

# Analysis and Modeling of H.264 Unconstrained VBR Video Traffic

## Harilaos Koumaras

*Business College of Athens (BCA), Computer Science Department,*

*4 Dimitressa Str., Athens, Greece*

*Email: koumaras@bca.edu.gr*

## Charalampos Skianis

*Institute of Informatics and Telecommunications NCSR «DEMOKRITOS», Patriarchou Gregoriou Str., Agia*

*Paraskevi, Attiki, 15310 Athens Greece*

*Email: skianis@iit.demokritos.gr*

## Anastasios Kourtis

*Institute of Informatics and Telecommunications NCSR «DEMOKRITOS», Patriarchou Gregoriou Str., Agia*

*Paraskevi, Attiki, 15310 Athens Greece*

*Email: kourtis@iit.demokritos.gr*

**Abstract.** In future communication networks, video is expected to represent a large portion of the total traffic, given that especially variable bit rate (VBR) coded video streams, are becoming increasingly popular. Consequently, traffic modeling and characterization of such video services is essential for the efficient traffic control and resource management. Besides, providing an insight of video coding mechanisms, traffic models can be used as a tool for the allocation of network resources, the design of efficient networks for streaming services and the reassurance of specific QoS characteristics to the end users. The new H.264/AVC standard, proposed by the ITU-T Video Coding Expert Group (VCEG) and ISO/IEC Moving Pictures Expert Group (MPEG), is expected to dominate in upcoming multimedia services, due to the fact that it outperforms in many fields the previous encoded standards. This paper presents both a frame and a layer (i.e. I, P and B frames) level analysis of H.264 encoded sources. Analysis of the data suggests that the video traffic can be considered as a stationary stochastic process with an autocorrelation function of exponentially fast decay and a marginal frame size distribution of approximately Gamma form. Finally, based on the statistical analysis, an efficient model of H.264 video traffic is proposed.

**Keywords:** H.264 video coding, Traffic analysis, Video modeling

## 1 Introduction

Multimedia applications and services have already possessed a major portion of the today traffic over computer and mobile communication networks. Among the various types of multimedia, video services (transmission of moving images and sound) are proven dominant for present and future broadband networks.

Raw video data has very high bandwidth and storage requirements making its transmission and storage impractical and economically unaffordable. For this reason, a lot of research has been

performed on developing techniques that exploit both temporal and spatial redundancy in video sequences, in order to succeed efficient data compression.

From the advent of video coding, two main encoding schemes were proposed and are still in use: The Constant Bit Rate (CBR) and the Variable Bit Rate (VBR) modes. The choice of VBR mode for video services over communication networks prevails against CBR mode due to a number of advantages such as

- Better video quality for the same average bit rate without the need to adjust the quantization parameters during the encoding as in CBR
- Shorter delay since the buffer size in the encoder side can be reduced without encountering an equivalent delay in the network
- Increased call-carrying capacity due to the fact that the bandwidth per call for VBR video may be lower than for equivalent quality of CBR source.
- Although a CBR transmission mode makes the network management easier, mainly due to the predictable traffic patterns, on the other hand it prevents a possible traffic gain via statistical multiplexing, which means that does not efficiently exploits the available capacity of the transmission channel as VBR does.

Efficient network utilization and constant picture quality can be achieved by VBR mode. However, when transmission and statistical multiplexing of VBR-coded video traffic is considered over a shared medium, like the Internet, the improvement in network utilization cannot be determined only by the compression ratio. VBR coding results in large fluctuations in bit rate and high correlation among the bit rates in successive time intervals due to the video content and the abrupt scene changes [28]. This complex nature of VBR-coded video traffic creates a challenge in the efficient design of communications/transmission networks and the associated traffic control. Therefore an accurate traffic study is necessary for the prediction of the network performance. A method of doing this is to perform real experiments using existing networks and actual sources. However, testing real networks is quite impractical, while performing tests with real video clips, although it is possible, the deduced results may be very video-content specialized and therefore not general and scalable. Thus, major surge of interest in the topic of VBR video traffic modeling has appeared, because it provides information on how VBR mode affects network performance and besides is a useful tool for traffic engineering of communication networks in order to optimize admission control, to perform short-term traffic forecast and optimize buffer lengths.

Generally speaking, an analysis of video traffic is required in order to develop an efficient video traffic model. Such models can be evaluated [1] by the fact that they must satisfy some criteria, namely: They must match certain characteristics of a real video sequence, such as probability density function, mean, variance, peak, autocorrelation etc. Moreover, the deduced generated video traffic must be similar to real video data, in order to be able to be used for predicting a desired performance metric (i.e. delay, buffer, size etc.). Furthermore, the proposed models should be simple and able to generate video traffic with low computational power.

Early studies in unconstrained VBR models examined various characteristics of VBR video traffic, such as differences in successive frame sizes and cluster lengths [2] or scene duration distributions [3]. Also recently introduced efficient modeling tools and techniques of VBR MPEG-1/H.261 coded video at frame and GOP level [4], [5].

Results from these and other works indicate that the frame sizes exhibits a bell-shape (e.g. [6], [7], [8]). Furthermore, in certain cases correlations in the video bit rate are found to decay

exponentially [6], [7], [9], [10], [11] while other studies [8], [12], [13] observe a more complex phenomenon, in which the correlation decay is rapid for the initial lags and then continues at a lower rate.

The most popular and widely used encoding algorithms are the ones developed by the Moving Picture Experts Group (MPEG) and the Video Coding Expert Group (VCEG) of the ITU. Recently these two organizations jointly developed a new codec, the H.264 or MPEG-4 Part 10 Advanced Video Coding (AVC) codec [14]. Featuring updated capabilities, the new codec can achieve 40-50% compression efficiency gain over today's optimized MPEG-2 codecs. Due to the advances of H.264 in comparison to earlier standards, e.g. H.263 [15] [16], it is expected that it will prevail in future networks and mobile application systems, making traffic modeling and characterization of H.264 video streams a useful tool for network managers and designers.

Following this trend, this work presents a detailed frame and layer (i.e. I, P and B) level analysis of H.264 video traffic and proposes an adequate traffic model.

The rest of the paper is organized as follows: Section 2 outlines the new characteristics and enhancements of the H.264 standard, Section 3 presents the statistical analysis of the H.264 video stream. Section 4 discusses on the video traffic modeling, presenting related work and a novel H.264 model. Finally, Section 5 concludes the paper.

