle ausführlich dargestellt. Dabei wurden Erweiterungen der bekannten Modelle eingeführt und die Anwendbarkeit der einzelnen Modelle auf die verschiedenen Problemstellungen diskutiert.

Neben der allgemein üblichen Beschreibung der Verkehrsströme mit Hilfe der Verteilung für die Zwischenankunftsabstände wurde auch die Beschreibung durch die Verteilung des Zählprozesses und daraus abgeleitete Kenngrößen, wie den „Index of Dispersion for Counts“ (IDC), eingeführt. Diese Beschreibung stellt eine Grundlage für die analytische Untersuchung einiger Quellflußkontrollmechanismen dar.

Im weiteren Verlauf der Arbeit wurden die zeitkontinuierlichen und zeitdiskreten Warteschlangenmodelle mit deterministischer Bedienzeit und geeignete Lösungsansätze für diese Modelle zusammengestellt. Solche Systemmodelle ergeben sich bei fast allen ATM-Untersuchungen, weil die übermittelten Dateneinheiten, nämlich die ATM-Zellen, eine einheitliche Länge aufweisen. Besonderes Augenmerk wurde auf Warteverlustsysteme mit einer zweiphasigen sporadischen Quelle als Ankunftsprozeß gelegt, weil dieses Quellenmodell eine ausreichend realistische Modellierung verschiedener Dienste erlaubt und das System gleichzeitig analytisch behandelbar bleibt.

Bei der Darstellung der zeitdiskreten Systemmodelle wurden das Phänomen der Gleichzeitigkeit von Ankunft und Bedienung systematisch behandelt und die Lösungsansätze entsprechend erweitert, um den Einfluß der verschiedenen Strategien für die Auflösung der Gleichzeitigkeit auf das Systemverhalten quantitativ beurteilen zu können. Anhand einiger Beispiele wurde dann die Anwendbarkeit der einzelnen Systemmodelle für die nachfolgenden Untersuchungen bewertet.

Die Quellen- und Systemmodelle und die entsprechenden Lösungsverfahren wurden im Kapitel 5 verwendet, um eine vergleichende Untersuchung verschiedener, bekannter Quellflußkontrollmechanismen durchzuführen, und um die zwei in Abschnitt 5.5 neu vorgeschlagenen Verfahren bezüglich ihrer Leistungsfähigkeit zu beurteilen.

Neben den im Kapitel 5 enthaltenen quantitativen Aussagen haben die Untersuchungen folgende allgemeinen Ergebnisse erbracht:

Es wurde gezeigt, daß bei der Quellflußkontrolle – und insbesondere bei der Überwachung der mittleren Zellrate einer Verbindung – einige prinzipielle Probleme auftreten. Eines dieser Probleme besteht darin, daß der Mittelwert der Zellrate, der von einer statistischen Quelle erst über einen bestimmten Zeitraum hinweg einigermaßen eingehalten wird, auf der Basis relativ kurzer Stichproben mit hoher Genauigkeit vorhergesagt werden soll. Da dies nur eingeschränkt möglich ist, müssen die Verfahren entweder überdimensioniert werden oder die Stichprobenintervalle müssen entsprechend lang gewählt werden, um eine akzeptabel geringe Verlustwahrscheinlichkeit für Quellen zu erhalten, welche die vereinbarten Parameter einhalten. Eine Überdimensionierung verringert dabei die Empfindlichkeit gegenüber einer wirklichen Verletzung der Parameter des Verkehrsvertrags, eine Verlängerung der Stichprobenintervalle, d.h. eine Erhöhung der Zählergrenze, verlängert die dynamische Reaktionszeit der Mechanismen. Beide Methoden vergrößern den Unterschied zwischen der im ungünstigsten Fall möglichen und der vereinbarten Verkehrscharakteristik.

In Anbetracht dieser grundsätzlichen Beschränkung des Quellflußkontrollkonzepts, die durch die statistischen Eigenschaften der Verkehrsquelle bestimmt wird, lassen komplexere Mechanismen nur eine begrenzte Steigerung der Effektivität der Quellflußkontrolle erwarten. Dabei

versprechen Verfahren, die wie die in Abschnitt 5.5 vorgeschlagenen speziell auf die Eigenheiten eines bestimmten Diensttyps abgestimmt sind, die besten Ergebnisse mit dem geringsten Implementierungsmehraufwand.

Ein weiteres Problem des Quellflußkontrollkonzepts liegt darin, daß bei der Festlegung des Verkehrsvertrags, d.h. beim Aufbau der Verbindungen, die Verkehrsparameter für die meisten Dienste nicht mit genügender Genauigkeit vorhergesagt werden können. Die Angabe der maximalen Zellrate mit der erforderlichen Genauigkeit ist in vielen Fällen möglich, während dies z.B. schon bei der mittleren Zellrate für viele Dienste unmöglich ist. Dadurch wird eine zusätzliche Überdimensionierung der Mechanismen nötig, was wiederum die Effektivität der Quellflußkontrolle weiter verringert.

Die Dimensionierung der Mechanismen erfordert demnach einen Kompromiß zwischen

- der Wahrscheinlichkeit für ein unberechtigtes Verwerfen von Zellen einer Verbindung, die sich an die vereinbarten Parameter hält,
- der Sensitivität bezüglich wirklicher Parameterverletzungen,
- der dynamischen Reaktionszeit,
- dem im ungünstigsten Fall in das Netz durchgelassenen Verkehr und
- dem Implementierungsaufwand,

wobei eine Erhöhung des Implementierungsaufwands wegen der prinzipiellen Beschränkungen des Konzepts nur eine begrenzte Verbesserung erbringt.

