7th ITC Seminar, New Jersey, October 1990

Paper 9.4.

Waiting Time Distribution of a Discrete SSMP/G/1 Queue and its Implications in ATM Systems

Wenyue Ding and Peter Decker

Aachen University of Technology
F.R. of Germany

In this paper a Special Semi-Markov Process (SSMP) for modeling ATM input processes at the cell level is introduced. The discrete SSMP/G/1 queue is analyzed yielding a recursive formula of the waiting time distribution for the calculation of which an accurate numerical algorithm is given. Finally, correlation effects of the interarrival time on queues is discussed.

1 Introduction

Successful design of the future Broadband ISDN (Integrated Service Digital Network) operated by the ATM (Asynchronous Transfer Mode) principle depends on adequate models of the traffic sources which include characteristic correlation properties. In this paper a class of Special Semi-Markov Processes (SSMP) is suggested for modeling these sources. It is shown that the SSMP can directly represent *discrete* processes, especially the input processes resulting from the ATM systems. As a first application the SSMP/G/1 queueing system is analyzed yielding a recursive formula for the stationary waiting time distribution. The correlation effect of input processes on the waiting time is discussed.

Some other kinds of special semi-Markov processes e.g. [10] [6], including the well-known Markov-modulated Poisson processes (MMPP), are restricted to continuous processes. Queues with general semi-Markov input processes and some specific service time distributions are analyzed in [3] [7] [8].

2 Characterisation of SSMP

A descriptive definition of a semi-Markov process (SMP) is given in [9]. An SMP is a stochastic process which moves from one to another of a countable number of states. The successive states visited form a Markov chain and the process stays in a given state a random length of time with a distribution function (d.f.) which may depend on this state as well as on the one to be visited next. The SSMP is an SMP whose sojourn time distribution in a given state depends *only* on the actual state. Hence, the transition distribution matrix [9] of this SSMP is given by

$$\mathbf{Q} = (Q_{ij}(x)) = (p_{ij}F_i(x)), \qquad 0 \le x < \infty, \qquad i,j = 1,2,\cdots,m \tag{1}$$

or in discrete-time domain

$$\mathbf{Q} = (Q_{ij}(k)) = (p_{ij}F_i(k)), \qquad k = 0,1,2,\cdots, \qquad i,j = 1,2,\cdots,m \tag{2}$$

where $\mathbf{P} = (p_{ij})$ denotes the transition probability matrix of the Markov chain, and F_i the sojourn time d.f. in state i. The arrival occurs immediately after the process has jumped to the next state.

Let P_i be the stationary probability of the matrix $\mathbf{P}$, the marginal d.f. of the interarrival time is given by

$$F(x) = \sum_{i=1}^{m} P_i F_i(x). \tag{3}$$

The h-th order transition probability matrix of $\mathbf{P}$ is given by the h-th power of $\mathbf{P}$:

$$\mathbf{P}^{(h)} = \mathbf{P}^h = (p_{ij}^{(h)}). \tag{4}$$

Then the h-th order autocorrelation coefficient of the interarrival time is

$$r_h = \frac{\sum_{i=1}^{m}\sum_{j=1}^{m} P_i \overline{X}_i (p_{ij}^{(h)} - P_j)\overline{X}_j}{\mathrm{Var}(X)}, \tag{5}$$

where $\overline{X}_i$ is the mean of the random variable having the d.f. $F_i(x)$ and $\mathrm{Var}(X)$ is the variance of $F(x)$. For the two-state case (m=2) this formula is reduced to

$$r_h = \frac{\kappa^h P_1 P_2 (\overline{X}_1 - \overline{X}_2)^2}{\mathrm{Var}(X)}, \tag{6}$$

where κ is the first order correlation coefficient of the Markov chain with the transition probability matrix $\mathbf{P}$ and is given by $\kappa = 1 - p_{12} - p_{21}$.

3 SSMP as traffic source models in ATM systems

In an ATM system the cell generating process from different call types can be easily characterized by the SSMP. Here are some simple examples:

- **Voice source**: a two-state SSMP could be used. State 1 corresponds to the silent period and additionally the talk spurt in case of one cell (at least in the theoretical case), state 2 to the talk spurt in case of more than one cell, Fig.1. The interarrival time distribution functions in the each state are chosen as follows

$$F_1(k) = \begin{cases} 0, & k < d \\ F_{\mathrm{sil}}(k-d), & k \geq d \end{cases} \qquad F_2(k) = \begin{cases} 0, & k < d \\ 1, & k \geq d \end{cases} \tag{7}$$

  where d describes the constant interval of two successive cells and $F_{\mathrm{sil}}(k)$ is the distribution of the silent length. In this model the silent length is not necessarily geometric distributed. The number of cells in the talk spurt has a geometric distribution, but this restriction can also be by-passed by taking more states for the talk spurt.

