Basic binary state-space model for time-varying communication channels: uncertainty and information capacity

Z.B. Krusevac, P.B. Rapajic and R.A. Kennedy

Abstract: This paper looks at capacity achieving detection strategies for information transfer over time-varying channels. The time-varying binary symmetric channel (TV-BSC) is identified as the basic binary state-space model. Separation of entropies principles and the TV-BSC model-based state-space approach are used to determine the performance bounds for coherent and non-coherent detection over time-varying communication channels. The mutual information rate over the TV-BSC, assuming channel estimation in the presence of channel noise, is shown to be below the channel information capacity because of lack of perfect channel knowledge. Furthermore, it is shown that TV-BSC model-based differential detection has a fundamental advantage over the channel estimation based detection since it theoretically preserves the TV-BSC information capacity when the observation interval approaches infinity. Simulation analysis corroborates the theoretical results, showing that multiple-symbol differential detection practically achieves the TV-BSC capacity in just a few symbol observation times.

1 Introduction

In practical mobile wireless communication systems, where the channel is highly non-stationary (relative to the data rate), the channel process is an additional underlying stochastic process at work in addition to channel noise and the information source. The channel stochastic process is not directly observable, but can only be observed through the two stochastic processes that produce the channel output sequence of observations [1]. Furthermore, the time-varying channel process is not completely observable given channel noise [2]. Thus, standard coding and detection solutions for information transmission over time-invariant channels, which are optimal in combating noise, may not be optimal assuming stochastic channel time-variations and lack of perfect knowledge of the channel. Therefore the time-varying communication channel, common in the mobile wireless communication systems, calls for fundamental information theory issues to be revisited.

The communication channel information capacity formula, $C = B\log(1 + S/N)$, where B is the channel bandwidth, S/N is the signal to AWGN (additive white Gaussian noise) power ratio, was given by Shannon [3]. The information rate R, which could be transmitted over the channel with vanishingly small probability of error, must satisfy the inequality $R \leq C$. This leads us to revisit the four explicit assumptions under which this formula was derived: infinite signal time delay (or very long signal delay), existence of a powerful channel coding/decoding scheme, stationary (time-invariant) channel model with bandwidth B and known communication channel parameters.

As far as wire-line communication channels can be considered time-invariant, the four mentioned assumptions listed above are closely satisfied. However, if the communication channel is time-varying, as in mobile wireless communication systems, the third and fourth assumptions require particular attention.

Non-stationary wireless communication channels violate the concept of signal spectrum and consequently the concept of signal bandwidth required in the third assumption [4]. Furthermore, stochastic process uncertainty of time-varying communication channels disregards the fourth assumption, since time-varying channel parameters estimation reduce the information transmission rate.

The real difficulty in establishing capacity results for time-varying channels stems from imperfect knowledge of the channel (channel uncertainty) rather than its time-varying nature, although it is the time variations in the presence of channel noise that render estimation difficult [5–7]. Thereby, assuming an uncertainty of the channel knowledge, the mutual information performance of standard (channel estimation based and differential) detection strategies needs further consideration.

In order to determine mutual information performance bounds for information transmission over time-varying channels, one has to be able to incorporate a knowledge of the propagation conditions to adequately relate received and transmitted signals. Model-based state-space system analysis, which is presented, uses the channel model, in state equation form, in addition to the channel output observation to examine the performance of different detection strategies. In fact, essential prerequisites for addressing information theoretic issues of time-varying channels are characterising properly the time-varying channel and developing an accurate, application-independent model of such a channel. However, the time variation of the channels proves to be very difficult to model and circumscribe [5].

doi:10.1049/iet-com:20060477

Paper first received 2nd August 2006 and in revised form 24th March 2007

Z.B. Krusevac and R.A. Kennedy are with the Department of Information Engineering, RSISE, the Australian National University, Canberra ACT 2601, Australia

P.B. Rapajic is with the Medway School of Engineering, University of Greenwich at Medway Pembroke, Chatham Maritime KENT, ME4 4TB, UK

E-mail: zarkok@rsise.anu.edu.au

As an alternative to modelling a communication channel at the physical level, model design can be formulated for the purpose of system performance analysis. When formulated in this way, as we do in this paper, the intention is to reveal the limits to information throughput and provide a theoretical framework for optimal solution at the application level. Further, this approach enables significant model complexity reduction which is better suited to provide insight into fundamental theoretical communication issues. A well-known example is the binary symmetric channel (BSC) which is relevant to time-invariant AWGN channels. That is, the essential qualitative elements of information theory for time-invariant AWGN channels can be efficiently explained using the BSC model. In [8], an intuitive approach is used to develop the time-varying binary symmetric channel (TV-BSC) as an analogy to the BSC for the time-varying case. We show that the TV-BSC has a strong information theoretical foundation and we use the TV-BSC model to provide mutual information performance analysis of time-varying communication channels.

The main contributions of this paper may be summarised as follows:

- We identify the TV-BSC as the basic binary state-space model. We use separation of entropies principles and the TV-BSC model-based state-space approach to determine the performance bounds for coherent and non-coherent detection over time-varying communication channels.
- We show that, in accordance with the separation of entropy principle, the mutual information rate over the TV-BSC, assuming channel estimation (coherent detection) and the presence of channel noise, is below the channel information capacity.
- We formally prove that multiple-symbol TV-BSC model-based differential detection, as a form of block-by-block maximum likelihood sequence detection, preserves the TV-BSC information capacity when the observation interval approaches infinity. This fundamental advantage of the differential detection is shown to be based on ability to exploit channel process innovation. Furthermore, we show that the differential scheme is very robust to a decrease in channel time correlation (quite the opposite trend to coherent detection methods), due to fact that the channel process entropy becomes predominant over the noise process entropy.

Notation: Throughout the paper, $p(\cdot)$ denotes the distribution of a random variable, $H(\cdot)$ denotes the entropy, $\mathcal{H}(\cdot)$ denotes the entropy rate, $I(\cdot\,;\cdot)$ denotes the mutual information and $\mathcal{I}(\cdot\,;\cdot)$ denotes the mutual information rate. Additionally, while X is a discrete-time stochastic process, X_k is the process random variable at time k with alphabet $\mathcal{X}$, x_k is a realisation of the random variable X_k and X^k is the sequence of the process random variables at time $n = 1$ through $n = k$, where $k \geq 1$.

2 Model-based state-space approach and time-varying binary symmetric channel

2.1 State-space time-varying channel modelling

The state-space channel model applies to general discrete and continuous time-varying channels, whose variation is governed by a stochastic process taking values over a state-space of time-invariant channels.

Variation of the continuous-time state-space channel is governed by a continuous-time stochastic process $S(t)$, $t > 0$ with state-space $\mathcal{C}$. Each channel state $c \in \mathcal{C}$ indexes a time-invariant continuous-time channel, and $S(t)$ is called the channel state at time t.