Ein einzelner Quellflußkontrollmechanismus ist in der Regel nicht ausreichend, um verschiedene Parameter des Verkehrsvertrages, z.B. die mittlere und die maximale Zellrate, gleichzeitig zu überwachen. Der Grund dafür liegt in den stark unterschiedlichen Zeitebenen, die dabei relevant sind. Es ist deshalb sinnvoll, mehrere einfache Quellflußkontrollmechanismen zu kombinieren, von denen jeder für die Überwachung eines bestimmten Parameters dimensioniert ist.

Die optimale Dimensionierung und die Effektivität der Verfahren hängt stark von der Charakteristik des überwachten Dienstes ab. Dienste mit einer gleichmäßigen, stromförmigen Charakteristik lassen sich effektiver überwachen als solche mit einer sporadischen, sehr büschelförmigen Charakteristik. Bei den letztgenannten Diensten ist die Effektivität besonders gering, wenn die Aktivitäts- und Pausenphasen sehr lang sind, weil in diesem Fall ein sehr langes Zeitintervall für eine genügend genaue Schätzung der mittleren Zellrate erforderlich ist.

Der im schlimmsten Fall von einem Quellflußkontrollmechanismus in das Netz durchgelassene Verkehr unterscheidet sich durch die Überdimensionierung und die erforderlichen hohen Zählergrenzen signifikant von demjenigen, der von einer gutartigen, statistischen Quelle

erzeugt wird und wirkt sich im Netz entsprechend ungünstiger aus. Diese Tatsache muß bei der Entscheidung über die Annahme einer neuen Verbindung und der damit verbundenen Ressourcenzuteilung in den einzelnen Netzelementen berücksichtigt werden, damit die erforderliche Dienstgüte garantiert werden kann.

Um die Effektivität der Quellflußkontrolle, die einen direkten Einfluß auf die Effektivität der Ressourcenzuteilung und -ausnutzung im Netz und damit auf die Wirtschaftlichkeit des Netzes hat, zu erhöhen, sind in einem zukünftigen ATM-Netz einige zusätzliche, dienstspezifische Maßnahmen wünschenswert. Für Dienste mit einer Ein/Aus-Charakteristik, einer hohen Büschelförmigkeit und einer langen Dauer der Aktivitäts- und Pausenphasen, wie z.B. Standbildübertragung oder interaktive Datenkommunikation, könnte der Einsatz einer dynamischen Ressourcenzuteilung Vorteile bringen. In diesem Fall könnten dann, ähnlich wie beim „Burst Switching“ [6, 7], die Verbindungen aufgebaut und z.B. die Adressen (VCI/VPI) zugeteilt werden, ohne daß jedoch der Verbindung in den Netzkomponenten ein bestimmter Anteil der Kapazität zugeordnet wird. Am Beginn einer Aktivitätsphase könnten dann durch geeignete Signalisierzellen für die Dauer der Aktivitätsphase die entsprechenden Netzressourcen reserviert werden. Die dadurch entstehende, zusätzliche Verzögerung in der Übermittlung der Nutzinformationen würde dabei in solchen Bereichen liegen, daß sie die Dienstgüte der nicht extrem echtzeitkritischen Dienste nicht signifikant verringert. Die Auslastung der Netzkomponenten könnte dadurch spürbar gesteigert werden, was den zusätzlichen Protokollaufwand rechtfertigen würde.

Bei der hochqualitativen Bewegtbildübertragung mit zeitlich variabler Bitrate wird die momentane Bitrate sowohl durch codierungsbedingte Effekte innerhalb von Bildern oder zwischen aufeinanderfolgenden Bildern, als auch durch Veränderungen im Bildausschnitt oder Szenenwechsel beeinflusst. Durch diese großen Dynamikunterschiede bei den Veränderungen ist eine effektive Quellflußkontrolle sehr schwierig. Die Effektivität der Quellflußkontrolle kann für solche Dienste, die außerdem noch extreme Echtzeitanforderungen stellen, durch Codierer verbessert werden, die den Zellstrom so begrenzen, daß er besser für die Quellflußkontrolle und für das statistische Multiplexen geeignet ist. Dadurch ergibt sich zwar eine Bildqualität, die nicht mehr konstant ist, die Qualitätsschwankungen können jedoch so gering gehalten werden, daß sie subjektiv nicht wahrnehmbar sind [127].

Zusammenfassend läßt sich sagen, daß die gegenüber kanalorientierten Konzepten erhöhte Flexibilität des ATM-Konzepts mit einem erhöhten Aufwand für die Signalisierung und die Verkehrssteuerung verbunden ist. Es ist weiterhin fraglich, ob sich durch das asynchrone Multiplexkonzept eine entscheidende Steigerung der Nutzauslastung der Netzressourcen

erreichen läßt. Trotzdem ist das ATM-Konzept aufgrund seiner inherenten Flexibilität das am besten für ein zukünftiges integriertes Breitbandkommunikationsnetz geeignete Konzept, weil es eine Integration der heutigen und der zukünftigen Dienste sowohl bei der Übertragung



als auch bei der Vermittlung erlaubt. Mit diesem Netzkonzept wird es Netzbetreibern und Privatanbietern in Zukunft möglich sein, neben den Massendiensten auch hochqualitative Dienste mit spezifischen Anforderungen bezüglich der erforderlichen Bandbreite selbst für kleinere Benutzergruppen kostengünstig anbieten zu können. Außerdem bietet dieses Konzept ideale Voraussetzungen, um eine Weiterentwicklung bereits bestehender Dienste, z.B. durch Ausnutzung neuartiger Codiervverfahren, ohne aufwendige Änderungen im Netz selbst zu unterstützen.

