Video Traffic Model for MPEG4 Encoded Video

C. H. Liew, C. Kodikara, A. M. Kondoz
Centre for Communication Systems Research
University of Surrey
Guildford, Surrey, GU2 7XH, UK
{c.liew,c.kodikara,a.kondoz}@surrey.ac.uk

Abstract— **To date, video traffic models in the literature have mostly considered the autocorrelation modeling of empirical video traffic for fixed source quantization parameters. Existing models also ignore the inter-dependencies between I-, P-, and B- frame types of MPEG4 coding, which have great impact on empirical queuing performance prediction accuracy. We propose a new Video Traffic Model (VTM) that is capable of generating output video traffic for wide range of quantization parameters in real time, while at the same time capturing the inter-dependencies between different frame types. The VTM performance is evaluated by means of packet loss rate prediction accuracy. Some existing models are implemented for performance benchmark. Simulation results show that the VTM captures empirical video traffic characteristic accurately and outperforms the existing models. Simulation results also show that the models that ignore the inter-dependencies between frame types can greatly underestimate the empirical packet loss rate.**

Keywords-component; video traffic model, mpeg, wireless video communication

I.   INTRODUCTION

Multimedia services are seen to be the revenue generating service for current and future networks. This can be observed from operator efforts to support services like video on demand and digital video broadcast (DVB-H). In order to study and evaluate the performance of video applications over wired and wireless communication systems, an accurate video traffic model is required. In this paper, we propose a Video Traffic Model (VTM) that is capable of synthetisizing MPEG4 video traffic in real time for a range of quantization parameters. This will enable use of VTM   for study of bandwidth adaptive video transmissions such as wireless video streaming, where video source rate is periodically and adaptively matched to the time-varying channel bandwidth by varying the quantization parameters in real time. In contrast, existing models, e.g. [1-3], have mostly modeled the video traffic characteristic for a fixed set of quatisation parameters (i.e. VBR video) and do not provide the adaptive capability of the proposed VTM. Secondly, the existing video traffic models have ignored inter-dependencies between I-, P-, and B- frame types of MPEG4 coding. Ignoring the inter-dependencies between I-, P- and B- frame types leads to underestimation of empirical traffic burstiness, and consequently the empirical packet loss rate prediction accuracy. In contrast, the VTM models the the inter-dependencies by using a Multinomial Method (MM) [4]. The inter-dependencies modeling using MM greatly improves the model accuracy in predicting the
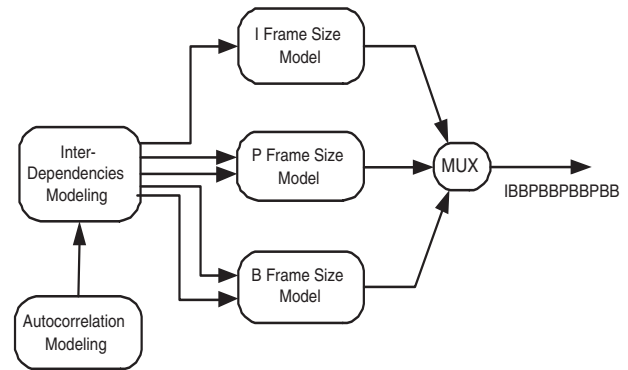

Figure 1.   The proposed Video Traffic Model (VTM)

empirical packet loss rate as will be shown later in this paper. Simulation has been performed to evaluate the proposed VTM as well as comparing it against the existing models. Three existing models, GOP-method [1], Nested-AR [2], FARIMA [3] have been implemented for performance comparisons. Nested-AR, FARIMA and GOP-method represent different cases of autocorrelation and inter-dependencies modeling. Simulation results show that the VTM captures the video traffic burstiness accurately and is superior to the existing models.

The proposed GVTM has three major parts as shown in Fig. 1. The first part consists of I-, P-, and B- frame size models. The second part models the inter-dependencies between different frame types using a Multinomial Method (MM). The final part models the autocorrelation structure of video traffic.

The rest of the paper is organized as follows. In Section II, we describe the frame size model. In Section III and VI, we discuss the inter-dependencies and autocorrelation modeling techniques. The summary for video traffic generation is presented in Section V. Finally, simulation results and discussions are presented in Section VI.