## **2 The H.264/AVC standard: Essential Issues and Current Status**

In 1998 the ITU-T VCEG issued a call for proposals (H.26L project), with main scope to double the coding efficiency in comparison to the already existing coding standards. In 2001, VCEG and ISO/IEC MPEG formed a Joint Video Team (JVT) in order to finalize the standard and submit for formal approval as H.264/AVC [14].

The new video coding standard known as H.264/MPEG-4 Advanced Video Coding (AVC), now in its fourth version, has demonstrated significant achievements in terms of coding efficiency, robustness to a variety of network channels and conditions, and breadth of applications [17]. Some essential indicative enhancements are:

- Variable block size support for motion compensation with luma block sizes down to 4x4, in conjunction with 4x4 level transformations.
- Quarter-sample motion vector accuracy.
- Extended reference frame selection for P frames, among various previously decoded frames.
- De-blocking filter within the motion-compensated prediction loop.
- New context-based adapted entropy coding methods: CAVLC and CABAC.

The main target of the aforementioned enhancements is the perceived quality improvement and the high-compression efficiency. With the expected wide breadth of applications, from videoconferencing and entertainment to streaming video and digital cinema, where the new coding standard is expected to be implemented, three basic feature sets (called profiles) were established to address these application domains:

- Baseline profile (BP): Designed to minimize complexity and provide high robustness and flexibility for use over a broad range of network environments and conditions.
- Main Profile (MP): Designed with an emphasis on compression coding efficiency capability.
- Extended Profile (XP): Designed to combine the robustness of the Baseline profile with a higher degree of coding efficiency and greater network robustness.

At present, the Baseline profile seems that it provides a good solution for its target application area. The JVT is working on incorporating a Scalable Video Coding (SVC) amendment into the

design of the existed H.264 standard. In terms of coding structure, a scalable bit stream will be composed of a base layer and one or more enhancement layer bit streams. The base layer will be conforming to one of the profiles of the prior H.264/MPEG-4 AVC design. Additional key issues are fidelity range extensions [18], [19], which addressed the issue of more demanding applications of H.264 in resolution, bits/sample and chroma sampling, and improvement on H.264 encoder performance [20].

### **3 Statistical Analysis of the H.264 encoded data**

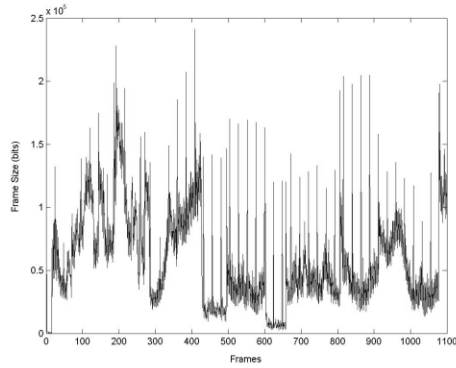
For the statistical analysis of H.264/AVC encoded data, the reference encoder JM is used, considering encodings without rate control and fixed quantization parameters for all test sequences. In H.264, the three common different frame modes are adopted, namely: Intra-frame (I), Predictive (P) and Bidirectional predictive (B), widely referred as I, P and B. In particular, the I frames are also called Intra frames, while B and P are known as Inter frames. The combination of successive types of frames forms a Group Of Pictures (GOP), whose length is mainly described by the distance of two successive I frames. In the described work, the frame rate is set constant at 25fps, coding GOP structure is set as IPBPBPBPB... and Intra-period adopts values between 3 and 12. Finally, a video segment from the film "Spider-man II" is used as reference signal. This segment consists of 18357 frames of YUV 4:2:0 format in 528x384 resolution.

#### **3.1 Frame Level Analysis**

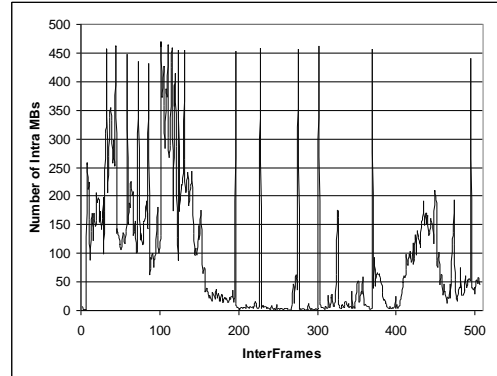
Focusing on the Frame Level analysis, Figure 1 illustrates the size of 1100 frames of an H.264 test signal (encoded with quantization scale 20 for all the frames and GOP length 12), where it can be noticed that the large frame sizes (periodical peaks in the figure) correspond to I frames, while the smaller ones are B frames and the intermediate frame sizes are P frames. Moreover, the periodicity that seems to appear in the peaks of I frames, corresponds to the distance of two successive I frames, which reveals the length of the used GOP. It is also noted that the frame size follows the spatial and temporal activity of the test signal, where more complex frames require more bits for their description, while static and simple frames are described by fewer bits.

Also another interesting observation is that inter-frames (especially P frames) present more intense fluctuation in comparison with the Intra frames. This stems from the fact that according to the content dynamics of the video signal, some Macro-Blocks (MBs) of the inter-frames may be intra-coded, which results in lower compression ratio and therefore higher frame sizes. Figure 2 depicts the total number of Intra MBs for the P frames of the total 1100 frames of Figure 1. It can be observed that the shape of the Intra MBs vs. Inter-frames graph (Figure 2) plays a major role in the form of the frame size graph (Figure 1). In other words, inter-frames appear to influence largely the actual video traffic.

A principal issue regarding the modeling of unconstraint Variable Bit Rate traffic is whether or not the encoded traffic can be considered as stationary process. In this respect, an encoding frame sequence from "Spider-man 2" was split in a moderate number of windows (actually four) and the empirical density function for the frame size was calculated from the samples of each window.

