FFT-based Identification of Gilbert-Elliott Data Loss Models

András Palkó
Department of Measurement and Information Systems
Budapest University of Technology and Economics
Budapest, Hungary
palkoandras96@gmail.com

László Sujbert
Department of Measurement and Information Systems
Budapest University of Technology and Economics
Budapest, Hungary
sujbert@mit.bme.hu

***Abstract*—Nowadays, radio or Internet-based communication is gaining popularity in various fields, e.g., in signal processing. Due to the not reliable real-time communication, some of the data are lost during the transmission. There are several stochastic models allowing the analysis of this phenomenon. The spectral properties of these models result in specific distortion in the signals' spectra. As the spectrum of the signal can easily be calculated via the fast Fourier transform (FFT), FFT-based identification methods of the data loss can be developed. In this paper a new identification method is proposed for the simpler cases of the Gilbert-Elliott model class. The paper summarizes the mathematical descripiton of data loss, and introduces the Gilbert-Elliott model family. The novelty of the paper is the identification method that is based on the autocorrelation function of the Gilbert-Elliott model. Practical results are presented on the validation of the identified model parameters, and the required number of FFT points. The theoretical results are supported by simulation and measurement examples.**

***Index Terms*—data loss, Gilbert-Elliott model, hidden Markov model, identification, FFT, autocorrelation**

I. Introduction

In traditional measurement and communication systems reliable, high-precision and fast data transmission is used. Recently, due to the technological development and quickening of the networking process (e.g., sensor networks, Internet of Things), cheaper devices and less reliable data transmission protocols are spreading. This can result in the partial damaging of the message during the transmission, which is commonly called as data loss.

Roughly speaking, data loss is a measurement or communication error caused by the not reliable procedures and equipment. It is a many-faced phenomenon: e.g. packets can be lost over computer networks, radio communication can be impaired by interference, or the success of the measurements can depend on external circumstances. These external circumstances can lead to invalid or missing samples. A typical example for a missing sample is a lost packet over the Internet, while a typical invalid sample is generated by an overdriven AD-converter. Additionally, in a distributed system, where multiple clock domains are used, synchronization issues can lead to multiplied or skipped samples.

When some data are lost, it is an obvious idea to retry the measurement or transmission, but there are cases when it is impossible. If a signal is measured in real-time, the lost samples cannot be replaced. To achieve reliable transmission, one needs to implement a complex state machine, which may not be economical. Furthermore, reliable transmission is impossible within a finite time interval, thus in real-time systems data loss must be specially handled.

A likely practical application arises during the development of real-time systems. In the early stages of the development process, these systems can be tested against data loss via simulation. To achieve this, the model of the data loss is needed. Usually stochastic models are used, some examples are the Gilbert-Elliott model [1], the hierarchical hidden Markov model [2] or a two-level model created for UDP channels [3].

On the other hand, identification of the data loss model is necessary, given a measurement of a data loss process. The identified model can be later used in the development process for simulations. In a previous work [4], an FFT-based identification method has been presented for the two-state Markov and the random independent models. In [5] we described the properties of the Gilbert-Elliott model family and a simple identification method.

In this paper, we introduce the autocorrelation of the Gilbert-Elliott model. Using this result, we present an improved identification method for the simpler models of the Gilbert-Elliott model family. This method utilizes the FFT and offers higher precision than the one described in [5]. While that former method only used some specific points of the PSD, the proposed one uses all of its information, resulting in more accurate estimates.

The paper is arranged as follows: section II gives a possible mathematical description of the data loss, and the Gilbert-Elliott model class is presented. Section III deals with the autocorrelation function and the proposed identification method. Section IV illustrates the procedure with some simulation and measurement examples. The paper concludes in section V.

The project was funded by the European Union, co-financed by the European Social Fund (EFOP-3.6.2-16-2017-00013). The research reported in this paper was supported by the BME-Artificial Intelligence FIKP grant of EMMI (BME FIKP-MI/SC).

