

Request Contention and Request Polling for the Upstream Media Access Control in ATM Access Networks

Joachim Charzinski

*Siemens AG, Public Communication Networks Group
Hofmannstr. 51, D-81359 München, Germany
Tel. +49 89 722 46803, Fax +40 89 722 26877
e-mail: joachim.charzinski@oen.siemens.de*

(The studies reported here were performed while the author was with the Institute of Communication Networks and Computer Engineering, University of Stuttgart, Germany)

Abstract

This paper presents a simulation study comparing different mechanisms used in the upstream channel of ATM access networks. In *request polling*, status information from the access stations is transmitted periodically in minislots dedicated to each access station. In *request contention*, the access stations use a contention based multiple access scheme with a multi-slot stack contention resolution algorithm for random access to the minislots. In addition, in both cases request information is piggy-backed to upstream ATM cells.

In a simulation study, *request polling* and *request contention* are compared using the same system parameters. It is shown that (1) the optimum stack parameter is greater than three if the round trip delay in the system is taken into account and (2) under most conditions request contention produces longer mean transfer delays and delay quantiles than request polling.

Keywords

ATM, Access Network, Media Access Control, Delay Quantile, Comparison, Request Contention, Stack Algorithm, Request Polling

1 INTRODUCTION

In a modern high-speed communication infrastructure, providing cost-effective access to a global broadband network is a crucial issue. Shared media ATM access networks have been under study for some time, mainly using Passive Optical Networks (PON) (Ballance et al. 1990, Glade & Keller 1993, Killat 1996) or the Hybrid Fibre Coax architecture (ATMF 1996). The bandwidth demands of residential users should not be under-estimated as even information

retrieval services like accessing the world wide web can create a high traffic volume for both directions of the network: Except for pure real-time media delivery services, most data transmission employs transport protocols using an acknowledgement mechanism to ensure successful transmission of all data packets. These acknowledgements – in addition to normal user traffic – can pose high bandwidth demands even on the “upstream” channel (from users to the public network) in an access network.

In this paper, two mechanisms employed in the shared media access control for the upstream transmission of cells in ATM access networks are compared in terms of the mean delay and the delay quantiles they introduce into the upstream traffic. In section 2, the investigated system models are introduced, the two media access control mechanisms are described briefly and the traffic scenarios used in the simulation study are defined. The choice of the stack parameter used in contention resolution is discussed in section 3 before the delay results are presented as a function of the ratio of time slots spent to support the MAC mechanisms.

2 SYSTEM MODELS

2.1 Reference Architecture

In the “downstream” direction (from public network to customers), the ATM time slot structure can be maintained on the access network. Each downstream cell is broadcast to every customer’s interface, where a filtering function can substitute all ATM cells not destined for the local interface by idle cells.

In the “upstream” direction (from customer to public network) of PON and HFC access networks, a media access control mechanism is needed to coordinate the times when different stations can access the shared medium. The network elements where the MAC (media access control) protocol is terminated will be called “access station” (AS) and “headend” (HE) to be independent of the actual network architecture used in a specific system. The role of access station can be performed by a network termination or an optical network unit (Killat 1996) and the headend functions can be found in a line termination (I.327 1993), the ATM Digital Terminal (ATMF 1997) or the DAVIC Access Node (DAVIC 1997).

We assume here that a single return channel is shared between the access stations using TDMA (time division multiple access). The issue of ranging has been discussed elsewhere (van Heijningen et al. 1994), and we can take the access network to be fully ranged, i.e. all access stations show the same reaction time 2τ between the headend issuing a permit and receiving the corresponding upstream ATM cell from the access station. The timing between the headend and one access station is depicted in Figure 1. Note that the sequence of permits issued by the headend is reproduced by the sequence of

upstream ATM cells arriving at the headend one round trip delay (2τ) later. The actual transmission time needed to emit an ATM cell on a high speed channel (one time slot of duration t_C) is usually at least an order of magnitude smaller than the round trip delay, so that it is not included in Figure 1 and we assume the transmission times of permits and cells and the associated hardware delays to be included in the round trip delay. For every time slot, the headend can issue a new permit to a different AS, even if the previous permit has not yet resulted in an upstream cell arriving at the head end. This mode of operation is also called “pipeline polling” (Quayle 1996). The destination of a permit is independent of the destination of the ATM cell it is attached to. Note that the whole transfer delay consists of a constant part (3τ) and a variable part ($T_{RW} + T_{PW}$).