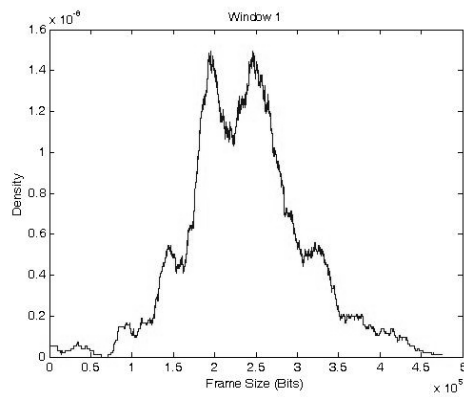


**Figure 1. The frame level analysis over a time-window of 1100 frames**

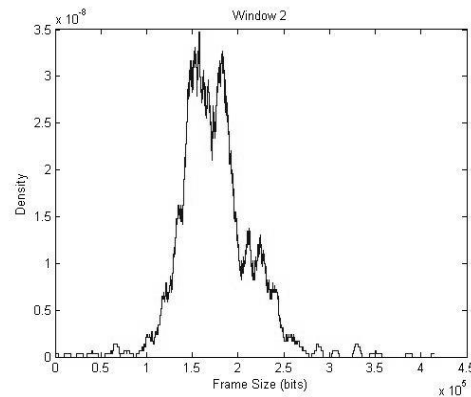


**Figure 2. The Intra MBs for the inter-frames (i.e. P) over a time of 1100 frames**

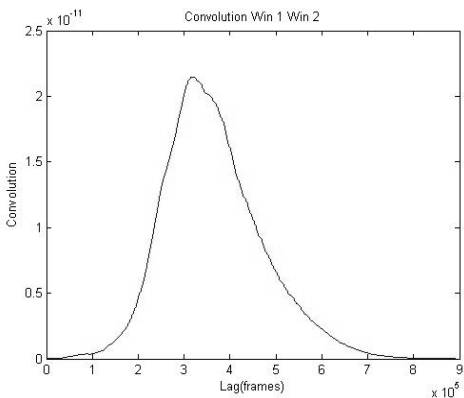
These windows densities, which are depicted in figures 3(a), (b), where found very similar, a property directly suggesting that the sequence is stationary [4], [29]. In order to expand further the second-order stationary [4], [29], the autocorrelations of these empirical densities were constructed for pairs of time windows, showing almost identical shape across window combinations (figures 3(c), (d)). Therefore, the aforementioned result about stationary is further reinforced.



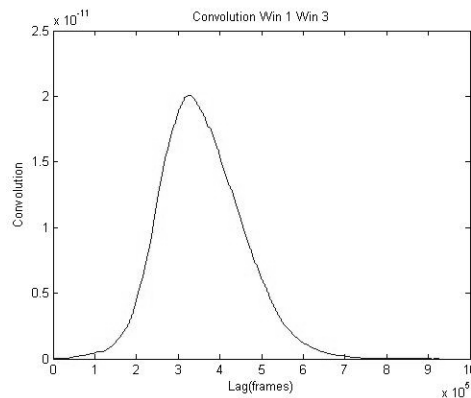
(a)



(b)

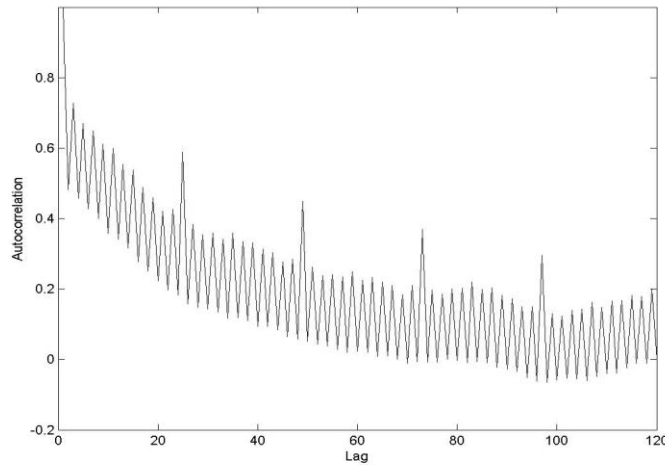


(c)



(d)

**Figure 3. Frame size histograms in different time windows (a), (b) and autocorrelations of such histograms (c), (d)**



**Figure 4. The autocorrelation of the 1100 frame sizes**

Figure 4 illustrates the autocorrelation function for the 1100 frames. It can be observed that the autocorrelation graph consist of periodic spikes that are superimposed on a decaying curve. The highest peaks correspond to the autocorrelation of the Intra frames of the video sequence, which are followed by 11 lower spikes before the next “Intra” peak. The lower spikes between two successive “Intra” peaks correspond to P frames, which are typically smaller than the I-frames. Finally, the wells between I and P peaks, correspond to the B frames of the test sequence, which are the smaller frames of all.

Based on the already discussed results, it can be deduced that the behavior of the H.264 encoded signal can be described as a superimposition of three different distributions, which result from three different frames modes (i.e. I/B/P). Therefore, elaborating each frame type separately is more efficient and produces more detailed description of the H.264 video traffic. The next section presents an I/B/P layer analysis of the encoded signal.

### **3.2 I/B/P Level Analysis**

For the I/P/B level analysis again the same video segment from the film “Spider-man II” is used as reference signal. In order to study the nature of the video stream, intra-frame period and quantization parameters are altered during the experiments. During each encoding process, video traces are captured, containing data on the type and the size of each encoded frame. As a result, frame statistics based on specific quantization scale and encoding settings are derived and depicted in Table 1 in the form of mean values and variances of I/P/B frame sizes. The notation (x,y,z)-l is used for the quantization scales of I,B,P frames and the selected intra-frame period.

From Table 1, it can be derived that higher encoding parameters, which cause coarser encoding quality, result in lower mean frame sizes and variations in comparison with lower quantization parameters, which produce better encoding quality. On the contrary, the alternation of Intra-frame period does not affect frame sizes, which remain practically constant. Table 1 depicts the mean value, the standard deviation and the min and max values of the same experimental sets expressed in Kbits.

Quantization Settings / Frame Types	I Frames (in Kbits)				B Frames (in Kbits)				P Frames (in Kbits)			
	Mean	$\sigma$	min	max	Mean	$\sigma$	min	max	Mean	$\sigma$	min	max
(10,10,10)-12	354.47	87.29	35.7	610.1	227.65	57.67	6.3	576.5	271.34	65.67	24.2	588.7
(20,20,20)-12	148.16	58.25	2.62	325.8	43.02	33.17	0.26	291.5	67.01	41.98	0.62	294.8
(30,30,30)-12	53.91	25.68	1.70	146.7	7.86	8.71	0.24	109.1	16.33	13.93	0.24	116.9
(20,20,20)-3	147.21	57.32	2.62	325.8	42.86	33.25	0.28	290.6	67.22	42.15	0.51	294.8
(20,20,20)-6	148.46	58.03	2.62	325.8	43.10	33.30	0.26	291.5	67.06	42.08	0.50	294.8

**Table 1: Frame statistics overview of the encoded signal in Kbits**

Finally, Table 2 contains the variation coefficients for the various quantization schemes, which represent a metric of the variation and the shape of the deduced frame size distribution. The variation coefficient is defined as  $S_x/\bar{x}$ , where  $S_x$  is the standard deviation and  $\bar{x}$  the mean value. Their deduced values are consistent with the aforementioned observations.