- **Video with variable bit rate**: every state $i, i = 1, 2, \cdots, m$ means a state with relatively constant bit rate λ_i, Fig.2. The diagonal element p_{ii} of $\mathbf{P}$ should be strongly greater than other element p_{ij}, since a scene of relatively constant moving contents lasts for a while. A geometric distribution for $F_i(k)$ may be sufficient. Another distribution type could be used to fit the slight variations around λ_i.

- **Data source**: many data sources can be approximated by renewal processes, which can be regarded as SSMP with only one state.

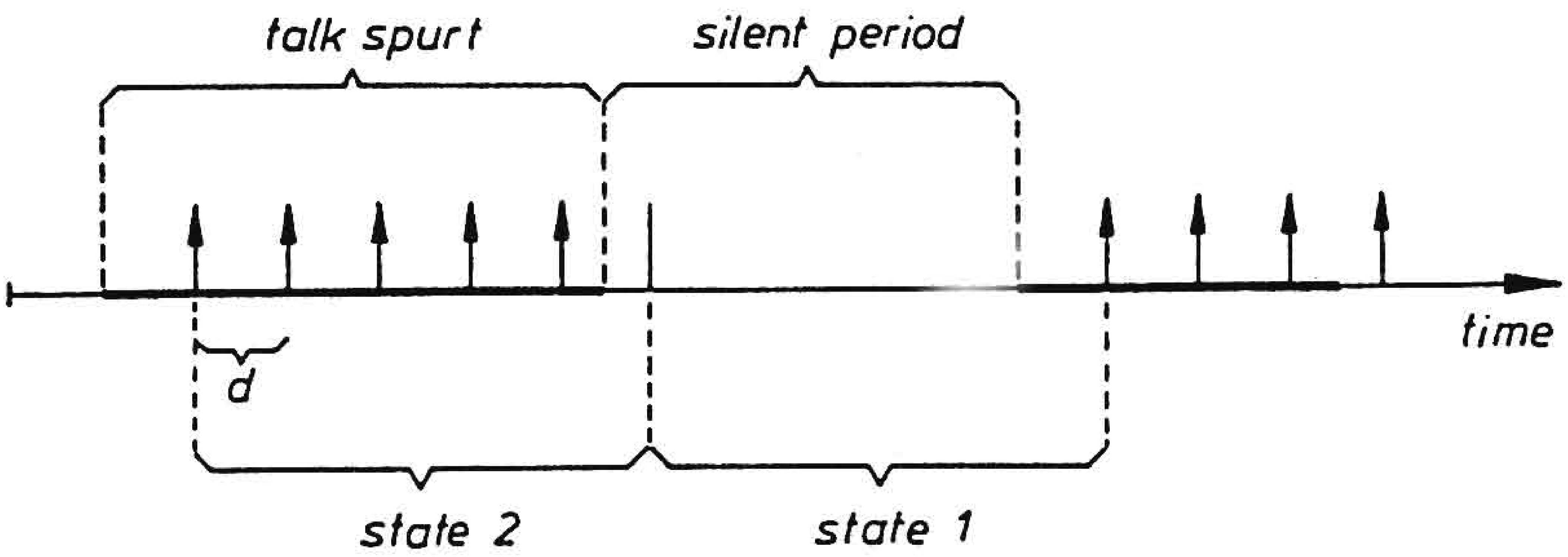

Fig. 1 A two-state SSMP model of an individual call with talk spurt and silent period.