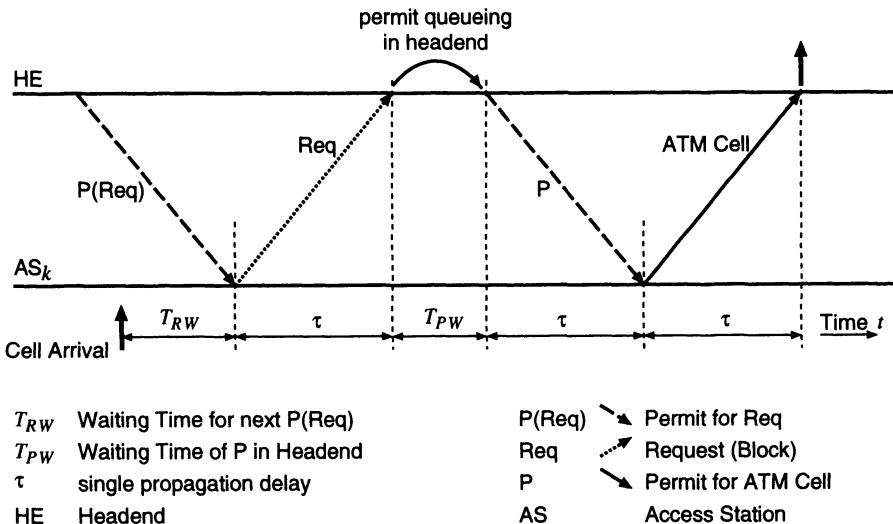


Figure 1 Requests, Permits and Waiting Time Contributions

Figures 1 (timing) and 2 (simulation model) show that an access station, after receiving an ATM cell from one of its sources, stores this cell in its cell queue and transmits a *request* to the headend at the next opportunity. After the request has arrived at the headend, the headend reacts by generating a permit for the respective access station and queueing it in its central permit queue, from which one permit is transmitted downstream in each time slot (except for the request block permits described later). When the access station receives a permit for an ATM cell, the cell waiting at the head of the station’s cell queue is transmitted upstream.

Requests to be transmitted to the headend can be piggy-backed on upstream ATM cells from the same access station. However, when an ATM cell arrives at an empty access station queue, there must be an additional way of trans-

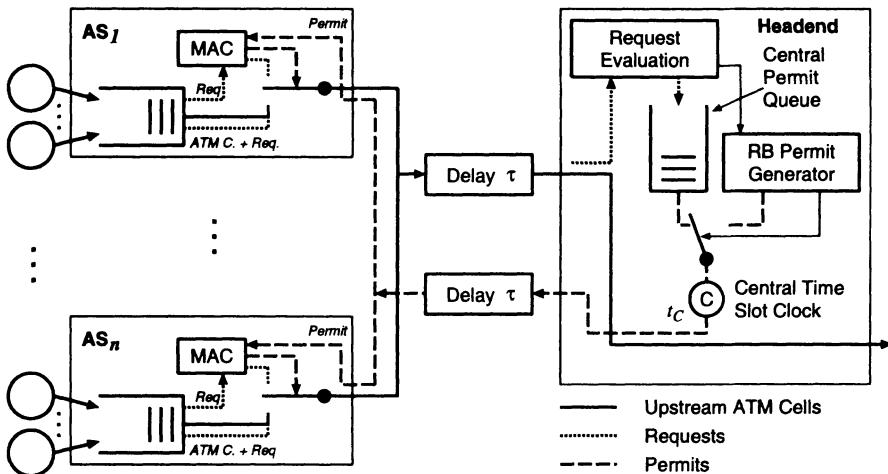
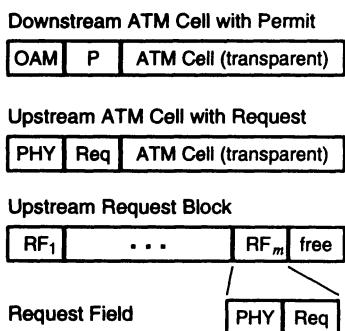


Figure 2 Simulation model

mitting a request. This is where so-called “request blocks” (RBs) are used. An RB has the duration of a normal ATM cell slot and it consists of m minislots, each giving one access station the opportunity to transmit a request. A request is simply the transmission of the value of the current queue length in the access station, which can be transformed into the number of newly arrived ATM cells by a simple algorithm in the headend (Killat 1996, Charzinski 1996).