Quantization Settings / Frame Types	I Frames Variation Coefficient	B Frames Variation Coefficient	P Frames Variation Coefficient
(10,10,10)-12	0.2463	0.2533	0.2420
(20,20,20)-12	0.3932	0.7710	0.6265
(30,30,30)-12	0.4763	1.1081	0.8530
(20,20,20)-3	0.3894	0.7758	0.6271
(20,20,20)-6	0.3909	0.7726	0.6275

**Table 2: Variation Coefficients of the various frame types**

In order to study the statistical behavior of the encoding stream, the Probability Density Functions (PDFs) for each frame type of the encoded signal at various quantization scales are drawn. Figure 5 depicts the representative case of (10-10-10)-12.

Observing that the derived graphs follow the expected bell-like shape, then the well adopted [4] method of moments is used in order to fit a gamma density (the equivalent of the negative binomial in the continuous domain) to the data output. The Gamma density is given by the following formula

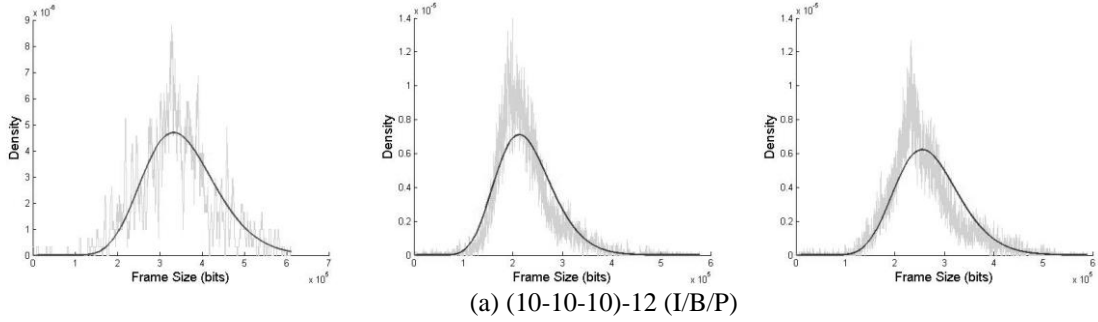
$$f_k = \frac{(x/\mu)^{p-1}}{\mu \Gamma(p)} e^{-x/\mu}, \mu, p > 0, x \geq 0, \Gamma(p) = \int_0^{\infty} t^{p-1} e^{-t} dt \quad (1)$$

Where  $\mu > 0$  is the scale parameter, and  $p > 0$  the shape parameter respectively.

The Gamma distribution has mean  $p\mu$  and variance  $p\mu^2$  and by equating to the mean and sample variance, denoted as  $m$  and  $v$  respectively, it can be deduced that  $\mu = v/m$  and  $p = m^2/v$ . Table 3 contains the Gamma distribution parameters for each quantization scale and Figure 5 illustrates the frame histograms in conjunction with the corresponding Gamma models.

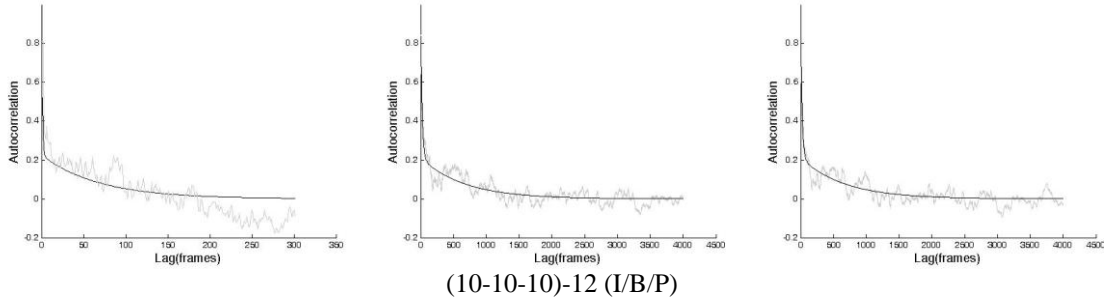
Quantization Settings / Frame Types	I Frames		B Frames		P Frames	
	p	$\mu$	p	$\mu$	p	$\mu$
(10,10,10)-12	16.487	21499	15.584	14608	17.071	15895
(20,20,20)-12	6.468	22905	1.682	25572	2.549	26287
(30,30,30)-12	4.406	12235	0.813	7857	1.376	11874
(20,20,20)-3	6.597	22316	1.661	25808	2.542	26446
(20,20,20)-6	6.545	22683	1.676	25720	2.540	26404

**Table 3: Gamma model statistics overview of the encoded signals**



**Figure 5. Representative frame size histograms and Gamma models**

As a next step, the autocorrelation function is derived for each frame type. Three representative graphs for the case of quantization scale 10-10-10 and Intra-frame period equal to 12, appear in figure 6, suggesting that the autocorrelation exhibits a reduced decay rate beyond the initial lags.



**Figure 6. Representative Autocorrelation Graphs for each scaling case of I/B/P frames**

It is observed that in the case of H.264 streams, the autocorrelation functions follow the same decaying shape, as in previously studied encoding formats, i.e. H.261 [4] and MPEG-1[5]. In principle, this phenomenon can be successfully captured by a weighted sum of two geometric terms

$$\rho_k = w\lambda_1^k + (1-w)\lambda_2^k \quad \text{with } |\lambda_2| < |\lambda_1| < 1 \quad (2)$$

where  $\lambda_{1,2}$  are the decay rates, which can be complex in general. Moreover, equation 2 corresponds to the general form of the autocorrelation function for AR(2) models [21], [6] which have been observed to match well some aspects of video conferencing traffic.

In order to further investigate the appropriateness of equation 2, the corresponding relevant parameters were estimated (in the domain of real-valued decay rates) through a least-squares fit



to the autocorrelation samples for the first 300 lags for the case of I frames and the first 4000 lags for the case of B and P frames, respectively. Numerical results appear in Table 4, while the graphs of the fitted models are compared to the sample autocorrelations in figure 6, demonstrating a satisfactory fit between the experimental and the modeling functions. Furthermore, in all cases the matched model represented in 2 captures the long-term trends of the autocorrelation decay.

