Bipartite HMM Model for Burst Errors Focused on the Generation of Gaps and Clusters

N. Maciel, Elaine C. Marques, M. Grivet and Ernesto L. Pinto

***Abstract*— A new Hidden Markov Model (HMM) for burst errors is proposed. This model is based on a compact representation of the error sequence in terms of succeeding pairs of clusters and gaps lengths. Its Markov chain has two classes of states associated to the generation of gaps and clusters lengths, respectively. The proposed model may be characterized by few parameters. An algorithm for ML ("Maximum Likelihood") estimation of these parameters on the grounds of the EM ("Expectation-Maximization") approach is derived. Some preliminary results of performance evaluation show that the model and the estimation algorithm here presented provide a flexible and efficient tool for capturing and reproducing statistics of interest in the context of burst errors modelling.**

***Keywords*— Burst Error, HMM, ML Estimation**

I. INTRODUCTION

In several communications scenarios, such as wireless communications systems, the statistical properties of the error process have great impact on the overall performance of the system [1], [2]. In particular, the occurrence of burst errors in the lower-level layers may severely degrade the performance of the upper layers of the protocol stack.

Burst errors can be originated in the propagation environment (fading), in the impulsiveness of noise, in interferences, or even in processing techniques with intrinsic memory mechanisms, such as decoding of convolutional codes and decision feedback equalization [3].

The development of accurate mathematical models that reproduce the statistical properties of error samples is of great interest to evaluate the effect of burst errors on the performance of higher-level protocols and also to develop effective countermeasures.

The most commonly used mathematical models to represent error processes with memory are based on hidden Markov chains (HMM - Hidden Markov Model) [2]. In general, these models are adjusted to empirical data by Maximum Likelihood (ML) estimation of its parameters, and the Baum-Welch algorithm (BW) is the main tool employed for this aim [4].

To the best of our knowledge all previous works in this field are based on a binary representation of the error sample. However, these samples frequently have long intervals without errors, which lead to long runs of a same bit in the binary representation. Besides being inefficient, this representation also leads to numerical difficulties for parameter estimation, since the most used estimation algorithm (BW) performs several computations for each symbol in the error sequence [4]. Some works have dealt with this problem by proposing alternative estimation algorithms tailored to efficiently process error binary data with long sequences of correct bits [5], [6].

We propose a different approach to HMM modelling of burst errors, which is rooted on a parsimonious representation of the error sample in terms of clusters (strings of errors between two correct decisions) and gaps (blocks of correct decisions between two errors). The basic idea is to develop HMM models whose output is a sequence of gaps and clusters lengths. Our expectation is to obtain flexible models with few parameters to estimate.

As a first attempt to materialize this idea, we propose in this paper a new HMM model for burst errors with two classes of hidden states that are responsible for the generation of gaps and clusters lengths, respectively. Only transitions between states of different classes are assumed to occur. The proposed model may be parameterized by a small set parameters characterizing the conditional probability distributions of gaps and clusters lengths, besides the state-transition probabilities. An EM algorithm for Maximum Likelihood estimation of those parameters is derived. Preliminary results of performance evaluation are presented. They have been obtained by assuming that the conditional distributions or gaps and clusters lengths are geometric distributions. These results show that the HMM model and the estimation algorithm herein presented are potentially useful and advantageous in the sense of providing flexible modelling of burst errors with a reduced set of parameters.

The paper is organized as follows. Section II introduces some concepts and definitions of statistical parameters usually adopted to characterize burst errors. The proposed model is presented in section III. The estimation of its parameters is addressed in section IV. Simulation results of performance evaluation are given in section V. Concluding remarks and directions for future works are presented in section VI.