II.   I-, P-, B- FRAME SIZE MODEL

In this section, we propose a generalized frame size model that can synthesize I-, P- and B- frame sizes for a quantization parameter range, $Q \in [1,31]$. First, we introduce a frame activity concept. Secondly, we study the I-, P-, and B- frame composition and its relative importance to the output frame size. We then establish a relationship between the frame activity and the output frame size for a given quantization parameter. Finally the overall frame size

model is presented.

A. Frame Activity

Frame activity, A, can be used to measure frame visual complexity. In general, an un-quantized and un-compressed picture frame with large frame activity would result in a large output frame size. In this paper, we have adopted a DCT method [5] for frame activity calculation. Here, we denote A_I, A_P, and A_B as the frame activities for I-, P-, and B- frames calculated before the quantization process in an MPEG4 encoder [6].

B. Frame Size Composition

The basic I-, P-, and B- frame size compositions are shown in Fig. 2. The header bits contain information for frame decompression and are found to be almost constant from the empirical trace. The texture bits contain the compressed pixel information as a result of video coding. We jointly consider the header bits and texture bits as Texture Size, T_χ, where $\chi \in \{I, P, B\}$. The motion vector bits contain information for temporal reference and are only present in P-, and B- frames. The motion vector bits are denoted using M_χ. In general T_χ is found to be dependent on the quantization parameter of an MPEG4 encoder while M_χ does not dependent on the quantization parameter.
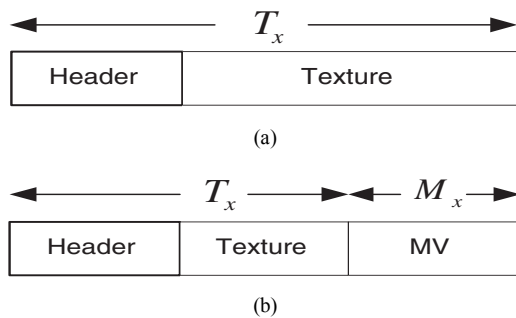

(a)

(b)

Figure 2. (a) I frame size composition. (b) P and B frame size composition

C. Relationship of Frame Activity to Texture Size

We have studied the relationship between the frame activity, A_χ, and the Texture size, T_χ, for the quantization range, $Q_\chi \in [1, 31]$, where $\chi \in \{I, P, B\}$. Fig. 3 portrays the plots of A_I against T_I. For the sake of brevity, we only plot curves for the quantization parameter 5, 15, and 25. We found that this family of curves can be closely fitted with a quadratic function. The corresponding quadratic curves are plotted in the same graph. Thus, for a given quantization parameter, Q_χ, the Texture size, T_χ, is related to the frame activity, A_χ, using the following equation:

$$T_\chi = C_0^{(\chi)} \times A_\chi^2 + C_1^{(\chi)} \times A_\chi + C_2^{(\chi)} \tag{1}$$

where $C_0^{(\chi)}$, $C_1^{(\chi)}$, and $C_2^{(\chi)}$ are the least square fitted coefficients for a given Q_χ. For I frame, the least square fitting is performed on the empirical trace for the quantization parameter, $Q_\chi \in [1, 31]$, and the calculated $C_0^{(\chi)}$, $C_1^{(\chi)}$, and $C_2^{(\chi)}$ are stored in an I- frame mapping table, TB_I. Same procedures are carried out for calculating the P and B frame mapping tables, TB_P and TB_B.

D. Marginal Distribution Modeling

The Cumulative Distribution Functions (CDFs) of the frame activity, A_I, A_P, A_B, and the motion vector size, M_P, M_B, are respectively found to be closely fitted to the Gamma distribution. The plots of Gamma fits are shown in Fig. 4.

