Scene-Based Characterization of VBR MPEG-Compressed Video Traffic*

Marwan Krunz† and Satish K. Tripathi
Institute for Advanced Computer Studies
Department of Computer Science
University of Maryland
College Park, MD 20742

Abstract

We present a comprehensive model for variable-bit-rate MPEG-coded video streams with I, P, and B frames. The model captures the bit-rate variations at multiple time scales. Long-term variations are captured by incorporating scene changes, which are most noticeable in the fluctuations of I frames. The size of an I frame is modeled by the sum of two random components: a scene-related component that reflects the overall level of scene activity and an AR(2) component that accounts for the fluctuations in I frames within a scene. Both components are generated by two autocorrelated random processes. The correlation structure for the modeled I subsequence is analytically derived. Two other random processes are used to model the sizes of P and B frames, respectively. The complete source model is then obtained by intermixing the three subsequences according to a given compression pattern. We derive the autocorrelation function for the composite process. The form of this function indicates that the model exhibits long-range dependence as it has an infinite sum. When using a repetitive compression pattern, the overall correlation structure is mainly dictated by the cross-correlations between the sizes of different types of frames, rather than the autocorrelations within each frame-type subsequence. Using standard statistical methods, we successfully fit the model to several empirical video traces. Synthetic and real traces are used to study the queueing performance for video streams at a statistical multiplexer. The results show that the model is sufficiently accurate in predicting the queueing performance for real video streams.

keywords: video modeling, MPEG, VBR, fitting, time series.

---

* This research was supported by the NSF grant # 9318933 and by a Shared University Research IBM grant.
†Corresponding author. Tel. (301) 405-2693. Email: krunz@cs.umd.edu.

1 Introduction

A BISDN/ATM network will transport a variety of streams that have different traffic patterns and performance requirements. Accurate characterization of these streams is essential for some of the services offered by BISDN (e.g., VBR service). By appropriately modeling the traffic, it is possible to dimension network resources while simultaneously satisfying the QoS requirements. Compared to other types of traffic, compressed video exhibits a very sophisticated behavior that is difficult to characterize. A video encoder generates a sequence of compressed frames whose sizes (in bits) vary depending on the activity of the scene as well as the compression technique. For constant-quality video the frame rate is fixed, and characterizing the variability of the bit rate amounts to modeling the frame size sequence.

In this paper, we investigate the modeling of MPEG-coded video movies. Many models were previously proposed for variable bit rate (VBR) video under various compression schemes [1, 3, 4, 6, 7, 8, 10, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25]. Since the VBR behavior of a video stream strongly depends on the compression technique, many of these models do not characterize MPEG-coded video. Moreover, some of them were specifically developed for low-motion video, including video-phone [15] and video-teleconferencing [6], and thus not applicable to video movies. Following the standardization of the MPEG algorithm and its wide acceptance, researchers started investigating the characteristics of MPEG-coded video streams [8, 25, 20, 7, 22]. The focus at first was on MPEG streams without interpolative (B) frames [8, 25, 20], which are believed to incur large encoding and decoding delays. Very few studies addressed the modeling of MPEG streams with all types of frames [7, 5]. In [7] Huang et al. proposed a self-similar fractal model for MPEG video, which captures the correlation structure and frame-size distribution for the three types of MPEG frames. While matching both statistics warrants a good approximation of the second-order properties, it may not be sufficient to capture the multiple-time-scale variations observed in real traces. Heyman et al. [5] attempted to provide appropriate fittings for the size of each type of MPEG frames, and suggested that additional work is still needed to model P and B frames.

Our work is being motivated by the need for a comprehensive characterization of MPEG-coded video streams that explicitly incorporates the multiple-time-scale variations of the bit rate. The proposed model can be used to generate synthetic MPEG streams to be used in many performance studies, including buffer dimensioning and bandwidth allocation at video servers and network nodes. Generating synthetic streams has two obvious advantages. First, performance studies can be carried out without requiring the actual traces (which may not be available). Second, a stochastic model encompasses many realizations (sample paths) which represent “structurally” similar but not identical synthetic streams. Hence, these streams are ideal for statistical multiplexing studies (a single empirical trace is hardly appropriate for generating multiple streams that have the same stochastic structure, but are not identical). Our model is comprehensive in the sense that it characterizes an MPEG stream that contains I, P, and B frames. In addition to approximating the marginal distribution and autocorrelation structure of the empirical data, the model captures the impact of scene changes on the bit rate variations.

The idea of incorporating scene changes in a traffic model is not particularly new. In the DAR(1) model, frame sizes are generated according to a finite-state Markov chain, where each state corresponds to a given frame size. Once the chain enters a given state, it stays in that state for an exponentially distributed period, which can be associated with a scene duration. Video frames within a scene have the same size. In [3] Frater et al. modified the DAR(1) model by using a different distribution for scene durations, which resulted in a more accurate prediction of the queueing performance. However, similar to the DAR(1) model, the "scenic" model in [3] does not incorporate the variations of frame sizes within scenes. Moreover, both models were developed for JPEG-like streams (i.e., MPEG streams with I frames only), whose empirical autocorrelation function decreases almost monotonically. Scene effects were also incorporated in [14] for the modeling of JPEG-like VBR video based on the TES approach (video was coded without motion compensation). For MPEG streams that contain I, P, and B frames, the empirical autocorrelation function has a complicated structure with multiple pseudo-periodic components resulting from the periodicity of the compression pattern and the large differences in the sizes of different types of frames. In [11] we examined the empirical statistics for I, P, and B frames. Some of our previous results will be used here to provide a complete model for MPEG video streams. Whereas conventional modeling studies were based on one or two empirical traces, our model was successfully fitted to four different traces. For brevity, we describe the fitting procedure for two of these traces, and provide a summary of the fitting results for all four traces.

The paper in structured as follows. In Section 2 we briefly describe the empirical sequences used in this study. In Section 3 we present the source model. Section 4 presents an algorithm for generating synthetic streams. Trace-driven simulations using both real and synthetic streams are conducted in Section 5. These simulations are intended to evaluate the model's accuracy in predicting the queueing performance for real streams and also to study the performance for multiplexed streams. The paper is concluded in Section 6.

2 Empirical MPEG Sequences

A standard MPEG encoder generates three types of compressed frames: I, P, and B. I frames are compressed using intraframe coding only, while the generation of P and B frames involve, in addition to intraframe coding, the use of motion compensation techniques (prediction techniques for P frames, and both prediction and interpolation techniques for B frames). As a result, I frames are generally the largest in size, followed by P frames, and finally B frames.

Four public domain traces were used in our study, which are listed in Table 1. These traces were captured by several researchers [4, 23, 11] using MPEG-I-compliant encoders (see references for compression details). The rationale for using several traces is to demonstrate the applicability of the model to compressed videos with diverse content. For brevity, we will describe the fitting procedure based on the first two traces, and summarize the fitting outcome for all traces in Table 4. Note that the first two traces differ significantly in their scene dynamics, with *Wizard of Oz* (Oz, for short) exhibiting less dynamics than *Star Wars*. Representative samples from the two traces are shown in Figure 1.

Trace	Length (frames)	Compression Pattern (N, M)
Star Wars [4]	174066	$(12, 3)$
Wizard of Oz [11]	41760	$(15, 3)$
Silence of the Lambs [23]	40000	$(12, 3)$
James Bond: Goldfinger [23]	40000	$(12, 3)$

Table 1: Empirical traces used in constructing the traffic model.

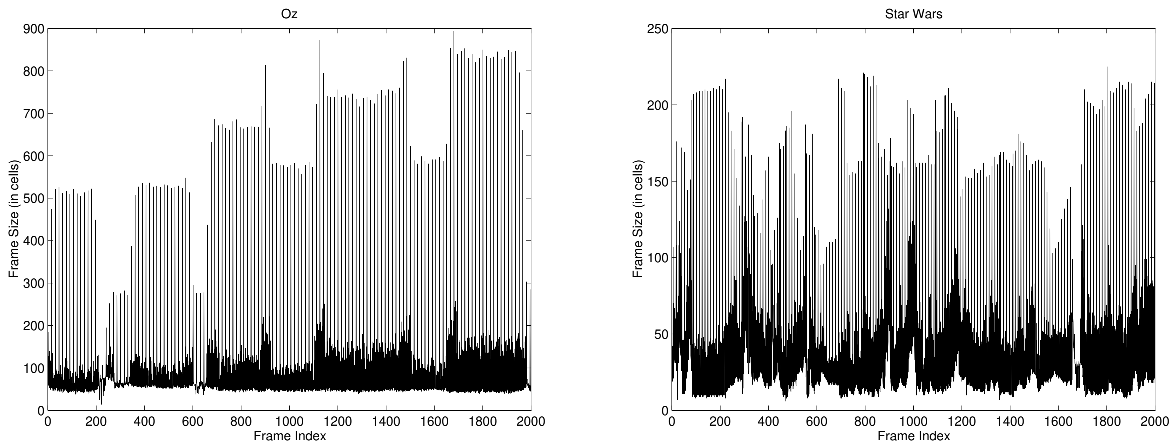


Figure 1: Representative samples from the empirical traces.

An important feature of typical MPEG encoders (both hardware and software) is the manner in which frame types are generated. Although not required by the MPEG standards, many available MPEG encoders, particularly those that target constant-quality compression, use the same Groupof-Pictures (GOP) pattern repeatedly in compressing all frames of a given video sequence. The GOP pattern specifies the number and temporal order of P and B frames between two successive I frames. Quite often, video streams are compressed using "regular" GOP patterns where the number of B frames between two reference frames (I or P) is fixed. In such cases, the GOP pattern is simply characterized by two parameters: the frame distance from an I frame to the next I frame (N), and the frame distance from an I frame to the following P frame (M). We will limit our attention to video sequences with "regular" GOP patterns. It should be emphasized that MPEG-compliant sequences *need not* use a "regular" GOP pattern, and not even a repetitive one. However, since many sequences are being coded with regular GOP patterns (often to simplify the codec design), developing a traffic model for such sequences has its merits.