II. BURST ERRORS

An error sample is usually represented as a sequence of bits zero and one that indicate correct decisions and the occurrence of errors, respectively. An *error cluster (EC)* is a sequence where the errors occur consecutively, and has a length equal to the number of ones [7]. A *gap (G)* is defined as a string of consecutive zeros between two ones, having a length equal to the number of zeros [1]. An *error-free burst (EFB)* is defined

N. Maciel, Elaine C. Marques, M. Grivet and Ernesto L. Pinto, Brazilian Army Technological Center (CTEx), Pontifical Catholic University of Rio de Janeiro (PUC) and Military Institute of Engineering (IME), Rio de Janeiro, RJ, E-mails: {nmaciel, melaine}@ctex.eb.br, marcogrivet@gmail.com, ernesto@ime.eb.br

as a sequence of zero with a length of at least η bits, where η is a positive integer [8]. An *error burst (EB)* is a sequence of zeros and ones starting and ending with a "1", and separated from neighboring error bursts by error-free bursts [8], [9]. Fig. 1 illustrates these definitions.

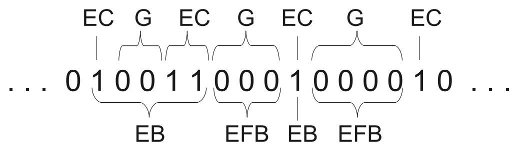

Fig. 1. An example of error sequence highlighting lengths of gaps (G), clusters (EC), bursts (EB) and error-free bursts (EFB), for $\eta = 3$.

Three commonly used burst error statistics are:

- the gap distribution, characterized by the probabilities of gap-lengths m_g, here denoted by $G(m_g)$ [1].
- the error cluster distribution, i. e., the probabilities of error-cluster lengths m_c, denoted by $C(m_c)$ [1].
- the autocorrelation function, denoted by $\rho(\Delta k)$,which is the conditional probability that the Δk bit following an error bit is also in error.

III. Proposed Model

Fig. 2 illustrates the proposed HMM model. It has two sets of hidden states associated with the generation of gaps and clusters, which are denoted by $\Omega_z \triangleq \{Z_1, Z_2, \ldots Z_M\}$ and $\Omega_u \triangleq \{U_1, U_2, \ldots U_N\}$, respectively.

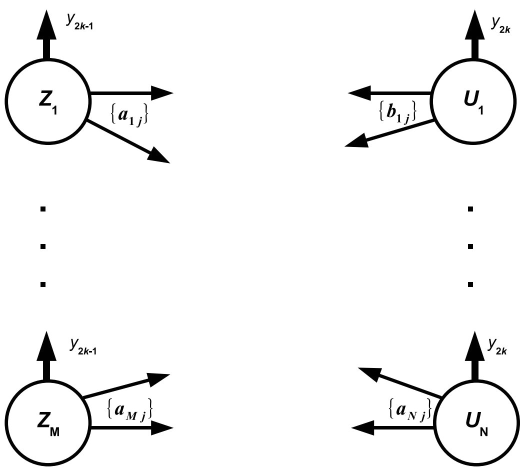

Fig. 2. The HMM model proposed.

The allowed transitions are between states of different classes only, so the state transition probability matrix is expressed as:

$$\mathbf{T} \triangleq \begin{pmatrix} \mathbf{0} & \mathbf{A} \\ \mathbf{B} & \mathbf{0} \end{pmatrix},$$

where $\mathbf{A} = [a_{ij}]_{M\times N}$, $\mathbf{B} = [b_{ij}]_{N\times M}$, $a_{ij} = P[e_{k+1} = U_j|e_k = Z_i]$ and $b_{ij} = P[e_{k+1} = Z_j|e_k = U_i]$.

Without loss of generality, we assume that the observed data begins with a gap. The observations at time indexes $2k-1$ and $2k$ correspond to lengths of succeeding gaps and clusters which are modelled as discrete random variables y_{2k-1} and y_{2k}. These random variables take values in the set {1, 2, 3,...} with conditional probabilities $P(Y_{2k-1}|e_{2k-1} = Z_i)$ and $P(Y_{2k}|e_{2k} = U_j)$.

The parameters of the conditional distributions of y_{2k-1} e y_{2k} form the vectors $\boldsymbol{\theta_z}$ and $\boldsymbol{\theta_u}$, respectively. Without loss of generality we also assume that these distributions are defined by a single parameter, so $\boldsymbol{\theta_z} = [\theta_z^i, i = 1, 2, \ldots, M]$ and $\boldsymbol{\theta_u} = [\theta_z^j, j = 1, 2, \ldots, N]$.