E. The Complete Frame Size Model

In this section, we present the complete frame size model based on the previous discussions. The frame size, S_χ, can be represented mathematically using:
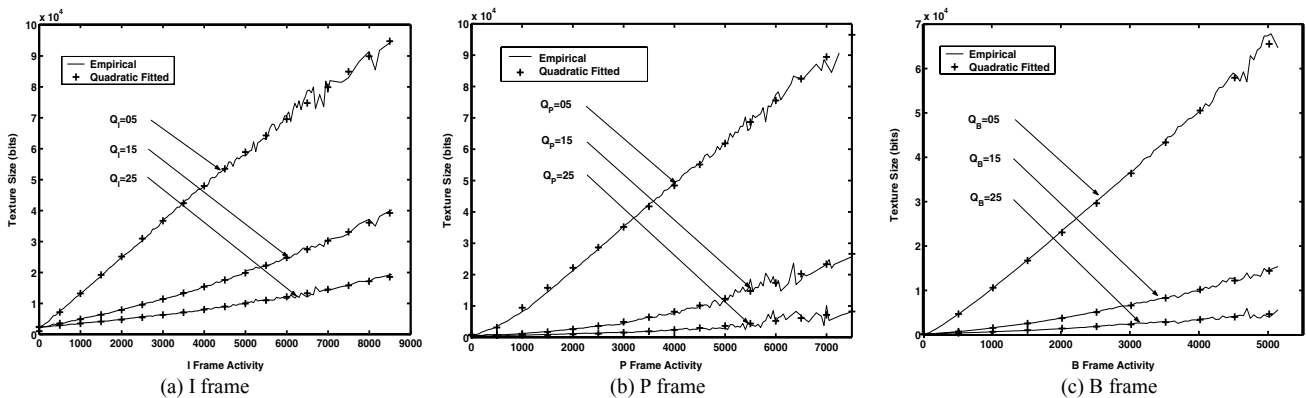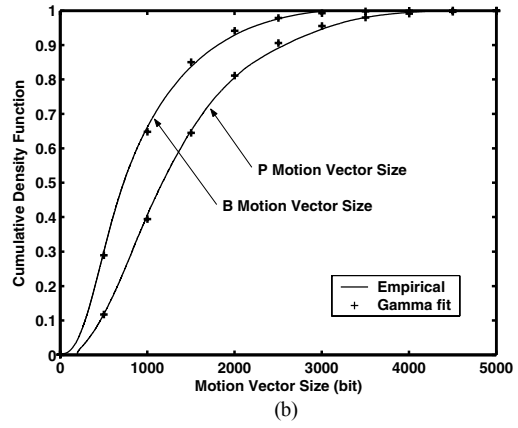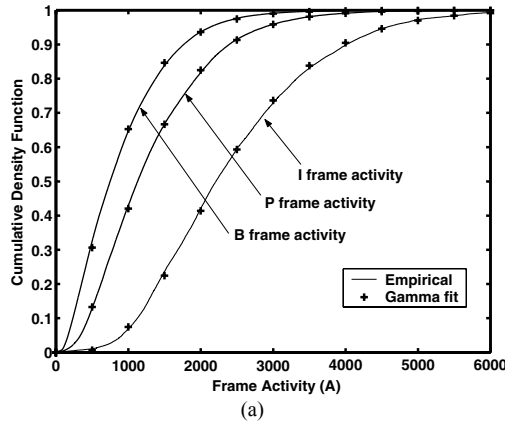

(a) I frame   (b) P frame   (c) B frame

Figure 3. Graphs shows the plot of frame activity against texture size and its corresponding quadratic fitting curves for different quantiser.

Figure 4. (a) I-, P-, B- frame activity CDF fitting. (b) P- and B- frame motion vector size CDF fitting

(a) I frame size model   (b) P and B frame size model
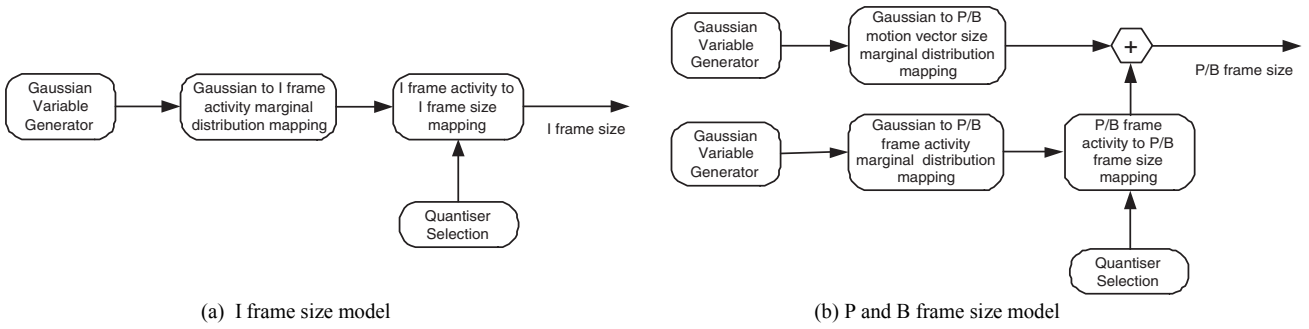
Figure 5. Frame size model for I-, P-, and B- frame.