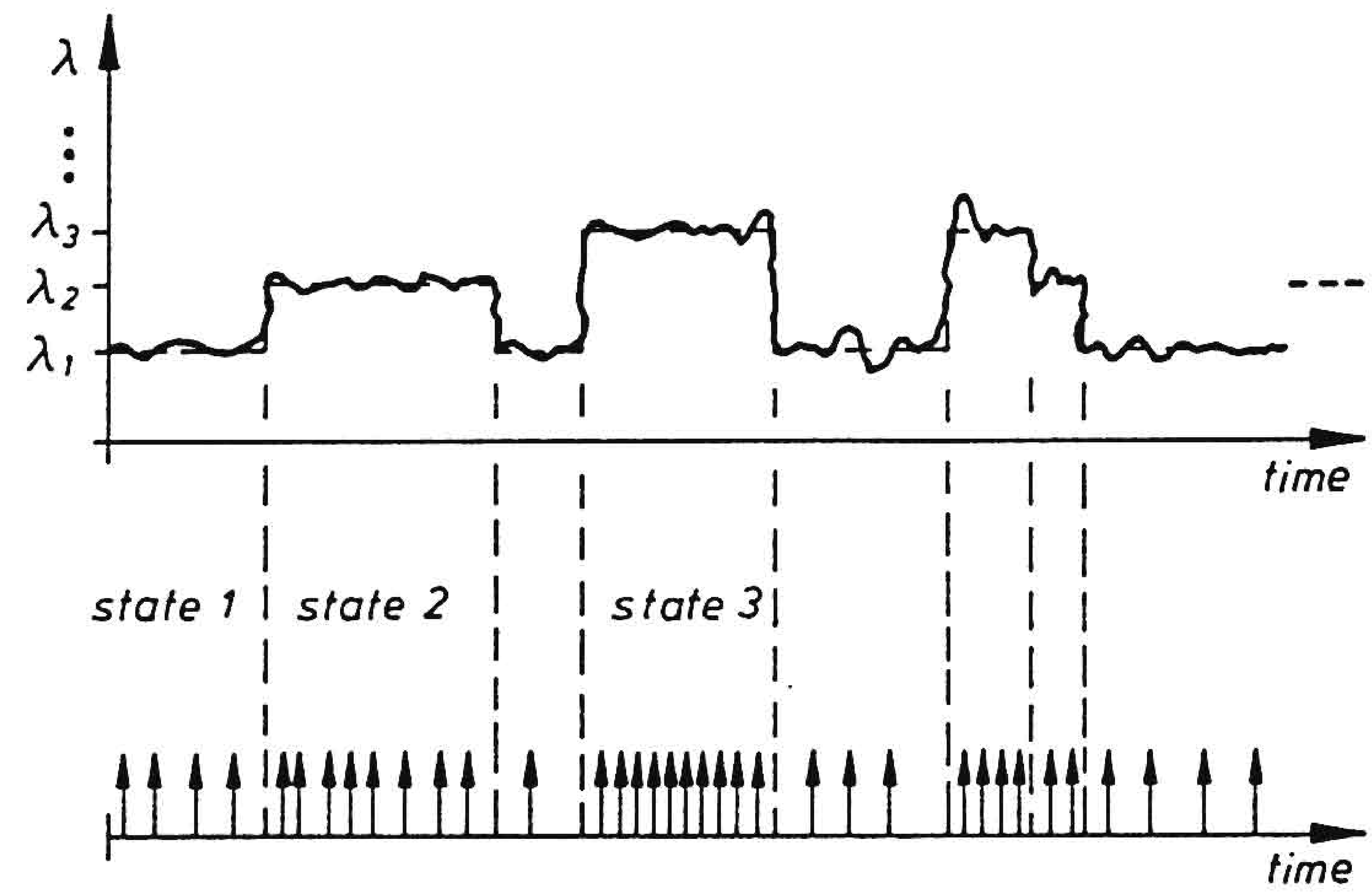

Fig. 2 An SSMP model for a variable bit rate video source.

- **The superposition** of several SSMP processes does not necessarily result an SSMP. But it could be approximated by an SSMP having only few states. Markov modulated Poisson processes (MMPP) are successfully applied for modeling superposition processes in many areas. A very close relation between SSMP and MMPP can be proven: an MMPP having two states is actually a continuous-time SSMP.

4 Waiting time distribution of the discrete SSMP/G/1 queue

In [2] a relation for the waiting time of SM/G/1 queue is given, which can only be solved for some particular cases. Our approach is based on the Lindley equation [5]. The following notations are used in the sequel:

A_n	:	time between $(n-1)$-st and n-th arrivals
B_n	:	service time of n-th job
W_n	:	waiting time of n-th job
S_n	:	state for n-th arrival $(S_n = 1, 2, \cdots, m)$
$C_{(n-1),i}$	:	$B_n - A_{n-1} \| (S_{n-1} = i)$
$P_S(i)$	:	$\Pr\{S = i\}$
$P_X(k)$	:	$\Pr\{X = k \text{ time slots}\}$
$P_{X_n S_n}(k, i)$	:	$\Pr\{X_n = k \text{ time slots} \wedge S_n = i\}$