The variation of the discrete-time state-space channel is determined by a discrete-time stochastic process S_n with state-space $\mathcal{C}$. The state-space is a set of discrete memoryless channels with common input and output alphabets, denoted by $\mathcal{X}$ and $\mathcal{Y}$, respectively.

In general, the state-space model places no restrictions on the stochastic process S_n. However, if the stochastic process S_n is Markov with stationary transition probabilities, and its state-space $\mathcal{C} = \{c_0, c_2, \ldots, c_{M-1}\}$ is finite, then the state-space model is called a finite-state Markov channel (FSMC).

2.2 Time-varying BSC model

The TV-BSC model, presented in [9], is a stationary two-state uniformly symmetric variable noise Markov channel. The states are designed as discrete memoryless BSC's with crossover probability p and $1-p$, $0 \leq p \leq 0.5$, for the non-inverting BSC state c_0 and the inverting BSC state c_1, respectively. The state transition probabilities are given by

$$p(S_{n+1} = c_1 | S_n = c_0) = p(S_{n+1} = c_0 | S_n = c_1) = q$$
$$p(S_{n+1} = c_0 | S_n = c_0) = p(S_{n+1} = c_1 | S_n = c_1) = 1 - q \quad (1)$$

where $0 \leq q \leq 0.5$. Because of the underlying Markov nature, the TV-BSC has the channel memory which depends on q and which is given by

$$\mu = 1 - 2 \cdot q \quad (2)$$

'Slow' TV-BSC exhibits a long memory ($\mu \to 1$) and its current channel state is very persistent ($q \to 0$). Similarly, a 'fast' TV-BSC frequently changes its state ($q \to 0.5$) and exhibits a short memory ($\mu \to 0$).

It is shown in [9] that the TV-BSC is unique as the basic FSMC model that shows a non-trivial influence on capacity because of the channel uncertainty. Furthermore, it is shown in [9] that the TV-BSC exhibits some non-trivial features of time-varying channels not captured in general by the Gilbert-Elliot channel [10], to which it bears some superficial resemblance.

The analysis in [9] uses an intuitive approach to develop the TV-BSC model. However, we show that the TV-BSC has a much more general information theoretic foundation. By defining $c_0 \triangleq 0$ and $c_1 \triangleq 1$, the TV-BSC can be analytically recognised as the basic binary time-varying state-space model, described by

$$y_n = S_n \oplus x_n \oplus \nu_n \quad (3a)$$
$$S_{n+1} = S_n \oplus \eta_n \quad (3b)$$

where x_n and y_n are the channel input and output binary sequence, respectively; ν_n and η_n are the channel (observation) and system noise process, respectively, and $\oplus$ is modulo-2 addition. Additionally, S_n, ν_n, $\eta_n \in \{0, 1\}$, $Pr(\nu_n = 1) = p$, $Pr(\eta_n = 1) = q$ and S_n, ν_n and η_n are statistically independent processes.

We use the TV-BSC model-based state-space approach to examine the mutual-information performance of different detection strategies for information transfer over time-varying channels. From the system analysis point of view, the principle objective of the model-based state-space approach is to exploit both, observation equation and the additional knowledge of non-stationary propagation conditions, in state equation form, in order to determine the mutual-information performance bounds.

3 Separation of entropies from time-varying channel output observation

Entropy rate of the channel output observation of the time-varying communication channel is determined by the information source entropy rate, channel process entropy rate and channel noise process entropy rate. Based on results from [2, 3, 8, 10, 11] the following three separation of entropies principles for the information transfer over time-varying channels can be identified.

1. If the channel process entropy is equal to zero (perfect channel state information (CSI) is available), by proper encoding of the information, errors of the information induced by a noisy channel can be reduced to any desired level without sacrificing the rate of the information transfer.
2. If the channel information source entropy is equal to zero (an infinitely long training sequence is transmitted), the channel process can be estimated with arbitrarily small probability of error. However, in this case, the information transmission rate is equal to zero.
3. If the noise process entropy is equal to zero (noiseless channel), the channel process is completely observable at the receiver, enabling capacity achieving ML (maximum-likelihood) information detection.

The first principle follows from the Shannon coding theorem. The second principle can be confirmed at the level of FSMCs, for instance, by using the decision-feedback estimator [11]. The third principle can be demonstrated by using optimal blind FSMC estimation [8].

The three separation of entropies principles reveal that if one of the entropy does not exist, the other two can be separately suppressed, allowing channel estimation and/or signal detection with overall capacity achieving mutual information performance.

However, an unknown time-varying channel process is not completely observable in the presence of channel noise, based on the channel output sequence of observations [2]. Furthermore, the basic Shannon channel coding theorem cannot be directly implemented because of lack of CSI. In other words, if all three stochastic processes (i.e. the information source, channel and noise) are unknown, their entropies cannot be separated and separately suppressed based only on the channel output observation sequence.

4 Channel estimation based detection

4.1 Channel process entropy and observable channel process entropy

Based on the state equation (3b), the entropy rate of the channel (state) process $\mathcal{H}(S)$, $S \triangleq (S_1, S_2, S_3, \ldots)$, of the TV-BSC model can be expressed as

$$\mathcal{H}(S) = h(q) \tag{4}$$

where $h(q) = -q\log_2 q - (1-q)\log_2(1-q)$ is the binary entropy function of q.

The channel process S is an underlying stochastic process that is not directly observable at the channel output [1]. The observable channel process entropy rate $\mathcal{H}_{\text{ob}}(S)$ of the channel process S at the receiver side can be expressed as

$$\mathcal{H}_{\text{ob}}(S) = C^{\text{CSI}} - C \tag{5}$$

where C is the information capacity of an unknown time-varying channel and C^{CSI} is the information capacity of the same channel with the perfect CSI assumption. In general, the observation of the channel process is affected by both the channel noise uncertainty and the information source uncertainty. Thereby, the observable channel entropy rate (5) is not necessarily the same as the entropy rate (4) of the channel process itself.

In order to determine the observable channel process entropy rate $\mathcal{H}_{\text{ob}}(S)$ of the TV-BSC, one can express the TV-BSC information capacity as [9]

$$C = 1 - \mathcal{H}(Z) \tag{6}$$

where $\mathcal{H}(Z)$ is the entropy rate of the TV-BSC error function $Z_n = X_n \oplus Y_n$ as $n \to \infty$, for an input distribution $p(X)$ which is uniform i.i.d. (independent identically distributed). Furthermore, the information capacity C^{CSI} of the TV-BSC with CSI is given as

$$C^{\text{CSI}} = 1 - \mathcal{H}(Z|S) \tag{7}$$

where $\mathcal{H}(Z|S)$ is the entropy rate of the TV-BSC error function $Z_n = X_n \oplus Y_n$ as $n \to \infty$, assuming CSI and the uniform i.i.d. input distribution $p(X)$.

