Abstract—A simple statistical method has been presented for evaluation of capacity for Gilbert-Elliott channels. Moreover, it has been shown that the Time-Varying Binary Symmetric Channel, as a basic model for some wireless communication channels, is a special case of Gilbert-Elliott channel. Using the simple statistical method, the capacity of this channel has been calculated over the range of parameters of channel.

I. INTRODUCTION

The Gilbert-Elliott channel [1] is an elementary finite state Markov channel (FSMC) with binary input-output alphabets, and binary states "good" and "bad". In each state the channel is characterized as a binary symmetric channel (BSC), with the crossover probabilities $p_G$ and $p_B$ ($p_G < p_B$) for states "good" and "bad", respectively. Due to the underlying Markov nature of the state process, the channel has memory that depends on the transition probabilities $g$ and $b$ between states, as shown in Figure 1. The memory in the channel inhibits a single letter formula for the capacity of the channel, however the capacity is obtained as a limiting expression [2].

In general, direct computation of the mutual information $I(X^n;Y^n)$ for channels with memory has exponential complexity in $n$, and since the sequence of information rate $1/n I(X^n;Y^n)$ usually converge slowly in $n$, direct computation of mutual information to find the channel capacity is not computationally feasible. Recently statistical method has been introduced for computation of information rate for finite state channels [3],[4],[5]. In this method a long random sequence $x^n$ is generated by a specific channel input distribution and the output sequence $y^n$ is obtained by channel simulation. The information rate is then approximated by evaluating

$$I(X^n;Y^n) \simeq \log p(x^n, y^n) - \log p(x^n) - \log p(y^n) \quad (1)$$

The probability of the sample $x^n$ is directly obtained from the input probability distribution, but $p(y^n)$ needs to be evaluated by the sum-product algorithms [6] operating on the trellis of channel.

For a special case of Gilbert-Elliott channels, namely the time varying binary symmetric channel (TV-BSC), in [7] this statistical method has been adopted for computation of channel capacity. In order to evaluate $p(y^n)$, a forward sum-product recursion on the joint source/channel trellis has been applied.

In this paper we show that the implementation of the statistical method for Gilbert-Elliott channel does not require sum-product algorithm, instead it can be implemented simply by iteratively generating a Bernoulli random variable (coin tossing) and modifying the probability of "success" for that variable based on the outcome in each iteration. We show that the random numbers generated by this method (we call it coin-tossing method) converge in probability to the capacity of Gilbert-Elliott channels. Numerical result confirms this convergence to the capacity which was also obtained by quantized evaluation of formula for capacity in [2]. This simplification in the statistical method for Gilbert-Elliott channel is a result of symmetry in the channel for any given state. Due to this symmetry, there is no need to generate the sequence $y^n$.

On the other hand, the TV-BSC, defined in [7], is described as a powerful tool for the information theoretic analysis of time varying channels. In this paper we show that there is no difference between the capacity analysis of the TV-BSC and Gilbert-Elliott channels. As a result the coin-tossing method of this paper can be applied for computation of channel capacity for TV-BSC. The results of this method coincide with the results of the statistical method [7] in which the sum-product algorithm has been applied.

The paper is organized as follows. In the next section we briefly describe the Gilbert-Elliott channel model and capacity evaluation by a limiting expression. In section III the coin-tossing method is presented and compared with the quantized method for iterative evaluation of capacity. In the last section the capacity of TV-BSC is investigated.
II. GILBERT-Elliott CHANNEL

The capacity of Gilbert-Elliott channel is obtained as the following limiting expression [2] (which doesn’t require maximization\(^1\)),

\[
C = 1 - \lim_{l \to \infty} E[h(q_l)],
\]

where \(h(.)\) is the binary entropy function. The random variable \(q_l \in [0, 1]\) is defined through the binary random process \(z^l\), as

\[
q_l = Pr(z_l = 1|z^{l-1}),
\]

where the process \(z^l\) (called error process) is in turn defined by two input and output processes \(x^l\) and \(y^l\) according to \(z_l = x_l \oplus y_l\). The process \(x^l\) is an iid process for any arbitrary distribution \(P(x)\), whereas \(y^l\) depend on the process \(x^l\) and state process \(s^l\) (independent of \(x^l\)) through

\[
Pr(y^l|x^l, s^l) = \prod_{i=1}^{l} Pr(y_i|x_i, s_i).
\]

The state process evolves as a stationary Markov process according to diagram in Figure 1, and \(Pr(y|x, s)\) in (4) is fixed and defined as a BSC for two states \(s = \text{"good"},\) and \(s = \text{"bad"}\).