The bi-variante sequence $\{(W_n, S_n), n = 0, 1, \cdots, \}$ is a Markov process, since the stochastic properties of (W_n, S_n) can be completely described by (W_{n-1}, S_{n-1}):

$$W_n = \begin{cases} W_{n-1} + B_n - A_{n-1}|S_{n-1}; & \text{if } W_{n-1} + B_n - A_{n-1}|S_{n-1} \geq 0; \\ 0 & \text{otherwise;} \end{cases} \tag{8}$$

$$P_{S_n}(j) = \sum_{i=1}^{m} p_{ij} P_{S_{n-1}}(i). \tag{9}$$

Our aim is to obtain a recursive formula for the joint probability density $P_{W_nS_n}(k,j)$. Because the state S_n is not determined by W_{n-1} we have

$$\begin{aligned} & \Pr\{W_n = k, S_n = j | W_{n-1} = l, S_{n-1} = i\} \\ = \; & \Pr\{W_n = k | W_{n-1} = l, S_{n-1} = i\} \Pr\{S_n = j | S_{n-1} = i\} \\ = \; & p_{ij} \Pr\{W_n = k | W_{n-1} = l, S_{n-1} = i\}. \end{aligned} \tag{10}$$

From eq.(8) with the random variable $C_{(n-1),i}$ defined above we obtain

$$\Pr\{W_n = k | W_{n-1} = l, C_{(n-1),i} = c\} = \begin{cases} \delta(k - l - c) & \text{if } l + c \geq 0; \\ \delta(k) & \text{otherwise.} \end{cases} \tag{11}$$

From this equation we have

$$\Pr\{W_n = k | W_{n-1} = l, S_{n-1} = i\} \tag{12}$$

$$= \sum_{c=-\infty}^{\infty} \Pr\{W_n = k | W_{n-1} = l, C_{(n-1),i}\} P_{C_{(n-1),i}}(c) \tag{13}$$

$$= \begin{cases} P_{C_{(n-1),i}}(k - l) & \text{if } k > 0; \\ \sum_{c=-\infty}^{0} P_{C_{(n-1),i}}(c - l) & \text{if } k = 0. \end{cases} \tag{14}$$

From eq.(10) we have

$$\Pr\{W_n = k, S_n = j | W_{n-1} = l, S_{n-1} = i\} = \begin{cases} p_{ij} P_{C_{(n-1),i}}(k - l) & \text{if } k > 0; \\ \sum_{c=-\infty}^{0} p_{ij} P_{C_{(n-1),i}}(c - l) & \text{if } k = 0. \end{cases} \tag{15}$$

The recursive formula for joint probability density $P_{W_nS_n}(k,j)$ is then obtained

$$P_{W_nS_n}(k,j) = \sum_{l=0}^{\infty} \sum_{i=1}^{m} \Pr\{W_n = k, S_n = j | W_{n-1} = l, S_{n-1} = i\} P_{W_{n-1}S_{n-1}}(l,i) \tag{16}$$

$$= \begin{cases} \sum_{l=0}^{\infty} \sum_{i=1}^{m} p_{ij} P_{C_{(n-1),i}}(k - l) P_{W_{n-1}S_{n-1}}(l,i) & \text{if } k > 0 \\ \sum_{l=0}^{\infty} \sum_{i=1}^{m} \sum_{c=-\infty}^{0} p_{ij} P_{C_{(n-1),i}}(c - l) P_{W_{n-1}S_{n-1}}(l,i) & \text{if } k = 0 \end{cases} \tag{17}$$

or in a compact form which we call the extended Lindley-equation [5]

$$P_{W_nS_n}(k,j) = \sum_{i=1}^{m} p_{ij} \pi_0 [P_{C_{(n-1),i}}(k) * P_{W_{n-1}S_{n-1}}(k,i)] \tag{18}$$

where the convolution operator $*$ is defined

$$x(k) * y(k) := \sum_{l=-\infty}^{\infty} x(k-l)y(l) \tag{19}$$

and π_0 denotes the discrete version of the sweeping up operator:

$$\pi_0(x(k)) = \begin{cases} 0, & k < 0 \\ \sum_{l=-\infty}^{0} x(l), & k = 0 \\ x(k), & k > 0 \end{cases} \tag{20}$$

Since the random variable $C_{(n-1),i}$ is the sum of B_n and $-A_{n-1}|(S_{n-1} = i)$, its probability density is then given by the convolusion:

$$P_{C_{(n-1),i}}(k) = P_{A_{n-1}|(S_{n-1}=i)}(-k) * P_{B_n}(k), \tag{21}$$

where the conditional probability for the interarrival time is given by

$$P_{A_{n-1}|(S_{n-1}=i)}(k) = F_i(k) - F_i(k-1) = P_{A,i}(k) \tag{22}$$

which is actually independent on n. Also $P_{B_n}(k)$ is the service time probability density and is assumed to be time-invariant:

$$P_{B_n}(k) = P_B(k) \tag{23}$$

Finally, we obtain

$$P_{C_{(n-1),i}}(k) = P_{C,i}(k) = P_{A,i}(k) * P_B(k) \tag{24}$$

Eq.(18) is reduced to the Lindley-equation [5] in case of renewal arrival processes, that means if $F_i(k) = F(k)$ or $p_{ij} = P_j$.