However, C^{CSI} is also the statistical average over the two states of the corresponding channel capacity as

$$\begin{aligned} C^{\text{CSI}} &= p(S_0 = c_0)\cdot C_{c_0} + p(S_0 = c_1)\cdot C_{c_1} \\ &= C_{\text{BSC}} \cdot \underbrace{(p(S_0 = c_0) + p(S_0 = c_1))}_{=1} \\ &= 1 - h(p) \end{aligned} \tag{8}$$

where C_{c0} and C_{c1} are the non-inverting and inverting states information capacity, respectively. Furthermore, $C_{c0} = C_{c1} = C_{\text{BSC}}$, where $C_{\text{BSC}} = 1 - h(p)$ is the information capacity of the BSC with the crossover probability p. Consistent with the stationary distribution we choose the initial state probabilities $p(S_0 = c_0) = p(S_0 = c_1) = 0.5$. Thus, based on (5) and (8), the information capacity of the TV-BSC is the information capacity of the BSC reduced by the observable channel entropy rate $\mathcal{H}_{\text{ob}}(S)$, which exists because of stochastic channel time variations. Thereby, the BSC can be recognised a special case (or the time-invariant equivalent) of the TV-BSC, when CSI is available.

Combining (5), (6), and (8) gives

$$\begin{aligned} \mathcal{H}_{\text{ob}}(S) &= \mathcal{H}(Z) - \mathcal{H}(Z/S) = \mathcal{H}(Z) - h(p) \\ &= \mathcal{H}(S) - \mathcal{H}(S/Z) = h(q) - \mathcal{H}(S/Z) \end{aligned} \tag{9}$$

where $\mathcal{H}(Z/S) = h(p)$ is the entropy rate of the noise process ν_n (3a). The second last equality follows from the fact that $\mathcal{H}(Z) - \mathcal{H}(Z/S) = \mathcal{H}(S) - \mathcal{H}(S/Z)$ [12].

Expression (9) shows that even though the channel process S_n exists independently of the channel noise process ν_n and, thereby, entropy rate $\mathcal{H}(S) = h(q)$ does not depend on p, the observable channel process entropy rate $\mathcal{H}_{\text{ob}}(S)$ is determined by both, channel process entropy $\mathcal{H}(S) = h(q)$ and noise process entropy rate $\mathcal{H}(Z|S) = h(p)$.

In the presence of channel noise $(0 < p \le 0.5)$, $\mathcal{H}(S|Z) > 0$ and consequently

$$\mathcal{H}_{\text{ob}}(S) < \mathcal{H}(S) = h(q) \tag{10}$$

However, in the noise-less situation $(p = 0)$, $\mathcal{H}(Z|S) = h(p) = 0$ and, thereby, $\mathcal{H}(S|Z) = 0$ which leads to

$$\mathcal{H}_{\text{ob}}(S) = \mathcal{H}(S) = h(q) \tag{11}$$

4.2 Mutual information performance of channel estimation based detection

Imperfect knowledge of the channel essentially affects mutual information performance of the standard channel estimation based decoding solutions which are optimal assuming perfect channel knowledge [5]. Here we use the model-based approach and the TV-BSC model, given by state-space model (3), to analyse mutual information performance of the channel estimation based detection.

Fig. 1 shows channel process entropy rate $\mathcal{H}(S) = h(q)$ and observable channel process entropy rate $\mathcal{H}_{ob}(S) = C^{CSI} - C$ of the TV-BSC. Channel state estimation reduces the channel process entropy rate $\mathcal{H}(S)$. However, as $\mathcal{H}_{ob}(S) \leq \mathcal{H}(S)$ then theoretically

$$\begin{aligned} \mathcal{I}_{estim} &= C^{CSI} - \mathcal{H}(S) = 1 - h(p) - h(q) \\ &\leq C = C^{CSI} - \mathcal{H}_{ob}(S) \end{aligned} \tag{12}$$

where $\mathcal{I}_{estim} = 1 - h(p) - h(q)$ is the mutual information rate, assuming channel estimation, as shown in Fig. 1. Furthermore, based on (10) and (11), one can write

$$\mathcal{I}_{estim} = 1 - h(p) - h(q) \begin{cases} < C & \text{for } 0 < p \leq 0.5 \\ = C & \text{for } p = 0 \end{cases} \tag{13}$$

If $q = 0$ the TV-BSC stays forever into the initial state S_0 and the channel process entropy rate $\mathcal{H}(S) = h(q) = 0$. Consequently, $\mathcal{H}(S/Z) = 0$ and $\mathcal{H}_{ob}(S) = 0$, which leads to

$$\mathcal{I}_{estim} = C^{CSI} \tag{14}$$

Equality in (13) follows from the third separation of entropies principle and (14) follows from the first principle. However, inequality in (13) confirms that separate time-varying channel estimation ($h(q)$ suppression) in the presence of channel noise, followed by the data detection ($h(p)$ suppression) cannot be capacity achieving.

Fig. 1 further shows that the channel estimation based detection performs worse as q increases and the channel becomes faster. If the TV-BSC is too fast (i.e. $q \to 0.5$, $\mu \to 0$), such that $h(q) \geq 1 - h(p)$, then the amount of information needed for the channel estimation is higher than the information capacity of the TV-BSC (denoted by the arrow in Fig. 1). Thereby, the channel estimation cannot be performed at all.

The optimal blind FSMC estimation, presented in [8], can be considered as a conformation, on an application level, of the mutual information performance analysis, presented here. It is shown in [8] that the blind channel estimation [13, 14] exploits implicit information which originates from the input signal redundancy (signal structure) [15]. Furthermore, it is shown that the maximum mutual information rate assuming optimal blind estimation in the presence of channel noise is strictly below the FSMC information capacity because of noisy channel estimation. Simulation analysis in [8] is performed using the Gilbert-Elliot channel.

In Fig. 2, we simulate mutual information performance of the optimal blind TV-BSC estimation to further confirm main outcomes of this analysis.

Fig. 2 shows that optimal blind output-feedback channel estimation in noisy environment achieves channel information capacity only if $\mu = 1$ (i.e. $q = 0$; channel stays forever in the initial state). However, for $0 < \mu < 1$ $(0 < q < 0.5)$, the maximum mutual information rate is strictly below the TV-BSC information capacity because of noisy channel estimation.

5 TV-BSC model-based differential detection

5.1 Mutual information performance of model-based differential detection

In order to examine whether there exists a basic optimal TV-BSC model-based encoding/detection strategy, we assume a two-state Markov source with all valid state transition probabilities, that is, $p_{i,j}^{(d)} \neq 0$, $i, j \in \{0, 1\}$, which is transmitted over the TV-BSC. We fix the channel transition structure (1) and calculate transition probabilities $p_{i,j}^{(d)}$ of the source which optimises the achievable mutual information rate over the channel, by repeating until convergence the following step [16]

$$p_{i,j}^{\{d\}} = \frac{\hat{\ell}_j}{\hat{\ell}_i} \frac{\hat{A}_{ij}}{\hat{W}_{max}} \quad i, j \in \{0, 1\} \tag{15}$$

where $\hat{A}_{ij} = \hat{A}_{ij}(p_{i,j}^{(d)})$ is the noisy adjacency matrix with its maximal eigenvalue $\hat{W}_{max}$ and the corresponding eigenvector $[\hat{\ell}_0, \hat{\ell}_1]'$.