IV. Estimation Algorithm

Following the EM approach to ML estimation, we consider:

- as **incomplete data**, the random sequence of outputs of length $2K$, denoted by $\mathbf{y} = (y_1, y_2, \ldots, y_{2K})$;
- as **complete data**, the sequence of outputs and the corresponding sequence of states, denoted by $\mathbf{x} = (\mathbf{y}, \mathbf{e})$, where $\mathbf{e} = (e_1, e_2, \ldots, e_{2K})$.

We also define $\boldsymbol{\theta_y} \triangleq [\boldsymbol{\theta_z}, \boldsymbol{\theta_u}]$ and $\boldsymbol{\theta} \triangleq [\mathbf{T}, \boldsymbol{\theta_y}]$.

The function to be maximized in the M step of the algorithm is expressed as:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}') &\triangleq E_{\mathbf{x}|\overline{\mathbf{Y}},\boldsymbol{\theta}'}\{\ln P(\mathbf{x}|\boldsymbol{\theta})\} \\ &= \sum_{\mathbf{X}} \ln P(\mathbf{X}|\boldsymbol{\theta}) P(\mathbf{X}|\overline{\mathbf{Y}}, \boldsymbol{\theta}') \end{aligned} \tag{1}$$

where $\overline{\mathbf{Y}}$ is a sample of the "incomplete data" and $\boldsymbol{\theta}'$ the current vector of parameter estimates, used to evaluate the above shown expectation.

It should be noted that

$$\begin{aligned} P(\mathbf{X}|\boldsymbol{\theta}) = P(\mathbf{Y}, \mathbf{E}|\boldsymbol{\theta}) &= P(\mathbf{E}|\boldsymbol{\theta}) P(\mathbf{Y}|\mathbf{E}, \boldsymbol{\theta}) \\ &= P(\mathbf{E}|\mathbf{T}) P(\mathbf{Y}|\mathbf{E}, \boldsymbol{\theta_y}). \end{aligned}$$

On the other hand,

$$P(\mathbf{X}|\overline{\mathbf{Y}}, \boldsymbol{\theta}') = \begin{cases} P(\overline{\mathbf{Y}}, \mathbf{E}|\boldsymbol{\theta}'), & \text{if } X = (\overline{Y}, \mathbf{E}). \\ 0, & \text{if } X \neq (\overline{Y}, \mathbf{E}) \end{cases}$$

Using the last two equations in (1) we obtain:

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}') &= \sum_{\mathbf{E}} \ln P(\mathbf{E}|\mathbf{T}) P(\overline{\mathbf{Y}}, \mathbf{E}|\boldsymbol{\theta}') + \\ &+ \sum_{\mathbf{E}} \ln P(\overline{\mathbf{Y}}|\mathbf{E}, \boldsymbol{\theta_y}) P(\overline{\mathbf{Y}}, \mathbf{E}|\boldsymbol{\theta}') \end{aligned} \tag{2}$$

Regarding (2), it should be noticed that:

- The two parcels on the right side can be maximized separately in its parameters;
- The maximization of the first parcel leads to calculations that are similar to those used to re-estimate the state-transition probabilities within the Baum-Welch algorithm.

The maximization of the second parcel on right side of (2) is considered in the following, where it is denoted by $Q_y(\boldsymbol{\theta_y}, \boldsymbol{\theta}')$. The problem to be addressed is therefore to maximize in $\boldsymbol{\theta_y}$ the function:

$$Q_y(\boldsymbol{\theta_y}, \boldsymbol{\theta}') \triangleq \sum_{\mathbf{E}} \ln P(\overline{\mathbf{Y}}|\mathbf{E}, \boldsymbol{\theta_y}) P(\overline{\mathbf{Y}}, \mathbf{E}|\boldsymbol{\theta}'). \tag{3}$$