5 Numerical algorithm for the stationary waiting time distribution

The stationary waiting time distribution may be solved by taking the limit of eq.(18) , with eq.(24)

$$P_{WS}(k,j) = \sum_{i=1}^{m} p_{ij}\pi_0[P_{C,i}(k) * P_{WS}(k,i)] \tag{25}$$

where $\lim_{n\to\infty} P_{W_nS_n}(k,i) = P_{WS}(k,i)$. The waiting time probability density and the d.f. are given by

$$P_W(k) = \sum_{i=1}^{m} P_{W_nS_n}(k,i) \tag{26}$$

and

$$F_W(k) = \sum_{l=0}^{k} P_W(l) \tag{27}$$

In case of a renewal arrival process the stationary probability density can be very efficiently and accurately solved by using the Cepstrum concept [1], which is unfortunately not straight forward applicable for other SSMP. We apply the directly iterative method whereby the FFT (fast Fourier transform) algorithm is used for efficient computation of the convolutions [1]. The calculation schema for one iteration is represented in Fig.3.

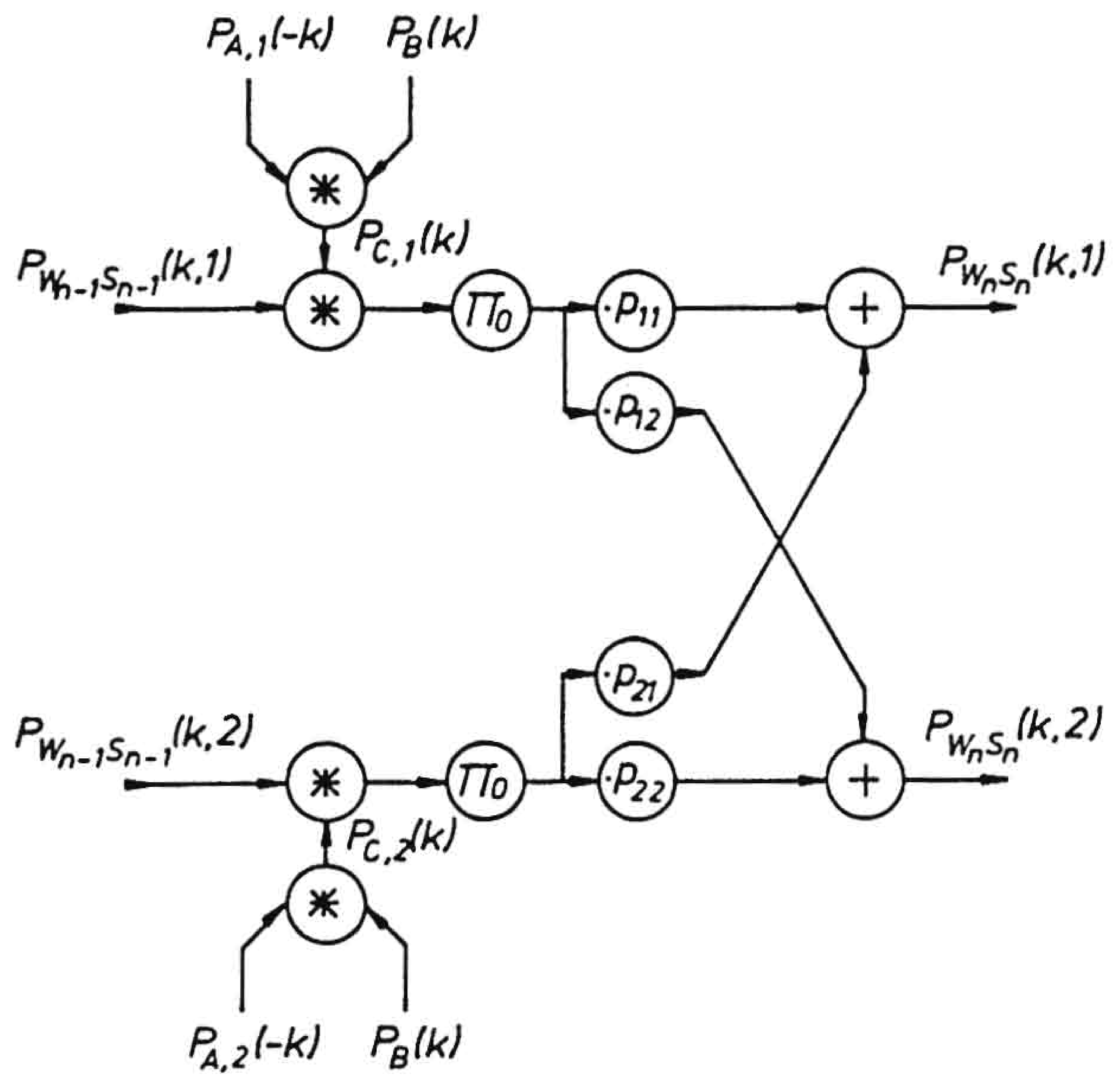

Fig. 3 Calculation schema for the waiting time probability density of a two-state SSMP/G/1 queue.