OAM	Operations, Administration and Maintenance + collision signal and ranging
PHY	Field for Physical Layer functions: synchronization, OAM and ranging (not in RBs)
P	Permit
Req	Request
RF _j	Request field from AS _j in a minislot
RB	Request Block with m minislots

Figure 3 Upstream and Downstream Slot Formats

As illustrated in Figure 3, each access station needs to transmit a synchronization preamble before transmitting an ATM cell or a request upstream in order to keep a minimum gap between successive transmissions and help the burst mode receiver in the headend synchronize to the specific access station’s transmitter (Killat 1996). Assuming that the downstream bandwidth is an integer multiple r of the upstream bandwidth, every r^{th} downstream cell has a valid permit attached. For the following studies, we assume that the preamble

together with the request field occupies three octets (Killat 1996), so that a slot consists of 56 octets and it can be divided into 18 minislots, of which $m = 16$ will be used for upstream request transmissions.

The following two subsections describe how these minislots are used for request polling or request contention.

2.2 Request Polling

In *request polling*, all access stations connected to the access network are polled periodically for requests. Thus, if there are more than m access stations, each one can transmit in one dedicated minislot of every $\lceil n_0/m \rceil$ th request block, with n_0 denoting the number of all access stations. Given the ratio ρ_P of time slots used for request blocks, the time between two request block permits is

$$t_P = t_C / \rho_P \quad (1)$$

and the time between two request block permits for the same access station is

$$t_{PE} = \lceil n_0/m \rceil t_P \quad . \quad (2)$$

The request polling mechanism has been analyzed for Poisson input traffic in (Charzinski 1997) where it was shown that there is an optimum value for the request block ratio ρ_P .

2.3 Request Contention

In *request contention*, only those access stations contend for the request minislots in a request block which have had a new ATM cell arrive to an empty queue. The mechanism used for collision resolution is the Multi-Slot q_W -ary Stack Algorithm described in (Bisdikian 1996) and (Golmie et al. 1996). As illustrated in Figure 4, instead of varying the polling period t_P , the number r_B of consecutive request blocks is varied with ρ_P here to be

$$r_B = \rho_P t_P / t_C \quad . \quad (3)$$

The reason is that in order for an access station to repeat a request in the following request block, it needs to have received information on whether the previous access to a request minislot resulted in a collision (Collision Signal CS in Figure 4). The state transition diagram for the access station is shown in the appendix.

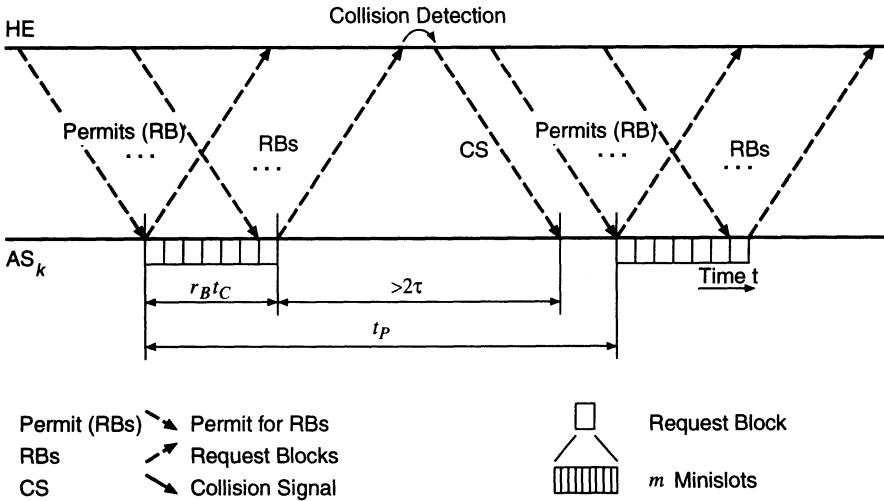


Figure 4 Timing Constraints using Request Contention

2.4 System Parameters and Traffic Scenarios

Table 1 defines four traffic scenarios. The data are divided into three categories: In the upper part of Table 1, the sources are defined. In the middle part, general system parameters are given, and in the lower part, the configuration of sources and access stations is described.