$$S_\chi = T_\chi + I_\chi \cdot M_\chi \qquad (2)$$

where $\chi \in \{I, P, B\}$, and I_χ is an indicator function:

$$I_\chi = \begin{cases} 1 & \chi \in \{P, B\} \\ 0 & \chi \in \{I\} \end{cases} \qquad (3)$$

Fig. 5 shows the I-, P-, and B- frame size models graphically. We use the I- frame size generation as an example to explain the frame generation process. First, a random Gaussian variable, X_1, is generated. Then, X_1 is mapped to A_I using a probability integral transform [7]:

$$A_I = F_\Gamma^{-1}(F_G(X_1)) \qquad (4)$$

where F_Γ^{-1} and F_G are the inverse CDF A_I and the CDF of and Gaussian variable respectively. Given the quantization parameter, Q_I, A_I is then mapped to T_I using (1) and $C_0^{(I)}$, $C_1^{(I)}$, $C_2^{(I)}$ are obtained from the I- frame mapping table TB_I (explained in Section II-C). Finally, the I- frame size, S_I, can be calculated using (2). P-, and B- frame size S_P and S_B can be calculated in a similar way.

TABLE I.   COVARIANCE MATRIX, Σ

Covariance Matrix	A_I	A_P	A_B	M_P	M_B
A_I	1.00	0.84	0.76	0.01	-0.17
A_P	0.84	1.00	0.92	0.41	0.14
A_B	0.76	0.92	1.00	0.46	0.29
M_P	0.01	0.41	0.46	1.00	0.79
M_B	-0.71	0.14	0.29	0.79	1.00

III. INTER-DEPENDENCIES MODELING

The I-, P-, and B- frame sizes of MPEG4 encoded video are highly inter-dependent as shown in Fig. 6. It can be observed that the I-, P-, and B- frame sizes follow a same trend. This represents a traffic burtiness to the network. This burstiness can be attributed to the inter-dependencies of underlying A_I, A_P, A_B, M_P, and M_B. The inter-dependencies can also be observed from the normalized covariance matrix, Σ, shown in Table I. It can be observed that the normalized covariance value can be as high as 0.92. In light of this observation, we have used a Multinomial Method (MM) [4] for modeling the inter-dependencies between A_I, A_P, A_B, M_P, and M_B. We used MM to generate 5 correlated Gaussian variables
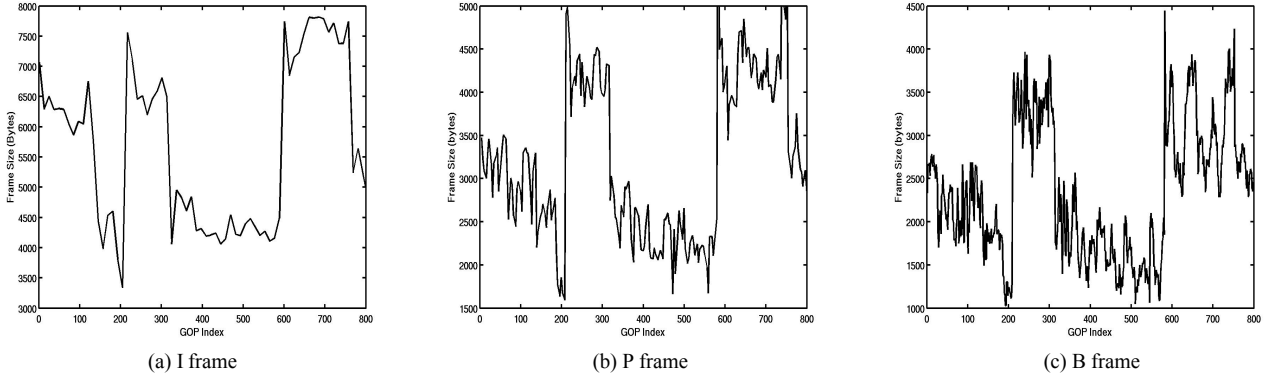
| (a) I frame | (b) P frame | (c) B frame |

Figure 6.   The plots show the I-, P-, and B- frame inter-frame dependency (i.e. following the same trend)