Maximization of $Q_y(\boldsymbol{\theta_y}, \boldsymbol{\theta}')$

We begin by expressing $P(\overline{\mathbf{Y}}|\mathbf{E}, \boldsymbol{\theta_y})$ as

$$P(\overline{\mathbf{Y}}|\mathbf{E}, \boldsymbol{\theta_y}) = \prod_{k=1}^{K} P[Y_{2k-1}|E_{2k-1}, \boldsymbol{\theta_z}] \prod_{k=1}^{K} P[Y_{2k}|E_{2k}, \boldsymbol{\theta_u}], \tag{4}$$

and using this expression in (3) to obtain

$$\begin{aligned} Q_y(\boldsymbol{\theta_y}, \boldsymbol{\theta}') \;&=\; \sum_{\mathbf{E}} \sum_{k=1}^{K} \ln P[Y_{2k-1}|E_{2k-1}, \boldsymbol{\theta_z}] P(\overline{\mathbf{Y}}, \mathbf{E}|\boldsymbol{\theta}') + \\ &+\; \sum_{\mathbf{E}} \sum_{k=1}^{K} \ln P[Y_{2k}|E_{2k}, \boldsymbol{\theta_u}] P(\overline{\mathbf{Y}}, \mathbf{E}|\boldsymbol{\theta}'). \end{aligned} \tag{5}$$

The two parcels on right side of (5) can be maximized separately, following similar approaches.

Considering only the first parcel and using an specific notation, the problem of interest consists on maximizing in $\boldsymbol{\theta_z}$ the function defined by

$$Q_z(\boldsymbol{\theta_z}, \boldsymbol{\theta}') \triangleq \sum_{\mathbf{E}} \sum_{k=1}^{K} \ln P[Y_{2k-1}|E_{2k-1}, \boldsymbol{\theta_z}] P(\overline{\mathbf{Y}}, \mathbf{E}|\boldsymbol{\theta}'). \tag{6}$$

It should be noted that this function may also be given by

$$\sum_{E_{2k-1}} \sum_{k=1}^{K} \ln P[Y_{2k-1}|E_{2k-1}, \boldsymbol{\theta_z}] G(\overline{\mathbf{Y}}, E_{2k-1}, \boldsymbol{\theta}') \tag{7}$$

in which $G(\overline{\mathbf{Y}}, E_{2k-1}, \boldsymbol{\theta}')$ is defined by:

$$\sum_{E_1} ... \sum_{E_{2k-2}} \sum_{E_{2k}} \cdots \sum_{E_{2K}} P[\overline{\mathbf{Y}}, E_1, E_2, \ldots, E_{2K}|\boldsymbol{\theta}'] \tag{8}$$

In fact, $G(\overline{\mathbf{Y}}, E_{2k-1}, \boldsymbol{\theta}') = P[\overline{\mathbf{Y}}, E_{2k-1}|\boldsymbol{\theta}']$, so the function to be maximized can be rewritten as:

$$Q_z(\boldsymbol{\theta_z}, \boldsymbol{\theta}') = \sum_{i=1}^{M} \sum_{k=1}^{K} \ln P[Y_{2k-1}|Z_i, \theta_z^i] P[\overline{\mathbf{Y}}, E_{2k-1} = Z_i|\boldsymbol{\theta}'] \tag{9}$$

At this point, it is worth to notice that the probabilities $P[\overline{\mathbf{Y}}, E_{2k-1} = Z_i|\boldsymbol{\theta}']$ can be efficiently computed, in a similar way to what is done in the BW algorithm.

To illustrate the maximization of $Q_z(\boldsymbol{\theta_z}, \boldsymbol{\theta}')$, i.e. the re-estimation of $\{\theta_z^i\}$ within this EM algorithm, we consider a model in which the lengths of "gaps" generated in the state Z_i are modelled by a geometric distribution of parameter θ_z^i given by:

$$\begin{aligned} P[Y_{2k-1}|Z_i, \theta_z^i] \;&=\; P[y_{2k-1} = Y_{2k-1}|Z_i, \theta_z^i] \\ &=\; {\theta_z^i}^{(Y_{2k-1}-1)}(1-\theta_z^i),\; i = 1, 2, ..., M \end{aligned}$$