Since the coding of B frames involve non-causal prediction, before transmitting a video sequence over the network, the compressed frames are rearranged according to the order of their decompression. Figure 2 shows an example of the temporal (i.e., display) and transmission orders of an MPEG sequence. Except for the first GOP, the transmission order appears as if it consists of replications of the GOP pattern (although the dependency structure is now different). Thus, the impact of the repetitive GOP pattern on the correlation structure of the VBR sequence is expected to exist in the transmitted sequence as well.

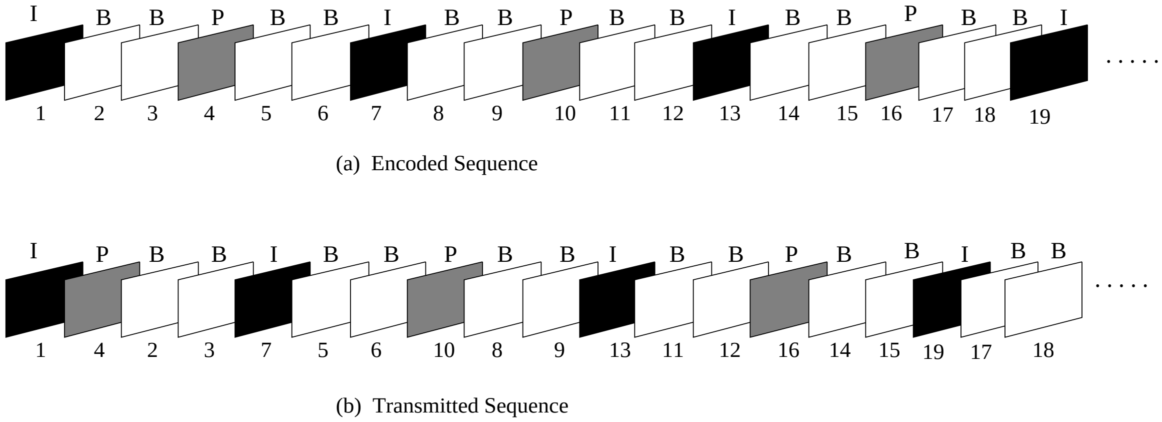
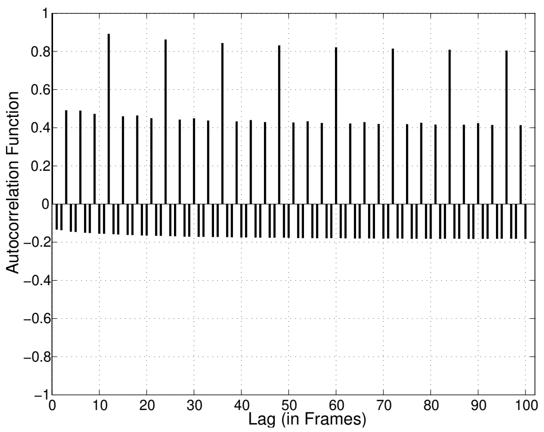
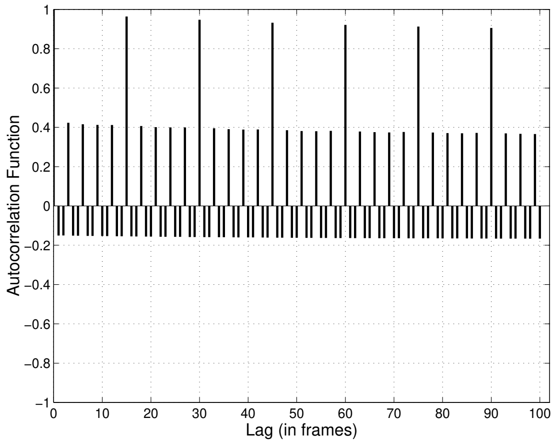


Figure 2: Encoding and transmission orders of an MPEG sequence.

The autocorrelation functions (ACF) for two traces are shown in Figure 3. The ACF exhibits a complicated structure that is strongly influenced by the form of the compression pattern. Two apparent periodic components are observed in the ACF: one at lags of kN and the other at lags of kM, where $k = 0, 1, 2, \ldots$. The latter component is mainly caused by correlations between P frames. Both components decay very slowly. Although the scene dynamics of a movie has certain impact on the correlation structure, it is the compression pattern that dictates the general shape of the ACF.



(a)



(b)

Figure 3: Autocorrelations in an MPEG-coded sequence: (a) *Star Wars* (b) *Wizard of Oz*.

3 Modeling a VBR MPEG Stream

3.1 Scene Length Distribution

It can be observed from Figure 1 that the fluctuations of I frames exhibit different behavior at different time scales (this is particularly obvious in *Oz* trace). Within small locality, I frames have relatively small fluctuations about some average level, which itself varies at a larger time scale. Such a behavior can be captured using a mean-nonstationary time series model. Alternatively, it can be captured by a modulated process in which the randomness occurs at multiple time scales. We opt for the latter approach, which has the advantage of rendering the model stationary.

The fluctuations in the overall level of the bit rate, which occur at the larger time scale, is often attributed to "scene" changes [14, 3]. A scene, in the visual sense, is loosely defined as a portion of the movie without sudden changes in view, but with some panning and zooming [14]. Incorporating a "scenic" component in a traffic model gives the model's predictions a physical meaning, making the synthetic realizations look more like real ones. Note that such a component need not correspond exactly to visual scenes, since what matters here is the VBR fluctuations rather than the visual description. In fact, it is impractical to base the model on visually defined scenes, which are difficult to measure.

Providing a scene-related component requires a functional definition of scene durations based on the bit rate variations. For this purpose, We use a heuristic [1] that is based on the fact that a 'significant' change in the sizes of two consecutive I frames is a strong indication of the start of a new scene. Thus, one can use the sizes of consecutive I frames in a trace to obtain the scene length

[1]A slightly different approach was used in [14] to determine scene boundaries in JPEG-based movies.

data. Let $\{\widehat{X}_I(n) : n = 1, 2, ...\}$ be the I frames subsequence in an empirical trace (throughout the paper, a '^' above a variable name indicates an empirical quantity). Suppose the kth I frame is the starting frame for the jth scene. Then the $(n+k+1)$th I frame indicates the start of the next scene if

$$\frac{\left|\widehat{X}_I(n+k+1) - \widehat{X}_I(n+k)\right|}{\left(\sum_{j=k}^{n+k} \widehat{X}_I(j)\right)/n} > T_1 \tag{1}$$

$$\text{and} \quad \frac{\left|\widehat{X}_I(n+k+2) - \widehat{X}_I(n+k)\right|}{\left(\sum_{j=k}^{n+k} \widehat{X}_I(j)\right)/n} > T_2 \tag{2}$$

where T_1 and T_2 are two thresholds. Both conditions must be satisfied to mark the start of a new scene. The first condition states that the examined I frame differs from the previous I frame by at least T_1 times the current mean of I frames in the underlying scene. This alone is not enough to provide a good mapping between the visual scenes and the computed ones, since within some visual scenes the bit rate increases abruptly for a fraction of a second and then returns to its normal level. Such a sudden increase is caused by flickering in the picture, and is not related to scene changes. The second condition is thus needed to ensure that the change in the bit rate is persistent. Although visual scenes could start at any time, determining scenes using I frames only has minor impact on the fitting accuracy since (as explained below) a scene lasts on average for several seconds, whereas the discrepancy in the starting time of a visual scene and the computed one is no more than a GOP period (1/2 sec).

It is clear that the thresholds T_1 and T_2 impact the correlations between the *computed* scenes; the larger their values the less correlated the scenes. We found that a very good one-to-one mapping between the visual scenes and the computed ones is obtained with $T_1 = 0.1$ and $T_2 = 0.15$. Within some tolerance about these values, the mapping is still sufficiently acceptable. Figure 4 shows a sample of the I frame subsequence in *Star Wars* and the corresponding computed scenes. Previous studies [9, 14] indicated that scene length can be modeled by an exponential (or geometric) distribution. The scene length histogram (not shown here) confirmed such result. Hence, we model the scene length by a geometric distribution with parameter q ($q < 1$). The simplicity of the geometric distribution allows us to derive tractable analytical results for the model. Let N_j be the length of the jth scene ($j = 1, 2, \ldots$). Then,

$$\Pr\{N_j = n\} = q^{n-1}(1-q), \text{ for } n = 1, 2, \ldots \tag{3}$$

The mean scene length is $E[N_j] = 1/(1-q)$. Sample statistics for the computed scene lengths are

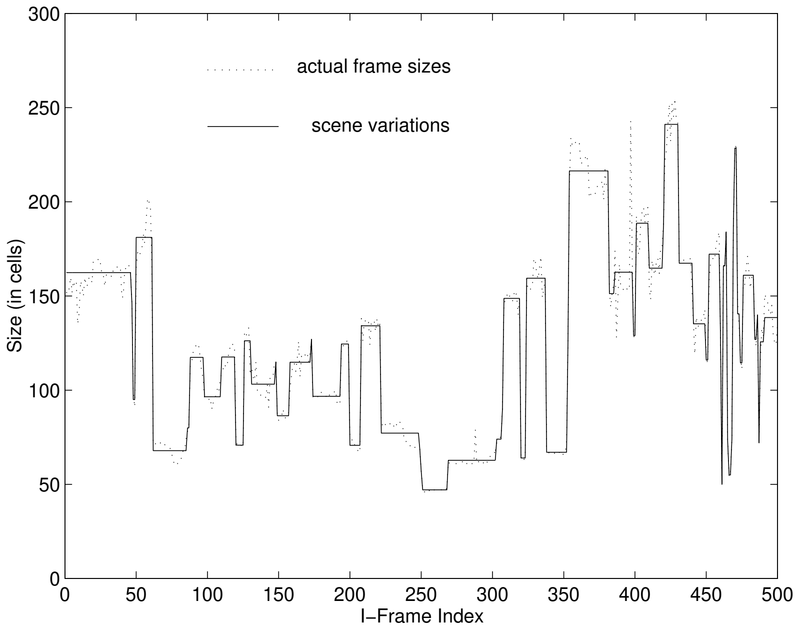


Figure 4: Segment of the I subsequence and the computed scenes (*Star Wars* trace).

given in Table 2. Note that *Oz* has a greater average scene length ($\sim$ 7 seconds) than that of *Star Wars* ($\sim$ 3 seconds). The average scene length can be used as a one-parameter indication of the overall level of activity of a movie (the higher the action of the movie, the smaller the average scene length), and movies can be classified according to this parameter.