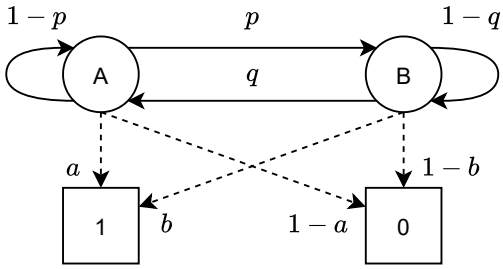

Fig. 1: Gilbert-Elliott data loss model

II. Preliminaries

A. Basic Definitions

Data loss can be described in discrete time with a so-called K_n availability indicator function:

$$K_n = \begin{cases} 1 & \text{if the sample is available at } n \\ 0 & \text{if the sample is lost at } n \end{cases} \tag{1}$$

With this indicator function we can define the μ data availability rate:

$$\mu = \mathbb{P}(K_n = 1) \tag{2}$$

where $\mathbb{P}(\cdot)$ is the probability operator. We can define further statistical properties, in this paper we will utilize the R_k autocorrelation function:

$$R_k = \mathbb{E}(K_n K_{n+k}) \tag{3}$$

where $\mathbb{E}(\cdot)$ is the expected value operator.

Data loss can easily be modeled as the product of the K_n data availability function and the x_{0n}discrete-time signal subject of data loss:

$$x_n = x_{0n} K_n \tag{4}$$

B. The Gilbert-Elliott Data Loss Model Family

1) Gilbert-Elliott Model: The Gilbert-Elliott data loss model is a two-state two-output hidden Markov model. In this paper, the states will be marked with A and B. The probabilities of a state transition from state A to state B and vice versa will be marked with p and q, respectively. The probabilities of getting an available sample in the two states will be a and b, respectively. Fig. 1 depicts this data loss model.

2) Simpler Models: The Gilbert-Elliott model is practically a model family. In this family simpler models are to be separated:

- **Gilbert-model [6]** When $a = 1$, the sample in state A is always available.
- **Complementary Gilbert-model** When $b = 0$, the sample in state B is always lost.
- **Two-state Markov model** When $a = 1$ and $b = 0$, the availability is a deterministic function of the state.
- **Random independent model** When $a = b$ or $p + q = 1$, the availability of different samples becomes independent.

As we can see, the Markov model is also a special case of the Gilbert and the complementary Gilbert models. Moreover, the random independent model is a special case of all of the mentioned models.

3) Basic Properties: The $\mathbf{P}$ transition probability matrix of the Gilbert-Elliott model is

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \tag{5}$$

We can get the π stationary state distribution from the $\pi\mathbf{P} = \pi$ equation:

$$\pi = \begin{bmatrix} \pi_{\mathsf{A}} & \pi_{\mathsf{B}} \end{bmatrix} = \begin{bmatrix} \frac{q}{p+q} & \frac{p}{p+q} \end{bmatrix} \tag{6}$$

The data availability rate of the Gilbert-Elliott model is

$$\mu = \pi_{\mathsf{A}} \mathbb{P}(1|\mathsf{A}) + \pi_{\mathsf{B}} \mathbb{P}(1|\mathsf{B}) = \frac{aq + bp}{p + q} \tag{7}$$

III. The Proposed Identification Method

The proposed identification method is an extension of the method described in [4]. This method is based on the autocorrelation of the Gilbert-Elliott model.

A. Autocorrelation of the Gilbert-Elliott model

The autocorrelation function of the Gilbert-Elliott model (for $k \geq 0$) is

$$R_k = \begin{cases} \frac{Z}{X} & k = 0 \\ \frac{Z^2}{X^2} + \frac{Y}{X^2}(1 - X)^k & k > 0 \end{cases} \tag{8}$$

where $X = p + q$, $Y = pq(a - b)^2$ and $Z = aq + bp$. For the derivation of the autocorrelation function see appendix A. Using the formula $\mu = \frac{Z}{X}$, we can rewrite the autocorrelation as

$$R_k = \mu^2 + \frac{Y}{X^2}(1 - X)^k + \left(\mu - \mu^2 - \frac{Y}{X^2}\right)\delta_k \tag{9}$$

where δ_k is the Kronecker delta.

We can determine the origins of the different terms by substituting the conditions for the simpler models into the autocorrelation expression. The constant μ^2 term simply comes from the nonzero mean of K_n.

If we take the conditions for the random independent model, the exponential term vanishes. When $a = b$ holds, $Y = 0$ is true, and in the $p + q = 1$ case, $(1 - X)^k$ becomes 0. Furthermore, the exponential term appears starting from the two-state Markov model, so we can conclude that the exponential term comes from the underlying Markov chain.