In this case, $Q_z(\boldsymbol{\theta_z}, \boldsymbol{\theta}')$ can be expressed as:

$$\begin{aligned} Q_z(\boldsymbol{\theta_z}, \boldsymbol{\theta}') \;&=\; \sum_{i=1}^{M} [\hat{m}(y, \overline{Y}, Z_i|\boldsymbol{\theta}') - \hat{P}(\overline{Y}, Z_i|\boldsymbol{\theta}')] \ln \theta_z^i + \\ &+\; \hat{P}(\overline{Y}, Z_i|\boldsymbol{\theta}') \ln(1-\theta_z^i) \end{aligned} \tag{10}$$

with

$$\hat{P}(\overline{Y}, Z_i|\boldsymbol{\theta}') = \sum_{k=1}^{K} P[\overline{\mathbf{Y}}, E_{2k-1} = Z_i|\boldsymbol{\theta}'],$$

and

$$\hat{m}(y, \overline{Y}, Z_i|\boldsymbol{\theta}') = \sum_{k=1}^{K} Y_{2k-1} P[\overline{\mathbf{Y}}, E_{2k-1} = Z_i|\boldsymbol{\theta}'].$$

By solving the equation

$$\frac{\partial}{\partial \theta_z^i} Q(\boldsymbol{\theta_z}, \boldsymbol{\theta}') = 0 \tag{11}$$

in θ_z^i, we obtain:

$$\theta_z^i = \frac{\hat{m}(y, \overline{Y}, Z_i|\boldsymbol{\theta}') - \hat{P}(\overline{Y}, Z_i|\boldsymbol{\theta}')}{\hat{m}(y, \overline{Y}, Z_i|\boldsymbol{\theta}')} \tag{12}$$

Equation (12) corresponds to the updating of the estimates of $\{\theta_z^i\}$ for $i \in \{1, 2, \ldots, M\}$.

If the geometric-distribution model is also used for conditional cluster-length distributions, a similar equation should be employed for parameter updating.

Summing up, the proposed estimation algorithm uses operations similar to those of the BW algorithm to update the estimates of state transitions probabilities, and uses expressions like (12) to update the estimates of the other parameters, i. e. the parameters of the conditional distributions of gaps and clusters. The computation of auxiliary variables may also be made for efficiency improvement, in an identical fashion to what is done in the BW algorithm [4].

V. Simulations and Results

In this section we present the results of two experiments performed to evaluate the ability of the proposed model to capture the statistical properties of burst-error samples.

A. Experiment 1

In this experiment, the target error sequence was produced by an instance of the proposed HMM model with a cluster state (U) and two gap states (Z1 and Z2). The transition probability from U to Z1 was set at 0.8 and the parameters of the conditional distributions of outputs were $\theta_{z1} = 0.98$, $\theta_{z2} = 0.96$ and $\theta_u = 0.02$.

We used a target error sequence of gaps and clusters lengths with one million samples (500000 gaps and 500000 clusters). It turned into a binary error sequence of length $23,010,273$ with a bit error rate of 0.022. The log likelihood of the tested instance of the proposed model for the target error sequence was $-2,447,117$.

In order to check the consistency of the proposed EM algorithm, we applied it to estimate the parameters of this instance of the proposed model.

After 500 iterations we obtained the estimates $\hat{\theta}_{z1} = 0.979378$, $\hat{\theta}_{z2} = 0.949944$, $\hat{\theta}_u = 0.0200363$, and a log likelihood of $-2,447,134$. These results indicate that the proposed algorithm may produce precise estimates of the model parameters.

In the continuation of this experiment, we generated 100 independent sequences one million samples by running the model with the estimated parameters. The burst error statistics presented in section II were empirically evaluated using these samples, as well as the original target error sequence. The