Trace	Number of Computed Scenes	Scene Length Statistic (in consecutive I frames)			
		Mean	Std. Deviation	Max.	Min.
Oz	199	14.0 (7 sec)	17.3	90	1
Star Wars	2452	5.92 (2.96 sec)	7.85	78	1

Table 2: Sample statistics for the scene length in the two traces ($T_1 = 0.1$, $T_2 = 0.15$).

We tested the hypothesis of scene lengths independence using the runs-up test [13] and found that at a 95% level of confidence, this hypothesis cannot be rejected. Hence, we assume that $\{N_j\}$ are *i.i.d.* with common geometric distribution.

3.2 Frame Size Distributions

Two important factors determine the size of an MPEG frame: the scene activity and the frame type. The effect of the scene activity occurs at a time scale of few seconds (the average scene length). Frame types are determined by the compression pattern, and hence, their impact on the bit rate occur at a smaller time scale (e.g., 1/30 sec). To capture the impact of the frame type, we use a composite model which consists of three random processes. Each process generates the sizes of one type of frames. The three processes are intermixed according to a given compression pattern to produce the overall frame size sequence. As we demonstrate later, the periodic nature of the compression pattern ensures that strong cross-correlations exist between frames of different types.

Although the effect of the scene activity exists in all types of frames, this effect is particularly obvious (and thus easier to capture) in the case of I frames. Because I frames are intra-coded, their sizes are (functionally) independent of the sizes of P and B frames. On the other hand, the size of a P or a B frame is the result of a complex interaction between scene activity and the sizes of other reference frames (I or P). In our model, we assume that scene activity affects the sizes of I frames only. Since I frames are the largest, incorporating the scene activity in their sizes results in capturing most of the actual impact of scene activity.

3.2.1 I Frames

By definition (see (1) and (2)), the sizes of I frames *within* a scene fluctuate slightly about some average value that represents the level of scene activity. Let $X_I(n)$ be the size of the nth I frame. We model $X_I(n)$ by the sum of two independent random variables:

$$X_I(n) = X_I^*(n) + \Delta_I(n) \tag{4}$$

where $X_I^*(n)$ reflects the level of activity of the scene, while $\Delta_I(n)$ represents the fluctuation of the nth I frame about $X_I^*(n)$.

First, we consider $X_I^*(n)$. This quantity is constant for all I frames in the same scene, and it varies from one scene to another. Hence, for the jth scene that started with the kth I frame:

$$X_I^*(k) = X_I^*(k+1) = \cdots = X_I^*(k+N_j-1) \stackrel{\triangle}{=} \widetilde{X}_I(j) \tag{5}$$

We assume $\{\widetilde{X}_I(j) : j = 1, 2, \ldots\}$ are *i.i.d.* with common *pdf* $f_{\widetilde{X}_I}$. This assumption was justified by the weak presence of correlations in the empirical sequence $\{\widehat{\widetilde{X}}_I(j)\}$, where $\widehat{\widetilde{X}}_I(j)$ is the sample average of the sizes of I frames in the jth scene of a given empirical trace. In [11] we tested three fits for the frame size distribution: Gamma, Weibull, and lognormal. We found the lognormal fit to be the most appropriate. Accordingly, we assume that $\widetilde{X}_I(j)$ is lognormally distributed with parameters $\widetilde{\mu}_I$ and $\widetilde{\sigma}_I$:

$$f_{\widetilde{X}_I}(u) = \begin{cases} \frac{1}{u\sqrt{2\pi\widetilde{\sigma}_I^2}} \exp\left[\frac{-(\ln u - \widetilde{\mu}_I)^2}{2\widetilde{\sigma}_I^2}\right], & u > 0 \\ 0, & otherwise \end{cases} \tag{6}$$

The mean ($m_{\widetilde{X}_I}$) and the variance ($v_{\widetilde{X}_I}$) of $\widetilde{X}_I$ are given by:

$$m_{\widetilde{X}_I} = \exp\left\{\widetilde{\mu}_I + \widetilde{\sigma}_I^2/2\right\} \tag{7}$$

$$v_{\widetilde{X}_I} = \exp\left\{2\widetilde{\mu}_I + \widetilde{\sigma}_I^2\right\}\left(\exp\left\{\widetilde{\sigma}_I^2\right\} - 1\right) \tag{8}$$

$\widetilde{\mu}_I$ and $\widetilde{\sigma}_I$ are estimated by their Maximum Likelihood Estimates (MLE) given in Table 4.

The following proposition gives the probability distribution for $X_I^*(n)$.

Proposition 1 *Let $f_{X_I^*}^{(n)}$ be the pdf for $X_I^*(n)$. Then*

$$f_{X_I^*}^{(n)} = f_{\widetilde{X}_I}, \textit{ for all } n. \tag{9}$$

Thus, $\{X_I^*(n)\}$ is a correlated identically distributed random sequence.

Proof: (by induction). It is clear that $f_{X_I^*}^{(1)} = f_{\widetilde{X}_I}$. Suppose that $f_{X_I^*}^{(n)} = f_{\widetilde{X}_I}$. The probability that the nth and the $(n+1)$th I frames are in the same scene is q. In this case, $X_I^*(n) = X_I^*(n+1)$, which implies that the two random variables are equal in distribution. With probability $1-q$, the $(n+1)$th I frame is the starting frame of a new scene, in which case $f_{X_I^*}^{(n+1)} = f_{\widetilde{X}_I}$ which is also equal to $f_{X_I^*}^{(n)}$. Thus, $f_{X_I^*}^{(n)} = f_{X_I^*}^{(n+1)}$. By induction on n, the assertion is true. □

Next, we consider $\Delta_I(n)$ which represents the variation of the size of the nth I frame about $X_I^*(n)$. For a given trace, we compute the empirical sequence $\{\widehat{\Delta}_I(n)\}$, where $\widehat{\Delta}_I(n) = \widehat{X}_I(n) - \widehat{\widehat{X}}_I(j_n)$, j_n being the scene index to which the kth I frame belongs. The sequence $\{\widehat{\Delta}_I(n)\}$ is then fitted by the stochastic process $\{\Delta_I(n)\}$. Notice that $\{\Delta_I(n)\}$ was assumed to be independent of $\{X_I^*(n)\}$, and thus invariant to scene changes.

The sample ACF for $\{\widehat{\Delta}_I(n)\}$ in each trace is shown in the LHS of Figures 5 and 6. Its shape suggests considering an autoregressive (AR) fit for $\{\Delta_I(n)\}$. In fitting an empirical sequence by an autoregressive model, the empirical partial autocorrelation function (PACF) is used to determine the number of autoregressive coefficients (i.e., the order of the autoregressive model) that are adequate for the fitting [2]. The sample PACF for the sequence $\{\widehat{\Delta}_I(n)\}$ in each trace (RHS of Figures 5-6) indicates non-negligible first and second autoregressive coefficients, suggesting an AR(2) fit. Accordingly, we characterize $\{\Delta_I(n)\}$ by an AR(2) process:

$$\Delta_I(n) = a_1\Delta_I(n-1) + a_2\Delta_I(n-2) + \varepsilon(n) \tag{10}$$

where $\{\varepsilon(n)\}$ is a sequence of *i.i.d.* random variables with mean zero and variance σ_ε^2. To determine an appropriate distribution for the residuals, let f_ε and f_Δ be the *pdf* for $\varepsilon(n)$ and $\Delta_I(n)$, respectively (note that $\{\Delta_I(n)\}$ is a stationary process). The shape of f_Δ can be inferred from the histogram of $\{\widehat{\Delta}_I(n)\}$, which is shown in Figure 7 for *Oz* trace (similar shape is observed in *Star Wars* trace). A normal fit seems quite appropriate. Hence, we let $f_\Delta = N(0, \sigma_\Delta^2)$. A normal distribution for $\{\Delta_I(n)\}$ arises naturally when the residuals are also normally distributed. Hence, we let $f_\varepsilon = N(0, \sigma_\varepsilon^2)$.

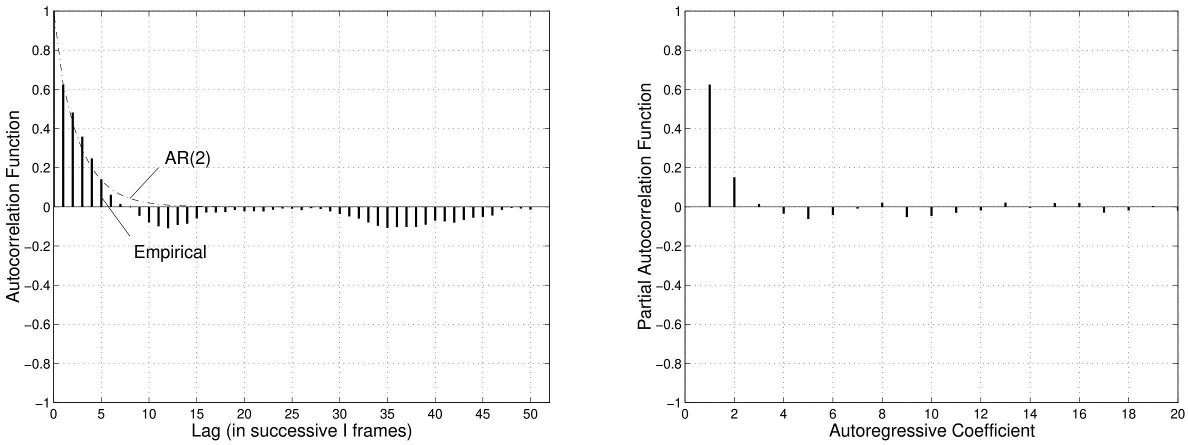


Figure 5: Autocorrelation and partial autocorrelation functions for $\{\widehat{\Delta}_I(n)\}$ in *Oz* trace.