The transition structure (15) converges towards the differential encoder structure, that is, $p_{i,j}^{(d)} \to 1/2$, for

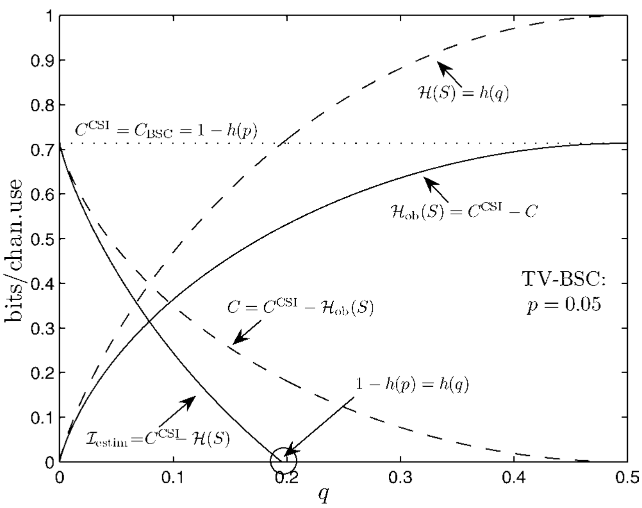

Fig. 1 *TV-BSC process entropy rate $\mathcal{H}(S) = h(q)$, observable TV-BSC process entropy rate $\mathcal{H}_{ob}(S)$ and mutual information rate $\mathcal{I}_{estim} = 1 - h(p) - h(q)$ assuming TV-BSC process estimation*

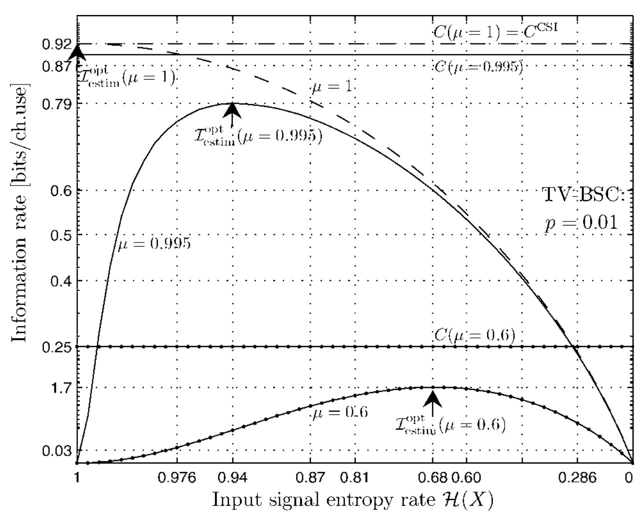

Fig. 2 *Mutual information performance for blind TV-BSC estimation*

$i, j \in \{0, 1\}$, and the cascade which consists of the differential encoder and TV-BSC preserves the TV-BSC information capacity (at worst, a tight lower bound) [16]. Thus, the expectation-maximisation version of the Arimoto-Blahut algorithm, for Markov process transmitted over a noisy finite-state machine channel (15), shows that the differential encoder is an optimal two-state Markov source, in a mutual information sense, which preserves the capacity of TV-BSC.

Generally, the idea of the model-based encoding/detection is to use the channel model, in state equation form, to find an optimal Markov source (encoder) in terms of the mutual information performance. The TV-BSC model-based encoding/detection still has the conventional, differential detection form. More complex channel models would have more complex, generalised forms of the model-based encoding/detection.

Differentially encoded sequence x_k, given by

$$x_k = b_k \oplus x_{k-1} \quad k = 1, 2, 3, \ldots \tag{16}$$

where b_k is kth information bit and x_0 is the reference bit, is transmitted over the TV-BSC, given by the state-space model (3a), (3b). Differential decoding of the received signal y_k is performed as follows

$$\begin{aligned} d_k &= y_k \oplus y_{k-1} = [S_{k-1} \oplus \eta_k \oplus b_k \oplus x_{k-1} \oplus v_k] \\ &\quad \oplus [S_{k-1} \oplus x_{k-1} \oplus v_{k-1}] \\ &= b_k \oplus v_k \oplus v_{k-1} \oplus \eta_k = b_k \oplus \varepsilon_k \oplus \eta_k \end{aligned} \tag{17}$$

Although the channel state process S_k in general exhibits the memory μ, given by (2), sequence d_k in (17) is determined by the innovation $\eta_k = S_k \oplus S_{k-1}$ of the channel process S_k, which is essentially i.i.d. (memoryless). This enables adoption of multiple-symbol differential detection [17], as a form of block-by-block maximum likelihood sequence detection, which is initially developed for time-invariant (memoryless) channels. The multiple-symbol differential detection exploits the distortion correlation from the sequence $\varepsilon_k = v_k \oplus v_{k-1}$, by using a sequence of $N + 1$ samples to detect jointly N transmitted symbols.

It is important to note here that the channel process innovation, which can be generalised as $\eta_k = f(S_n|S_{n-1})$, is the key concept for the model-based differential detection over time-varying channels. In fact, we use the term 'differential' to emphasise the ability of the algorithm to exploit the channel process innovation in order to provide capacity achieving performance. Thereby, the model-based differential detection significantly overcomes the meaning of the conventional differential detection.

In order to analyse mutual information performance of the differential encoding/detection scheme, we define an equivalent channel struture, Fig. 3, for a cascade which consists of the differential encoder (16), the TV-BSC (3a), (3b) and differential decoder (17).