Since the aforementioned model of autocorrelation functions approximates successfully the long term decay rate, which is the most important for queuing design, there is no point for further pursuing the issue towards more complex models.

Autocorrelation Type /Quantization Scale	I	B	P
<b>10-10-10-12</b>	w = 0.2248 $\lambda_1=0.9855$ $\lambda_2=0.3884$	w = 0.2132 $\lambda_1=0.9985$ $\lambda_2=0.9526$	w = 0.2014 $\lambda_1=0.9985$ $\lambda_2=0.9467$
<b>20-20-20-12</b>	w = 0.2472 $\lambda_1=0.9828$ $\lambda_2=0.7216$	w = 0.1859 $\lambda_1=0.9987$ $\lambda_2=0.9658$	w = 0.1712 $\lambda_1=0.9987$ $\lambda_2=0.9594$
<b>30-30-30-12</b>	w = 0.2011 $\lambda_1=0.9819$ $\lambda_2=0.7417$	w = 0.2650 $\lambda_1=0.9981$ $\lambda_2=0.9457$	w = 0.1923 $\lambda_1=0.9984$ $\lambda_2=0.9465$
<b>20-20-20-3</b>	w = 0.1931 $\lambda_1=0.9973$ $\lambda_2=0.9093$	w = 0.1686 $\lambda_1=0.9980$ $\lambda_2=0.9683$	w = 0.1590 $\lambda_1=0.9982$ $\lambda_2=0.9482$
<b>20-20-20-6</b>	w = 0.2124 $\lambda_1=0.9942$ $\lambda_2=0.8377$	w = 0.1746 $\lambda_1=0.9988$ $\lambda_2=0.9672$	w = 0.1648 $\lambda_1=0.9986$ $\lambda_2=0.9566$

**Table 4: The parameters of the autocorrelation model**

Having performed I/P/B level analysis for various encoding schemes of H.264 video traffic and showing satisfactory matching between the experimental data and the corresponding quantitative mathematical data, a video traffic modeling of H.264 unconstraint video traffic can be performed.

## 4 Video Traffic Modeling

### 4.1 Related Work

During the last years, video traffic modeling has been an active research area, where several efficient models for VBR traffic have been proposed in the literature. Broadly speaking, these models can be classified into the following categories [1]:

i. Auto-regressive models

The Discrete Auto Regressive(1) –DAR(1) model is the most primitive model that was proposed for modeling of video conference traffic. In [7], such a model is tested for 10-second long sequences based on the AR process:

$$x(n) = a_1 x(n-1) + e(n) \quad (3)$$

where  $x(n)$  is the bit rate of the coded video during the  $n^{\text{th}}$  frame,  $e(n)$  is a Gaussian process with variance  $\sigma^2$  – usually described as residuals- and  $a_1$  is the autocorrelation coefficient at lag-1, when the sequence is considered as stationary. Moreover, this model can be considered as a continuous-state, discrete-time Markov process, which is a special case of the general form

$$x(n) = \sum_{i=1}^p a_i x(n-i) + e(n) \quad (4)$$

where  $p$  is the order of the AR process. In this statement, the residuals  $e(n)$  are assumed uncorrelated and normally distributed, while Xu *et al* proposed a Gamma distribution residual process instead of a normal distribution in the AR process, analyzing a first-order Gamma AR (GAR) model of the form:

$$\phi_e(s) = a^3 + 3a^3(1-a)\frac{\beta}{\beta+2} + 3a(i-a)^2\left(\frac{\beta}{\beta+s}\right)^2 + (1-a)^3\left(\frac{\beta}{\beta+s}\right)^3 \quad (5)$$

where  $\phi_e(s)$  is the Laplace transform of the residual process,  $\alpha$  is the AR coefficient and  $\beta$  is the scale parameter. Although the GAR model outperformed in some cases the DAR(1), the fact that the residual generation is too complicated if the order of the AR is increased discourages its use and practical implementation.

Moreover, the Gamma Beta AR (GBAR) process has been proposed in [23], where it is assumed that the AR coefficients  $A_n$  are Beta distributed and the residuals  $B_n$  are Gamma distributed. More specifically, it is proposed that

$$X_n = A_n X_{n-1} + B_n \quad (6)$$

Although the GBAR model is shown to be more efficient than GAR model, it is not suitable for studying admission control algorithms particularly in ATM networks.

Finally, a general AR model has been described by [24], where firstly the model decomposes the given Gamma process into a weighted sum of a number of  $x^2(1)$  sequences. Afterwards, each sequence element is obtained by squaring a Gaussian process, which is efficiently generated by using an AR model from the given covariance matrix.

ii. Markov models

Furthermore, in the literature many Markov chains based models have been presented, which use states for representing bit-rate range. These models demonstrate that the current state of a Markov process depends only on its previous state, and not on any additional previous states. This can be briefly described by the Markovian property: a stochastic process  $X_k$  with state space  $S=\{1,2,3,\dots\}$  is Markovian if for every  $n$  and all states  $i_1, i_2, \dots$  satisfies the Markov property

$$P[X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_1 = i_1] = P[X_n = i_n | X_{n-1} = i_{n-1}] \quad (7)$$

A MC based model has been proposed in [25], which can also capture multiple video activity levels and scene changes. The transition matrix  $\mathbf{P}$  of this Markov chain is given by

$$\mathbf{P} = \rho \mathbf{I} + (1-\rho) \mathbf{Q} \quad (8)$$

Where  $\rho$  is the autocorrelation coefficient,  $\mathbf{I}$  is the identity matrix and each row of  $\mathbf{Q}$  consists of the negative binomial probabilities. The frame sizes remain constant during a scene but vary from one scene to another according to negative-binomial probabilities.

iii. Long range dependent models

Other works, in an attempt to track more closely the complexities observed in the long range autocorrelation decay, employ more elaborate means such as the autoregressive moving-average process (ARMA), which adds a moving average giving

$$x(n) = a_0 + \sum_{i=1}^p a_i x(n-i) + \sum_{j=0}^q b_j e(n) \quad (9)$$

where  $p$  is the order of the AR part and  $q$  is the order of moving average part. Such a model was described in [26].

iv. Hybrid models that combine both Markov and regression models

Also, more complicated models that combine both Gaussian AR(1) models with parameters modulated according to the transitions of a MC have been proposed in the literature. [12], [27], [28].