## Literaturverzeichnis

- [1] M. H. Ackroyd: Computing the Waiting Time Distribution for the G/G/1 Queue by Signal Processing Methods. *IEEE Transactions on Communications*, Vol. 28, No. 1, S. 52-58, Januar 1980.
- [2] G. B. Adams III, D. P. Agrawal, H. J. Siegel: A Survey and Comparison of Fault-Tolerant Multistage Interconnection Networks. *IEEE Computer*, Vol. 20, No. 6, S. 14-27, Juni 1987.
- [3] J. L. Adams: The Virtual Path Identifier and its Applications for Routing and Priority of Connectionless and Connection-Oriented Services. *International Journal of Digital and Analog Cabled Systems*, Vol. 1, No. 4, S. 257-262, 1988.
- [4] J. L. Adams: Signalling and ATM in the Age of Mobility. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 11.1, Paris, Oktober 1989.
- [5] H. Ahmadi, W. E. Denzel: A Survey of Modern High-Performance Switching Techniques. *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 7, S. 1091-1103, September 1989.
- [6] S. R. Amstutz: Burst Switching - An Introduction. *IEEE Communications Magazine*, Vol. 21, S. 36-42, November 1983.
- [7] S. R. Amstutz: Burst Switching - An Update. *IEEE Communications Magazine*, Vol. 27, No. 9, S. 50-57, September 1989.
- [8] D. Anick, D. Mitra, M. M. Sondhi: Stochastic Theory of a Data-Handling System with Multiple Sources. *The Bell System Technical Journal*, Vol. 61, No. 8, S. 1871-1894, Oktober 1982.
- [9] G. J. Anido, A. W. Seeto: Multipath Interconnection: A Technique for Reducing Congestion Within Fast Packet Switching Fabrics. *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 9, S. 1480-1488, Dezember 1988.
- [10] O. Baireuther: Bildfernsprechversuchsnetz - 140-Mbit/s-Verbindungsnetz der Bigfon-Projekte. *ntz, Band 38, Heft 4*, S. 214-218, 1985.

- [11] H. Bakker, G. H. Domann, H. Herbrig, M. Hilbig, H. J. Matt: BERKOM: The first Broadband ISDN Exchange and Transmission System in Berlin. *International Journal of Digital and Analog Cabled Systems*, Vol. 2, No. 1, S. 13-28, 1989.
- [12] R. Ballart, Y.-C. Ching: SONET: Now it's the Standard Optical Network. *IEEE Communications Magazine*, Vol. 29, No. 3, S. 8-15, März 1989.
- [13] V. Benes: *Mathematical Theory of Connecting Networks*. Academic Press, New York, 1965.
- [14] B. Bocek, F. Kirtz, K. Koeck, F. Mikley, M. Schäfer: Breitbandvermittlungen für das Breitbandvorläufernetz der Deutschen Bundespost. *ANT Nachrichtentechnische Berichte*, Heft 4, S. 66-73, 1987.
- [15] C. Blondia: The N/G/1 Finite Capacity Queue. *Commun. Statist.-Stochastic Models* 2, Vol. 5, 1988.
- [16] C. Blondia: A Discrete-Time Markovian Arrival Process. *RACE R1022 Dokument PRLB\_123\_0015\_CC\_CD*, August 1989.
- [17] C. Blondia, T. H. Theimer: A Discrete-Time Model for ATM Traffic. *RACE R1022 Dokument PRLB\_123\_0018\_CC\_CD / UST\_123\_0022\_CC\_CD*, Oktober 1989.
- [18] A. Bricca, R. Carli: Impact of Multiconnection Services on Layer 3 Signalling Protocol. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 11.3, Paris, Oktober 1989.
- [19] U. Briem: *Analyse von Warteschlangenmodellen mit korreliertem Eingangsprozeß*. Diplomarbeit Nr. 997 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1990.
- [20] H. Bruneel: Queueing Behaviour of Statistical Multiplexers with Correlated Inputs. *IEEE Transactions on Communications*, Vol. 36, No. 12, S. 1339-1341, Dezember 1988.
- [21] R. G. Bubenik, J. S. Turner: Performance of a Broadcast Packet Switch. *IEEE Transactions on Communications*, Vol. 37, No. 1, S. 60-69, Januar 1989.
- [22] W. R. Byrne, A. Papanicolaou, M. N. Ransom: World-Wide Standardization of Broadband ISDN. *International Journal of Digital and Analog Cabled Systems*, Vol. 1, No. 4, S. 181-192, 1988.
- [23] CCITT: Recommendations G.70x: „Synchronous Digital Hierarchy“. *Blue Book*, Vol. III.4, Genf, 1989.
- [24] CCITT: Recommendations I.xxx: „Integrated Services Digital Network (ISDN)“. *Blue Book*, Vol. III.7-III.9, Genf, 1989.
- [25] CCITT: Recommendation I.113: „Vocabulary of Terms for Broadband Aspects of ISDN“. *Blue Book*, Vol. III.7, Genf, 1989.