Name		ND100	ND10	TS100	TS10
Source Period	t_S/t_C	100	10	100	10
Burstiness	b	1	1	2	2
Mean Burst Length	c_B	–	–	10	20
Minislots per RB	m	16	16	16	16
Propagation delay	τ/t_C	18	18	18	18
Total Acc. Stations	n_0	128	128	128	128
Active Acc. Stations	n	80	8	70	10
Sources per Station	b_S	1	1	2	1
Offered Load in %	A_R	80	80	70	50
Multiplexing Gain	$b \cdot A_R$	0,8	0,8	1,4	1,0

Table 1 Traffic Scenarios and System Parameters

In two scenarios, the sources have a low cell rate (one cell every 100 slots), so that the mechanism of requests coupled to upstream ATM cells cannot normally be used. In the other two scenarios, the sources have a high cell rate (one cell every 10 slots), so that coupled requests can almost always be used. Two scenarios (ND) use periodic traffic and two use talkspurt-silence (TS) sources with a geometrically distributed number of equidistant cells in a burst and negative exponentially distributed silence periods. The burstiness b denotes the ratio of peak to long-term mean cell rate. A burstiness of $b=2$ is enough to ensure that normally the last ATM cell has left the access station before the next burst arrives, so that the access station becomes idle between bursts.

The single propagation delay is assumed to be 18 cell slots, which roughly corresponds to $50\mu\text{s}$ or a distance of 10 km after ranging between access stations and headend at 155.52 Mbit/s. The total number of access stations $n_0 = 128$ is not varied, so that the ratio $t_{PE}/t_P = 8$ is fixed for all scenarios.

A target offered load of $A_R = 80\%$ was chosen for the ND scenarios. The offered load in the TS scenarios was determined by limiting the burst scale loss probability in a bufferless multiplexer to 10^{-9} using (Roberts 1991, (7.1.5)).

3 RESULTS

Before the results for the ND and TS scenarios are compared in sections 3.2 and 3.3, the influence of the stack parameter q_W (see section 2.3) is investigated in section 3.1. All results have been acquired using discrete event simulation with 10^7 cells. The confidence intervals shown are for a 95 % confidence level. For mean values they have been computed from 10 part test results using a Student- t test. Delay quantiles have been extracted from the empirical delay distribution measured during the whole simulation and their confidence intervals have been computed according to the method given in (Sachs 1992). For simulating the periodic ND scenarios, the phase shift method described e.g. in (Panken 1997) was used.

3.1 Dimensioning the Stack Parameter

For scenario ND100, Figure 5 gives the mean end-to-end delay (EED) and Figure 6 gives the ($1-10^{-5}$) quantile of EED as a function of the stack parameter q_W . The request block period is $t_P = 50 t_C$ and there are $r_B = 4$ consecutive request blocks used in each period, leading to a ratio of $\rho_P = 12.5\%$ of request blocks per upstream slot. Only integer values of q_W are possible; the discrete points have only been interconnected to increase the readability of the diagram.

For the long round trip delay τ considered here, the optimum value of q_W

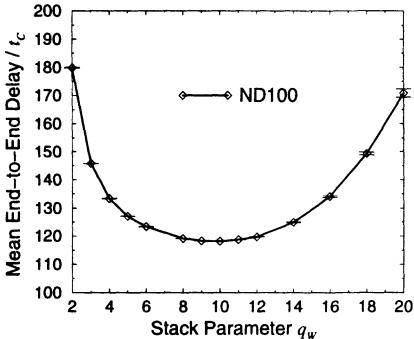


Figure 5 ND100: Mean Delay

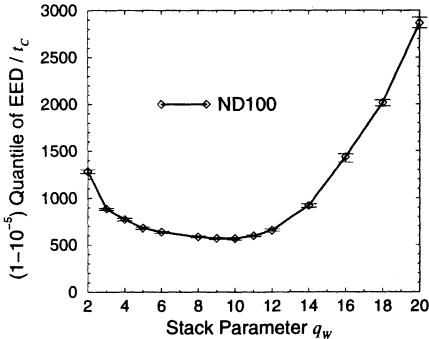


Figure 6 ND100: Delay Quantile

is around 10, which is much more than the value of 3 that has been shown to be the optimum for systems with immediate collision feedback (Bisdikian 1996, Merakos & Bisdikian 1988). For the following investigations, results for both $q_w = 3$ and $q_w = 10$ will be given.