$X = (X_1, X_2, X_3, X_4, X_5)$, which will respectively be mapped to A_I, A_P, A_B, M_P and M_B such that they are also correlated. The MM is described as follows: Let $X = (X_1, X_2, X_3, X_4, X_5)$ be a correlated vector calculated using

$$X = LZ \qquad (5)$$

where L is a 5×5 weighting matrix and $Z = (Z_1, Z_2, Z_3, Z_4, Z_5)$ are zero mean and unity variance Gaussian variables. It can be proved that X has zero mean and LL^T covariance [4]. In order to generate X with the same cross correlation to the empirical trace, we equate the covariance matrix, Σ, from Table I to LL^T and solve for matrix L using Cholesky Decomposition [8]. The L matrix is then used in (5) to generate the correlated vector X.

IV. AUTOCORRELATION MODELING

The autocorrelation structures of I, P, and B frames are due to the autocorrelation structure of underlying A_I, A_P, A_B, M_P, and M_B. We have used a Spatial Renewal Process (SRP) [9] to model the AutoCorrelation Functions (ACFs) of A_I, A_P, A_B, M_P, and M_B. We refer the reader to [9] for detail discussions of SRP. The steps utilizing SRP for ACF modeling are as follows:

1. Calculate the background frame activity process for all the uncompressed raw video frames. Note that this calculation is performed on unprocessed video frames, i.e. without motion compensation in the MPEG4 encoder chain [6].

2. Calculate the ACF of background frame activity process. Use SRP to model the ACF.

3. Generate the sequence Z_1 using SRP. Z_2-Z_5 are randomly generated as zero mean and unity Gaussian variables.

4. $Z = (Z_1, ..., Z_5)$ is mapped to $X = (X_1, ..., X_5)$ using (5). Note that in this case X_1 derives its ACF directly from Z_1 whereas X_2-X_5 derive their ACFs from Z_1 by means of inter-dependencies modeling in (5).

5. Finally, X_1, X_2, X_3, X_4, and X_5 are mapped to A_I, A_P, A_B, M_P, M_B using (4).

The ACFs modeling of A_I, A_P, A_B, MV_P, and MV_B are thus achieved.

V. SUMMARY OF VTM TRAFFIC GENERATION

The traffic generation procedures are as follows:

1. Decide current frame type, χ, in the GOP, IBBPBBPBBPBB. Decide the quantization parameter, Q_χ, based on certain criteria, e.g. output frame bit rate matching the channel bandwidth.

2. Given the frame type, χ, and Q_χ, initialize the mapping coefficients $C_0^{(\chi)}$, $C_1^{(\chi)}$, and $C_2^{(\chi)}$ from the χ mapping tables, TB_χ, as discussed in Section II.C.

3. Generate Z_1 using SRP. SRP the models autocorrelation structure of background frame activity process (See Section IV). Generate Z_2-Z_5 randomly as zero mean, unity variance Gaussian variables.

4. Map $(Z_1, ..., Z_5)$ to $(X_1, ..., X_5)$ using (5). (5) models the cross correlation structure in $(X_1, ..., X_5)$. Go to Step 5 for I frame, Step 6 for P frame and Step 7 for B frame.
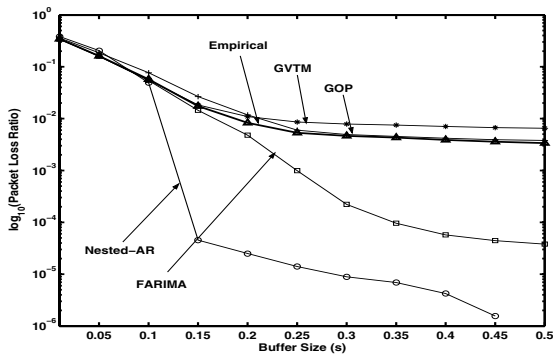
Figure 7. Plot of empirical packet loss rate prediction of VTM and existing models.

5. Map Gaussian variable X_1 to I frame activity A_I using (4). Then map I frame activity A_I to frame size S_I using (1), (2) and mapping coefficients $C_0^{(I)}$, $C_1^{(I)}$, $C_2^{(I)}$. See Fig. 5(a).