The equivalent channel state at time instant k, $S_k^{(e)}$ can be defined as

$$S_k^{(e)} = c_{i,j}^{(e)} = [v_k = i, v_{k-1} = j] \quad i, j \in \{0, 1\}$$

The transition structure of the equivalent channel is given by (Fig. 3)

$$\begin{aligned} q^{(e)}_{(2\cdot i+j),(2\cdot m+n)} &= p\left(S_k^{(e)} = c_{m,n}^{(e)} | S_{k-1}^{(e)} = c_{i,j}^{(e)}\right) \\ &= \begin{cases} p(v_k = m) = p_m & \text{for } i = n \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{18}$$

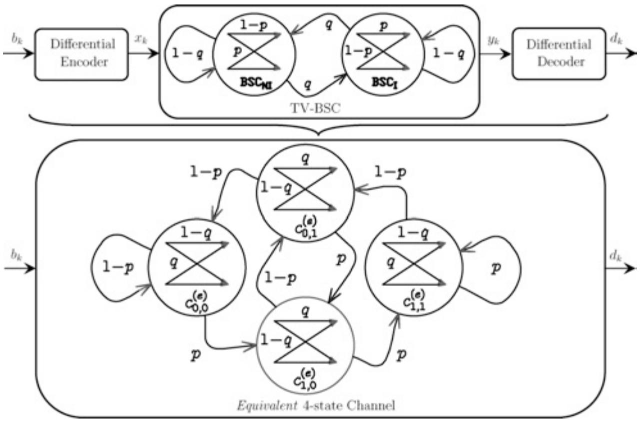

Fig. 3 *Equivalent Markov channel as the cascade of differential encoder, TV-BSC and differential decoder*

where $p_m \triangleq p(v_k = m) = 1 - p$ for $m = 0$, $p_m \triangleq p(v_k = m) = p$ for $m = 1$ and $i,\ j,\ m,\ n \in \{0,\ 1\}$. Consistent with the stationary distribution, the vector of initial channel state probabilities $Q_0^{(e)} = [p(S_0^{(e)} = c_{i,j}^{(e)})]$, i, $j \in \{0, 1\}$ is the solution of the equation $Q^{(e)^{\mathrm{T}}} \cdot Q_0^{(e)} = Q_0^{(e)}$.

However, by balancing probabilities [12], the $(2{\cdot}i + j)$th element of $Q_0^{(e)}$ can be found as

$$[Q_0^{(e)}]_{2\cdot i+j} = p(S_0^{(e)} = c_{i,j}^{(e)}) = p_i \cdot p_j \tag{19}$$

where $i, j \in \{0, 1\}$.

Additionally, the equivalent channel state law is given by (Fig. 3)

$$\begin{aligned} p_{m,n}^{(e)} &= p\left(z^{(e)} = m | S_k^{(e)} = c_{i,j}^{(e)}\right) \\ &= p[\eta_k = i \oplus j \oplus m] = q_\ell \end{aligned} \tag{20}$$

where $z^{(e)} \triangleq b_n \oplus d_n$ is the error functions for the equivalent channel, $n = 2 \cdot i + j$, $l = i \oplus j \oplus m$ and m, i, $j = \{0, 1\}$. Furthermore, $q_\ell \triangleq p(\eta_k = \ell) = 1 - q$ *for* $\ell = 0$, $q_\ell \triangleq p(\eta_k = \ell) = q$ for $\ell = 1$.

The following theorem confirms that the differential detection scheme preserves the information capacity of the TV-BSC channel, as the observation interval $N \to \infty$.

Theorem: The information capacity of the equivalent channel, which is the cascade of the differential encoder, TV-BSC and differential decoder, is equal to the information capacity of the original TV-BSC

$$\begin{aligned} C^{(e)} &= 1 - \lim_{N\to\infty} \frac{1}{N} H\left(Z^{(e)^N}\right) \\ &= 1 - \lim_{N\to\infty} \frac{1}{N} H(Z^N) = C \end{aligned} \tag{21}$$

that is, the differential encoding/detection scheme is information lossless.

Proof: For a stationary stochastic process [12]

$$\lim_{N\to\infty} \frac{1}{N} H(Z^N) = \lim_{N\to\infty} H(Z_N | Z^{N-1}) \tag{22}$$

By the chain rule [12], $H(Z^N) = \sum_{i=1}^{N} H(Z_i|Z^{i-1})$, we have

$$\lim_{N\to\infty} H(Z_N|Z^{N-1}) = \lim_{N\to\infty}\left(\sum_{i=1}^{N} H(Z_i|Z^{i-1}) - \sum_{i=1}^{N-1} H(Z_i|Z^{i-1})\right) = \lim_{N\to\infty} (H(Z^N) - H(Z^{N-1})) \tag{23}$$

Combining (22) and (23), (6) becomes

$$C = 1 - \lim_{N\to\infty} \frac{1}{N} H(Z^N) = 1 - \lim_{N\to\infty} \frac{1}{N+1} H(Z^{N+1}) = 1 - \lim_{N\to\infty} (H(Z^{N+1}) - H(Z^N)) \tag{24}$$

Similarly, the equivalent channel is uniformly symmetric, variable noise channel and its information capacity is

$$C^{(e)} = 1 - \lim_{N\to\infty} \frac{1}{N} H\left(Z^{(e)^N}\right) = 1 - \lim_{N\to\infty} \left(H\left(Z^{(e)^N}\right) - H\left(Z^{(e)^{N-1}}\right)\right) \tag{25}$$

for an input distribution, that is uniform i.i.d.

Lemma 1 in the Appendix proves that $H(Z^{N+1}) = 1 + H(Z^{(e)^N})$. Consequently

$$H(Z^{N+1}) - H(Z^N) = H\left(Z^{(e)^N}\right) - H\left(Z^{(e)^{N-1}}\right) \tag{26}$$

Thus

$$C = C^{(e)} \tag{27}$$

□

A special case implementation of above analysis should also be mentioned. By assuming $q = 0$, the TV-BSC stays forever into initial state and becomes a (time-invariant) BSC. According to (21), one can conclude that differential detection over an unknown BSC achieves the channel information capacity (and so the performance of optimal coherent detection). However, the $q = 0$ assumption means that the channel is constant indefinitely. From the information theory point of view, the indefinitely constant channel assumption is equivalent to the channel state information assumption. Thus, if the channel is time-invariant then there is no fundamental advantage of using the differential over coherent detection and vice-versa.

5.2 Simulation analysis

Fig. 4 depicts the mutual information rate $\mathcal{I}_N^{(e)}$ over state transition probability q, for symbol-by-symbol and N-symbol differential detection ($N = 2, 3$) over the TV-BSC, given by

$$\mathcal{I}_N^{(e)} = 1 - \frac{1}{N} H\left(Z^{(e)^N}\right) \tag{28}$$

assuming an input distribution that is uniform i.i.d. [11]. In order to calculate the entropy $H(Z^{(e)^N})$, the distribution $p(Z^{(e)^N}|S_0^{(e)})$ is calculated recursively by using backward iterative procedure formulated in [1, 18]. Analysis from Fig. 4 confirms that $N = 2$-symbol differential detection of binary phase shift keying already provides a noticeable improvement in mutual information performance in comparison with symbol-by-symbol differential detection. Furtermore, with observation times only on the order of a few symbol intervals, multiple-symbol differential detection practically achieves the channel information capacity ($N = 3$ at higher SNR ($p = 0.05$)).