3.2 Constant Cell Rate Traffic

Figure 7 shows the $(1-10^{-5})$ quantile of end-to-end delay in the ND100 scenario as a function of the ratio ρ_P of upstream slots used for request blocks. As above, the discrete points have been interpolated to increase the readability of this and the following diagrams. Using request contention (RC) on the same number of minislots leads to much longer delay quantiles than request polling (RP). Increasing the stack parameter q_w from 3 to 10 helps in reducing the delay, but leads to instability when ρ_P is less than about 10 %. Due to the inherent stability of an $n*D/D/1$ periodic service system even at the load limit, the delay quantile reaches a minimum at $\rho_P = 20\%$ where the total upstream slot utilization $\rho_P + A_R$ reaches 1.

In scenario ND10, the high source rate ensures that in steady state, all new ATM cell arrivals can be communicated to the headend via requests coupled to upstream ATM cells from the same access station. Therefore the difference between request polling and request contention is much smaller than in the ND100 scenario discussed above. The remaining difference is due to the different distribution of service interruptions for request blocks in the headend permit queue: As shown in Figure 4, multiple request blocks must be blocked together to effectively increase the number of minislots available for contention resolution. This leads to service interruptions in the headend queue lasting for $r_B = \rho_P t_P / t_C$ time slots. On the other hand, for request polling the request blocks are distributed over the whole period t_{PE} , leading to much shorter service interruptions at a time.

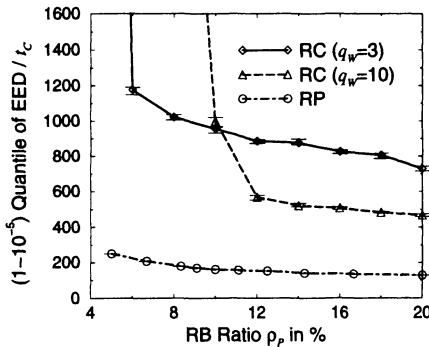


Figure 7 ND100: Delay Quantile

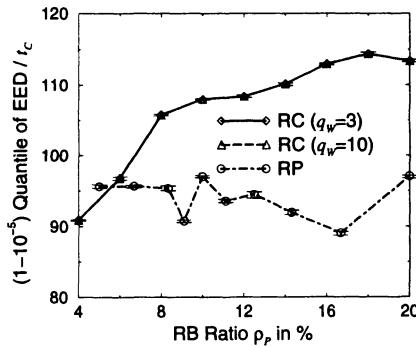


Figure 8 ND10: Delay Quantile

3.3 Variable Cell Rate Traffic

The results for request polling in Figures 9 through 12 show that the optimum RB ratio ρ_P is around 10–15 %. For smaller values of ρ_P , the waiting time T_{RW} for a request block dominates whereas for larger ρ_P the headend queueing delay T_{PW} increases as the load limit is approached in the headend queue.