6. Map Gaussian variables X_2 and X_4 to P frame activity A_P and motion vector size M_P using (4). Then map P frame activity A_P to texture T_P using (1) and mapping coefficients $C_0^{(P)}$, $C_1^{(P)}$, $C_2^{(P)}$. Sum texture T_P and motion vector size M_P to obtain frame size S_P. See Fig. 5(b)

7. Map Gaussian variables X_3 and X_5 to B frame activity A_B and motion vector size M_B using (4). Then map B frame activity A_B to texture size T_B using (1) and mapping coefficients $C_0^{(B)}$, $C_1^{(B)}$, $C_2^{(B)}$. Sum texture size T_B and motion vector size M_B to obtain frame size S_B. See Fig. 5(b).

8. Repeat 1-7 for the total number of required frames.

VI. SIMULATION RESULTS AND DISCUSSION

We validate the proposed VTM by means of empirical packet loss rate prediction accuracy. We have implemented Nested-AR [2], FARIMA [3], GOP-method [1] for performance comparisons. Nested-AR, FARIMA and GOP-method represent three different cases. Nester-AR models the ACF of I frame while ignoring the ACFs of P and B frames. Nested-AR also ignores the inter-dependencies between different frame types. FARIMA models the ACFs of I-, P-, and B- frames, but ignores the inter-dependencies between different frame types. GOP-method, models the ACFs of I, P, and B frames, and also the inter-dependencies between frame types. Several video sequences have been tested. However, we only include the result from film sequence "Lord of the Rings: The Two Towers" due to space limitation. We have set the quantization parameters of VTM

for the I-, P-, and B- frames to fixed values of 5, 10 and 15 so that it can be easily compared to the existing VBR video traffic models. VTM, Nested-AR, FARIMA and GOP are respectively used to generate synthetic traffic to a FIFO queue at different bandwidth utilizations. The packet loss rate for all the models are recorded and compared to the empirical packet loss rate. We present the results for the bandwidth utilization of 40% in Fig. 7. VTM is shown to predict empirical packet loss rate accurately and is superior over the Nested-AR and FARIMA. The GOP-based method has similar performance to VTM, but it lacks the flexibility of VTM being able to reproduce video traffic for different quantization parameters in real time. Fig. 7 also shows that the model that ignores inter-dependencies between frame types, e.g. Nested-AR and FARIMA, can greatly underestimate the empirical packet loss rate. Similar results are found for the bandwidth utilizations 60% and 80%.

VII. CONCLUSION

In this paper, a new Video Traffic Model (VTM) that can generate video traffic for quantization parameter range [1,31] is proposed. This is important for study of adaptive video transmission where rate adaptation is achieved by varying quantization parameters at the encoder. The VTM also considers inter-frame dependencies between I-, P-, and B-frames. Simulation results show that the VTM accurately predicts empirical traffic characteristics and outperform the existing video traffic models that ignore the inter-frame dependencies.

ACKNOWLEDGMENT

REFERENCES

[1] O. Rose, "Statistical Propeties of MPEG Video Traffic and Their Impact on Traffic Modelling in ATM Systems", Report No. 101, Feb 1995.

[2] D. Liu, E. I. Sara, and W. Sun, 哲 ested auto-regressive processes for mpeg-encoded video traffic modeling,_ IEEE Trans. Circuits Syst., vol. 11, pp. 169–183, Febuary 2001.

[3] N. Ansari, H. Liu, Q. Shi, and H. Zhao, "On modeling mpeg video traffics," IEEE Trans. Broadcast., vol. 48, pp. 337–347, December 2002.

[4] E. M. Scheuer and Stoller, "On the generation of normal random vectors," Technometrics, vol. 4, pp. 278–281, May 1962.

[5] W. J. Kim, J. W. Yi, and S. D. Kim, "A bit allocation method based on picture activity for still image coding," IEEE Trans. Image Processing, vol. 8, pp. 974–977, July 1999.

[6] M. Ghanbari, Video Coding: An Introduction to Standard Codecs, 1st ed. IEE, 1999.

[7] A. M. Mood and A. G. Graybill, Introduction to The Theory of Statistics, 3rd ed. McGraw-Hill, 1974.

[8] R. K. Bock, W. Krischer, "The Data Analysis Briefbook", Version 16, April 1998.

[9] T. Taralp, M. Devetsikiotis, and I. Lambadaris, "Efficient fractional gaussian noise generation using the spatial renewal process," in Proc. of the IEEE International Conference on Communications, Atlanta, USA, June 1998, pp. 1456–1460.