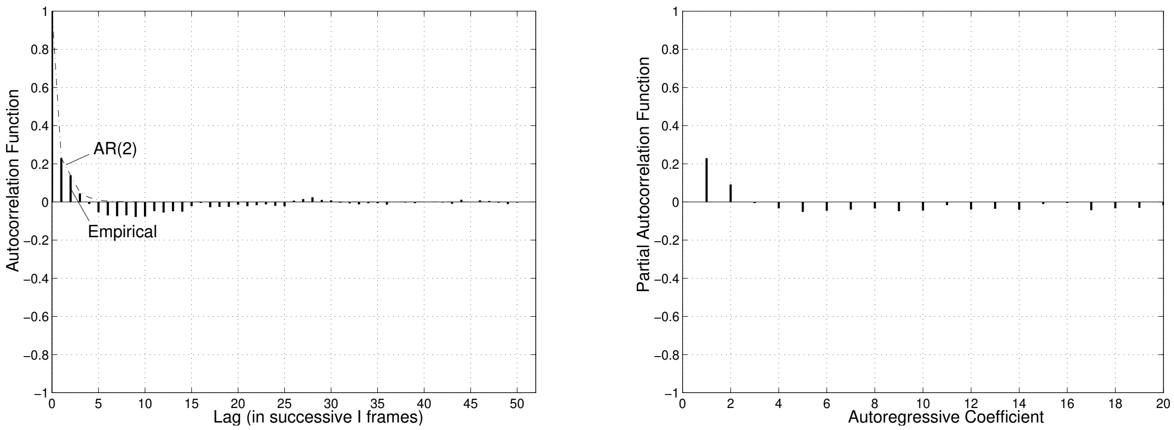


Figure 6: Autocorrelation and partial autocorrelation functions for $\{\widehat{\Delta}_I(n)\}$ in *Star Wars* trace.

At this point, it is instructive to briefly compare our model for I frames to other similar models that were previously proposed. In [6] an AR(2) model was suggested for a video-teleconferencing stream generated from an H.261-compliant encoder. Video-conferencing exhibits much less scene dynamics than movies, thus only the AR(2) component is needed. Similar to our model, a video sequence in the DAR(1) model [9] consists of several scenes with exponentially distributed *i.i.d.* durations. A negative binomial distribution, and later a discretized lognormal distribution [5], was suggested for the average frame size in the DAR(1) model. Fluctuations within a scene are not incorporated. The "scenic" model of Frater et al. [3] is similar to the DAR(1) model except for scene durations which have a *pdf* of the form $p(x) = a/(x^n + b^2)$ ($n \approx 2$ for *Star Wars*). It was argued that such a *pdf* results in a better fitting for the autocorrelation function. Although empirically this might be true, the use of such a distribution renders the underlying model nonstationary, thus preventing any tractable analysis. Variations within scenes were not accounted for in the "scenic" model. Our belief is that detailed characterization of scene durations is important when considering a sequence of I frames only. When I frames are intermixed with P and B frames, the significant differences between the sizes of different types of frames along with the periodicity of the GOP pattern are the major factors that impact the correlation structure. Our model for I frames can be regarded as a generalization of both the AR(2) model for video-teleconferencing and the DAR(1) model. Table 3 summarizes the scene-related differences between our model for I frames and other models.

Feature	Ours	DAR(1)	Frater et al. model [3]	Video-teleconferencing model [6]
Compression	MPEG	H.261 and JPEG	JPEG	H.261
Scene durations	geometric	exponential	$a/(x^n + b^2)$	—
Intra-scene variations	AR(2)	—	—	AR(2)

Table 3: Scene-based time series models for I frames.

Estimates of a_1 and a_2 in (10) can be obtained using Yule-Walker equations [2]:

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{11}$$

where r_k is the sample autocorrelation in $\{\widetilde{\Delta}_I(n)\}$ at lag k. These estimates are given in Table 4. Using these estimates, we fitted the sequence $\{\widetilde{\Delta}_I(n)\}$ in each trace to an AR(2) model and computed the experimental residuals. The ACF for the residuals in the *Oz* trace is shown in Figure 8. The absence of correlations among the residuals is a strong evidence of the appropriateness of our model (similar behavior was also observed in other traces). An estimate of σ_ε^2 is given by the the sum-of-squares of the experimental residuals divided by the number of I frames. The variance of $\Delta_I(n)$ is given by:

$$\sigma_\Delta^2 = \frac{(1-a_2)\sigma_\varepsilon^2}{(1+a_2)\left[(1-a_2)^2 - a_1^2\right]} \tag{12}$$

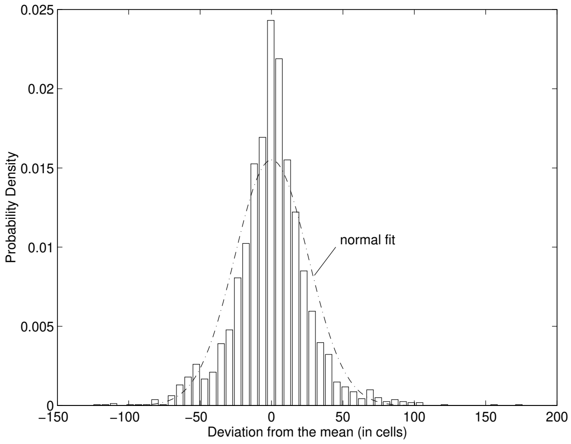


Figure 7: Histogram for the fitted sequence $\{\widehat{\Delta}_I(n)\}$ in *Oz* trace.

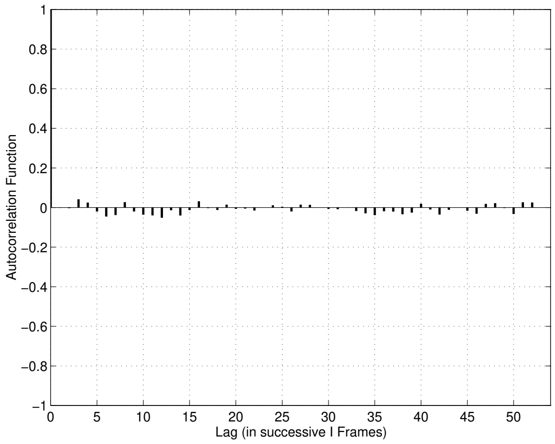


Figure 8: ACF for the empirical residuals of the AR(2) fit (*Oz* trace).

Let κ_m be the autocorrelation in $\{\Delta_I(n)\}$ at lag m ($m > 0$). Then [2],

$$\kappa_m = \frac{A(1-B^2)A^m - B(1-A^2)B^m}{(A-B)(1+AB)} \tag{13}$$

where A and B are the roots of the characteristic equation $x^2 - a_1 x - a_2 = 0$. In all the empirical traces that we examined (including *Oz* and *Star Wars*), the roots were always real, implying that κ_m is a mixture of damped exponentials. Real roots occur when $a_1^2 + 4a_2 \geq 0$.

Since $\{X_I^*(n)\}$ and $\{\Delta_I(n)\}$ are stationary and independent of each other, $\{X_I(n)\}$ is also a stationary process. The random variables $X_I(1), X_I(2), \ldots$, are identically distributed with common *pdf* f_{X_I} that is given by the convolution of $f_{\widetilde{X}_I}$ and f_Δ.

Proposition 2 *Let $\rho_I(m)$ denote the autocorrelation in $\{X_I(n)\}$ at lag m. Then,*

$$\rho_I(m) = \frac{q^m v_{\widetilde{X}_I} + \kappa_m \sigma_\Delta^2}{v_{\widetilde{X}_I} + \sigma_\Delta^2}. \tag{14}$$

Proof: For $m > 0$,

$$\rho_I(m) \triangleq \frac{E\left[(X_I(1) - m_{X_I})(X_I(1+m) - m_{X_I})\right]}{v_{X_I}} \tag{15}$$

$$= \frac{E\left[X_I^*(1)X_I^*(1+m)\right] + E\left[\Delta_I(1)\Delta_I(1+m)\right] - m_{X_I}^2}{v_{X_I}} \tag{16}$$

(since $\{X_I^*(n)\}$ and $\{\Delta_I(n)\}$ are independent, and $E\left[\Delta_I(n)\right] = 0$). The variance of $X_I(1)$ is given by $v_{X_I} = v_{\widetilde{X}_I} + \sigma_\Delta^2$. Consider the numerator of the RHS of (16). The probability that the first I frame and the $(1+m)$th I frame belong to the same scene is q^m. Thus, $X_I^*(1) = X_I^*(1+m)$ with probability q^m. Otherwise, $X_I^*(1)$ and $X_I^*(1+m)$ are independent with identical distribution $f_{\widetilde{X}_I}$. Therefore,

$$E\left[X_I^*(1)X_I^*(1+m)\right] = \begin{cases} E\left[X_I^*(1)^2\right], & w.p.q^m \\ E\left[X_I^*(1)\right]E\left[X_I^*(m+1)\right], & w.p.(1-q^m) \end{cases} \tag{17}$$

Also, $E\left[\Delta_I(1)\Delta_I(1+m)\right] = \kappa_m \sigma_\Delta^2$. Thus, (16) becomes

$$\rho_I(m) = \frac{q^m E\left[X_I^*(1)^2\right] + (1-q^m)E^2\left[X_I^*(1)\right] + \kappa_m \sigma_\Delta^2 - m_{X_I}^2}{v_{\widetilde{X}_I} + \sigma_\Delta^2} \tag{18}$$

which reduces to the desired form after some manipulations. □

From this proposition, the ACF for the sequence $\{X_I(n)\}$ decreases at a rate that is determined by the sum of two terms : $q^m v_{\widetilde{X}_I}$ and $\kappa_m \sigma_\Delta^2$. The first term, which is the result of correlations due to scene effect, decreases at a geometric rate. The second term, which results from correlations in $\{\Delta_I(n)\}$, decreases as a sum of damped exponentials. Typically, the first term is dominant, and $\rho_I(m)$ approximately decreases at a geometric rate. *Thus, the generating process for I frames is short-range dependent.* We will see later that when the three processes that generate I, P and B frames are intermixed, the composite model exhibits long-range dependence.