In order to analyse how quickly $\mathcal{I}_N^{(e)}$, given by (28), approaches the TV-BSC information capacity C, given by (6), by increasing the observation interval, we define the normalised mutual information rate.

$$\tau = \frac{\mathcal{I}_N^{(e)}}{C} \tag{29}$$

Since $\mathcal{I}_N^{(e)}$ in (29) is normalised by the channel information capacity C, the normalisation factor is different for channels with different capacities. Thus, a greater τ means that mutual information rate approaches closer to the channel capacity, but it does not mean a higher absolute mutual information rate. Fig. 5 shows the normalised mutual information rate τ for multiple-symbol differential detection over TV-BSC for two different speeds of channel time-variations $q = 0.01$ and $q = 0.05$ and two different channel state crossover probabilities, $p = 0.05$ and $p = 0.1$. For the same state transition probability q, a longer observation time (more symbols for detection) is needed to approach the channel information capacity in more noisy conditions

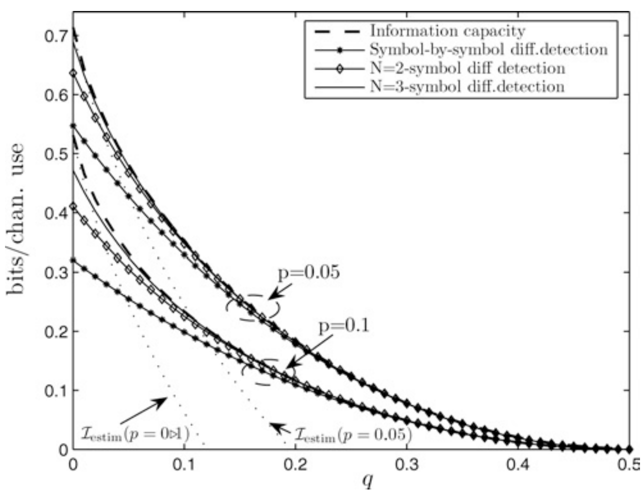

Fig. 4 *Mutual information rate $\mathcal{I}_N^{(e)}$ (28) over the state transition probability q, for symbol-by-symbol and N-symbol differential BPSK ($N = 2, 3$) over the TV-BSC*

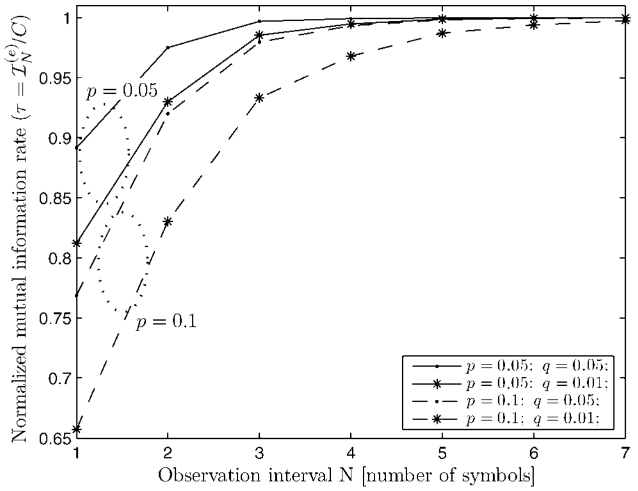

Fig. 5 *Normalised mutual information rate $\tau = \mathcal{I}_N^{(e)}/C$ for multiple-symbol differential detection over the TV-BSC*

($p = 0.1$ against $p = 0.05$ in Fig. 5). For instance, for $p = 0.1$ and $q = 0.01$, seven-symbol observation interval is needed to achieve 0.99 of the channel capacity, Fig. 5. This observation is clearly related to a higher phase distortion correlation ε_k in (17) and exists independently of the channel time variations. The same phenomenon can be observed for time-invariant channels [17].

For the same channel state crossover probability p, faster TV-BSC needs a shorter observation time (less symbols for detection) to approach channel information capacity compared to slower TV-BSC ($q = 0.05$ against $q = 0.01$ in Fig. 5). The reason is that η_k in (17) becomes more dominant over ε_k for faster channel. Consequently, the mutual information penalty of truncating the phase distortion correlation from ε_k, by using a shorter observation interval is smaller compared to slower channel.

Thus, the differential detection scheme is capacity achieving and very robust to an increase in channel time-variation rate (fading rate increase) which are quite opposite trends to coherent detection methods.

6 Conclusions

This paper introduces three key concepts for determining mutual information performance bounds for different detection strategies over time-varying channels. These concepts are separation of entropy principles, model-based differential detection and equivalent channel model. While the separation of entropy principles is essential for determining performance of the coherent detection, the equivalent channel enables us to calculate the mutual information performance of the model-based differential detection. The model-based differential detection, as shown here, uses the state-space approach to exploit channel process innovation in order to provide capacity achieving performance. Thereby, this concept has a much more general meaning than the conventional (multiple-symbol) differential detection.

This paper shows that there is a fundamental advantage of model-based differential detection over channel estimation based detection when the channel is not static (time-varying) and when there is noise–a distinction which is not present in the time-invariant case. This is established by showing a differential scheme which is capacity achieving. In contrast, for channel estimation based detection, it is shown that the lack of perfect channel knowledge, because of noisy channel estimation, essentially reduces the achievable mutual information rate below the channel information capacity.

7 Acknowledgment

This work was partially founded by Australian Research Council Discovery Project number DP 0346430.

8 References

1 Rabiner, L.R.: 'A tutorial on hidden Markov models and selected applications in speech recognition,', *Proc. IEEE*, 1989, **77**, (2), pp. 257–286
2 Gallager, R.G.: 'Information theory and reliable communication' (Wiley, New York, 1968)
3 Shannon, C.E.: 'A mathematical theory of communication', *Bell Syst. Tech. J.*, 1948, **27**, pp. 379–423, 623–656
4 Wiener, N.: 'Extrapolation, interpolation, and smoothing of stationary time series' (M.I.T. Press, Cambridge, MA, 1949)
5 Medard, M.: 'The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel', *IEEE Trans. Inf. Theory*, 2000, **46**, (3), pp. 933–946
6 Lapidoth, A., and Narayan, P.: 'Reliable communication under channel uncertainty', *IEEE Trans. Inf. Theory*, 1998, **44**, (6), pp. 2148–2177
7 Telatar, T.E., and Tse, D.N.C.: 'Capacity and mutual information of wideband multipath fading channels', *IEEE Trans. Inf. Theory*, 2000, **46**, (4), pp. 1384–1400
8 Krusevac, Z.B., Rapajic, P.B., and Kennedy, R.A.: 'Optimal implict channel state estimation for finite state Markov communication channels'. Proc. IEEE Int. Symp. on Inform. Theory (ISIT'06), Seattle, Washington, USA, July 2006
9 Krusevac, Z.B., Rapajic, P.B., and Kennedy, R.A.: 'On channel uncertainty modelling–an information theoretic approach'. Proc. IEEE Glob. Telecomm. Conf. (Globecom), Dallas, TX, USA, November 2004, pp. 410–414
10 Mushkin, M., and Bar-David, I.: 'Capacity and coding for Gilbert-Elliott channels', *IEEE Trans. Inf. Theory*, 1999, **35**, (6), pp. 1277–1290
11 Goldsmith, A.J., and Varaiya, P.: 'Capacity, mutual information, and coding for finite-state Markov channels', *IEEE Trans. Inf. Theroy*, 1996, **42**, (3), pp. 868–886
12 Cover, T.M., and Thomas, J.A..: 'Elements of information theory' (Wiley, New York, 1991, 1st edn.)
13 Johnson, R., Schniter, P., Endres, T.J., Behm, J.D., Brown, D.R., and Casas, R.A.: 'Blind equalization using the constant moulus criterion: a review', *Proc. IEEE*, 1998, **86**, (10), pp. 1927–1950
14 Benvenister, A., and Goursat, M.: 'Blind equalizers', *IEEE Trans. Commun.*, 1984, **32**, pp. 871–833
15 Alajaji, F., Phamdo, N., Farvardin, N., and Fuja, T.: 'Map detection of binary sources over discrete Markov channels'. Proc. IEEE Int. Symp. on Inf. Theory (ISIT), Trondheim, Norway, June 1994
16 Kavčić, A.: 'On the capacity of Markov sources over noisy channels'. IEEE Global Commun. Conf. (GLOBECOM), San Antonio, TX, November 2001, pp. 2997–3001
17 Divsalar, D., and Simon, M.K.: 'Multiple-symbol differential detection of MPSK', *IEEE Trans. Commun.*, 1990, **38**, (3), pp. 300–308
18 Bahl, L.R., Coke, J., Jelinek, F., and Raviv, J.: 'Optimal decoding of linear codes for minimizing symbol error', *IEEE Trans. Inf. Theory*, 1974, **20**, (3), pp. 284–287