As (3) shows the random variable \(q_l\) has a probability mass function depending on the distribution of the random sequence \(z^{l-1}\). It is shown [2] that the following recursion exists between the sequence of random variables \(q_l, l = 0, 1, \ldots\)

\[
q_l(z^{l-1}) = v(z_{l-1}, q_{l-1}(z^{l-2})),
\]

where assuming \(p_G < p_B, p_G \neq 0, p_B \neq 1, p_G \leq q \leq p_B\), the function \(v(., .)\) is

\[
v(z, q) \triangleq \begin{cases} 
  p_G + b(p_B - p_G) + \mu(q - p_G)(1 - p_B)/(1 - q), & z=0; \\
  p_G + b(p_B - p_G) + \mu(q - p_G)p_B/q, & z=1.
\end{cases}
\]

and \(\mu = 1 - g - b\). The initial value for this recursion is

\[
q_0 = Pr(z_0 = 1) = Pr(s_0 = 0)p_G + Pr(s_0 = 1)p_B = (g p_G + b p_B)/(g + b),
\]

in which the assumption

\[
Pr(s_0 = G) = g/(g + b), \quad Pr(s_0 = B) = b/(g + b),
\]

for the initial distribution of the state process is to ensure stationarity of this process. We have \(p_G \leq q_0 \leq p_B\) and the recursion (5) always maintain the relation \(p_G \leq q_l \leq p_B\) for any \(l\).

Computation of (2) for a finite \(l\) requires evaluation of the probability mass function on \(q_l\), using recursive equation (5).

\(^1\)This expression doesn’t refer to a maximization problem. The freedom on the choice of iid input distribution for evaluating this expression indicates that the expression is constant over the range of iid distributions, and it is an upper bound on the information rate, which depends on input distribution. Only for uniform iid distribution the information rate measures up to this expression.

It can be shown that the sequence \(\{q_l\}_{l=0}^{\infty}\) is a Markov process with transition probability of

\[
Pr(q_l = \alpha|q_{l-1} = \beta) = \begin{cases} 
  1 - \beta, & \alpha = v(0, \beta); \\
  \beta, & \alpha = v(1, \beta).
\end{cases}
\]

and initial distribution

\[
Pr(q_0 = (g p_G + b p_B)/(g + b)) = 1.
\]

Although the conditional distribution (9) is only a two point probability mass, the distribution of \(q_l\) is a 2\(^l\) point probability mass, and in the limit \(l \to \infty\) it converges to a probability density function over \([p_G, p_B]\).

III. STATISTICAL EVALUATION OF CAPACITY

A problem for computation of (2) is the exponential increase in the complexity for evaluation of probability distribution \(q_l\). To avoid this computational complexity, in [2] the range of \(q\) \((p_G \leq q \leq p_B)\) is quantized to a fixed number of levels, and in each recursion \(q\) is mapped to this finite set of numbers. Although this method limits the computational complexity, the precision could be sacrificed. An example of variation of distribution of \(q_l\) for different values \(l\) is shown in Figure 2.

In relation to statistical evaluation of information rates, in [4] it has been shown that if a random sequence \((x^n, y^n, s^n)\) is generated according to the joint probability distribution

\[
p(x^n, y^n, s^n) = p(s_0) \prod_{i=1}^{n} p(x_i, y_i, s_i | s_{i-1}),
\]

where \(p(x_i, y_i, s_i | s_{i-1})\) is not depending on \(i\), then

\[
-\frac{1}{n} \log p(x^n) \rightarrow \lim_{k \to \infty} \frac{1}{k} H(x^k)
\]

\[
-\frac{1}{n} \log p(y^n) \rightarrow \lim_{k \to \infty} \frac{1}{k} H(y^k)
\]

\[
-\frac{1}{n} \log p(x^n, y^n) \rightarrow \lim_{k \to \infty} \frac{1}{k} H(x^k, y^k).
\]

Fig. 2. Development of probability distributions of \(q_l, l = 3, 6, 10\) for the case \(p_G = 0.2, p_B = 0.6, g = 0.01, b = 0.05\) and quantization 0.01.
where \( \rightarrow \) is convergence in probability\(^2\). The limits in the right hand side are the entropy rates, which are positive numbers for each process or joint processes. From (11) \(-1/n \log p(y^n|x^n)\) converges in probability to \(\lim_{k \to \infty} -1/kH(y^k|x^k)\). It is also shown [4] that for finite state Markov channels, including the Gilbert-Elliott channel, the condition of (10) is satisfied, assuming \(x^n\) be Markov or iid.