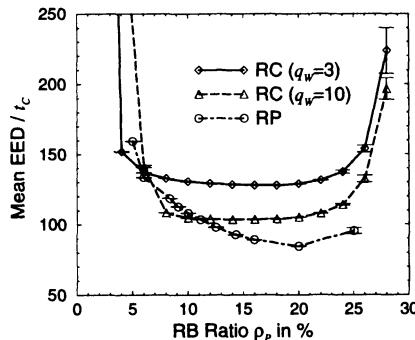


Figure 9 TS100: Mean Delay

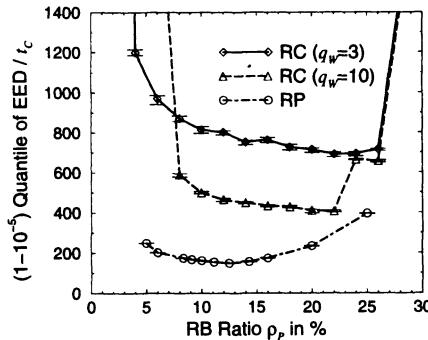


Figure 10 TS100: Delay Quantile

While request contention can lead to lower mean delays than request polling when there is little bandwidth spent for request blocks (Figures 9 and 11), this situation is reversed when ρ_P is increased or when the delay quantiles in Figures 10 and 12 are considered. In scenario TS100 with its many low rate on/off sources and a mean offered load of $A_R = 70\%$, request contention shows instability both at low and high values of ρ_P , regardless of the stack parameter q_w . On the other hand, request polling shows a minimum delay quantile at around $\rho_P \approx 12\%$ in scenario TS100 (Figure 10) and is not so prone to become instable. The results for scenario TS10 shown in Figures 11 and 12 clearly indicate that even if the number of stations contending for access to the request minislots is low (only 10 out of 128 stations are active)

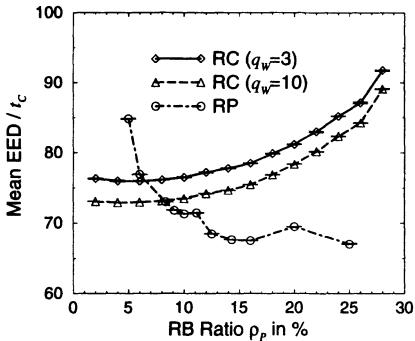


Figure 11 TS10: Mean Delay

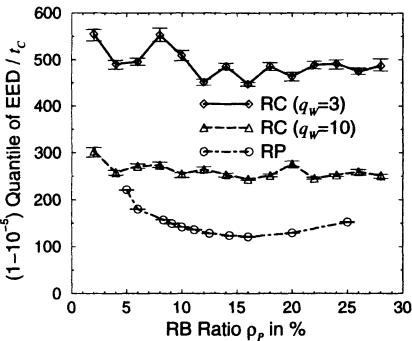


Figure 12 TS10: Delay Quantile

and request contention is only being used at each beginning of a burst, the mean delay and the $(1-10^{-5})$ delay quantile can be much lower under request polling than under request contention if a sufficient number of minislots is available.

4 CONCLUSIONS

A simulation study has been presented evaluating two mechanisms used in the upstream media access control of ATM access networks. The comparison shows that even under conditions favourable for contention based mechanisms, i.e. a small proportion of active stations and contention based access only being used at the beginning of each data burst, request polling leads to lower mean delays and also to lower delay quantiles when *the same number* of upstream minislots is dedicated for request transmission. In practice, the difference will be even greater because synchronization of a burst mode receiver to an unknown transmitter takes longer than if the transmitter phase and amplitude is known before. Therefore the synchronization preambles for request contention will have to be longer than for request polling and the number of minislots per request block will be smaller, thus further increasing the delay with request contention.

As a side result, it was shown that the optimum stack parameter for request contention is more than 3 if the access stations have to wait for one round trip delay to get feedback information on the success of previous random access to a minislot.

REFERENCES

- J.W. BALLANCE, R.F. LEE, P.H. ROGERS, M.F. HALLS, "A B-ISDN Local Distribution System based on a Passive Optical Network." *Proc. GLOBECOM'90*, 1990, pp. 305.4.1–305.4.5