- [26] CCITT: Recommendation I.121: „Broadband Aspects of ISDN“. *Blue Book, Vol. III.7*, Genf, 1989.
- [27] CCITT: Recommendations I.2xx: „Service Capabilities“. *Blue Book, Vol. III.7*, Genf, 1989.
- [28] CCITT: Recommendation I.311: „B-ISDN General Network Aspects“. *Temporary Documents of the Study Group XVIII Meeting in Geneva*, Juni 1989.
- [29] CCITT: Recommendation I.320: „ISDN Protocol Reference Model“. *Blue Book, Vol. III.8*, Genf, 1989.
- [30] CCITT: Recommendation I.324: „ISDN Network Architecture“. *Blue Book, Vol. III.8*, Genf, 1989.
- [31] CCITT: Recommendation I.411: „ISDN User-Network Interfaces – Reference Configurations“. *Blue Book, Vol. III.8*, Genf, 1989.
- [32] CCITT: Recommendation I.440: „ISDN User-Network Interface Data Link Layer – General Aspects“. *Blue Book, Vol. III.8*, Genf, 1989. (identisch mit Q.920)
- [33] CCITT: Recommendation I.441: „ISDN User-Network Interface Data Link Specification“. *Blue Book, Vol. III.8*, Genf, 1989. (identisch mit Q.921)
- [34] CCITT: Recommendation I.451: „ISDN User-Network Interface Layer 3 Specification“. *Blue Book, Vol. III.8*, Genf, 1989. (identisch mit Q.931)
- [35] CCITT: Recommendation X.25: „Interface Between Data Terminal Equipment (DTE) and Data Circuit-Terminating Equipment (DCE) for Terminals Operating in the Packet Mode and Connected to Public Data Networks by Dedicated Circuit“. *Blue Book, Vol. VIII.2*, Genf, 1989.
- [36] CCITT: Temporary Document No. 74 of the Study Group XVIII Meeting in Seoul, Januar 1989.
- [37] CCITT: Temporary Documents of the Study Group XVIII Meeting in Geneva, Juni 1989.
- [38] CCITT: Temporary Documents of the Study Group XVIII Meeting in Geneva, Januar 1990.
- [39] A. Q. Christensen: *Untersuchung des N/G/1-s Systems mit den Methoden der Matrixgeometrie*. Diplomarbeit Nr. 935 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1989.
- [40] E. Cinlar: *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, N.J., 1975.
- [41] C. Clos: A Study of Non-blocking Switching Networks. *Bell System Technical Journal, Vol. 32, No. 2*, S. 406-424, März 1953.

- [42] J. Cosmas: The Characterization of Video Codecs as General Modulated Deterministic Processes. *RACE R1022 Dokument QMC\_123.0022-CD-CC*, September 1989.
- [43] J.-P. Coudreuse, M. Servel: PRELUDE: An Asynchronous Time-Division Switched Network. *Proceedings of the IEEE ICC-87*, S. 769-773, Juni 1987.
- [44] J.-P. Coudreuse: ATM: A Contribution to the Debate on Broadband ISDN. *International Journal of Digital and Analog Cabled Systems*, Vol. 1, No. 4, S. 213-221, 1988.
- [45] D. R. Cox, P. A. W. Lewis: *The Statistical Analysis of Series of Events*. Methuen & CO LTD, London, 1966.
- [46] J. N. Daigle, J. D. Langford: Models for Analysis of Packet Voice Communications Systems. *IEEE Journal on Selected Areas in Communications*, Vol. 4, No. 6, S. 847-855, September 1986.
- [47] F. Denissen, E. Desmet, G. H. Petit: The Policing Function in an ATM Network. *Proceedings of the 1990 International Zurich Seminar on Digital Communications*, S. 131-144, Zürich, März 1990.
- [48] M. Devault, J.-Y. Cochennec, M. Servel: The "Prelude" ATD Experiment: Assessments and Future Prospects. *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 9, S. 1528-1537, Dezember 1988.
- [49] G. H. Dobrowski, M. Kerner, D. R. Spears, D. S. Wilson: Evolving the Network Toward B-ISDN. *Proceedings of the XIII International Switching Symposium (ISS)*, Vol. I, S. 1-7, Stockholm, Mai 1990.
- [50] G. Domann: BERKOM-Testnetz und BISDN/IBFN-Systemkonzept. *Elektrisches Nachrichtenwesen*, Band 62, Nr. 3/4, S. 325-332, 1988.
- [51] F. Van den Dool: Policing and Load Control and related Functions. *RACE R1022 Dokument DNL\_311&321\_008-CD-CC*, Juni 1988.
- [52] P. Douglas, M. Vetterli: Statistical Analysis of the Output Rates of Two Variable Bit-rate Video Coders. *Proceedings of the Third International Workshop on Packet Video*, Morristown, N.J., März 1990.
- [53] Z. Dziong, J. Choquette, K.-Q. Liao, L. Mason: Admission Control and Routing in ATM Networks. *Proceedings of the ITC Specialist Seminar on „Traffic Theories for New Telecommunications Services“*, Paper 15.3, Adelaide, September 1989.
- [54] A. E. Eckberg, Jr.: The Single Server Queue with Periodic Arrival Process and Deterministic Service Times. *IEEE Transactions on Communications*, Vol. 27, No. 3, S. 556-562, März 1979.
- [55] A. E. Eckberg, D. T. Luan, D. M. Lucantoni: Bandwidth Management: A Congestion Control Strategy for Broadband Packet Networks - Characterizing the Throughput-