If we calculate the autocorrelation of the function $K_n - \mu$, we get the same result as (9), only without the constant μ^2 term. In this case, the $k = 0$ element of R_k is equal with the variance of K_n. The variance has two components: one comes from the exponential term, and the other comes from the term with δ_k. This is because there are multiple sources of randomness in the model: the state transitions and the random outputs in each state. As we have seen the exponential part comes from the state transitions, and if we express the coefficient of δ_k with the model parameters, we get

$$\mu - \mu^2 - \frac{Y}{X^2} = \pi_{\mathsf{A}} a(1 - a) + \pi_{\mathsf{B}} b(1 - b) \tag{10}$$

which means that the part of the variance associated with δ_k is the sum of the variances in each state weighted by the stationary distribution.

B. Overview

The steps of the proposed method are summarized below.

1) Handle the block-based data loss, decimate K_n if needed.
2) Calculate $\hat{\mu} = \text{mean}\,(K_n)$, $K'_n = K_n - \hat{\mu}$.
3) Obtain the $\hat{R}_k$ autocorrelation estimate.
4) Get the LPC (linear prediction filter) coefficients.
5) Examine the got coefficients, check for random independent model, get $\hat{X}$ and $\hat{Z}$ estimates.
6) Fit the $C(1-\hat{X})^k$ exponential to $\hat{R}_k$, calculate the $\hat{Y}$ estimate, check for two-state Markov model.
7) Calculate the model parameters.

Fig. 2 illustrates the method. Boxes marked with yellow contain the handling of the block-based data loss, red color shows the steps from the method described in [4], while blue color indicates the extension for the Gilbert and complementary Gilbert models.

Fig. 3 demonstrates the computation dataflow. Here $\hat{S}\,(f)$ denotes the PSD estimate.

C. The Proposed Method in Detail

Step 1. The case of the block-based data loss is to be handled. The data loss is block-based when the data are grouped into fixed size blocks and these blocks are either fully available or fully lost. In this case, we only need to know the availability of the blocks, so the indicator function can be decimated by the block size. For this step, the block size needs to be known, which is usually an a priori information, e.g., from the communication protocol. If for some reason the block size is unknown, it can be estimated as the greatest common divisor of the sequence lengths of the available and the lost samples.

Step 2. The value $\hat{\mu}$ is calculated as the mean of K_n, then subtracted from it:

$$K'_n = K_n - \hat{\mu} \tag{11}$$

The value of $\hat{\mu}$ will be used later for getting the $\hat{Z}$ parameter estimate. The subtraction removes the constant μ^2 term from the autocorrelation as its presence can impair the dynamic model fitting in step 4.

Step 3. The $\hat{R}_k$ autocorrelation of K'_n is calculated. It can be performed efficiently by getting the IFFT of the PSD of K'_n. With this way of estimation the autocorrelation, it is necessary for the autocorrelation to decay in the DFT block. Fortunately, as the decay is exponential, this condition is not critical.

Step 4. The $k > 0$ part of the autocorrelation function is approximated by an autoregressive system. In theory, the Gilbert-Elliott model class gives autocorrelation functions with no zeros and at most one poles. It has no pole if it is random independent and one pole otherwise. The model fitting can be done by using an LPC which determines the coefficients of a forward linear predictor by minimizing the prediction error in the least squares sense [7]. As the result of the LPC estimation we get the a_1, a_2, …, a_L coefficients, where L is the order of the model.

Step 5. The LPC coefficients are investigated. If the $a_3, a_4, \ldots$ coefficients are not negligible, the data loss does not belong to the Gilbert-Elliott model family. If the coefficient a_2 is negligible al well, we have a random independent data loss with data availability probability $\hat{\mu}$. Else, we can get the parameter estimate

$$\hat{X} = 1 + a_2 \tag{12}$$

From (7) we can get the next parameter estimate:

$$\hat{Z} = \hat{\mu}\hat{X} \tag{13}$$

Step 6. Estimation of the magnitude of the exponential term in the autocorrelation. This can be done by least squares fitting the $E_k = C(1-\hat{X})^k$ function to $\hat{R}_k$ for $k = 1, \ldots, M$, where M is the length of the autocorrelation. We get the value of constant C as:

$$C = \frac{\sum_{k=1}^{M} E_k \hat{R}_k}{\sum_{k=1}^{M} E_k^2} \tag{14}$$

If $\hat{R}_0 \simeq C$, then the coefficient of the δ_k term is negligible in $\hat{R}_k$, which means we have a two-state Markov model based data loss. As C is the estimate of the magnitude of the exponential term of the autocorrelation, we can get the parameter estimate

$$\hat{Y} = C\hat{X}^2 \tag{15}$$

Step 7. Calculation of the model parameters from the $\hat{X}$, $\hat{Y}$ and $\hat{Z}$ estimates. Unfortunately, as the Gilbert-Elliott model has four parameters, the model parameters are ambiguous. However, for the simpler models, this is an unambiguous problem. The Gilbert model parameter estimates can be calculated as

$$\hat{p} = \frac{\hat{X}(\hat{X}-\hat{Z})^2}{(\hat{X}-\hat{Z})^2+\hat{Y}} \quad \hat{q} = \frac{\hat{X}\hat{Y}}{(\hat{X}-\hat{Z})^2+\hat{Y}} \quad \hat{b} = \frac{\hat{Z}(\hat{X}-\hat{Z})-\hat{Y}}{\hat{X}(\hat{X}-\hat{Z})} \tag{16}$$

The complementary Gilbert model parameter estimates can be obtained as

$$\hat{p} = \frac{\hat{X}\hat{Y}}{\hat{Z}^2+\hat{Y}} \quad \hat{q} = \frac{\hat{Z}\hat{X}}{\hat{Z}^2+\hat{Y}} \quad \hat{a} = \frac{\hat{Z}^2+\hat{Y}}{\hat{Z}\hat{Y}} \tag{17}$$

Finally the two-state Markov model parameter estimates are

$$\hat{p} = \hat{X} - \hat{Z} \quad \hat{q} = \hat{Z} \tag{18}$$

D. Remarks

Steps 4 and 5 are a significant improvement over the method described in [5]. While the former method simply sampled the PSD is some specific points, the new one fits a model to the autocorrelation. It means that the new method uses all the information contained in the PSD, in contrast to only some specific values of it. This acts as a noise reduction which makes the precision of the parameter estimates much higher. Consequently, now we have a more reliable check for belonging to the Gilbert-Elliott model family, opposed to