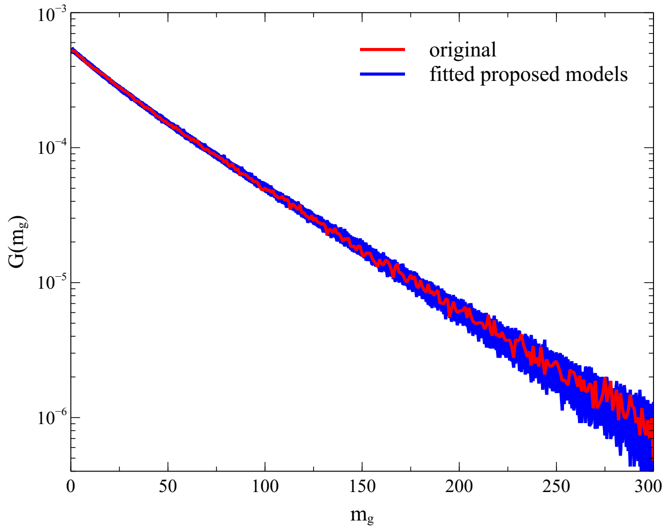

Fig. 3. Estimates of the gap distribution for Experiment 1.

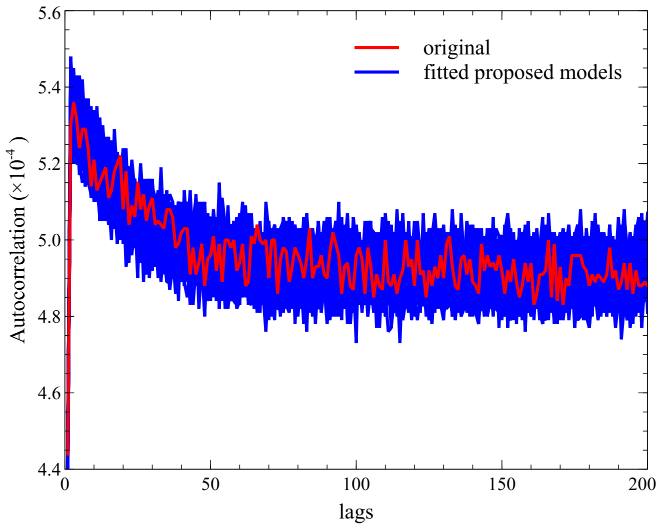

Fig. 5. Estimates of the error autocorrelation function for Experiment 1.

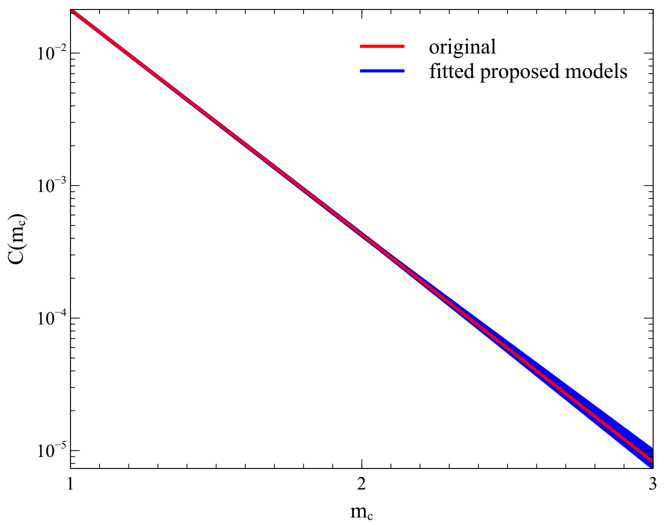

Fig. 4. Estimates of the cluster distribution for Experiment 1.

results so obtained are shown in Fig. 3, 4 and 5, in which the curves in red have been correspond to results obtained with the target error sequence and the curves in blue have been obtained from the error sequences generated with the estimated model.

On the basis of results of this experiment we can say that the estimation algorithm seems to work properly.

B. Experiment 2

The results of Experiment 1 motivated us to investigate the ability of the proposed model to capture the burstiness of a target error sequence produced by a fading channel. With this aim we generated a binary error sequence of length $10,000,000$ produced by a time-varying Rayleigh channel with Jakes' Doppler spectrum. The SNR was fixed at $20dB$ and the normalized maximum Doppler shift f_DT at 10^{-3}.

For the sake of comparison, we used this sequence to estimate the parameters of one instance of the proposed model and a Fritchman model [7]. The state-space of an N-state Fritchman model is partitioned into two groups: the first group is composed by k states that only generate error-free outputs, and the second group is composed of $N-k$ states that generate errors only. The allowed state transitions are between states of different groups, besides self-transitions.