- Burstiness Filter. *Proceedings of the ITC Specialist Seminar on „Traffic Theories for New Telecommunications Services“*, Paper 4.4, Adelaide, September 1989.
- [56] P. L'Ecuyer: Efficient and Portable Random Number Generators. *Communications of the ACM*, Vol. 31, No. 6, S. 742-749 und S. 774, 1988.
- [57] J. Filipiak: A Structural Model of Traffic Management and Control in Broadband ISDNs. *IEEE Communications Magazine*, Vol. 27, No. 5, S. 25-31, Mai 1989.
- [58] J. Filipiak: Structured System Analysis Methodology for Design of an ATM Network Architecture. *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 8, S. 1263-1273, Oktober 1989.
- [59] G. Gallassi, G. Rigolio, P. Vaccari, L. Verri: Resource Allocation in ATM Networks. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 22.6, Paris, Oktober 1989.
- [60] G. Gallassi, G. Rigolio, L. Fratta: ATM: Bandwidth Assignment and Bandwidth Enforcement Policies. *Proceedings of the IEEE Globecom '89*, Paper 49.6, Dallas, November 1989.
- [61] E. Garetti, R. Melen, A. Tonietti: Efficiency of the Leaky-Bucket Mechanism in the Framework of ATM Resource Allocation. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 21.2, Paris, Oktober 1989.
- [62] L. R. Goke, G. J. Lipovski: Banyan Networks for Partitioning Multiprocessor Systems. *Proceedings of the First Annual Symposium on Computer Architecture*, S. 21-28, 1973.
- [63] P. Gonet, P. Adam, J. P. Coudreuse: Asynchronous Time-Division Switching: The Way to Flexible Broadband Communication Networks. *Proceedings of the 1986 International Zurich Seminar on Digital Communications*, S. 141-148, Zürich, März 1986.
- [64] A. Gravey, G. Hebuterne: Simultaneity in Discrete-Time Single Server Queues with Bernoulli Inputs. *Interner Bericht, CNET, Lannion*, 1989.
- [65] D. Gross, C. M. Harris: *Fundamentals of Queueing Theory*. John Wiley & Sons, New York, 1974.
- [66] R. Grünenfelder: Cell Traffic Characterization in a B-ISDN according to the Slotted ATD Mode. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 23.10, Paris, Oktober 1989.
- [67] S. Halfin: Batch Delays versus Customer Delays. *Bell Systems Technical Journal*, Vol. 62, No. 7, S. 2011-2015, September 1983.
- [68] E. F. Haselton: New Switching Concept Integrates Voice and Data Bursts. *Proceedings of the IEEE GLOBECOM '84*, S. 126-129, Atlanta, Dezember 1984.
- [69] G. Hebuterne, A. Gravey: A Space-priority Queueing Mechanism for Multiplexing ATM-Channels. *Proceedings of the ITC Specialist Seminar on „Traffic Theories for New Telecommunications Services“*, Paper 7.4, Adelaide, September 1989.

- [70] H. Heffes, D. M. Lucantoni: A Markov Modulated Characterization of Packetized Voice and Data Traffic and Related Statistical Multiplexer Performance. *IEEE Journal on Selected Areas in Communications*, Vol. 4, No. 6, S. 856-868, September 1986.
- [71] W. Held: *Untersuchung von Prioritätsmechanismen für ATM-Netze*. Diplomarbeit Nr. 964 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1989.
- [72] C. F. Hemrick, L. J. Lang: Introduction to Switched Multi-Megabit Data Service, (SMDS), an early Broadband Service. *Proceedings of the XIII International Switching Symposium (ISS)*, Vol. II, S. 1-8, Stockholm, Mai 1990.
- [73] P. Henrici: Fast Fourier Methods in Computational Complex Analysis. *Siam Review* 21, S. 481-527, 1979.
- [74] M. G. Hluchyi, M. J. Karol: Queueing in Space-Division Packet Switching. *Proceedings of the IEEE INFOCOM '88*, S. 334-343, New Orleans, 1988.
- [75] J. M. Holtzman: Characteristics of Superpositions of Traffic Streams. *Proceedings of the 4th ITC Specialist Seminar on ISDN Traffic Issues*, Brüssel, Mai 1986.
- [76] A. Huang, S. Knauer: STARLITE: A Wideband Digital Switch. *Proceedings of the IEEE GLOBECOM '84*, S. 121-125, Atlanta, Dezember 1984.
- [77] S. S. Huang: Modeling and Analysis for Packet Video. *Proceedings of the IEEE GLOBECOM '89*, S. 881-885, Dallas, November 1989.
- [78] M. N. Huber, E. P. Rathgeb, T. H. Theimer: Banyan Networks in an ATM-Environment. *Proceedings of the Ninth International Conference on Computer Communication*, S. 167-174, Tel Aviv, Oktober 1988.
- [79] M. N. Huber: *Ein Netzknotenkonzept für integrierte Durchschalte- und Paketvermittlung*. 49. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift) Universität Stuttgart, Institut für Nachrichtenvermittlung und Datenverarbeitung, 1990.
- [80] J. Y. Hui: Resource Allocation for Broadband Networks. *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 9, S. 1598-1608, Dezember 1988.
- [81] IEEE: Draft International Standard 802.2: „Local Area Networks, Part 2, Logical Link Control“, 1986.
- [82] IEEE: Draft Proposed Standard 802.6/D7: „Distributed Queue Dual Bus (DQDB) Metropolitan Area Network (MAN)“, Mai 1989.
- [83] ISO 7498: „Information Processing Systems: Open Systems Interconnection, Basic Reference Model“, 1984.
- [84] V. B. Iversen, A. L. Kragh, B. F. Nielsen: *The Exact Evaluation of the M/D/1/n Queue*. Institute of Mathematical Statistics and Operations Research, Technical University of Denmark, Lyngby, Research Report No. 10/1988, 1988.