In comparison with the aforementioned models and methods, this paper will introduce a novel model that exploits three discrete DAR(1) processes for the generation of each I, B and P stream in conjunction with a Markov based decision mechanism for bit rate range and scene change simulation. Therefore, due to its novelty and hybrid nature, the proposed model is called by the authors as Markov Modified Model.

## 4.2 The proposed H.264 Markov Modified Model

We discussed in the previous sections the statistical properties of I, B and P frame types. This detailed traffic study and mathematical quantitative approach of H.264 VBR video is necessary

for understanding the properties of its characteristics, which will be used for generating synthetic H.264 traffic by an appropriate video model of H.264 Unconstraint VBR traffic. Thus, in order the proposed model to be efficient, it must fulfill the following criteria [1]:

- It must adopt specific statistical characteristics of the real video traffic, like PDF and ACF.
- The characteristics of the synthetic video model must be similar to the ones of the real video, so that it can be used instead of real video traffic for predicting desired performance metrics. This means that the proposed model must simulate successfully the long range dependencies (LRD) and short range dependencies (SRD) of the real video traffic.
- It must be as simple as possible and able to generate synthetic video traffic with low computational complexity.
- It must generate efficient synthetic video traffic for a wide range of video sources ranging from static video content up to very active.

According to the aforementioned criteria and based on the statistical analysis and quantitative mathematical approach of section 3, it is possible to generate sequences of stationary discrete random variables with Markov properties using the discrete autoregressive process of order 1, DAR(1).

The Markov properties are reflected by the fact that the distribution of  $F_n$ , of the size of the  $n^{\text{th}}$  frame, only depends on  $F_{n-1}$ . Such process is specified by the stationary marginal distribution of  $F_n$  and several other parameters which, independently of the marginal distribution, determine the correlation structure of the sequence.

According to the DAR(1) model, the first order autoregressive form is given by

$$F_n = V_n F_{n-1} + (1 - V_n) Y_n \text{ for } n=1, 2, \dots \quad (10)$$

where  $\{V_n\}$  are independent and identically distributed (iid) binary random variables with  $P(V_n = 1) = 1 - P(V_n = 0) = \alpha$  with  $0 \leq \alpha < 1$  and  $\{Y_n\}$  are iid random variables with a marginal distribution  $\pi$ .

In other words the model defines the current observation to be a mixture of two independent random variables: It is either the last observation with probability  $\alpha$ , or another independent sample from the same distribution. It is a very simple and general model since  $\pi$  is the distribution of any random variable and the correlation structure is independent of  $\pi$ . The autocorrelation function (ACF) of  $\{F_n\}$  as defined by (3) is given by  $\rho_F(k) = \alpha^k, k = 0, 1, \dots$  and  $\{F_n\}$  is a Markov chain with transition probability matrix given by  $aI + (1-a)Q$ , where  $I$  is the identity matrix and  $Q$  is a matrix of whose rows are the distribution  $\pi$ .

This approach is used independently for the study of the I, B and P frames trying to capture the fact that a video sequence consists of various scenes with different spatial and temporal activity levels and thus a candidate video traffic model should capture this *inter* and *intra* scene state. More specifically, during a scene, the sizes of the same frame type remain typically constant, while on the contrary follow different sizes over scenes changes.

In this respect, two discrete processes for frame generation are considered:

- The first process simulates the intra-scene state, which means that the frame size, referring to frames of the same type, retains the characteristics of the previous frame.

- The second process, which models the inter-scene state, the frame size is generated using a AR(1) process of the form  $x(n)=\alpha_1x(n-1)+e(n)$ , based on the size of the previous frame size, where  $\alpha_1$  is the autocorrelation parameter at lag-1 and  $e(n)$  a residual following the normal distribution.

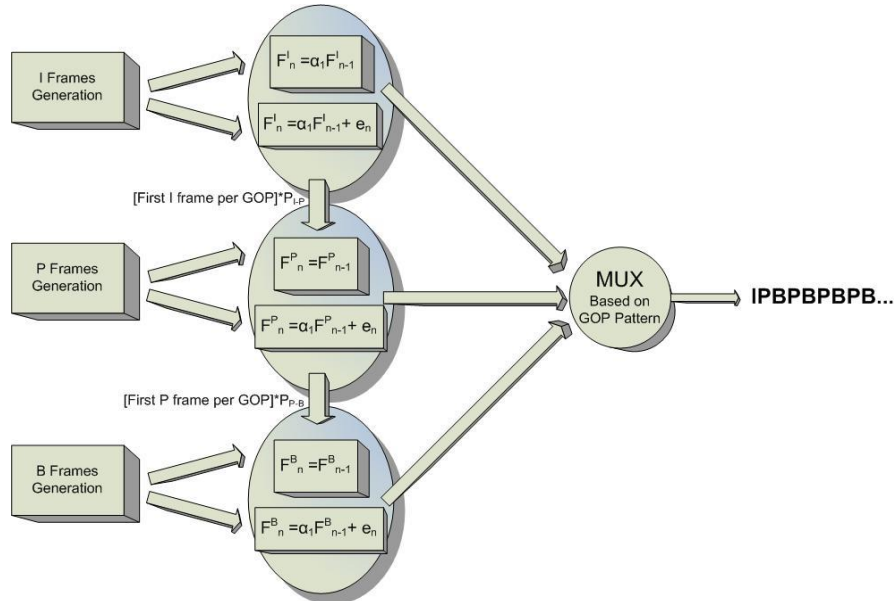
Especially for the case of I frames, which are strongly responsible for the *inter*-GOP correlation of the video stream, which is well characterized by the ACF of the I-frames, the first process is considered as  $I(n)=\alpha_1I(n-1)$ , where  $\alpha_1$  is the autocorrelation coefficient at lag-1, in order to better capture this phenomenon.

Although *inter*-GOP correlation, described by the ACF of the I frames, is an important measure, another aspect of video traffic is the correlation between I/P/B frames within the same GOP (the *intra*-GOP correlation). If uncorrelated I, P and B component generation is considered, then the estimated aggregate sequence will not represent successfully the burstness of the real data, resulting in poor estimation of the network resources and poor simulation of the *intra*-GOP dependence. So, a candidate model must adapt the correlation properties between the various frame types within every GOP.