In this experiment we considered a 5-state Fritchman model with 3 error-free states and 2 error states. In respect of the proposed model, we considered an instance with 3 gap-generating states and 2 cluster states. These two models have 12 parameters to be estimated. In the case of the Fritchman model, the BW algorithm was used for parameter estimation.

Table I shows the average processing time for each iteration obtained in this experiment. In this table "T_F" and "T_P" respectively denote the processing time of the BW algorithm, for the Fritchman model, and the processing time of the proposed EM algorithm, for the model here proposed. We can see that the processing time spent with the proposed model is much smaller that the one necessary for estimating the Fritchman model.

TABLE I

AVERAGE PROCESSING TIME FOR EACH INTERACTION IN EXPERIMENT 2.

T_F	T_P
150 seconds	7 seconds

Figures 6, 7 and 8 show the estimates of the burst-error statistics under consideration that have been obtained in this experiment. The procedure adopted to obtain these estimates was similar to that described in the previous section, except for the fact that only one sequence of outputs was generated with both the Fritchman model and the one here proposed. These figures show that the two models give rise to similar fits of burst-error statistical parameters.

These preliminary results show that the proposed model and algorithm are promising tools for burst error modelling. It should be noticed that several instances of the proposed model should be tried in order to well modelling error sequences, by

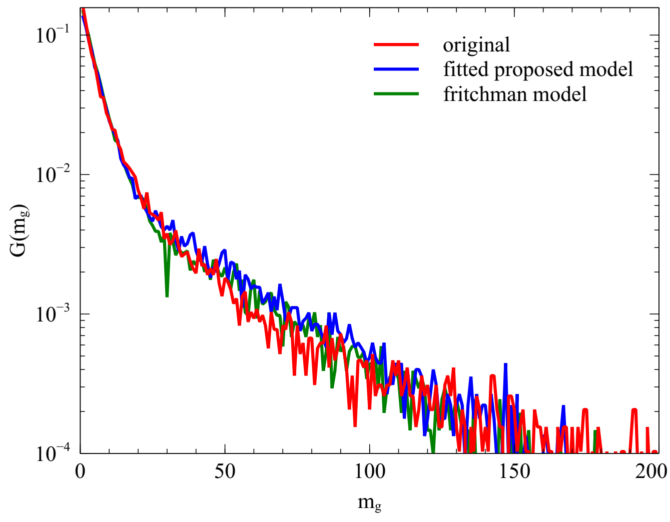

Fig. 6. Estimates of the gap distribution for Experiment 2.

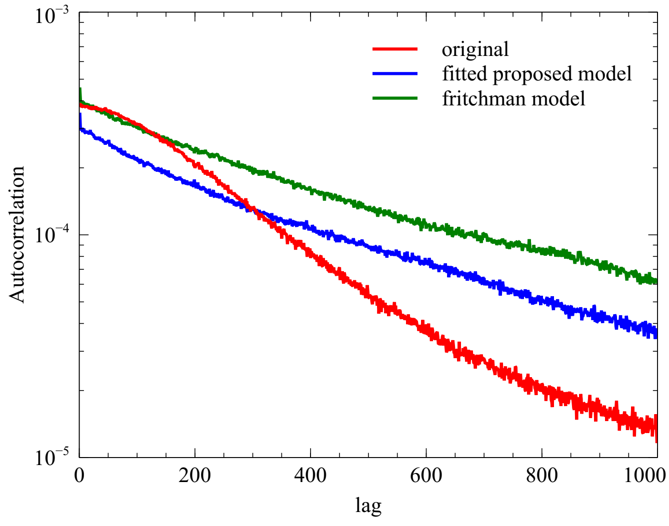

Fig. 8. Estimates of the error autocorrelation function for Experiment 2.