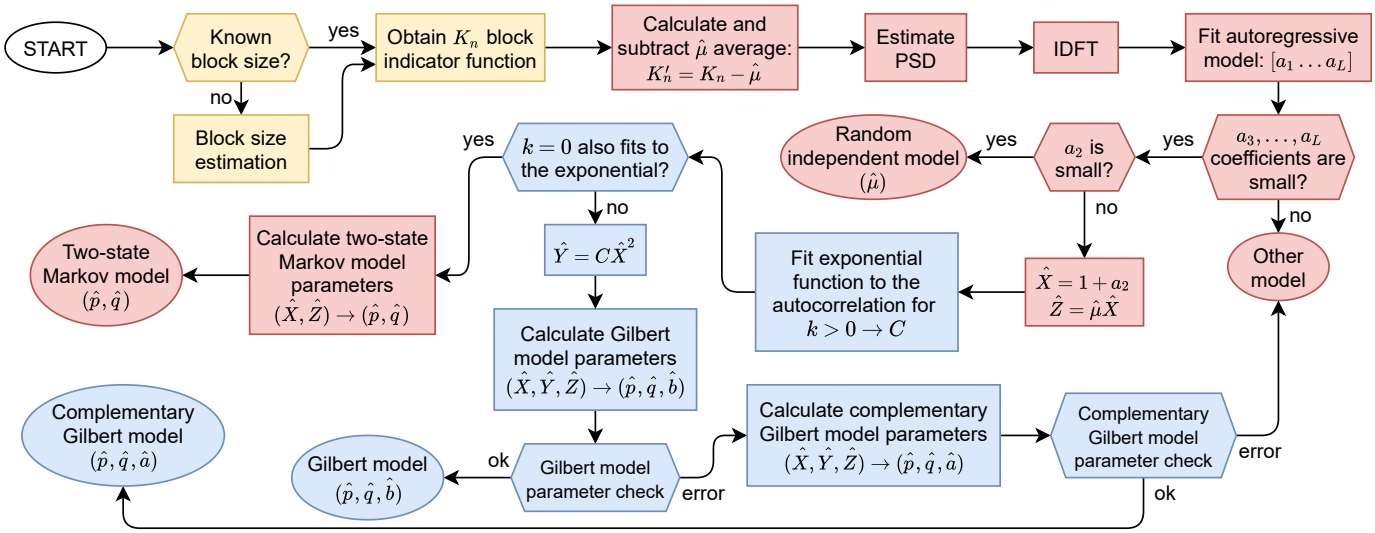

Fig. 2: The proposed method

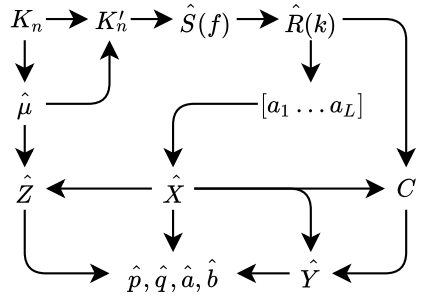

Fig. 3: The computation dataflow

TABLE I: Lower bound for X

$N \backslash r$	2	3	4	5
64	0.134	0.194	0.25	0.302
256	0.0353	0.0525	0.0694	0.086
1024	0.00895	0.0134	0.0178	0.0222
4096	0.00225	0.00337	0.00487	0.00561
16384	0.000562	0.000843	0.00112	0.0014
65536	0.000141	0.000211	0.000281	0.000351

forcing identification and only checking the model parameters afterwards for the obvious errors.

The DFT length required for the PSD can be estimated by the decay of the autocorrelation function. Provided that the exponential term can be neglected if $(1 - X)^k < 10^{-r}$, the length of the DFT is at least

$$N_{\min} = -\frac{2r}{\lg(1 - X)} \tag{19}$$

Based on our experience, usual DFT lengths (e.g., 1024 points) are satisfactory for the calculation of the PSD. Conversely, for a given N DFT length and r decay exponent, the following condition for the value of X can be given:

$$X \in \left[1 - 10^{-\frac{2r}{N}}, 1 + 10^{-\frac{2r}{N}}\right] \tag{20}$$

Table I shows the lower bound of this condition for some r and N values.

The choice of M in (14) is important. We should include that part of the autocorrelation where its value is high, but exclude the part where it decayed. If the decayed part is also included, on that part the model is fit to the noise, which is not a problem in itself, but it reduces the weight of the initial part which carries the important information. If M is too low, then too much of the initial part is left out.

Another possibility of estimating X and Y would be to take the logarithm of $\hat{R}_k$ and then perform a linear regression for $k > 0$. After removing the mean value from K_n, $\hat{R}_k$ should decay to 0 while staying positive. As noise can cause $\hat{R}_k$ to be negative, and the logarithm is sensitive to changes with these small values, this linear regression should also be performed to the first M values of $\hat{R}_k$. This could be the subject of future research.

IV. Examples

A. Simulation Results

1) The Simulation Environment: A simulation environment was created in MATLAB to test the proposed method. A series of simulations were executed with the Gilbert model with the following input variables:

- p, q, b Gilbert model parameters
- N_{DFT} DFT length
- N_{B} number of DFT blocks

During the simulation $N_{\text{DFT}}N_{\text{B}}$ samples of the indicator function of a Gilbert model with p, q, b parameters were generated. The initial state distribution was the stationary one. The autocorrelation estimate was calculated using FFT, with N_{B} not overlapping blocks of N_{DFT} samples.

The order of the resulted LPC model was 10 (or 7 – the maximum possible – for $N_{\text{DFT}} = 16$). The negligibility of the got a_3, a_4, ...coefficients was decided by comparing their Euclidean norm with 0.1. After calculation of the $\hat{p}$, $\hat{q}$, $\hat{b}$ parameter estimates, they were checked for being in the $[0, 1]$ range. If any of the parameters were outside this range or

the a_3, a_4, ...coefficients were not negligible, the parameter estimates were marked as invalids.

$M = 1000$ simulations were executed for each $[p, q, b, N_{\mathrm{DFT}}, N_{\mathrm{B}}]$ variable configuration. From this, the P_{invalid} probability of invalid results was calculated. Furthermore, from the valid results the mean and standard deviation of the parameter estimation errors were calculated.