The necessary number of iterations is clearly dependent on the starting density $P_{W_0S_0}(k,i)$, on the minimal error required, and on the parameters of the queue. So if the load $\rho = \overline{B}/\overline{A}$ ($\overline{A}$: mean interarrival time, $\overline{B}$: mean service time) or the correlation r_h eq.(6) is high, one needs more iterations for the same accuracy. To ensure the error not exceeding a prescribed limit we choose two iteration sets, one with a starting d.f. which is definitely under the stationary d.f. and the other definitely over. The former approaches up to the stationary waiting time distribution, while the latter approaches down to it. This makes possible to give an error measure about the calculated waiting time distribution. A universal, but not efficient choise of the two starting functions could be

$$P_{W_0S_0,u}(k,i) = P_i\ \delta(k), \quad P_{W_0S_0,o}(k,i) = P_i\ \delta(k-K) \tag{28}$$

where K is the maximal distribution length. If information about the waiting time distribution is a priori available, more efficient starting functions could be used.

One error measure can be defined as the maximal relative error of the complementary d.f. $G_W(k) = 1 - F_W(k)$ and given by

$$\Delta = \max\{\frac{|G_{W,o}(k) - G_{W,u}(k)|}{G_W(k)}\} \text{ for } k = 0,1,\cdots,K \text{ and } G_W(k) > G_{\text{min}} \tag{29}$$

where $G_W(k) = 0.5[G_{W,o}(k)+G_{W,u}(k)]$ and G_{min} is the prescribed minimal level of $G_W(k)$. The iteration is stopped, if the error Δ is less than a prescribed value.

To illustrate how this algorithm works, we choose the well-known M/M/1 queue having the load $\rho = 0.7$. However the interarrival time d.f. und the service time d.f. are discretized to enable the use of our method. The results after 10th, 20th, 30th, 40th iterations from both sides and the exact result are shown in Fig.4.

6 Correlation effect on queues

We now give a numerical example of the two-state SSMP/D/1 queue where the marginal distribution of the interarrival time is kept fixed for a given load ρ. The service time has a normalized constant length of 1. The interarrival time in state 1 or 2 is geometric or binomial distributed, respectively:

$$P_{A,1}(k) = (1-q_1)q_1^k, \quad k = 0, 1, \cdots \tag{30}$$

$$P_{A,2}(k) = \binom{\varepsilon}{k}(1-q_2)^k q_2^{\varepsilon-k}, \quad k = 0, 1, \cdots, \varepsilon \tag{31}$$

Following parameters are chosen to be fixed

$$\varepsilon = 32, \quad P_1 = 0.9, \quad P_2 = 0.1.$$

For a given load ρ the parameters q_1 and q_2 are chosen under the *balanced mean* condition $P_1\overline{A}_1 = P_2\overline{A}_2$ and given by

$$q_1 = \frac{1}{1+2\rho P_1}, \quad q_2 = 1 - \frac{1}{2\rho P_2 \varepsilon} \tag{32}$$

For a given ρ the only parameter which can be varied is the correlation measure κ eq.(6). The mean waiting time depending on κ is shown in Fig.5. We can conclude from our results that the positive correlation have negative effect on the queueing performance. This negative effect increases for increasing correlation and increasing load. Thus when designing an ATM system the correlation of input processes should be generally taken into account and the renewal assumption or renewal approximation of these processes must be very carefully examined.