Generating the sequences of \(x^n, y^n\) for a Gilbert-Elliott channel and based on the relation \(z_l = x_l \oplus y_l\) generating a sequence of \(z^n\) is the same as generating the sequence \(z^n\) according to (3), using recursion (5) (ie: both \(z^n\) have the same probability distribution). A given \(z^n\) corresponds to a set of \(x^n, y^n\), where \(-1/n \log p(y^n|x^n)\) for all those elements converge to the \(\lim_{k \to \infty} -1/kH(y^k|x^k)\), thus \(-1/n \log p(z^n)\), where \(z^n\) is generated by (3) converges to \(\lim_{k \to \infty} -1/kH(y^k|x^k) = \lim_{k \to \infty} -1/kH(z^k)\).

Based on this convergence, here we develop a simple statistical method for computation of capacity for the Gilbert-Elliott channels. The numerical results match with the statistical method of [7] and the quantized probability distribution method of [2].

Considering that the Gilbert-Elliott channel is “indecomposable” [8], the channel capacity is given by

\[
C = \lim_{n \to \infty} \max \frac{1}{n}I(x^n; y^n),
\]

(12)

where maximization is over all possible probability distributions on input n-sequences. We have

\[
I(x^n; y^n) = H(y^n) - H(y^n|x^n) = H(y^n) - H(z^n).
\]

Since the Gilbert-Elliott channel is a special case of “uniformly symmetric variable noise” channel [9], its capacity is achieved at the uniform iid distribution [9, Theorem 5.1]. At this distribution \(H(y^n) = n \log(|\mathcal{Y}|) = n\). On the other hand \(H(z^n)\) doesn’t depend on \(P(x^n)\), thus from (12) and (13),

\[
C = 1 - \lim_{n \to \infty} \frac{1}{n}H(z^n).
\]

(14)

Now the entropy rate \(\lim_{n \to \infty} 1/nH(z^n)\) can be approximated by generating a long random sequence of \(z^n\) and evaluating \(-1/n \log p(z^n)\) for that sequence. According to (3), the random sequence of \(z^n\) can be generated recursively. By knowing the first \(l - 1\) elements of the sequence, ie: \(z^n|z^{l-1}\), then \(z_l \in \{0, 1\}\) can be generated as a Bernoulli \(q_l\), and \(q_l\) is recursively updated by (5). According to the preceding discussion, the sequence of \(f(n)\) generated as follows, converge in probability to \(\lim_{n \to \infty} 1/nH(z^n)\).

\[
f(n) \triangleq -\frac{1}{n} \log p(z^n)
= -\frac{1}{n} \sum_{i=1}^{n} \log p(z_i|z^{i-1})
= -\frac{1}{n} \sum_{i=1}^{n} [z_i \log(q_i) + \bar{z}_i \log(1 - q_i)],
\]

(15)

where \(\bar{z} = 1 - z\), and \(z_i\) is generated as Bernoulli \(q_i\). Consequently the capacity of the Gilbert-Elliott channel can be approximated by \(C = 1 - f(n)\) for sufficiently large \(n\).

The last expression in (15) formulates the coin-tossing method, which means that generate a Bernoulli \(q_l\), if the result is “success” add the term \(-\log(q_l)\), else add \(-\log(1 - q_l)\) (ie: add the information associated with the events of “success” or “fail”), update \(q_i\) to \(q_{i+1}\) according to the result by \(v(0, q_i)\) or \(v(1, q_i)\) in (6), and continue.