- [85] R. Jain, S. A. Routhier: Packet Trains – Measurements and a New Model for Computer Network Traffic. *IEEE Journal on Selected Areas in Communications*, Vol. 4, No. 6, S. 986–995, September 1986.
- [86] H. Jans: *Verkehrsanalyse von Vermittlungssteuerungen mit taktgesteuerter Ein-/Ausgabe und Prioritäten*. 37. Bericht über verkehrstheoretische Arbeiten (Dissertationschrift) Universität Stuttgart, Institut für Nachrichtenvermittlung und Datenverarbeitung, 1983.
- [87] B. Käs, R. Kleinewillinghöfer-Kopp, O. Osterbo, H. Pettersen: Admission control – Basic Properties and First Results. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 22.1, Paris, Oktober 1989.
- [88] M. J. Karol, M. G. Hluchyi: Using a Packet Switch for Circuit-Switched Traffic: A Queueing System with Periodic Input Traffic. *Proceedings of the ICC '87*, S. 1677–1682, Juni 1987.
- [89] M. J. Karol, M. G. Hluchyi, S. P. Morgan: Input Versus Output Queueing on a Space-Division Packet Switch. *IEEE Transactions on Communications*, Vol. 35, No. 12, S. 1347–1356, Dezember 1987.
- [90] L. Katzschner, R. Scheller: Probability of Loss of Data Traffic with different Bit Rates Hunting one common PCM Channel. *Proceedings of the 8th International Teletraffic Congress (ITC)*, S. 525.1–525.8, Melbourne, 1976.
- [91] J. Keilson: The Ergodic Queue Length Distribution for Queueing Systems with Finite Capacity. *Journal of the Royal Statistical Society, Ser. B*, Vol. 28, S. 190–201, 1966.
- [92] D. G. Kendall: Stochastic Processes Occuring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. *Annals of Mathematical Statistics* 24, S. 338–354, 1953.
- [93] U. Killat: Einige Anmerkungen zu Begriffen und Konzepten der ATM-Technik. *Informatik-Fachberichte 205: Tagungsband zur GI/ITG Fachtagung „Kommunikation in verteilten Systemen“*, S. 479–493, Stuttgart, Februar 1989.
- [94] L. Kleinrock: *Queueing Systems, Vol. 1: Theory*, John Wiley & Sons, New York, 1975.
- [95] L. Kleinrock: *Queueing Systems, Vol. 2: Computer Applications*, John Wiley & Sons, New York, 1976.
- [96] W. Kowalk, R. Lehnert: The 'Policing Function' to Control User Access in ATM Networks - Definition and Implementation. *Proceedings of the ISSLS'88*, Boston, September 1988.
- [97] H. Kröner: Comparative Performance Study of Space Priority Mechanisms for ATM Networks. *Proceedings of the IEEE INFOCOM 90*, S. 1136–1143, San Francisco, Juni 1990.



- [98] H. Kröner, Thomas H. Theimer: Queueing Models for ATM Systems - A Comparison. *Proceedings of the 7th International Teletraffic Congress Seminar*, Morristown, N.J., Oktober 1990.
- [99] C. P. Kruskal, M. Snir: The Performance of Multistage Interconnection Networks for Multiprocessors. *IEEE Transactions on Computers*, Vol. 32, No. 12, S. 1091-1098, Dezember 1983.
- [100] P. J. Kühn: Wartezeitprobleme der Daten- und Nachrichtenverkehrstheorie. *Vorlesung an der Universität Stuttgart*.
- [101] P. J. Kühn, T. Raith, P. Tran-Gia: Methodik der stationären Systemsimulation (in [100]).
- [102] P. J. Kühn: Input Process Model GMDP. *RACE R1022 Dokument UST\_123-0003\_CC\_CD*, Juni 1988.
- [103] J. J. Kulzer, W. A. Montgomery: Statistical Switching Architectures for Future Services. *Proceedings of the XI International Switching Symposium (ISS)*, Paper 43-A-1, Florenz, 1984.
- [104] D. V. Lindley: The Theory of Queues with a Single Server. *Proc. of the Cambridge Philosophical Society*, 48, S. 277-289, 1952.
- [105] J.-R. Louvion, P. Boyer, A. Gravey: A Discrete-Time Single Server Queue with Bernoulli Arrivals and Constant Service Time. *Proceedings of the 12th International Teletraffic Congress (ITC)*, S. 2.4B.2.1-2.4B.2.7, Turin, Juni 1988.
- [106] D. M. Lucantoni, V. Ramaswami: Efficient Algorithms for Solving the Non-Linear Matrix Equations Arising in Phase Type Queues. *Commun. Statist.-Stochastic Models 1*, No. 1, S. 29-51, 1985.
- [107] D. M. Lucantoni, K. S. Meier-Hellstern, M. F. Neuts: A Single Server Queue with Server Vacations and a Class of Non-Renewal Arrival Processes. *Working Paper 88-008*, Systems and Industrial Engineering Department, University of Arizona, 1988.
- [108] E. Lutz: *Mittlere Wartezeit und Wartezeitverteilungsfunktion für Warte- und Warte-verlustsysteme des Typs M/G/1 mit geordneter und zufälliger Abfertigungsreihenfolge*. Diplomarbeit Nr. 431 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1973.
- [109] K. A. Lutz: Considerations on ATM Switching Techniques. *International Journal of Digital and Analog Cabled Systems*, Vol. 1, No. 4, S. 237-243, 1988.
- [110] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson, J. D. Robbins: Performance Models of Statistical Multiplexing in Packet Video Communications. *IEEE Transactions on Communications*, Vol. 36, No. 7, S. 834-844, Juli 1988.

- [111] A. C. McKellar, E. G. Coffman Jr.: Organizing Matrices and Matrix Operations for Paged Memory Systems. *Communications of the ACM*, Vol. 12, No. 3, S. 153-165, 1969.
- [112] R. J. McMillen: A Survey of Interconnection Networks. *Proceedings of the IEEE GLOBECOM '84*, S. 105-113, Atlanta, 1984.
- [113] S. E. Minzer, D. R. Spears: New Directions in Signaling for Broadband ISDN. *IEEE Communications Magazine*, Vol. 27, No. 2, S. 6-14, Februar 1989.
- [114] S. E. Minzer: Broadband ISDN and Asynchronous Transfer Mode. *IEEE Communications Magazine*, Vol. 27, No. 9, S. 17-24, September 1989.
- [115] K. H. Möhrmann: Codierung von Videosignalen für die digitale Übertragung. *Telcom report 10*, Heft 6, S. 340-345, 1987.
- [116] R. W. Muise, T. J. Schonfeld, G. H. Zimmermann III: Experiments in Wideband Packet Technology. *Proceedings of the 1986 Zurich Seminar on Digital Communications*, S. 135-139, Zürich, 1986.
- [117] A. N. Netravali, J. O. Limb: Picture Coding: A Review. *Proceedings of the IEEE*, Vol. 68, No. 3, S. 366-407, März 1980.
- [118] M. F. Neuts: A Versatile Markovian Point Process. *Journal of Applied Probability*, Vol. 16, S. 764-779, 1979.
- [119] M. F. Neuts: *Matrix-Geometric Solutions in Stochastic Models*. The John Hopkins University Press, 1981.
- [120] M. F. Neuts: *Structured Stochastic Matrices of M/G/1 Type and their Applications*. Marcel Dekker, Inc., New York, 1989.
- [121] M. Nomura, T. Fujii, N. Ohta: Basic Characteristics of Variable Rate Video Coding in ATM Environment. *IEEE Journal on Selected Areas in Communication*, Vol. 7, No. 5, S. 752-760, Juni 1989.
- [122] NTG-Empfehlung 0903: „Nachrichtenverkehrstheorie - Begriffe“. *Nachrichtentechnische Zeitschrift*, Band 37, Heft 7, S. 465-474 und Heft 8, S. 539-552, 1984.
- [123] H. Ohnishi, T. Okada, K. Noguchi: Flow Control Schemes and Delay/Loss Tradeoff in ATM Networks. *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 9, S. 1609-1616, Dezember 1988.
- [124] P. Pancha, M. E. Zarki: Modeling Video Sources for Resource Allocation in ATM based BISDN. *Proceedings of the Third International Workshop on Packet Video*, Morristown, N.J., März 1990.
- [125] J. H. Patel: Performance of Processor-Memory Interconnections for Multiprocessors. *IEEE Transactions on Computers*, Vol. 30, No. 10, S. 771-780, Oktober 1981.