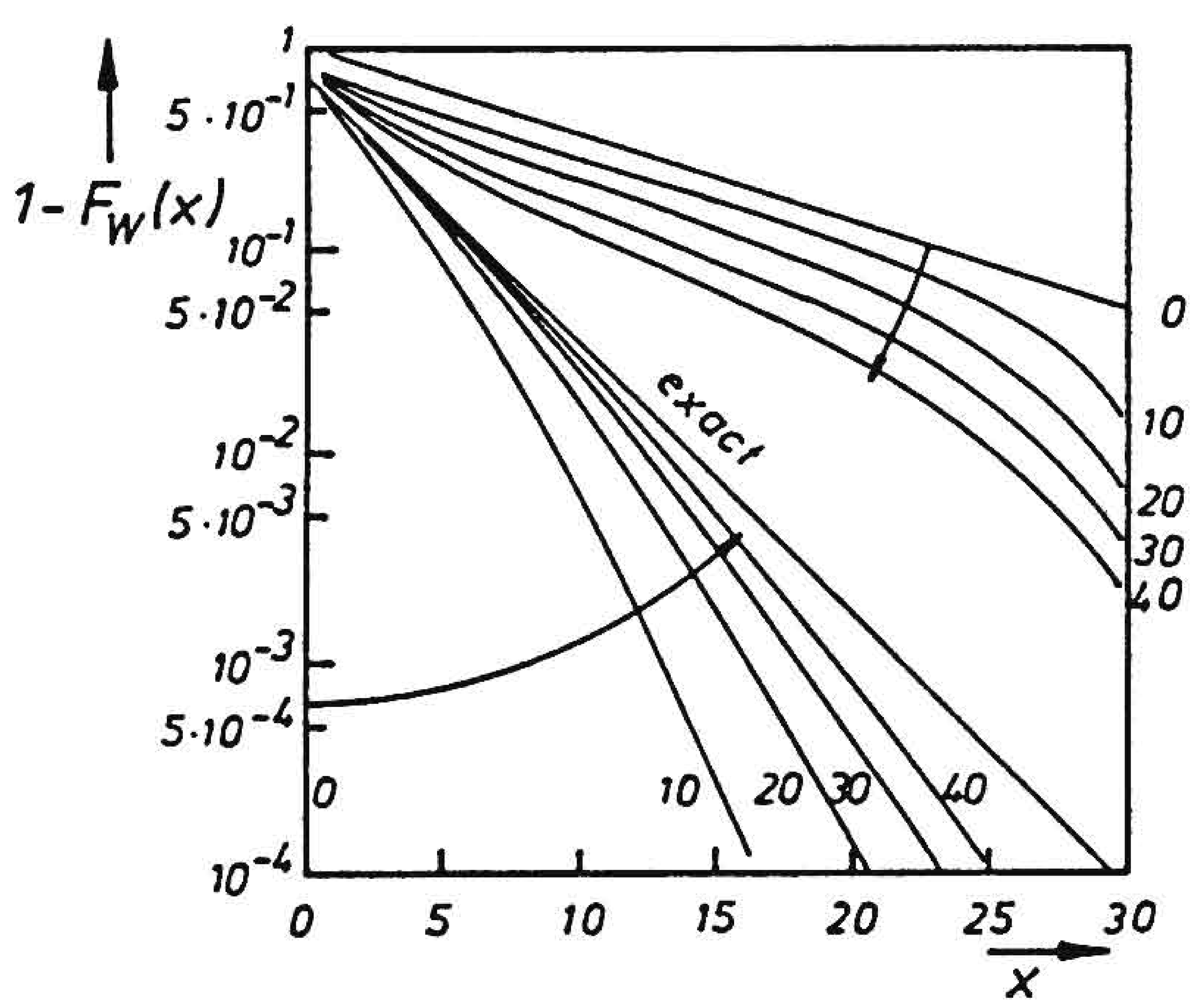

Fig. 4 The compl. waiting distributions of the discretized M/M/1 queue after 10th, 20th, 30th, 40th iterations from both sides.