9 Appendix

Lemma 1: For input distributions $p(X)$ and $p(B)$ of x_n and b_n, respectively, that are uniform i.i.d., the entropy $H(Z^{N+1})$ of the TV-BSC and the entropy $H(Z^{(e)^N})$ of the equivalent channel are related as

$$H(Z^{N+1}) = 1 + H\big(Z^{(e)^N}\big) \tag{30}$$

Proof: By using (19) for the initial channel state probability and (18) and (20), the distribution of the error function sequence $Z^{(e)^N}$ of the equivalent channel, after N time instants can be expressed

$$\begin{aligned} p\big(Z^{(e)^N}\big) &= \sum_{i=0}^{1}\sum_{j=0}^{1} p\big(Z^{(e)^N} | S_0^{(e)} = c_{i,j}^{(e)}\big) p\big(S_0^{(e)} = c_{i,j}^{(e)}\big) \\ &= \sum_{i=0}^{1}\sum_{j=0}^{1}\left[\sum_{k=0}^{1}\sum_{\ell=0}^{1} p\big(Z^{(e)^N}, S_N^{(e)}\right. \\ &\left. = c_{k,\ell}^{(e)} | S_0 = c_{i,j}^{(e)}\big)\right] p_i \cdot p_j \\ &= \sum_{i=0}^{1} p_i \left[\sum_{k=0}^{1}\left[\sum_{\ell=0}^{1} p(Z^{(e)^N}, S_N\right.\right. \\ &\left.\left. = c_{k,\ell}^{(e)} | S_0^{(e)} = c_{i,0}^{(e)})\right]\right] \end{aligned} \tag{31}$$

where the last equality follows from the fact that the transition from the initial channel state

$S_0^{(e)} = c_{i,j}^{(e)} = c_{i,0}^{(e)}$ does not depend on j by (18) and $\sum_{j=0}^{1} p_j = 1$.

However, expression $\sum_{\ell=0}^{1} p(Z^{(e)^N}, S_N^{(e)} = c_{k,\ell}^{(e)} | S_0^{(e)} = c_{i,0}^{(e)})$ in (31) can be calculated by using backward recursion, based on (18) and (20). The recursion starts

$$\sum_{\ell=0}^{1} p(Z^{(e)^N}, S_N^{(e)} = c_{k,\ell} | S_0^{(e)} = c_{i,0})$$
$$= \sum_{\ell=0}^{1} p\left(Z_N^{(e)}, Z^{(e)^{N-1}}, S_N^{(e)} = c_{k,\ell}^{(e)} | S_0^{(e)} = c_{i,0}^{(e)}\right)$$
$$= \sum_{\ell=0}^{1}\sum_{n=0}^{1}\sum_{t=0}^{1} p\left(Z^{(e)^{N-1}}, S_{N-1}^{(e)} = c_{n,t}^{(e)} | S_0^{(e)} = c_{i,0}^{(e)}\right) \cdot p\left(Z_N^{(e)} | S_N^{(e)} = c_{k,\ell}^{(e)}\right) \cdot p\left(c_{k,\ell}^{(e)} | c_{n,t}^{(e)}\right)$$
$$= p_k \cdot \sum_{\ell=0}^{1} q_{(Z^{(e)}|c_{k,\ell}^{(e)})} \cdot \sum_{t=0}^{1} p\left(Z^{(e)^{N-1}}, S_{N-1}^{(e)} = c_{\ell,t}^{(e)} | S_0^{(e)} = c_{i,0}^{(e)}\right) \tag{32}$$

and ends by

$$\sum_{\ell=0}^{1} p(Z_1^{(e)}, S_1^{(e)} = c_{k,\ell}^{(e)} | S_0^{(e)} = c_{i,j}^{(e)})$$
$$= \sum_{\ell=0}^{1} p(Z_1^{(e)} | S_1^{(e)} = c_{k,\ell}^{(e)}) p(S_1^{(e)} = c_{k,\ell}^{(e)} | S_0^{(e)} = c_{i,j}^{(e)})$$
$$= p(Z_1^{(e)} | S_1^{(e)} = c_{k,i}^{(e)}) p(S_1^{(e)} = c_{k,i}^{(e)} | S_0^{(e)} = c_{i,j}^{(e)})$$
$$= p_k \cdot q_{(Z^{(e)}|c_{k,i}^{(e)})} \tag{33}$$

where $q_{(Z^{(e)}|c_{k,i}^{(e)})} = p(Z^{(e)} | S^{(e)} = c_{k,\ell}^{(e)})$ is the equivalent channel state law for the channel state $c_{k,i}^{(e)}$, given by (20).