- [126] B. Pauwels: Terminal-To-Network Communication in an ATM-Based ISDN. *International Journal of Digital and Analog Cabled Systems*, Vol. 2, No. 1, S. 49-59, 1989.
- [127] F. Pereira, L. Masera: 'Constant Quality' Coding in ATM Networks: From Videophone to HDTV. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 31.1, Paris, Oktober 1989.
- [128] K. J. Pütz: Datenanwendungen in öffentlichen Breitbandnetzen - Möglichkeiten und Grenzen von ATM. *Informatik-Fachberichte 205: Tagungsband zur GI/ITG Fachtagung „Kommunikation in verteilten Systemen“*, S. 463-478, Stuttgart, Februar 1989.
- [129] V. Ramaswami: The N/G/1 Queue and its Detailed Analysis. *Advances in Applied Probability* 12, S. 222-261, 1980.
- [130] V. Ramaswami, G. Latouche: Modeling Packet Arrivals from Asynchronous Input Lines. *Proceedings of the 12th International Teletraffic Congress (ITC)*, Paper. 4.1A.5, Turin, Juni 1988.
- [131] V. Ramaswami, W. Willinger: Efficient Traffic Performance Strategies for Packet Multiplexers. *Proceedings of the ITC Specialist Seminar on „Traffic Theories for New Telecommunications Services“*, Paper 4.2, Adelaide, September 1989.
- [132] C. Rasmussen, J. Sorensen: A simple Call Acceptance Procedure in an ATM Network. *Proceedings of the ITC Specialist Seminar on „Traffic Theories for New Telecommunications Services“*, Paper 15.4, Adelaide, September 1989.
- [133] E. P. Rathgeb: *Algorithmen zur numerischen Analyse des zeit-diskreten GI/G/1-Systems im Frequenzbereich*. Diplomarbeit Nr. 769 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1985.
- [134] E. P. Rathgeb, T. H. Theimer, M. N. Huber: Buffering Concepts for ATM Switching Networks. *Proceedings of the IEEE GLOBECOM '88*, S. 1277-1281, Hollywood, Florida, November 1988.
- [135] E. P. Rathgeb, T. H. Theimer, M. N. Huber: ATM Switches - Basic Architectures and their Performance. *International Journal of Digital and Analog Cabled Systems*, Vol. 2, No. 4, S. 227-236, 1989.
- [136] J. Riordan: *Stochastic Service Systems*. John Wiley & Sons, New York, 1962.
- [137] J. Roberts: Jitter due to an ATM Multiplex - Application to Peak Rate Policing. *COST 224 - TD(89)035 -*, August 1989.
- [138] M. H. Rossiter: A Switched Poisson Model for Data Traffic. *A.T.R.*, Vol. 21, No. 1, S. 53-57, 1987.
- [139] J. Van Ruymbeke: Service Coding for Broadband Services. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 11.2, Paris, Oktober 1989.

- [140] D. Schmidt: *Iterative Algorithmen zur numerischen Analyse des zeitdiskreten GI/G/1-Systems im Zeitbereich*. Studienarbeit Nr. 743 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1985.
- [141] H. Schneider: The Concept of Virtual Paths and Virtual Channels in ATM-Networks. *Proceedings of the 1990 International Zurich Seminar on Digital Communications*, S. 63-73, März 1990.
- [142] F. C. Schoute: Simple Decision Rules for Acceptance of Mixed Traffic Streams. *Proceedings of the 12th International Teletraffic Congress (ITC)*, Paper. 4.2A.5, Turin, Juni 1988.
- [143] P. Sen, B. Maglaris, N.-E. Rikli, D. Anastassiou: Models for Packet Switching of Variable-Bit-Rate Video Sources. *IEEE Journal on Selected Areas in Communication*, Vol. 7, No. 5, S. 865-869, Juni 1989.
- [144] J. Shin, J. Kim: An Analysis of Packet Video Multiplexing by a 3-State Discrete-Time Markov Chain Model. *Proceedings of the Third International Workshop on Packet Video*, Morristown, N.J., März 1990.
- [145] M. Sidi, W.-Z. Liu, I. Cidon, I. Gopal: Congestion Control Through Input Rate Regulation. *Proceedings of the IEEE Globecom '89*, Paper 49.2, Dallas, November 1989.
- [146] R. Sieglen: *Zeitdiskrete Analyse einer Überlastabwehrstrategie in Vermittlungssystemen*. Studienarbeit Nr. 924 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1987.
- [147] A. de Smedt: ATM Signalling Channel Allocation and Meta-Signalling Issues. *Proceedings of the 3rd RACE 1022 Workshop*, Paper 12.6, Paris, Oktober 1989.
- [148] K. Sriram, W. Whitt: Characterizing Superposition Arrival Processes in Packet Multiplexers for Voice and Data. *IEEE Journal on Selected Areas in Communications*, Vol. 4, No. 6, S. 833-846, September 1986.
- [149] R. E. Staehler, J. J. Mansell, E. Messerli, G. W. R. Luderer, A. K. Vaidya: Wide-band Packet Technology for Switching Systems. *Proceedings of the XII International Switching Symposium (ISS)*, S. 448-454, Phoenix, März 1987.
- [150] S. Sumita, T. Ozawa: Achievability of Performance Objectives in ATM Switching Nodes. *Proceedings of the International Seminar on Performance of Distributed and Parallel Systems*, S. 45-56, Kyoto, Dezember 1988.
- [151] R. Suri: Infinitesimal Perturbation Analysis for General Discrete Event Systems. *Journal of the ACM*, Vol. 34, No. 3, S. 686-717, Juli 1987.
- [152] T. H. Theimer, E. P. Rathgeb, M. N. Huber: Performance Analysis of Buffered Banyan Networks. *Proceedings of the International Seminar on Performance of Distributed and*