2) Comparison with the "Three-Point" Method: An identical set of simulations were executed for the method described in [5]. That method is termed as the Three-Point method, because that method used only three specific values of the PSD. To demonstrate the improvement over the Three-Point method, the $p = 0.05, q = 0.03, b = 0.01, N_{\mathrm{DFT}} = 256$ cases were selected with N_{B} going from 1 to 1000 in 1–2–5 steps.

Fig. 4 shows the comparison of the two methods. As we can see on the topmost plot, this method has much better probability to acquire a valid estimation. We got valid results from all of the simulations with 20 or more DFT blocks, while the Three-Point one did not reach even 80% valid probability until 100 DFT blocks.

The lower three plots show the estimation errors of p, q and b. The lines show the mean error, the length of the error bars corresponds to one standard deviation. When only one DFT block was available, the Three-Point method did not produce valid results in any of the 1000 repetitions, thus there are no error data for these cases. A significant improvement can be seen in the accuracy of the p and q transition probability estimates, where the decrease in the mean and the standard deviation of the error is about an order of magnitude. A smaller, but still definite improvement can be seen in the accuracy of the b parameter estimate.

3) Accuracy with "Short" DFTs: Fig. 5 shows the results of simulations with parameters $p = q = 0.01$, $b = 0.2$, $N_{\mathrm{B}} = 1000$ and N_{DFT} going from 16 to 4096 by the powers of 2. The meaning of the error bars in the lower plot is the same as in Fig. 4.

As the figure shows, there are deterministic errors in the parameter estimates for the shorter DFTs. They are originated in the too small DFT length, which means that the autocorrelation did not decay in the DFT block. For $X = 0.02$ and $r = 2$, according to (19) we would need at least 456 DFT points. This length is drawn as a green line in the figure. As we can see, it clearly separates the plot to two parts with high and low errors.

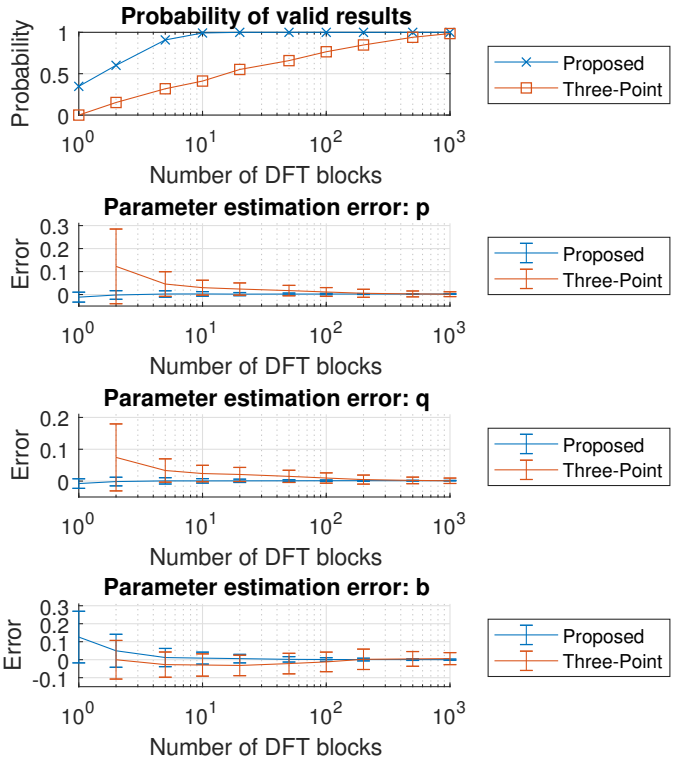

Fig. 4: Simulation – comparison: the proposed method requires less DFT blocks to obtain more precise estimates

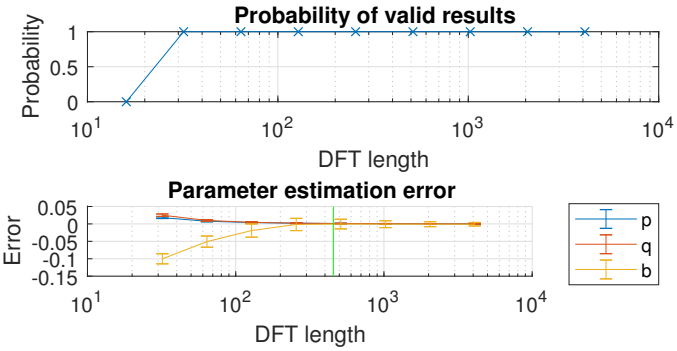

Fig. 5: Simulation – DFT length: using too small DFT length, the autocorrelation does not decay, which results in a biased estimate

B. Measurements

The performance of the proposed method is demonstrated by the measurement data used in [5]. These data were acquired by periodically sending UDP packets between two computers. The indicator function was obtained from the identifiers of the received packets. Several measurements were performed with varying packet size, sending interval, means of connection, parallel network activity, etc..