3.2.2 P and B Frames

As mentioned earlier, our methodology is based on characterizing the frame size subsequence for each frame type, and then using the compression pattern to compose the complete frame size sequence. It is shown in the next section that constructing the frame size sequence in this manner captures the average amount of correlations between different types of frames. Let $\{X_P(n)\}$ and $\{X_B(n)\}$ be the frame size subsequences for P and B frames, respectively. We assume that $\{X_P(n)\}$ (also $\{X_B(n)\}$) is a sequence of *i.i.d.* random variables with common *pdf* f_{X_P} (f_{X_B} for B frames). Although in a real MPEG stream positive correlations do exist between the sizes of P frames (similarly, between B frames), these correlations are negligible compared to the correlations between different types of frames (which are induced by the structure of the compression pattern). In [11] we found that f_{X_P} and f_{X_B} can be appropriately characterized by lognormal distributions with parameters (μ_P,σ_P) and (μ_B,σ_B), respectively. In fact, a gamma or a Weibull *pdf* can also be used, with slightly less accurate fitting (see [11] for details). Histograms for $\{\widetilde{X}_P(n)\}$ and $\{\widetilde{X}_B(n)\}$ along with their fitting distributions are shown in Figures 9 and 10 for *Oz* trace. The fits in these figures are based on the MLE of the parameters of these distributions.

Table 4 summarizes the outcome of the fitting procedure for the four empirical traces. It contains estimates of the nine parameters required to specify the model. As expected, the estimated values differ from one trace to another. For the parameters that are associated with the lognormal fits, their values depend strongly on the dimensions of the captured frames. The values for other parameters (a_1, a_2, and σ_ε) depend, as well, on the dynamics of the scenes within a movie. Investigating the visual factors that impact the values of the model parameters will be addressed in a future paper.

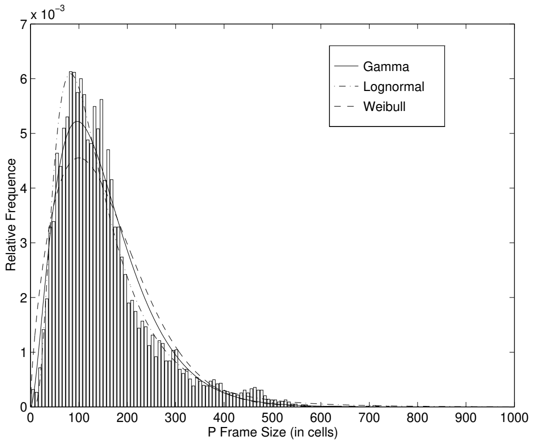


Figure 9: Histogram and fitting *pdf*s for $\{\widehat{X}_P(n)\}$ in *Oz* trace.

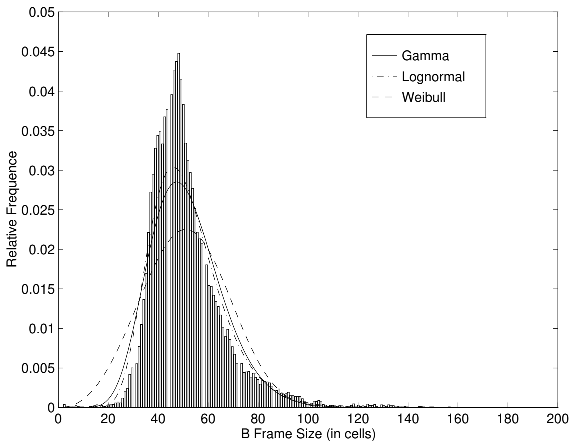


Figure 10: Histogram and fitting *pdf*s for $\{\widehat{X}_B(n)\}$ in *Oz* trace.

Trace	$\widetilde{\mu}_I$	$\widetilde{\sigma}_I$	a_1	a_2	σ_ε	μ_P	σ_P	μ_B	σ_B
Oz	5.9651	0.4832	0.5300	0.1512	19.8	4.8294	0.6443	3.9039	0.2746
Star Wars	4.9147	0.3861	0.2088	0.0917	8.9644	3.9324	0.5843	2.8049	0.5242
Lambs	4.5813	0.3852	0.3965	0.1516	4.36	2.5884	0.8721	1.9861	0.6937
Bond	5.3050	0.3663	0.4271	0.0380	10.3	4.5656	0.5111	3.2572	0.3761

Table 4: Estimates of the various parameters used in the model. The MLE are used for $\widetilde{\mu}_I$, $\widetilde{\sigma}_I$, μ_P, σ_P, μ_B, and σ_B. Yule-Walker estimates are used for a_1 and a_2. The estimate of σ_ε^2 is given by the sum-of-squares of the experimental residuals divided by the number of I frames.

3.3 The Autocorrelation Structure for the Composite Model

In this section, we derive the autocorrelation function for the sequence $\{X_n\}$ where X_n is the size of the nth frame (of an arbitrary type) in the model. The sequence $\{X_n\}$ is the composition of the three subsequences that were described in the previous sections.

Before deriving the ACF, we need to check the stationarity of $\{X_n\}$. It is clear that if $\{X_n\}$ is generated in such a way that the first frame is always an I frame, the second a B, the third a B, and so on, then $\{X_n\}$ is nonstationary since the frame size distribution varies with n. However, if we let $\{X_n\}$ start in an arbitrary frame location in the GOP pattern and from that point continue to generate frames according to the compression pattern, then we can show that $\{X_n\}$ is stationary.

This scenario is quite realistic considering that MPEG streams at an ATM multiplexer are admitted at different times. Thus, when the first frame (which is an I frame) of a newly admitted stream arrives at the multiplexer, the other ongoing streams could be sending other types of frames. From the perspective of the new stream (and the traffic control that is activated when new connections are established), the other streams look as if they were initialized at different locations of their compression patterns.

Let S_n be a discrete random variable which describes the location of the nth frame within the compression pattern. The state space for S_n is $\Omega_S = \{1, 2, \ldots, N\}$. S_n determines the frame type: if $S_n = 1$, then the type is I; if $S_n = 1 + M,\ 1 + 2M, \ldots,\ 1 + (N - M)$, then the type is P; otherwise, it is B. Given $S_n = i$, the distribution of X_n is stationary, and depends only on i. For $i = 1, \ldots, N$, let $F_i(u) \triangleq \Pr[X_n \leq u / S_n = i]$, and let $f_i(u)$ be the corresponding *pdf*. It is easy to see that $\{S_n\}$ is an irreducible and homogeneous Markov chain, with transition probabilities:

$$p_{ij} \triangleq \Pr[S_n = j / S_{n-1} = i] = \begin{cases} 1 & \text{if } j = i + 1 \text{ and } i = 1, \ldots, N - 1, \text{ or } i = N \text{ and } j = 1 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

All states of the chain are positive recurrent with mean recurrence time equals to N. Thus, there is a unique stationary distribution Π such that $\Pi \mathcal{P} = \Pi$. Let $\Pi = [\pi_1\ \pi_2\ \ldots\ \pi_N]$. Then, $\pi_i = 1/\text{mean recurrence time } = 1/N$ for all i. The mean, variance, and distribution of X_n are given in the following proposition.

Proposition 3 *If $S_1 \sim \Pi$, then the random process $\{X_n\}$ is first-order stationary. In this case, the pdf for X_n is given by:*

$$f_X(u) = \frac{1}{N} \sum_{i=1}^{N} f_i(u) = \frac{1}{N} \left(f_{X_I}(u) + (N/M - 1) f_{X_P}(u) + (N - M) f_{X_B}(u) \right). \tag{20}$$

The mean and variance of X_n are given by:

$$\begin{aligned} m_X \triangleq E[X_n] &= \frac{1}{N} \left(m_{X_I} + (N/M - 1) m_{X_P} + (N - M) m_{X_B} \right) && (21) \\ v_X \triangleq E\left[(X_n - m_X)^2 \right] &= \frac{1}{N} \Big(v_{X_I} + m_{X_I}^2 + (N/M - 1)(v_{X_P} + m_{X_P}^2) + (N - M)(v_{X_B} + m_{X_B}^2) \Big) - m_X^2. \end{aligned}$$

Proof: If $S_1 \sim \Pi$, then the Markov chain $\{S_n\}$ is stationary, i.e., $\Pr[S_k = i] = \pi_i = 1/N$ for all i and k.

$$\Pr[X_n \leq u] = \sum_{i=1}^{N} \Pr[X_n \leq u / S_n = i] \Pr[S_n = i] \tag{22}$$

$$= \frac{1}{N}\sum_{i=1}^{N} F_i(u). \tag{23}$$

Taking the derivative of (23), we obtain (20), which shows that $\{X_n\}$ is first-order stationary. The expressions for the mean and variance follow in a straightforward manner. □

It is possible, but algebraically tedious, to show that $\{X_n\}$ is also second-order stationary. We skip the proof and proceed to provide the ACF of $\{X_n\}$.

Proposition 4 *Let ρ_k be the ACF for $\{X_n\}$ at lag k, and let $h(i,k) \stackrel{\triangle}{=} [(i+k-1) \bmod N] + 1$. Then,*

$$\rho_k = \frac{(1/N)\sum_{i=1}^{N}\left(E\left[X_1/S_1 = i\right] E\left[X_1/S_1 = h(i,k)\right]\right) - m_X^2}{v_X}, \text{ for all } k,\ k \neq N,\ 2N, \ldots \tag{24}$$

$$= \frac{(1/N)\sum_{i=2}^{N} E^2\left[X_1/S_1 = i\right] + (1/N)E\left[X_1X_{1+k}/S_1 = 1\right] - m_X^2}{v_X} \text{ for } k = N,\ 2N, \ldots \tag{25}$$

Proof: see appendix. □

All the quantities in the RHS of (24) and (25) are known. The middle term in the numerator of (25) reflects the contribution of the correlations between I frames only, and is given by:

$$E\left[X_1X_{1+k}/S_1 = 1\right] = q^m E\left[(X_I^*)^2\right] + (1-q^m)E^2\left[X_I^*\right] + \kappa_m\sigma_\Delta^2 \tag{26}$$

where $m = k/N$ (see the appendix for more details). The following remarks can be made about the ACF in (24) and (25). First, at lags that are not multiples of N, the first term in the numerator of (24) can take only two possible values: one value when $k-1$ is a multiple of M (i.e., $k = M+1,\ 2M+1, \ldots$), and the other value for all other k, where $k \neq$ multiple of N. Secondly, at lags which are multiples of N, the middle term in (25), which is given in (26), decreases almost exponentially with k to its asymptotic limit $(1/N)E^2[X_I^*] = (1/N)E^2[X_1/S_1 = 1]$. The other two terms in the numerator do not change with k. Thus, as k goes to infinity at increments of N, ρ_k converges to a positive value given by:

$$\lim_{k\to\infty} \rho_{kN} = \frac{\frac{1}{N}\sum_{i=1}^{N} E^2\left[X_1/S_1 = i\right] - m_X^2}{v_X} \tag{27}$$