*Parallel Systems*, S. 57-72, Kyoto, Dezember 1988. (Eine erweiterte und überarbeitete Version erscheint in *IEEE Transactions on Communications*).

- [153] T. H. Theimer, C. Blondia: A Discrete-Time Model for ATM Traffic. *RACE R1022 Dokument PRLB\_123.0018\_CD\_CC / UST\_123.0022\_CC\_CD*, Oktober 1989.
- [154] T. H. Theimer: How to Compute the Moments of a GMDP. *RACE R1022 Dokument UST\_123.0023\_CD\_CC*, November 1989.
- [155] A. Thomas, J. P. Coudreuse, M. Serval: Asynchronous Time-Division Techniques: An Experimental Packet Network Integrating Videocommunication. *Proceedings of the XI International Switching Symposium (ISS)*, Paper 32-C-2, Florenz, Mai 1984.
- [156] P. Tran-Gia: A Renewal Approximation for the Generalized Switched Poisson Process. in: *Mathematical Computer Performance and Reliability*, G. Iazeolla, P. J. Courtois, A. Hordijk (editors), Elsevier Science Publishers B.V. (North-Holland), S. 167-179, 1984.
- [157] P. Tran-Gia: *Zeitdiskrete Analyse verkehrstheoretischer Modelle in Rechner- und Kommunikationssystemen*. 46. Bericht über verkehrstheoretische Arbeiten (Habilitationsschrift), Universität Stuttgart, Institut für Nachrichtenvermittlung und Datenverarbeitung, 1988.
- [158] P. Tran-Gia, H. Ahmadi: Analysis of a Discrete-Time  $G^{(X)}/D/1-s$  Queueing System with Applications in Packet-Switching Systems. *Proceedings of the IEEE INFOCOM'88*, New Orleans, 1988.
- [159] R. C. F. Tucker: Accurate Method for Analysis of a Packet-Speech Multiplexer with Limited Delay. *IEEE Transactions on Communications*, Vol. 36, No. 4, S. 479-483, April 1988.
- [160] J. S. Turner: Design of an Integrated Services Packet Network. *Ninth Data Communications Symposium, ACM Sigcomm Computer Communication Review*, Vol. 15, No. 4, S. 124-133, September 1985.
- [161] J. S. Turner: New Directions in Communications. *Proceedings of the 1986 International Zurich Seminar on Digital Communications*, S. 25-32, Zürich, März 1986.
- [162] J. S. Turner: The Challenge of Multipoint Communication. *Proceedings of the 5th ITC Seminar on „Traffic Engineering for ISDN Design and Planning“*, Lake Como, Italy, Mai 1987.
- [163] J. S. Turner: Design of a Broadcast Packet Switching Network. *IEEE Transactions on Communications*, Vol. 36, No. 6, S. 734-743, Juni 1988.
- [164] A. Weber: *Analyse von Mehrkanaldurchschaltungen in dreistufigen Zeit-Raum-Zeit-Koppelanordnungen*. Diplomarbeit Nr. 885 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1988.

- [165] B. X. Weis: Cell Alignment in ATM Systems via Header Error Check: Performance Evaluation and Hardware Considerations. *Proceedings of the 1990 International Zurich Seminar on Digital Communications*, S. 115-131, März 1990.
- [166] M. Wippenbeck: *Zeitdiskrete Analyse des GI/D/1-s Systems*. Studienarbeit Nr. 936 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1989.
- [167] M. Wizgall: ECR 900: The Pan-European Digital Mobile Communication System. *Proceedings of the International Telecom Symposium (ITS) '88*, Taipei, Taiwan, September 1988.
- [168] R. W. Wolff: Poisson Arrivals See Time Averages. *Operations Research*, Vol. 30, No. 2, S. 223-231, März-April 1982.
- [169] Y. S. Yeh, M. G. Hluchyi, A. S. Acampora: The Knockout Switch: A Simple, Modular Architecture for High-Performance Packet Switching. *IEEE Journal on Selected Areas in Communications*, Vol 5, S. 1274-1283, Oktober 1987.
- [170] D. M. Young: *Iterative Solution of Large Linear Systems*. John Wiley & Sons, New York, 1968.
- [171] C. Yuan, J. A. Silvester: Queuing Analysis of Delay Constrained Voice Traffic in a Packet Switching System. *IEEE Journal on Selected Areas in Communications*, Vol. 7, No. 5, S. 729-738, Juni 1989.