Combining (31) with (32) and (33), one can write

$$p(Z^{(e)^N}) = \sum_{i=0}^{1} p_i \sum_{k=0}^{1} p_k \sum_{\ell=0}^{1} q_{(Z^{(e)}|c_{k,\ell}^{(e)})} \cdot p_\ell \sum_{t=0}^{1} q_{(Z^{(e)}|c_{\ell,t}^{(e)})} \cdot p_t \cdot \ldots \cdot \sum_{v=0}^{1} q_{(Z^{(e)}|c_{m,v}^{(e)})} \cdot p_v \cdot q_{(Z^{(e)}|c_{v,i}^{(e)})}$$
$$= \sum_{k=0}^{1} p_k \sum_{\ell=0}^{1} q_{(Z^{(e)}|c_{k,\ell}^{(e)})} \cdot p_\ell \sum_{t=0}^{1} q_{(Z^{(e)}|c_{\ell,t}^{(e)})} \cdot p_t \cdot \ldots \cdot \sum_{v=0}^{1} q_{(Z^{(e)}|c_{m,v}^{(e)})} \cdot p_v \sum_{i=0}^{1} p_i \cdot q_{(Z^{(e)}|c_{v,i}^{(e)})} \tag{34}$$

Furthermore, for the original TV-BSC

$$\sum_{i=0}^{1} p(z_{N+1} = m, Z^N | S_0 = i) = \sum_{i=0}^{1}\sum_{k=0}^{1} p(z_{N+1} = m, Z^N, S_N = k | S_0 = i) = \sum_{i=0}^{1}\sum_{k=0}^{1} p(Z^N, S_N = k | S_0 = i) \cdot p(z_{N+1} = m | S_N = k)$$
$$= \sum_{i=0}^{1}\sum_{k=0}^{1} p(Z^N, S_N = k | S_0 = i) \sum_{\ell=0}^{1} p(z_{N+1} = m | S_{N+1} = \ell) \cdot p(S_{N+1} = \ell | S_N = k)$$
$$= \sum_{i=0}^{1}\sum_{k=0}^{1} p(Z^N, S_N = k | S_0 = i) \sum_{\ell=0}^{1} p_{m\oplus\ell} q_{\ell\oplus k}$$
$$= \sum_{\ell=0}^{1} p_{m\oplus\ell} \sum_{k=0}^{1} q_{\ell\oplus k} \left[\sum_{i=0}^{1} p(Z^N, S_N = k | S_0 = i)\right] \tag{35}$$

However, $\sum_{i=0}^{1} p(Z^N, S_N = k | S_0 = i)$ in (35) can be calculated by using backward recursion. It starts

$$\sum_{i=0}^{1} p(Z^N, S_N = k | S_0 = i) = \sum_{i=0}^{1}\sum_{n=0}^{1} p(Z_N, Z^{N-1}, S_N = k, S_{N-1} = n | S_0 = i) = \sum_{i=0}^{1}\sum_{n=0}^{1} p(Z^{N-1}, S_{N-1} = n | S_0 = i) \cdot p(Z_N | S_N = k) \cdot p(S_N = k | S_{N-1} = n)$$
$$= p_{(Z,k)} \cdot \sum_{n=0}^{1} q_{k\oplus n} \sum_{i=0}^{1} p(Z^{N-1}, S_{N-1} = n | S_0 = i) \tag{36}$$

and ends by

$$\sum_{i=0}^{1} p(Z_1, S_1 = k | S_0 = i)$$
$$= \sum_{i=0}^{1} p(Z_1 | S_1 = k) p(S_1 = k | S_0 = i)$$
$$= p_{(Z,k)} \underbrace{\sum_{i=0}^{1} q_{k\oplus i}}_{=1} = p_{(Z,k)} \tag{37}$$

where $p_{(Z,k)} = p(Z|S = k)$ is the TV-BSC law at the state $S = k$. Combining (35) with (36) and (37), for any $m \in \{0, 1\}$, one can obtain

$$\sum_{i=0}^{1} p(z_{N+1} = m, Z^N | S_0 = i) = \sum_{\ell=0}^{1} p_{m\oplus\ell} \times \sum_{k=0}^{1} q_{\ell\oplus k} \cdot p_{(Z,k)} \cdot \ldots \cdot \sum_{v=0}^{1} q_{r\oplus v} \cdot p_{(Z,v)} \tag{38}$$

However, because of constellation symmetry ($p_i = 1 - p$ for $i = 0$, $p_i = p$ for $i = 1$; $q_i = 1 - q$ for $i = 0$, $q_i = q$ for $i = 1$), (34) and (38) are the same combinations of the same p_i and q_j multiplications. It leads to the following equality

$$\sum_{Z^N}\left[\sum_{i=0}^{1} p(z_{N+1} = m, Z^N | S_0 = i) \cdot \log_2\left(\sum_{i=0}^{1} p(z_{N+1} = m, Z^N | S_0 = i)\right)\right]$$
$$= \sum_{Z^{(e)^N}}\left[p(Z^{(e)^N}) \log_2 p(Z^{(e)^N})\right] \tag{39}$$

for any $m \in \{0,1\}$. Consequently

$$\sum_{Z^{N+1}}\left[\sum_{i=0}^{1} p(Z^{N+1}|S_0 = i)\cdot \log_2\left(\sum_{i=0}^{1} p(Z^{N+1}|S_0 = i)\right)\right]$$
$$= \sum_{m=0}^{1}\sum_{Z^N}\left[\sum_{i=0}^{1} p(z_{N+1} = m, Z^N|S_0 = i)\right.$$
$$\left.\cdot\ \log_2\left(\sum_{i=0}^{1} p\big(z_{N+1} = m, Z^N|S_0 = i\big)\right)\right]$$
$$= \sum_{m=0}^{1}\sum_{Z^{(e)^N}}\left[p\big(Z^{(e)^N}\big)\log_2 p\big(Z^{(e)^N}\big)\right]$$
$$= -\sum_{m=0}^{1} H\big(Z^{(e)^N}\big) = -2\cdot H\big(Z^{(e)^N}\big) \qquad (40)$$

Finally, since the initial TV-BSC state probability $p(S_0 = i) = 0.5$ for any $m \in \{0, 1\}$, we have

$$H(Z^{N+1}) = -\sum_{Z^{N+1}} p(Z^{N+1})\cdot \log_2(p(Z^{N+1}))$$
$$= -\sum_{Z^{N+1}}\left[\frac{1}{2}\sum_{i=0}^{1} p(Z^{N+1}|S_0 = i)\right.$$
$$\left.\cdot\ \log_2\left(\frac{1}{2}\sum_{i=0}^{1} p(Z^{N+1}|S_0 = i)\right)\right]$$
$$= -\frac{1}{2}\cdot \log_2\left(\frac{1}{2}\right)\cdot \underbrace{\sum_{i=0}^{1}\underbrace{\sum_{Z^{N+1}} p(Z^{N+1}|S_0 = i)}_{=1}}_{=2}$$
$$-\frac{1}{2}\underbrace{\sum_{Z^{N+1}}\left[\textstyle\sum_{i=0}^{1} p(Z^{N+1}|S_0 = i)\cdot \log_2\left(\sum_{i=0}^{1} p(Z^{N+1}|S_0 = i)\right)\right]}_{-2\cdot H(Z^{(e)^N})\text{ by (40)}}$$
$$= \log_2(2) + H\left(Z^{(e)^N}\right) = 1 + H\left(Z^{(e)^N}\right) \qquad \square$$