For this reason, in our case the correlation coefficients are calculated between the first neighboring I-P, I-B and P-B frames of each GOP structure. Table 5 shows the corresponding results for various encoding schemes, showing that there is high correlation between the neighboring first I-P frames of each GOP and lower between the P-B, and even lower between I-B.

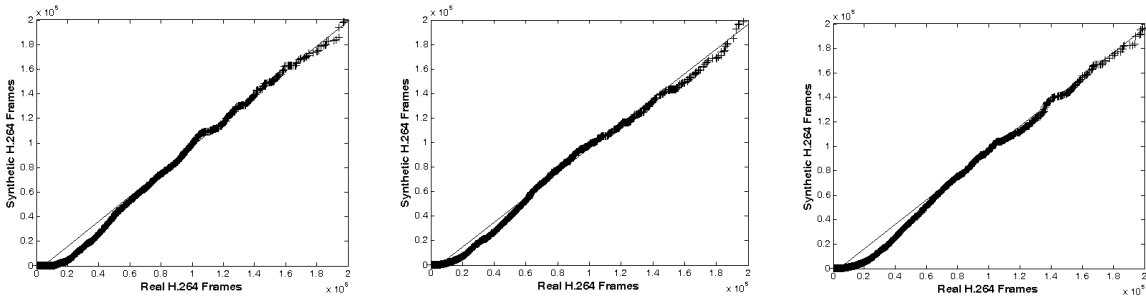
Quantization Scale	I-P Correlation Coefficient	I-B Correlation Coefficient	P-B Correlation Coefficient
(10-10-10)-12	0.6412	0.1750	0.1721
(20-20-20)-12	0.6250	0.0653	0.0948
(30-30-30)-12	0.5180	-0.0381	0.0916

**Table 5: Correlation Coefficient of Intra/Inter frames**



**Figure 7. Block Diagram of the proposed model**

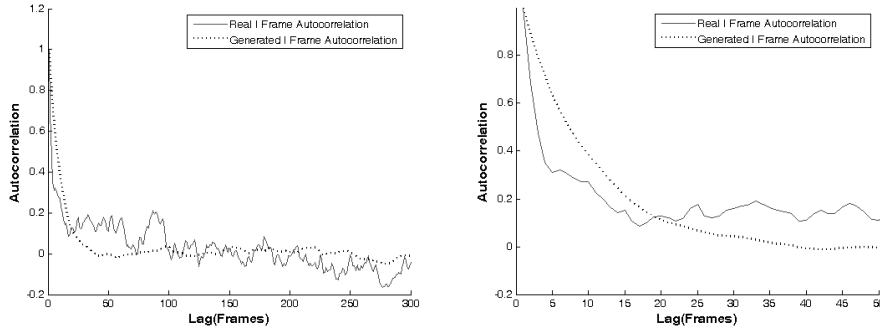
Figure 7 depicts the proposed algorithm for the generation of synthetic H.264 traffic. The proposed model, except from the use of the aforementioned DAR(1) models, it also exploits the *I-P* correlation coefficient and the *P-B* correlation coefficient for the generation of P and B streams respectively. More specifically, based on the GOP length and structure that the generated traffic has, the proposed algorithm, during a GOP initialization, propagates the first I and P frame size, multiplied by the corresponding correlation coefficient of Table 5, as feed to the respective P and B DAR processes. By this way, it manages to captures the *intra*-GOP dependences of the actual traffic, which is an important issue for the modeling of real traffic video characteristics.



**Figure 8. Q-Q plots of real and generated H.264 traffic**

In the proposed model, the MUX component of the described block diagram is responsible for the appropriate multiplexing of I, B and P streams according to a specific GOP pattern.

For the case of (20-20-20)-12, various synthetic traffics of 3 min duration each were generated by the proposed model and compared against various actual samples. Finally the corresponding Q-Q plots are depicted in Figure 8, reporting the good behavior of the proposed model.



**Figure 9. ACF of actual I-frame sizes and that of synthetic for the case of LRD and SRD**

Finally, Figure 9 shows the ACF of the actual I-frame sizes and that of synthetic traffic. As observed, the proposed model captures satisfactorily the *inter*-coded correlation of the actual H.264 traffic both at the cases of LRD and SRD.

Therefore, it has been shown that the latest H.264 video encoding standard can be satisfactorily modeled by adapting already known video traffic techniques on the characteristics of the new standard. This is expected because the synthesis of the statistical distribution of H.264 traffic remains similar to previous standards. More specifically, H.264 encoded traffic remains a superimposition of three different distributions that come from three different frames modes (i.e. I/B/P), exactly like earlier standards. Therefore, elaborating on this triple model separately makes possible the use of already proposed models and techniques, like the DAR(1) model and Markov

chains. All the new features of the H.264 standard aim at increasing compression efficiency in conjunction with improvement of the deduced perceived quality level for a specific encoding bit rate. Thus, H.264 by featuring variable block size support down to 4x4, quarter-sample motion vector accuracy, extended reference frame selection for P frames, de-blocking filter within the motion-compensated prediction loop and new context-based adapted entropy coding methods: CAVLC and CABAC, it succeeds better compression efficiency in comparison to the previous standards, without altering the distribution characteristics of the generated stream and the encoding scheme, which remains similar to the previous MPEG standards.

## 5 Conclusions

This paper reports on an experimental study of H.264 encoded video streams where additional statistical analysis established general results about the video traffic. The experiments covered cases with different quantization scales and GOP lengths, showing that the derived data can be expressed as superimposition of three discrete frame contributions. In this respect, the density functions of the I/B/P frame sizes were derived and it was shown that they can be successfully represented by Gamma distributions. Moreover, I/B/P autocorrelation functions were drawn, showing that they exhibit an exponentially decaying shape. Finally, a novel GOP adaptive model for generating and simulating H.264 VBR traffic is presented and evaluated by comparing real and generated video traffic.

## Acknowledgement

Part of this work was undertaken in the context of the ICT project ADAMANTIUM FP7-ICT/214751 project, which is partially funded by the Commission of the European Union.