The convergence of the coin-tossing method is shown in Figure 3, which shows convergence to the same value that can be obtained by more computationally demanding method of evaluating probability distribution in each iteration, depicted in Figure 4.
For a BSC (or Gilbert-Elliott channel with $p_G = p_B$), $q_l$ is fixed (the recursion (5) maintains $q_l = p_G = p_B$), and Equation (15) reduces to

$$f(n) = -1/n \sum_{i=1}^n [\delta_i \log(q_l) + \bar{\delta}_i \log(1 - q_l)]$$

$$= -[1/n \sum_{i=1}^n \delta_i] \log(q_l) - [1/n \sum_{i=1}^n \bar{\delta}_i] \log(1 - q_l)$$

$$\Rightarrow -q \log(q_l) - (1 - q) \log(1 - q_l)$$

$$= h(q)$$

(16)

where $\Rightarrow$ is convergence in probability, and the convergence is due to the weak law of large numbers (note that the expectation of Bernoulli$q_l$) is $q_l$. This verifies the convergence of $1 - f(n)$ to the capacity of BSC, ie: $C = 1 - h(q)$.

On the other hand for $p_G < p_B$, the convergence of the $f(n)$ in (15) to $\lim_{n \to \infty} E[h(q_l)]$, (thus $1 - f(n)$ to the capacity) can be explained by the markovity of the process $\{q_l\}_{n=0}^\infty$ over the state space of $[p_G, p_B]$. If the limiting distribution of this Markov process be $\psi(.)$, and we partition the range $[p_G, p_B]$ to $m$ infinitesimal intervals with centers at $q_j$ and size $\delta_j$, then any sufficiently long sample process (with length $n$) would pass through the state partition $\delta_j$ a proportion of time $(n_j/n)$ almost equal to $\delta_j \psi(q_j)$. Therefore (15) can be written as

$$f(n) =$$

$$-1/n \sum_{i=1}^n [\delta_i \log(q_l) + \bar{\delta}_i \log(1 - q_l)]$$

$$=-1/n \sum_{j=1}^m \sum_{i=1}^{n_j} [z_{ij} \log(q_{ij}) + \bar{z}_{ij} \log(1 - q_{ij})]$$

$$=-\sum_{j=1}^m \sum_{i=1}^{n_j} [1/n_j \delta_{ij} \log(q_{ij}) + 1/n_j \bar{\delta}_{ij} \log(1 - q_{ij})]$$

$$\Rightarrow -\sum_{j=1}^m \delta_j \psi(q_j)[q_j \log(q_j) + (1 - q_j) \log(1 - q_j)]$$

$$\approx \int \psi(q) [q \log(q) + (1 - q) \log(1 - q)] dq$$

$$= \lim_{l \to \infty} E[h(q_l)].$$

where subscript for $q$ has been used for different purposes, indicating the probability of success for the $i$-th randomly generated variable, the center of $j$-th interval, or the $l$-th distribution on $q$.

Numerical investigation shows that a better convergence per computation can be achieved if instead of one sequence $z^n$, $k$ sequences of $[z^n]_{j=1}^k$ be generated in parallel, and $\lim_{n \to \infty} 1/n H(z^n)$ be estimated by the average of $1/n \log p(z_j)$, $j = 1, \ldots, k$, ie: $\lim_{n \to \infty} 1/n H(z^n)$ be estimated by

$$f'(n, k) = -1/(nk) \sum_{i=1}^n \sum_{j=1}^k [\delta_{ij} \log(q_{ij}) + \bar{\delta}_{ij} \log(1 - q_{ij})],$$

(17)

for sufficiently large $n, k$.

For example in Figure 7 we see that the average of $k = 10$ samples converges more than 10 times faster than each sample.

IV. THE TIME VARYING BINARY SYMMETRIC CHANNEL

The Gilbert-Elliott channel described in Section 1 is a one state Markov channel model where the states "good" and "bad" are assigned to the BSCs having minimum and maximum crossover probabilities, respectively. All analytical results are correct for any value $0 < p_G < p_B < 1$. In fact for any two-state channel with state set $S = \{0,1\}$, where conditioned on each state the channel is a BSC with crossover probabilities $\epsilon_0, \epsilon_1$ ($0 < \epsilon_0, \epsilon_1 < 1$), and state transition probabilities $p_0, p_1$, the analytical result of the previous sections holds. Equation (6) for such a channel is

$$v(z, q) \triangleq$$

$$\{ \epsilon_0 p_0 + \epsilon_1 p_1 + [\mu q(1 - \epsilon_0 - \epsilon_1) + \mu \epsilon_0 \epsilon_1]/(1 - q), \quad z = 0;$$

$$\epsilon_0 p_0 + \epsilon_1 p_1 + \mu(\epsilon_0 + \epsilon_1) - \mu \epsilon_0 \epsilon_1/q, \quad z = 1. \}$$