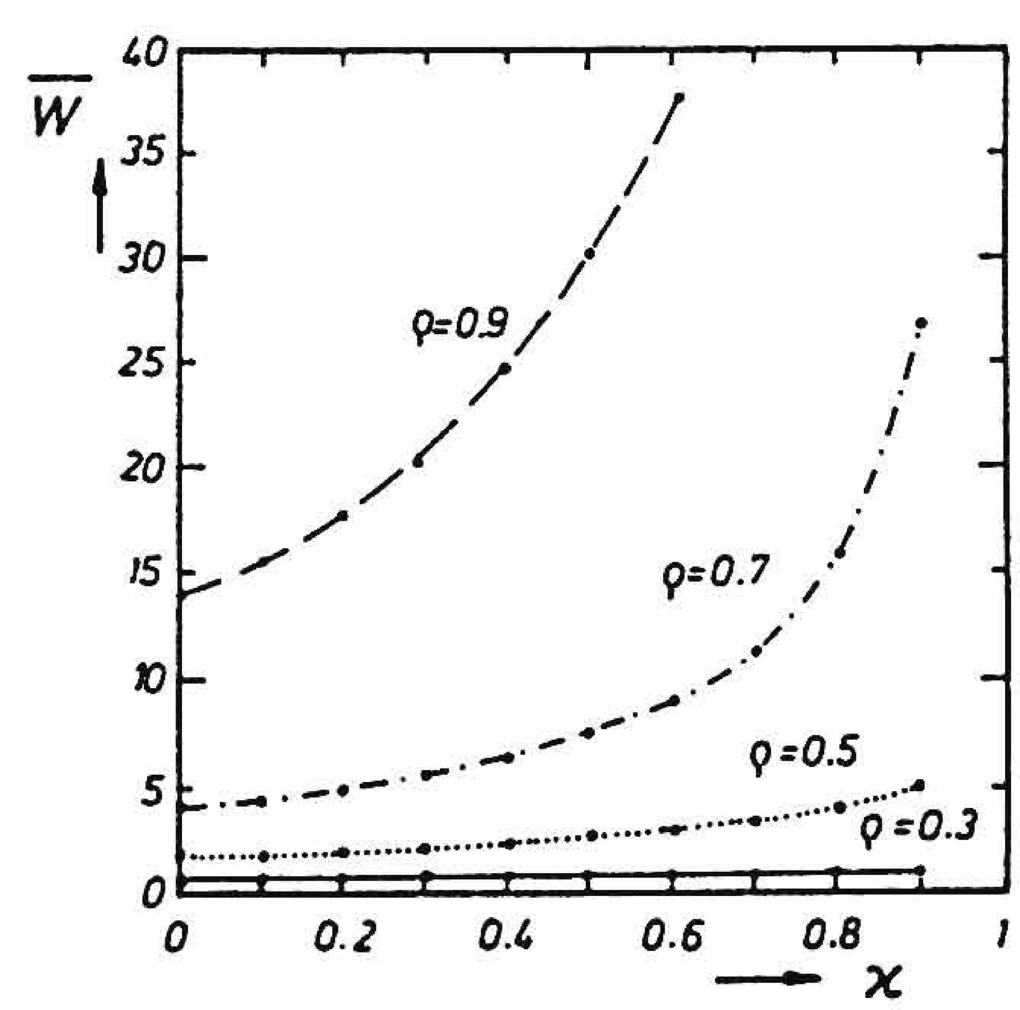

Fig. 5 Mean waiting $\overline{W}$ of a two-state SSMP/D/1 queue in dependence of correlation κ eq.(6).

7 Final remarks

This paper presents first queueing results of the SSMP model. Future research will be done to solve some loss systems in which the assumption of an infinite queue size can not be made. Formula very close to the extended Lindley equation for these systems can be obtained and the lost probability can be calculated, which is extremely important for some data types in B-ISDN. Lost probability is discussed by Tran-Gia and Ahmadi for renewal batch arrival [11].

Acknowledgement

The authors thank Dr. R. Perera for helpful discussions and Dr. K. Meier-Hellstern for literature hints.

References

[1] Ackroyd, M.H., *Computing the waiting time distribution for the G/G/1 queue by signal processing methods.* IEEE Trans. Commun. **28** (1980), 52-58.

[2] Arjas, E., *On the use of a fundamental identity in the theory of semi-Markov queues.* Adv. Appl. Prob. 4 (1972), 271-284.

[3] Çinlar E., *Queues with semi-Markovian arrivals.* J. Appl. Prob. 4 (1967), 365-379.

[4] Decker P., *Time-discrete numerical methods for a SM/G/1 queue (in german).* Diploma thesis at the Institute of Teleprocessing, Aachen University of Technology, Aachen, 1990.

[5] Lindley, D.V., *The theory of queues with a single server.* Proc. of the Cambridge Philosophical Society **48** (1952), 277-289.

[6] Lucantoni, D.M., *New results on the single server queue with a batch Markovian arrival process.* To appear in Stochastic Models, 1991.

[7] Neuts, M. F., *Markov chains with applications in queueing theory, which have a matrix-geometric invariant probability vector.* Adv. Appl. Prob. **10** (1978), 185-212.

[8] Neuts, M. F., *Matrix-geometric solutions in stochastic models.* The Johns Hopkins University Press, Baltimore, 1981.

[9] Pyke, R., *Markov renewal processes: definitions and preliminary properties.* Ann. Math. Stat. **32** (1961), 1231-1242.

[10] Ramaswami, V., *The N/G/1 queue and its detailed analysis.* Adv. Appl. Prob. **12** (1980), 222-261.

[11] Tran-Gia, P., Ahmadi, H., *Analysis of a discrete-time* $G^{[X]}/D/1\text{-}S$ *queueing system with applications in packet-switching systems.* Proc. INFOCOM '88, New Orleans, Mississipi, March 27-31, 1988, Paper 9A.1.