- M. GLADE, H. KELLER, "Novel Algorithm for Time Division Multiple Access in Broadband ISDN Passive Optical Networks." *Int. J. of Digital and Analog Comm. Systems* Vol. 6 (1993) pp. 55–62
- U. KILLAT (Ed.), *Access to B-ISDN via PONs – ATM Communication in Practice*, Wiley/Taubner 1996
- THE ATM FORUM, "The UPSTREAMS Protocol for HFC Networks, Revision 1.", ATM Forum/IEEE Doc. 95-1435R1, Jan. 1996
- P. VAN HEIJNINGEN, T. MOSCH, A. VAN OOVEN, L. D'ASCOLI, K. DE BLOK, P. SOLINA, "Out-of-Band Ranging Method for ATM over PON Access Systems." *Proc. ECOC'94*, Florence, Italy, 1994
- ITU-TSS, ITU-T Recommendation I.327 *B-ISDN Functional Architecture*. March 1993.
- THE ATM FORUM TECHNICAL COMMITTEE RBB WG, *RBB Baseline Document Draft*. Doc. BTD-RBB-001.02, April 1997
- DIGITAL AUDIO-VISUAL COUNCIL (DAVIC), Specification Part 02 Version 1.2 *System Reference Models and Scenarios*. Rev. 4.0 Geneva, Switzerland, 1997
- A. QUAYLE, "Broadband passive optical network media access control protocols." in *Proc. SPIE Photonics East 1996 Conf. on All-Optical Comm. Systems: Architecture, Control and Network Issues II* Vol. 2919, Boston, MA, USA, Nov. 1996, pp. 268–278
- J. CHARZINSKI, "A new approach to ATM access networks." in *Proc. of SPIE Photonics East 1996 Conf. on Broadband Access Systems* Vol. 2917, Boston, 1996, pp. 108–119
- J. CHARZINSKI, "Performance Analysis of a Multiple Access Mechanism for ATM Access Networks." in *Proc. IEEE ATM'97 Workshop*, Lisboa, Portugal, May 1997, pp. 477–485
- C. BISDIKIAN, "Performance Analysis of the Multi-slot n -ary Stack Random Access Algorithm (msSTART)." *IEEE Project 802.14 Working Group Doc. IEEE802.14-96/117*, May 1996
- N. GOLMIE, S. MASSON, G. PIERIS, D. SU, "Performance evaluation of MAC protocol components for HFC networks." in *Proc. SPIE Photonics East 1996 Conf. on Broadband Access Systems* Vol. 2917, Boston, 1996, pp. 120–130
- J.W. ROBERTS, EDITOR, *Performance evaluation and design of multiservice networks*. Final report of the COST 224 project, Office for Official Publications of the European Communities, Luxembourg, 1991
- L. SACHS, *Angewandte Statistik – Anwendung statistischer Methoden*. Springer-Verlag, Berlin, 7. Auflage 1992, S. 338 (in German)
- F.J.M. PANKE, *Design and performance evaluation of multiple-access protocols for ATM-based passive optical networks* Dissertation, Wiskunde en Informatica, Katholieke Universiteit Nijmegen. Thesis Publishers, Amsterdam, NL, 1997
- L.F. MERAKOS, C. BISDIKIAN, "Delay Analysis of the n -Ary Stack Random-

Access Algorithm." *IEEE Trans. Information Theory*, Vol. 34, No. 5, Sep. 1988, pp. 931–942

APPENDIX 1 STATE TRANSITION DIAGRAM FOR REQUEST CONTENTION

Figure 13 defines the MAC process in an access station for request contention. The annotation $A[B]/C$ along an arrow is to indicate that this transition between two states happens when event A occurs and if condition B is true. During the transition, action C is performed.

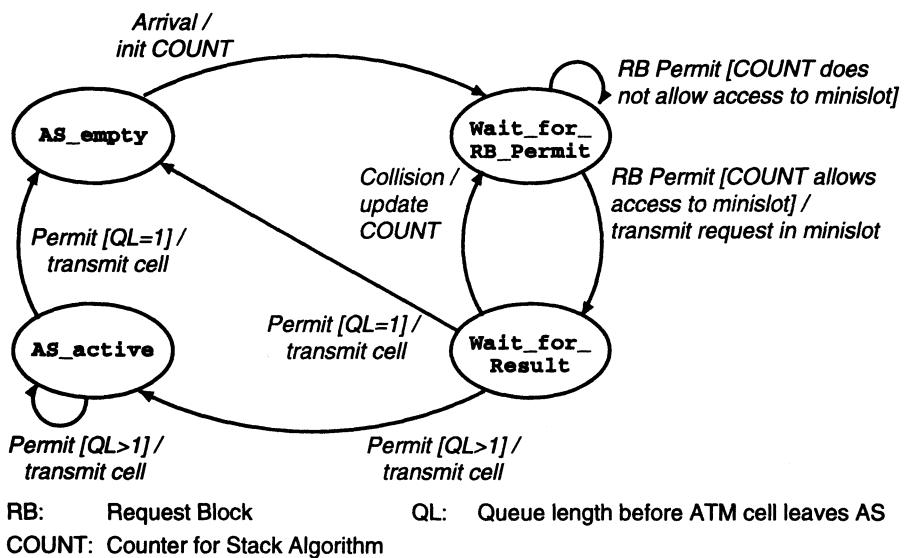


Figure 13 State Transition Diagram for an Access Station with Request Contention