Fig. 6 illustrates the accuracy of the proposed method using the data of a measurement from the previously described set. This measurement incorporated a shared Internet connection of a mobile phone, the packets were sent between computers physically about 250 km apart. Both the Three-Point and the proposed method identified this data loss as a two-state Markov model. It can be clearly seen that the proposed method gave a more accurate estimate.

Fig. 7 shows the PSD of the indicator function from another measurement (wired Internet connection, 250 km distance, sending many packets in a short time), and the PSD of its identified model using the Three-Point method. The mentioned method identified this data loss as a Gilbert model, however, comparing the PSDs, it can be seen that this model is a poor fit disregarding the two bumps in the PSD at higher frequencies. On the contrary, the proposed method recognizes that this is not a Gilbert-Elliott type model.

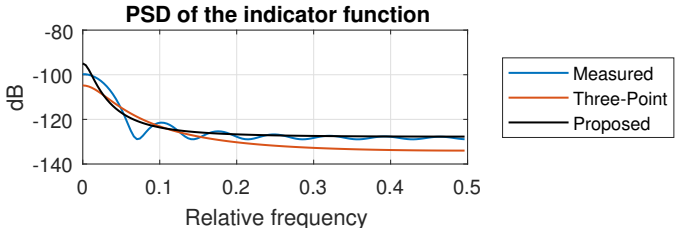

Fig. 6: Measurement – performance: the proposed method offers a better fit to the measurement data

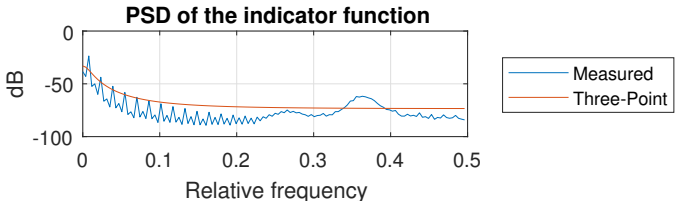

Fig. 7: Measurement – other model: the Three-Point method disregards the bumps at higher frequencies, while the proposed method recognizes that this is not a Gilbert-Elliott type model

V. Conclusion

In this paper a new identification method for the simpler cases of the Gilbert-Elliott model class was presented. The paper started with the mathematical description of the data loss, then the Gilbert-Elliott model class was introduced. Next, the main contribution of the paper was presented. This is the proposed identification method which estimates the autocorrelation, then fits a dynamic model. The autocorrelation function of the Gilbert-Elliott model is of key importance. Up to the knowledge of the authors it has not been shown before. Examining the coefficients it can be determined which model of the Gilbert-Elliott family is a good hypothesis and the values of the model parameters are calculated. Comparing the method to the previous one, the usage of all of the information contained in the PSD results in higher precision, assuming the same observation interval. On the contrary, less time is needed for the same accuracy. These improvements in precision and model selection ability were demonstrated by simulation and measurement examples.

Appendix

A. Derivation of the Autocorrelation Function

The derivation needs the expression of the powers of the transition probability matrix. Based on [8] and [9] they can be expressed as

$$\mathbf{P}^k = \frac{\left(\lambda_2\lambda_1^k - \lambda_1\lambda_2^k\right)\mathbf{I} + \left(\lambda_2^k - \lambda_1^k\right)\mathbf{P}}{\lambda_2 - \lambda_1} \tag{21}$$

where $\lambda_1 = 1-p-q = y$ and $\lambda_2 = 1$ are the eigenvalues of $\mathbf{P}$. With some calculation, we can get

$$\mathbf{P}^k = \frac{1}{p+q}\begin{bmatrix} q+py^k & p-py^k \\ q-qy^k & p+qy^k \end{bmatrix} \tag{22}$$

Now the autocorrelation function can be expressed by definition:

$$R_k = \mathbb{E}\left(K_n K_{n+k}\right) = \mathbb{E}\left(K_0 K_k\right) \tag{23}$$