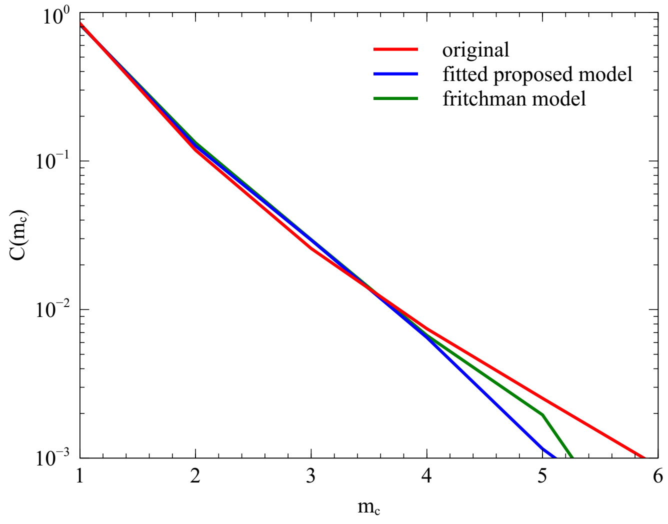

Fig. 7. Estimates of the cluster distribution for Experiment 2.

adopting different combinations of gap-generating and cluster-generating states, as well as using other distributions for modelling the lengths of gaps and clusters produced in each state.

VI. Conclusion

A new HMM model for burst errors has been proposed, which is based on the generation of succeeding pairs of gaps and clusters lengths and may be characterized by a set of few parameters to be adjusted to data. A simple algorithm for ML estimation of those parameters has been derived by following the EM approach. Some preliminary numerical results showed that the proposed model and the estimation algorithm may provide a flexible and effective tool for capturing and reproducing several statistics of interest for burst error modelling. In particular, it was verified that the model and estimation algorithms here proposed are able to produce reasonably good fits to data in a much smaller time than a Fritchmam model with the same number of parameters adjusted by the Baum-Welch algorithm. We stress that we have just started to exploit the potential of this model and there are several possibilities for amendments on its structure and/or in the distributions of gaps and clusters to be used. We have therefore good reasons to think that there is room for improvements and much better numerical results may be obtained on the grounds of the proposed methodology for burst error modelling. A more in-depth investigation of the performance characteristics of this modelling approach, as well as other applications to the burst-error processes generated in communication systems of current interest will be pursued in future works.

Acknowledgements

The authors would like to express their thanks to the Brazilian Council for Scientific and Technological Development (CNPq) for the partial financial support to this work.

References

[1] S. TSai, "Markov Characterization of the HF Channel," *IEEE Transactions on Communication Technology*, v.17, pp. 24-32, February 1969.
[2] O.S. Salih, C.-X. Wang and D. I. Laurenson, "Three Layered Hidden Markov Models for Binary Digital Wireless Channels," *IEEE International Conference on Communications*, pp. 1-5, June 2009.
[3] C. X. Wang and Wen Xu, "A New Class of Generative Models for Burst-Error Characterization in Digital Wireless Channels," *IEEE Transactions on Communications*,v.55, pp. 453-462, March 2007.
[4] J. Bilmes, "A Gentle Tutorial on the EM Algorithm Including Gaussian Mixtures and Baum-Welch," *ICSI Technical Report*, TR-97-021 1997.
[5] Fengqin Yang, Changhai Zhang and Ge Bai, "A Novel Genetic Algorithm Based on Tabu Search for HMM Optimization," *Fourth International Conference on Natural Computation*, v.4, pp. 57-61, 2008.
[6] S. Phon-Amnuaisuk, "Estimating HMM Parameters Using Particle Swarm Optimisation," *Proceedings of the EvoWorkshops 2009 on Applications of Evolutionary Computing*, 2009.
[7] B. D. Fritchman, "A Binary Channel Characterization Using Partioned Markov Chains," *IEEE Transactions on Information Theory*, v.13, pp. 221-227, April 1967.
[8] J. Garcia-Frias and Pedro M. Crespo, "Hidden Markov Models for Burst Error Characterization in Indoor Radio Channels," *IEEE Transactions on Vehicular Technology*, v.46, pp.1006-1020 1997.
[9] CCITT, "Terms and Definitions", *Blue Book*, v.1.