(18)

where $\mu = 1 - p_0 - p_1$. Equation (18) is the same as (6), by $p_G = \min\{\epsilon_0, \epsilon_1\}, p_B = \max\{\epsilon_0, \epsilon_1\}$, and $p_0, p_1$ be the corresponding $b$ or $q$. Equation (18) clearly shows that there is no discrimination between states. Hence, although the physical interpretation of channels with different values of $\epsilon_0, \epsilon_1$ could be different, all accept the same capacity analysis. In particular when ($\epsilon_0$ and/or $\epsilon_1$) > 0.5 the interpretation of the channel features are different. In this case a greater $\epsilon$ doesn’t necessarily mean a "bad" state or the process $z^l$ is not necessarily the "error process".

A special case of the above mentioned class of channels is the "time varying binary symmetric channel", TV-BSC. This channel is defined in [7] as the same model as a Gilbert-Elliott channel (Figure 1) with $\epsilon_1 = 1 - \epsilon_0$ and $p_0 = p_1$, thus the channel is characterized by two parameters $\epsilon, p$. The states of channel instead of "good" and "bad", are "inverting" and "non-inverting". Knowing the state of the channel in the receiver, the quality of channel in terms of probability of error will be the same regardless of the state. However it is assumed that like Gilbert-Elliott channel, no direct channel state information is available at the receiver.

Despite these differences in physical interpretation of the states for TV-BSC as compared to Gilbert-Elliott channels, the capacity of TV-BSC is still corresponds to the maximum of information rates in (12). The channel is still a uniformly symmetric variable noise [9], hence this maximum is achieved by uniform iid input distribution (the optimum distribution obtained by the stochastic Arimoto-Blahut algorithm in [7] is also uniform iid). Therefore capacity can be described by the entropy of $z^n$ as in (14), and the remaining steps to the formula (2) will be the same (see proof of [2, proposition 4]). The main speciality for TV-BSC is that the conditional probability $P(z|s)$ is according to the following table. Nevertheless, all of the discussions on the capacity of Gilbert-Elliott channel are valid for any conditional probability distribution $P(z|s)$.

| $s$  | $P(z|s)$ |
|-----|---------|
| $s = 0$ | $1 - \epsilon$ |
| $s = 1$ | $\epsilon$ |

The TV-BSC is an elementary model for some wireless communication channels. Basic characteristics of this model.
can be highlighted by representing the channel as the concatenation of an random inverter and a BSC, as in Figure 5. The random inverter has two state of invert, and non-invert where the movement between states is a Markov process. The random inverter represents fading and memory in channels, whereas the BSC represents white noise in channels.

The coin-tossing method can be used to evaluate the capacity of TV-BSC for different values of $\epsilon, p$. For this channel, the recursive formula (6) for updating $q_i$ reduces to

$$v(z, q) = \begin{cases} p + \mu(1 - \epsilon)/(1 - q), & z=0; \\ 1 - p - \mu(1 - \epsilon)/q, & z=1, \end{cases} \tag{19}$$

where $\mu = 1 - 2p$. The capacity of TV-BSC for the range $0 < \epsilon, p < 1$ has been obtained by this method, and shown in Figure 6. The capacity for specific values $(\epsilon, p) \in \{(0.01, 0.01), (0.01, 0.1), (0.2, 0.01)\}$ are identical to the values obtained in [7], where the sum-product algorithm has been applied for implementing the statistical method. The convergence of coin-tossing method for $\epsilon = p = 0.01$ by both Equations (15) and (17) $(k = 10)$ is shown in figures 7.

V. CONCLUSION

In this paper a simple statistical method for computation of capacity for Gilbert-Elliott channels has been introduced. In each iteration we only need to generate a bernoulli random variable, and use a simple recursive formula. It is proved that the sequence of numbers generated by this method converges in probability to the capacity of channel. Furthermore, The time varying binary symmetric channel has been shown to have the same capacity formula as the Gilbert-Elliott channels, and the new statistical method has been applied for evaluation of capacity for this channel.

REFERENCES