Although the ACF for $\{X_I(n)\}$ decreases exponentially, the ACF for the composite sequence $\{X_n\}$ does not converge to zero. The nonzero limit is the consequence of the cross-correlations between

Initialization:	set *Scene_max* := desired number of scenes
	set values for compression pattern: N, M
	set scene length parameter: q
	set the values for $\widetilde{\mu}_I$, $\widetilde{\sigma}_I$, μ_P, σ_P, μ_B, σ_B, a_1, a_2, and σ_ε^2
	$X \stackrel{\triangle}{=}$ frame size (in cells)

```
Loop:   for j = 1 to Scene_max do
            generate the length of scene j: N_j ~ Geometric(p)
            generate X~_I(j) ~ lognormal(μ~_I, σ~_I)
            for i = 1 to N_j do
                for k = 1 to L do
                    if k = 1, then                /* an I frame */
                        X = X~_I(j) + Δ_I(i), where
                        Δ_I(i) = â_1 Δ_I(i − 1) + â_2 Δ_I(i − 2) + ε(i)
                    else
                        if remainder(k/M) = 1, then          /* a P frame */
                            X ~ lognormal(μ̂_P, σ̂_P)
                        else                      /* a B frame */
                            X ~ lognormal(μ̂_B, σ̂_B)
                        end if
                    end if
                end for
            end for
        end for
```

Figure 11: Generating a synthetic stream based on the model.

different types of frames. These correlations are *induced* by periodically generating frames according to the GOP pattern. Thus, although $\{X_I(n)\}$ is Markovian (exhibits short-range dependence), and $\{X_P(n)\}$ and $\{X_B(n)\}$ are *i.i.d.* sequences, the *composite model* exhibits long-range dependence since $\sum_{k=0}^{\infty} \rho_k = +\infty$. Enhancing the model by incorporating the autocorrelations in $\{X_P(n)\}$ and $\{X_B(n)\}$ will have minor impact on the overall correlation structure.

4 Generating Synthetic MPEG Streams

The algorithm in Figure 11 can be used to generate synthetic MPEG streams based on the proposed model. An example of a synthetic stream generated based on the fitting of *Oz* trace is shown in Figure 12. Synthetic streams generated based on the model are "structurally" similar to the empirical traces; a property which could not be obtained without incorporating scene dynamics.

The frame size histogram is shown in Figure 13 (note that no closed-form solution is available for the frame size distribution). The analytical ACF of the model is plotted in Figure 14 using the parameters' values from the the fitting of *Oz* and *Star Wars*. By comparing these figures to their counterparts in the original traces, we see that the model captures the general shapes of the frame size distribution and ACF of an MPEG stream. However, correlation values in the model are slightly less than the empirical ACF, mainly because of our simplifying assumptions regarding the correlations in each of the three subsequences. Nevertheless, the model captures the induced correlations that result from the periodicity of the GOP pattern and the large differences in the sizes of different types of frames, and these correlations seem to have the major impact on the overall correlation structure.

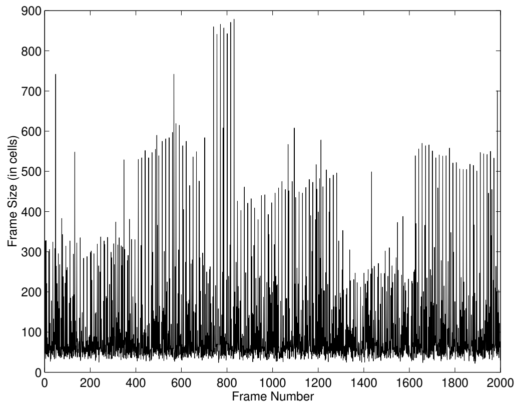


Figure 12: Example of a synthetic trace.

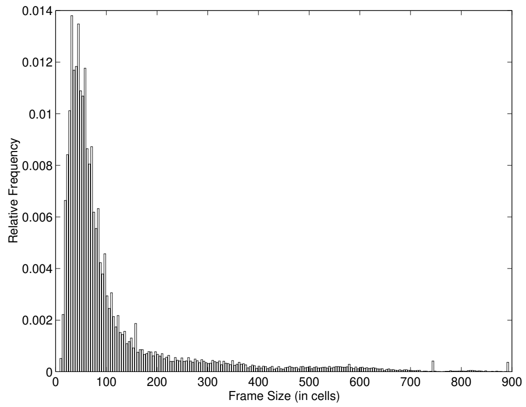


Figure 13: Model-based frame size histogram.

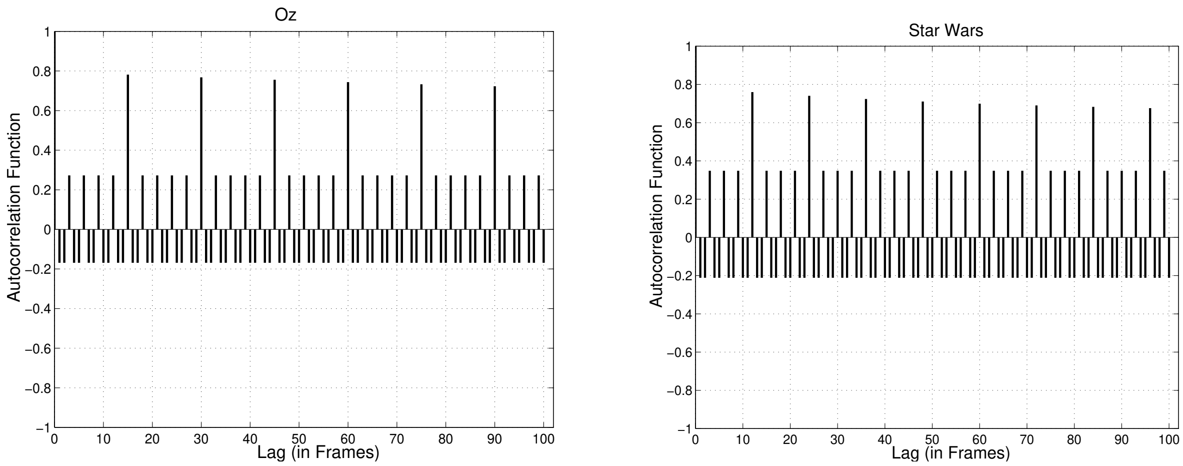


Figure 14: Analytical ACF for the frame size sequence (using MLE for the parameter values).

5 Queueing Performance at an ATM Multiplexer

Traffic models are commonly used to predict the performance at a statistical multiplexer and to allocate buffer and bandwidth resources. To verify the appropriateness of our model in predicting the queueing performance at a multiplexer, we conducted simulation experiments of a finite FIFO queue whose input consists of either real video sources or synthetic sources generated according to the model. Our primary measure of interest is the cell loss rate. Due to space limitations, only the results for *Oz* trace and its synthetic counterparts are reported.

The first set of experiments are related to the queueing performance for a single stream (either real or synthetic). Although in practice statistical multiplexing involves many streams that are admitted into a common queue, it is important to first study the single-stream case for the purpose of testing the appropriateness of our *source* model. Figures 15 and 16 depict the cell loss rate and the average queue length, respectively, at two loads: $U = 40\%$ and $U = 60\%$. Twenty simulation

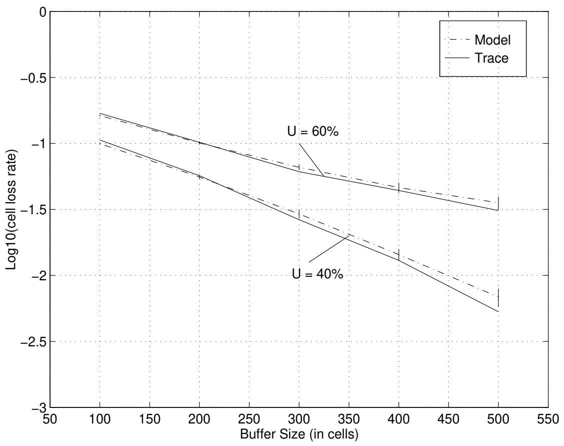


Figure 15: Cell loss rate versus buffer size for one stream (*Oz* trace).

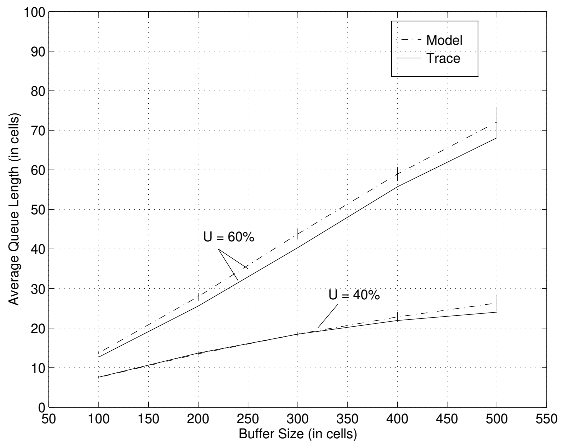


Figure 16: Average queue length versus buffer size for one stream (*Oz* trace).

replications were used to obtain the model-related results (the 95% confidence intervals are indicated by the short vertical lines). For small buffers, the model predictions are very accurate. As buffer size increases, the model starts to slightly overestimate the loss rate. We attribute this behavior to the persistent correlations in the model (Proposition 4). In general, the overflow probability is known to be adversely affected by the accumulated correlations. Because of the persistent correlations in the model, as the buffer size increases, the cumulative effect of the model correlations exceeds that of the empirical correlations, and the model starts to overestimate the loss performance. For practical buffer sizes, the discrepancy is small. Note that even for a light load of 40%, the cell loss rate (both actual and predicted) under a single video stream is quite high. This is a consequence of the bursty and correlated nature of MPEG video traffic. Smaller loss rates can be obtained by using extremely large buffers or by further reducing the load. Unfortunately, the empirical traces are not long enough to give statistically credible results at very small loss rates, say below 10^{-4}. The reason is that a cell loss rate of 10^{-4} implies the loss of a few hundreds of cells from a given empirical trace (e.g., 716 cells in *Star Wars*). Cell losses are strongly correlated as they mainly occur in few (sometimes less than ten) large I frames. in a given trace. Because of the strong correlations between successive I frames within a scene, the errorred frames are also correlated, so the several hundreds of lost cells come from a handful of scenes (for 10^{-4} loss rate, losses occurred in 3 scenes in the *Oz* trace). Longer video traces are needed to investigate very low loss rates for a single stream. At loads above 60%, the loss rate is very high, and the model predictions are expectedly very accurate (not shown here).