The latter equation holds because of the time invariance of the Gilbert-Elliott model. Considering that $K_n \in \{0,1\}$ we can write

$$R_k = \sum_{\mathsf{S}_1,\mathsf{S}_2\in\{\mathsf{A},\mathsf{B}\}} \pi_{\mathsf{S}_1}\mathbb{P}\left(1|\mathsf{S}_1\right)\mathbb{P}\left(\mathsf{S}_1 \overset{k}{\to} \mathsf{S}_2\right)\mathbb{P}\left(1|\mathsf{S}_2\right) \tag{24}$$

where $\mathbb{P}\left(1|\mathsf{S}_i\right)$ is the probability of getting a 1 output in state S_i and $\mathbb{P}\left(\mathsf{S}_i \overset{k}{\to} \mathsf{S}_j\right)$ is the probability of transitioning in k steps from state S_i to S_j. Noticing that the former probabilities are the a and the b parameters of the Gilbert-Elliott model and the latter probabilities are the elements of $\mathbf{P}^k$, the autocorrelation function can be expressed as:

$$\begin{aligned} R_k &= \pi_{\mathsf{A}} a \mathbf{P}^k_{\mathsf{AA}} a + \pi_{\mathsf{A}} a \mathbf{P}^k_{\mathsf{AB}} b + \pi_{\mathsf{B}} b \mathbf{P}^k_{\mathsf{BA}} a + \pi_{\mathsf{B}} b \mathbf{P}^k_{\mathsf{BB}} b = \\ &= \frac{(aq+bp)^2 + pq\,(a-b)^2\,y^k}{(p+q)^2} \end{aligned} \tag{25}$$

The above expression is correct for $k > 0$, but incorrect for $k = 0$. For $k = 0$ the availability probability was counted twice, hence the result is incorrect. The correct result can be obtained directly from the definition:

$$R\left(0\right) = \mathbb{E}\left(K_0^2\right) = \mathbb{E}\left(K_0\right) = \mu \tag{26}$$

as $K_n \in \{0,1\}$. Putting together, the autocorrelation function is

$$R_k = \begin{cases} \frac{aq+bp}{p+q} & k = 0 \\ \frac{(aq+bp)^2+pq(a-b)^2y^k}{(p+q)^2} & k > 0 \end{cases} \tag{27}$$

With the substitutions $p+q \mapsto X$, $pq\,(a-b)^2 \mapsto Y$ and $aq+bp \mapsto Z$, the correlation function is

$$R_k = \begin{cases} \frac{Z}{X} & k = 0 \\ \frac{Z^2}{X^2} + \frac{Y}{X^2}\left(1-X\right)^k & k > 0 \end{cases} \tag{28}$$

References

[1] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," in *The Bell System Technical Journal*, vol. 42, Sept. 1963.

[2] S. Fine, Y. Singer, and N. Tishby, "The hierarchical hidden markov model: Analysis and applications," *Machine Learning*, vol. 32, no. 1, pp. 41–62, Jul 1998.

[3] M. Ellis, D. P. Pezaros, T. Kypraios, and C. Perkins, "A two-level markov model for packet loss in udp/ip-based real-time video applications targeting residential users," *Computer Networks*, vol. 70, pp. 384 – 399, 2014.

[4] L. Sujbert and G. Orosz, "Frequency domain identification of data loss models," in *Acta IMEKO*, vol. 6, no. 4, Dec. 2017, pp. 61–67.

[5] A. Palkó and L. Sujbert, "Time and frequency domain description of Gilbert-Elliott data loss models," in *2019 20th International Carpathian Control Conference (ICCC)*, May 2019, pp. 1–6.

[6] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, no. 5, pp. 1253–1265, 1960.

[7] L. Jackson, *Digital Filters and Signal Processing 2nd ed.* Kluwer Academic Publishers, 1989.

[8] Formula for matrix raised to power n. Accessed: 2018-10-18. [Online]. Available: https://www.freemathhelp.com/forum/threads/55028-Formula-for-matrix-raised-to-power-n?p=227129&viewfull=1#post227129

[9] Using Cayley-Hamilton theorem to calculate matrix powers. Accessed: 2018-10-18. [Online]. Available: https://www.physicsforums.com/threads/using-cayley-hamilton-theorem-to-calculate-matrix-powers.671687/