## 6 References

- [1] A. Alheraish, "Autoregressive video conference models", *International Journal of Network Management*, Vol. 14, pp329-337, 2004.
- [2] H.S. Chin, J.W. Goodge, R. Griffiths and D.J. Parish, "Statistics of video signals for viewphone-type pictures", *IEEE Journal on Selected Areas in Communications*, Vol.7, No.5, pp826-832, 1989.
- [3] W. Verbiest, L. Pinnoo and B. Voeten, "The impact of the ATM concept on video coding", *IEEE Journal on Selected Areas in Communications*, Vol.6, No.9, pp1623-1632, 1988.
- [4] C. Skianis, K. Kontovasilis, A. Drigas and M. Moatsos, "Measurement and Statistical Analysis of Asymmetric Multipoint Videoconference Traffic in IP Networks", *Kluwer, Telecommunication Systems*, Vol.23, pp95-122, 2003.
- [5] N.D. Doulamis, A.D. Doulamis, G.E. Konstantoulakis and G.I. Stassinopoulos, "Efficient Modeling of VBR MPEG-1 Coded Video Sources", *IEEE Transactions on Circuits Systems Video Technology* Vol.10, No.1, pp93-112, 2000.
- [6] D.P. Heyman, A. Tabatabai and T.V. Lakshman, "Statistical analysis and simulation study of video teleconference traffic in ATM networks", *IEEE Transactions on Circuits Systems Video Technology*, Vol.2, No.1, pp49-59, 1992.
- [7] B. Maglaris, D. Anastassiou, P. Sen, G. Karlsson and J.D. Robbins, "Performance models of statistical multiplexing in packet video communications", *IEEE Transactions on Communications*, Vol.36, No.7, pp834-843, 1988.
- [8] M. Nomura, T. Fujii and N. Ohta, "Basic characteristics of variable rate video coding in ATM environment", *IEEE Journal on Selected Areas in Communications*, Vol.7, No.5, pp752-760, 1989.
- [9] D.M. Cohen and D.P. Heyman, "Performance modeling of video teleconferencing in ATM networks", *IEEE Transactions on Circuits Systems Video Technology*, Vol.3, No.6, pp408-422, 1993.

- [10] B.G. Haskell, "Buffer and channel sharing by several interframe picturephone coders", *Bell Systems Technical Journal*, Vol.51, No.1, pp261–289, 1972.
- [11] D.M. Lucantoni, M.F. Neuts and A.R. Reibman, "Methods for performance evaluation of VBR video traffic models", *IEEE/ACM Transactions on Networking*, Vol.2, No.2, pp176–180, 1994.
- [12] G. Ramamurthy and B. Sengupta, "Modeling and analysis of a variable bit rate video multiplexer", in: *Proc. of the 7th Internat. Teletraffic Congress Seminar*, Morristown, NJ., 1990.
- [13] R.M. Rodriguez-Dagnino, M.R.K. Khansari and A. Leon-Garcia, "Prediction of bit rate sequences of encoded video signals", *IEEE Journal on Selected Areas in Communications*, Vol.9, No.3, pp305–314, 1991.
- [14] T. Wiegand, G. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC Video Coding Standard", *IEEE Transactions on Circuits and Systems for Video Technology*, Special Issue in H.264, 2003.
- [15] B. Girod, E. Steinbach, N. Faerber. "Performance of the H.263 Video Compression Standard", *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*. Special issue on recent development in video: Algorithms, Implementation and Applications. Springer Editions. no. 17. pp.101-111, 1997.
- [16] E. Steinbach, N. Faerber, B. Girod. "Standard Compatible Extension of H.263 for robust video transmission in Mobile environments". *IEEE Transactions on Circuits and Systems for Video Technology*. Vol.7, no.6, pp.872-881, 1997.
- [17] G. J. Sullivan, P. Topiwala, and A. Luthra, "The H.264/AVC Advanced Video Coding Standard: Overview and Introduction to the Fidelity Range Extensions", *SPIE Annual Conference on Applications of Digital Image Processing XXVII*, Special Session on Advances in the New Emerging Standard H.264/AVC, pp454-474. 2004.
- [18] Y. Su, M. T. Sun. "Encoder Optimization for H.264/AVC Fidelity Range Extensions". *Proceedings of VCIP 2005*, Beijing, China, 2005.
- [19] G. Sullivan, "The H.264/MPEG-4 AVC video coding standard and its deployment status", *Proceedings of VCIP 2005*, Beijing, China, 2005.
- [20] Y.L. Lai, Y.Y. Tseng, C. W. Lin, Z. Zhou, M.T. Sun, "H.264 Encoder Speed-UP via Joint Algorithm/Code-Level Optimization" in *Proceedings of VCIP 2005*, Beijing, China, 2005.
- [21] D.R. Cox and H.D. Miller, *The Theory of Stochastic Processes*, Chapman & Hall, London, 1965
- [22] S. Xu, Z. Hung, "A gamma autoregressive video model on ATM networks". *IEEE Transactions on Circuits and Systems for Video Technology*, Vol 8, no. 4, pp.138-142, 1996.
- [23] E. McKenzie, "Autoregressive moving average processes with negative binomial and geometri mariginal distributions", *Advances in Applied Probability*, Vol. 18, pp.679-705, 1986.
- [24] QT. Zhang, "A general AR-based technique for the generation of arbitrary gamma VBR video traffic in ATM networks". *IEEE Transactions on Circuits and Systems for Video Technology*. Vol. 9(10), pp. 1130-1137, 1999.
- [25] P. Sen, B. Maglaris, N. E. Rikli and D. Anastasiou, "Models for packet switching of variable bit rate video in ATM networks", *IEEE Journal on Selected Areas in Communications*, Vol. 7, no. 5, pp.865-869, 1989.
- [26] R. Grunfelder, J. P. Cosmas, S. Manthorpe, A. Odinma-Okafor. "Characterization of Video Codecs as Autoregressive Moving Average processes and Related Queuing System Performance". *IEEE Journal on Selected Areas in Communications* Vol. 9(3): 284–293, 1991.
- [27] C. Shim, I. Ryoo, J. Lee and S. Lee, "Modeling and call admission control algorithm of variable bit rate video in ATM Networks", *IEEE Journal on Selected Areas in Communications*, Vol.12(2) pp.332–344, 1994.
- [28] F. Yegenoglu, B. Jabbari and Y.-Q. Zhang, "Motion-classified autoregressive modeling of variable bit rate video", *IEEE transactions on Circuits Systems Video Technology*, 3(1) pp42–53, 1993.
- [29] M. Haag, "Stationary and Nonstationary Random Processes", Version 2.1, online available at <http://cnx.rice.edu/content/m10684/latest/>, 2002.