In computing the loss rate for one stream, we used a simple approximation that significantly reduced the simulation time. Since cells within a frame are evenly distributed, the approximate number of lost cells from the nth frame with size X_n can be obtained from the length of the queue at the time of arrival of the nth frame (Q_n), the buffer size (W), and the service rate (C). The approximation is used in the algorithm in Figure 17, which was used to obtain the cell loss rate. We verified its goodness by comparing the results to exact ones obtained from cell-level simulations. Some examples are shown in Figure 18. For *Oz* trace (41760 frames), it took less than 30 seconds on a Sparc10 workstation to compute the approximate loss rate, compared to about two hours for the exact cell-level simulations.

```
Q_1 := 0
lost_cells := 0
for j = 1 to last_frame do
  if X_j < C then /* underflow */
     Q_{j+1} := max{Q_j + X_j - C, 0}
  else
     T := Q_j + X_j - C - W
     if T > 0          /* cell losses */
        lost_cells := lost_cells + T
        Q_{j+1} := W
     else
        Q_{j+1} := T + W
     end if
  end if
end for
```

Figure 17: Algorithm for approximating the loss rate for one stream.

Load	Buffer Size (cells)	Cell Loss Rate	
		Exact	Approximate
60%	200	1.02×10^{-1}	1.02×10^{-1}
	400	3.48×10^{-2}	3.51×10^{-2}
	700	1.52×10^{-2}	1.56×10^{-2}
40%	200	5.70×10^{-2}	5.69×10^{-2}
	400	9.27×10^{-3}	9.27×10^{-3}
	700	1.60×10^{-5}	1.49×10^{-5}

Figure 18: Exact and Approximate cell loss rates for *Oz* trace (one stream).

Next, we examined the model's capability to predict the loss performance for multiplexed video streams. One difficulty in doing such a study is that the loss performance for multiplexed MPEG sources with I, P, and B frames strongly depends on the manner in which the GOP periods of different sources are aligned. For example, if we multiplex two sources that have the same GOP pattern, then the highest loss rate is observed when I frames from both sources arrive simultaneously at the multiplexer. Because of the periodicity of the GOP pattern, I frames from the two sources continue to overlap, causing large cell losses. On the other hand, if an I frame from the first source and a B frame from the other source (say the first B frame in the GOP) overlap in time, the loss rate is reduced by several orders of magnitude. This implies that the measured loss rates in different simulations of *arbitrarily* multiplexed MPEG streams cannot be assumed to be *i.i.d.* random variables, and thus confidence intervals are not meaningful. It is more appropriate to study the loss performance under a particular "arrangement" of the multiplexed streams, since different arrangements result in completely non-comparable results. The impact of stream alignment on bandwidth allocation was studied in [12] within a deterministic framework. Since our focus is on evaluating the model's accuracy rather than studying the effect of video scheduling, we will limit our simulation experiments (for both real and synthetic data) to the "optimal" arrangement of multiplexed video sources that was shown in [12] to give the minimum required bandwidth (note, however, that the study in [12] was based on a deterministic traffic envelop that provided time-varying upper bounds on the bit rates). According to this optimal arrangement, the multiplexed streams are aligned such that a GOP of the first source is one frame lag from a GOP of the second source, a GOP of the second source is one frame lag from a GOP of the third source, and so on. Frame boundaries are aligned. An example of the temporal structure for optimally arranged five sources with common GOP pattern *IBBPBB* is shown in Figure 19.

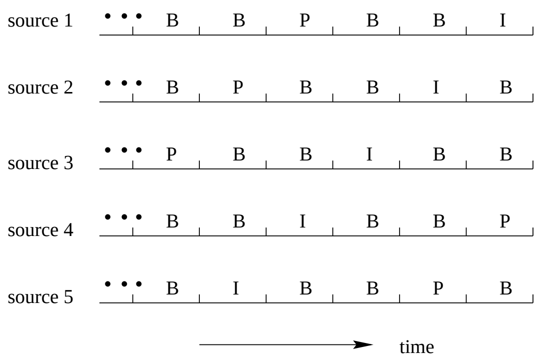


Figure 19: Example of temporal structure of multiplexed streams with "optimal" arrangement.

Source #	Starting Frame
1	1
2	8341
3	16682
4	25023
5	33364

Figure 20: Starting frames for five multiplexed streams using real data.

In all the following experiments, the multiplexed streams were arranged according to the aforementioned approach. Table 5 shows the results from multiplexing five MPEG sources, using real and synthetic data. Streams are aligned "optimally" as shown in Figure 19. To produce fives real sources from *Oz* trace, we used a commonly employed approach in which the empirical trace is arranged as a circular list. A real source starts at some frame location in the list and continues until it completes the circle. One problem with this approach is that the extracted streams may not be independent, especially if their starting frames are not far apart. Unfortunately, no other alternative is available if multiplexing of structurally similar real sources is to be studied. To reduce the possibility of correlations between different sources, we chose to space out the starting times of different sources as much as possible, as shown in Figure 20 (typically, the starting frames are chosen randomly [6]). The loss rates for the synthetic streams in Table 5 represent the averages among ten different simulation replications. Sample variances of the loss rates for the synthetic data ranged from 1% to 10% of the mean values. Hence, the confidence intervals are sufficiently tight (they are not shown for brevity).

Buffer Size (in cells)	U = 80%		U = 70%		U = 60%	
	Real	Model	Real	Model	Real	Model
100	9.92E−2	6.72E−2	6.04E−2	3.85E−2	2.53E−2	1.70E−2
200	8.10E−2	5.08E−2	4.32E−2	2.60E−2	1.37E−2	9.51E−3
300	6.56E−2	3.89E−2	3.05E−2	1.74E−2	7.34E−3	4.98E−3
400	5.18E−2	2.97E−2	2.12E−2	1.15E−2	3.98E−3	2.61E−3
500	3.97E−2	2.24E−2	1.41E−2	7.55E−3	2.13E−3	1.37E−3
600	2.95E−2	1.67E−2	9.09E−3	4.93E−3	1.06E−3	7.04E−4
700	2.12E−2	1.23E−2	5.76E−3	3.17E−3	4.56E−4	3.56E−4
800	1.49E−2	8.84E−3	3.53E−3	1.99E−3	1.42E−4	1.73E−4

Table 5: Cell loss rate when five streams are multiplexed.

From this table, the model is shown to provide good predictions of the performance for real sources (i.e., within an order of magnitude). As explained before, the model underestimates the loss rate for small buffers, and as the buffer size increases, the gap becomes smaller. For large buffers, the model slightly overestimates the performance. In both cases, the model's predictions are quite acceptable.

Given the demonstrated appropriateness of the model, it can be used to study the multiplexing of large number of streams and determine the required amounts of resources to guarantee a desired level of loss performance. In this case, the model offers the means to generate statistically similar but independent sources. We present some simulation examples that show the practical use of the model. Figure 21 depicts the cell loss rate versus buffer size for 10 multiplexed video streams at three loads. Each simulation experiment was repeated 10 times, each time with 10 different synthetic sequences (95% confidence intervals are indicated by the short vertical lines). Each synthetic stream consisted of 200,000 frames. With this number of frames, statistically credible results can be obtained for very small loss rates. Figure 22 depicts the cell loss rate versus the number of multiplexed streams at a given load and buffer size (W). The values are for the averages among ten replications (confidence intervals were sufficiently tight, and are not shown for brevity). In general, increasing the number of multiplexed streams while fixing the load and buffer size results in a smaller loss rate, with the effect being more pronounced at lighter loads (notice that buffer size is not normalized to the number of streams). However, this trend is not always true. In particular, a sudden increase in the loss rate is observed in all curves when the number of multiplexed streams goes from 15 to 16. This increase is related to the manner in which the streams were arranged. According to the "optimal" arrangement, I frames of a given source are one frame period apart from I frames of the subsequent source. With a GOP period of 15 frames ($N = 15$ for Oz sources), a maximum of 15 sources can be arranged without overlapping I frames of different sources. After that, the overlap is inevitable, causing the observed increase in the loss rate when 16 sources are multiplexed. Some less obvious effects due to the overlap of P frames can also be noticed. A more elaborate study of the impact of stream alignment on the loss rate will be provided in a future paper.

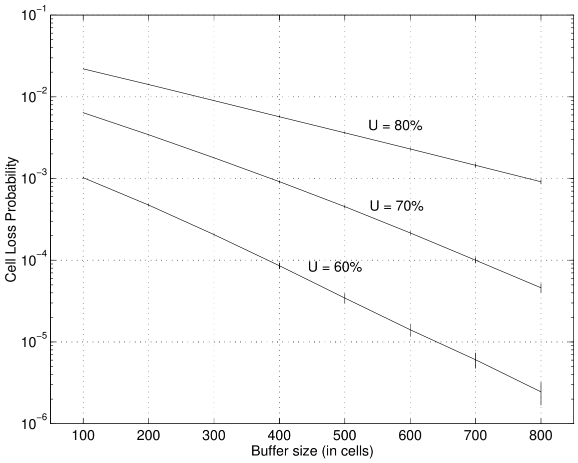


Figure 21: Cell loss rate for ten multiplexed streams based on the model.

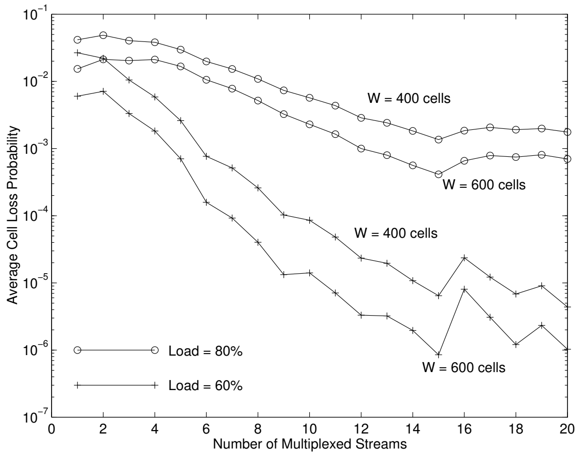


Figure 22: Cell loss rate versus the number of multiplexed streams (based on model)

6 Summary

We presented a comprehensive traffic model for MPEG-coded video streams with all types of MPEG frames. The model was successfully fitted to four different empirical video sequences. In addition to approximating the marginal distributions for frame sizes and the overall correlation structure, the model uses a scene-related component to capture bit rate variations at multiple time scales. We found that the fluctuations of I frames can be characterized by a modulated process in which the local variations are modeled by an AR(2) process that varies around a scene-related random mean. Two autocorrelated processes were needed to characterize I frames. The sizes of P and B frames were modeled by two *i.i.d.* random processes with lognormal marginals. We derive the analytical form for the autocorrelation function of the composite model. This form indicates the existence of persistent correlations that are attributed to the periodicity of the GOP pattern and the significant differences between the sizes of different types of frames. The sum of these correlations (over all lags) is infinite. Thus, although the individual random subsequence are Markovian, the mixing according to a fixed GOP renders the model long-range dependent. This, somehow surprising, result can be extended to stationary processes with different marginals that are intermixed according to a fixed periodic pattern. Simulations were used to show the appropriateness of the model in predicting the queueing performance for real video sources. Model-based synthetic streams were used to study the performance of multiplexed MPEG streams. For such studies, the model provides a much easier and less expensive approach to generating synthetic traces than having to obtain real ones.

Appendix
Proof of Proposition 4

To show that the ACF for $\{X_n\}$ is given by (24) and (25), we consider two cases:

Case 1: $k \neq N,\ 2N, \ldots$

$$\rho_k \triangleq \frac{E\left[(X_1 - m_X)(X_{1+k} - m_X)\right]}{v_X} \tag{28}$$

$$= \frac{E\left[X_1 X_{1+k}\right] - m_X^2}{v_X} \tag{29}$$

Consider the term $E\left[X_1 X_{1+k}\right]$.

$$\begin{aligned} E\left[X_1 X_{1+k}\right] &= \int_0^\infty \Pr\left[X_1 X_{1+k} > u\right] du \\ &= \int_0^\infty \sum_{i=1}^{N} \Pr\left[X_1 X_{1+k} > u / S_1 = i\right] \cdot \Pr\left[S_1 = i\right] du \\ &= \sum_{i=1}^{N} \pi_i E\left[X_1 X_{1+k} / S_1 = i\right] \end{aligned}$$

$$= \frac{1}{N}\sum_{i=1}^{N} E\left[X_1 X_{1+k}/S_1 = i\right] \quad (30)$$

The two conditional events $[X_1 \leq x_1/S_1 = i]$ and $[X_{1+k} \leq x_2/S_1 = i]$ are (conditionally) independent for $k \neq$ a multiple of N. Hence,

$$E\left[X_1 X_{1+k}\right] = \frac{1}{N}\sum_{i=1}^{N} E\left[X_1/S_1 = i\right] E\left[X_{1+k}/S_1 = i\right] \quad (31)$$

$$= \frac{1}{N}\sum_{i=1}^{N} E\left[X_1/S_1 = i\right] E\left[X_{1+k}/S_{1+k} = h(i,k)\right] \quad (32)$$

$$= \frac{1}{N}\sum_{i=1}^{N} E\left[X_1/S_1 = i\right] E\left[X_1/S_1 = h(i,k)\right] \quad (33)$$

which proves the assertion.

Case 2: $k = mN$ **for** $m = 1, 2, \ldots$
In this case, $E\left[X_1 X_{1+k}\right]$ can be written as

$$E\left[X_1 X_{1+k}\right] = \frac{1}{N}\sum_{i=2}^{N} E\left[X_1 X_{1+k}/S_1 = i\right] + \frac{1}{N} E\left[X_1 X_{1+k}/S_1 = 1\right] \quad (34)$$

For $i = 2, 3, \ldots, N$,

$$E\left[X_1 X_{1+k}/S_1 = i\right] = E\left[X_1/S_1 = i\right] E\left[X_{1+k}/S_1 = i\right] = E^2\left[X_1/S_1 = i\right] \quad (35)$$

(the first equality holds because each of the subsequences $\{X_P(n)\}$ and $\{X_B(n)\}$ are *i.i.d.*. The second equality holds because, in this case, $S(1+k) = S(1+mN) = S(1)$). The term $E\left[X_1 X_{1+k}/S_1 = 1\right] = E\left[X_I(1) X_I(1+m)\right]$ which is given in the proof of Proposition 2. □

References

[1] J. Beran, R. Sherman, M. S. Taqqu, and W. Willinger. Long-range dependence in variable bit-rate video traffic. *IEEE Transaction on Communications*, 43:1566–1579, 1995.

[2] G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forcasting and Control.* Holden-Day, Inc., 1976. (Revised Edition).

[3] M. R. Frater, J. F. Arnold, and P. Tan. A new statistical model for traffic generated by VBR coders for television on the Broadband ISDN. *IEEE Trans. on Circuits and Systems for Video Technology,*

4(6):521–526, Dec. 1994.

[4] M. W. Garrett and W. Willinger. Analysis, modeling, and generation of self-similar VBR video traffic. In *Proceedings of the SIGCOMM'94 Conference (Computer Communications Review)*, pages 269–280, Sept. 1994.

[5] D. Heyman, , E. Tabatabai, and T. Lakshman. Statistical analysis of MPEG2-coded VBR video traffic. In *Proc. of the Sixth International Workshop on Packet Video*, 1994.

[6] D. P. Heyman, A. Tabatabai, and T. V. Lakshman. Statistical analysis and simulation study of video teleconferencing traffic in ATM networks. *IEEE Trans. on Circuits and Systems for Video Technology*, 2(1):49–59, Mar. 1992.

[7] C. Huang, M. Devetsikiotis, I. Lambadaris, and A. Kaye. Modeling and simulation of self-similar variable bit rate compressed video: A unified approach. In *Proc. of SIGCOMM '95*, 1995.

[8] B. Jabbari, F. Yegengolu, Y. Kuo, S. Zafar, and Y.-Q. Zhang. Statistical characterization and block-based modeling of motion-adaptive coded video. *IEEE Trans. on Circuits and Systems for Video Technology*, 3(3):199–207, June 1993.

[9] P. A. Jacobs and P. A. W. Lewis. Time series generated by mixtures. *Journal of Time Series Analysis*, 4(1):19–36, 1983.

[10] K. Koga, Y. Iijima, I. Kazumoto, and T. Ishiguro. Statistical performance analysis of an interframe encoder for broadcast TV signals. *IEEE Transaction on Communications*, 29:1868–1875, Dec. 1981.

[11] M. Krunz, R. Sass, and H. Hughes. Statistical characteristics and multiplexing of MPEG streams. In *Proc. of the IEEE INFOCOM '95 Conference*, pages 455–462, Boston, Apr. 1995.

[12] M. Krunz and S. K. Tripathi. Impact of video scheduling on bandwidth allocation for multiplexed MPEG streams. *Multimedia Systems Journal*, 1996. (to appear).

[13] A. M. Law and W. D. Kelton. *Simulation Modeling and Analysis.* McGraw-Hill, Inc., second edition, 1991.

[14] A. A. Lazar, G. Pacifici, and D. E. Pendarakis. Modeling video sources for real-time scheduling. Technical Report 324–93–03, Columbia University, Department of Electrical Engineering and Center for Telecommunications Research, Apr. 1993.

[15] B. Maglaris et al. Performance models of statistical multiplexing in packet video communications. *IEEE Journal on Selected Areas in Communications*, 36(7):834–844, July 1988.

[16] N. M. Marafih, Y.-Q. Zhang, and R. L. Pickholtz. Modeling and queueing analysis of variable-bit-rate coded video sources in ATM networks. *IEEE Trans. on Circuits and Systems for Video Technology*, 4(2):121–128, Apr. 1994.

[17] B. Melamed. TES: A class of methods for generating autocorrelated uniform variates. *ORSA Journal on Computing*, 3(4):317–329, 1991.

[18] B. Melamed and D. Pendarakis. A TES-based model for compressed Star Wars video. In *Proceedings of IEEE GLOBCOM '94*, pages 120–126, 1994.

[19] B. Melamed, D. Raychaudhuri, B. Sengupta, and J. Zdepski. TES-based video source modeling for performance evaluation of integrated networks. *IEEE Transaction on Communications*, 42(10):2773–2777, Oct. 1994.

[20] P. Pancha and M. El Zarki. Bandwidth-allocation schemes for variable-bit-rate MPEG sources in ATM networks. *IEEE Trans. on Circuits and Systems for Video Technology*, 3(3):190–198, June 1993.

[21] G. Ramamurthy and B. Sengupta. Modeling and analysis of a variable bit rate video multiplexer. In *Proc. of IEEE INFOCOM '92*, volume 2, pages 817–827, 1992.

[22] D. Reininger, B. Melamed, and D. Raychaudhuri. Variable bit rate MPEG video: Characteristics, modeling and multiplexing. In *Proc. of the 14th Int. Teletraffic Congress*, pages 295–306, Antibes Juan-les-Pins, France, June 1994.

[23] O. Rose. Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems. In *Proceedings of the 20th Annual Conference on Local Computer Networks*, Minneapolis, MN, 1995.

[24] P. Skelly, M. Schwartz, and S. Dixit. A histogram-based model for video traffic behavior in an ATM multiplexer. *IEEE/ACM Transactions on Networking*, 1(4):446–459, Aug. 1993.

[25] F. Yegengolu, B. Jabbari, and Y.-Q. Zhang. Motion-classified autoregressive modeling of variable bit rate video. *IEEE Trans. on Circuits and Systems for Video Technology*, 3(2):42–53, Feb. 1993.