# Equivalent channel for capacity analysis of differential detection over time-varying communication channels

Zarko B. Krusevac
Australian National University
Canberra, ACT 2052 Australia
zarko.krusevac@rsise.anu.edu.au

Predrag B. Rapajic
University of Greenwich at Med. Pemb.
KENT ME4 4TB United Kingdom
P.Rapajic@greenwich.ac.uk

***Abstract*— This paper introduces the concept of equivalent channel for mutual information performance analysis of multiple-symbol differential MPSK ($M$-phase shift keying) over time-correlated, time-varying flat-fading communication channels. It is proven that differential detection scheme theoretically preserves the channel information capacity when the observation interval approaches infinity. A state space approach is used to model time correlation of time varying channel phase. It is shown that the differential decoding implicitly uses a sequence of innovations of the channel process time correlation and this sequence is essentially uncorrelated (i.i.d). It enables utilization of multiple-symbol differential detection, as a form of block-by-block maximum likelihood sequence detection for capacity achieving mutual information performance.**

## I. Introduction

*Background and Motivation* - This paper uses a model-based, state-space approach for mutual information performance analysis of multiple-symbol differential detection over time-correlated, time-varying communication channels. In order to calculate the information capacity of the differential encoding/decoding scheme, this paper defines an *equivalent* FSM (finite-state Markov) structure, which is based on the state-space finite-state channel model and analytical expressions for the differential encoder and decoder.

The differential detection creates dependency between consecutive receiver outputs [1], providing possibility to use the correlation between the phase distortion experienced by different transmitted PSK symbols. Conventional *symbol-by-symbol* differential detection suffers from a performance penalty (additional required SNR at a given bit error rate [2]) when compared to ideal (perfect carrier phase reference) coherent detection [1]. However *multiple-symbol* differential detection [2]–[4] exploits the phase distortion correlation by using a sequence of $N+1$ samples to detect jointly $N$ transmitted symbols. In [5] it is shown that, assuming a constant channel phase, there is no a fundamental advantage, in terms of the achievable information rates, of using multiple-symbol differential PSK or coherent PSK.

Motivated by the rather encouraging performance of the multiple-symbol differential detection over the additive white Gaussian noise (AWGN) channel [2], [3], [5], error performance of multiple-symbol differential detection of PSK signaling over time-correlated, time-varying flat-fading Rayleigh channels is considered in [6], [7]. However, when analyzing differential detection over time-varying flat-fading channels, the literature limits attention to two extreme cases of modeling, either assuming fading channel gain time variations are not correlated, representing the most rapidly time-varying case [4], or the time-variations are sufficiently slow that they are virtually time-invariant over the observation interval as in the block fading case [4], [8]. While independent fading model underestimates the channel information capacity, the block model does not enable an analysis of channel process time correlation effects upon the mutual information performance. In addition, given an infinitely long block, the block-model degenerates to the time invariant channel and, hence, an overall information capacity analysis when observation interval approaches infinity does not make sense. Thus, in order to analyze mutual information performance related to channel process time correlation, one needs more realistic models which capture the time-varying channel time correlation properties.

This paper introduces and implements an autoregressive (AR) state space model which is superior in modeling the time correlation properties of time-varying fading channels than either the independent fading model or the block fading model, commonly find in the literature.

*Contributons* -

1) The concept of equivalent FSM channel is introduced which enables capacity analysis of multiple-symbol differential MPSK ($M$-phase shift keying) over time-correlated, time-varying communication channels.
2) It is proven that the multiple-symbol differential detection theoretically preserves the channel information capacity when the observation time approaches infinity.
3) Simulation analysis confirms theoretical findings by showing that multiple-symbol differential ML detection of BPSK and QPSK practically achieves the channel information capacity with observation times only on the order of a few symbol intervals.

## II. Signal Model

If a flat fading process is slow enough, so that it is essentially constant over symbol intervals, the $k$th matched filter output at the receiver side, for MPSK transmission, can be represented as [1]

$$r_k = g_k u_k + n_k \tag{1}$$

where $r_k$ is the received signal, $g_k$ is a correlated channel fading process, $u_k$ is the transmitted MPSK signal and $n_k$ is an i.i.d. complex-valued Gaussian noise process (AWGN). The information symbol $x_k$ takes values in $\{0, 1, \ldots, M-1\}$ and is mapped to the transmitted MPSK signal as

$$u_k \triangleq \sqrt{\mathcal{E}_s}\exp(j2\pi x_k/M) \tag{2}$$

where $\mathcal{E}_s$ is the energy per symbol. In general, the actual realization of flat-fading gain $g_k$ in (1)

$$g_k \triangleq X_k + jY_k = a_k \exp(j\theta_k) \tag{3}$$

is unknown to the receiver *a priori*. $a_k$ and $\theta_k$ denote the channel amplitude and phase, respectively. This leads to a statistical characterization of the fading channel.

## III. State Space Time-varying Channel Phase Model

In order to use more realistic models, which are more consistent with real propagation conditions than independent or block fading models, our mutual information analysis assumes an autoregressive finite-state Markov model to capture the correlated nature of the time variations of $\theta_k$. This provides the following advantages: 1) simplification; 2) ease of computer modeling; 3) a simplified algebraic description of the channel phase dynamics; but most importantly 4) the model falls in a class for which one can facilitate the use of the forward-backward algorithm enabling significant information theoretic results to be brought to bare on the problem.

For the amplitude, $a_k$, we assume it is an independent (uncorrelated) time-varying fading channel amplitude process. Our approach is supported by the BER performance analysis in [9] and mutual information performance in [10] which show that PSK receivers which rely on a simple MMSE symbol-by-symbol amplitude estimation combined with forward-backward phase estimation on the finite-state Markov phase model, perform only slightly worse then having perfect amplitude knowledge (amplitude CSI) at the receiver.

The time-varying flat-fading channel phase is partitioned into $M$ equiprobable, non-overlapping intervals. Each partition corresponds to an FSM channel state which can be identified with $m \in \{0, 1, \ldots, M-1\}$ as follows [9]

$$\theta_k \in \Omega(m) \triangleq \Big[\frac{2\pi m - \pi}{M}, \frac{2\pi m + \pi}{M}\Big) \Longleftrightarrow s_k = m, \tag{4}$$

where $\theta_k$ and $s_k \in \{0, 1, \ldots, M-1\}$ are channel phase and corresponding channel state, at time instant $k$, respectively.

Assuming an AR(1) $M$-state $M$-ary symmetric Markov model for time correlated, time-varying flat-fading channel phase and based on (1), (2) and (3), the state-space channel model becomes

$$y_k = s_k \oplus x_k \oplus v_k \tag{5a}$$

$$s_k = s_{k-1} \oplus \eta_k \tag{5b}$$

where $y_k$ is the received signal phase, $\eta_k$ is an $M$-ary i.i.d. process noise, $\oplus$ and $\ominus$ are modulo-$M$ addition and subtraction, respectively and $v_k$ is $M$-ary phase noise.

The channel state law for each particular channel state $i \in \{0, 1, \ldots, M-1\}$, is modeled as an $M$-ary symmetric channel. For a given channel state $s_k = i$ at the time instant $k$, crossover probabilities $p_{n,m}(i)$, that $y_k = m$ is received if $x_k = n$ is sent, $n, m \in \{0, 1, \ldots, M-1\}$, is given by

$$\begin{aligned} p_{n,m}(i) &= p(y_k = m | x_k = n, s_k = i) \\ &= p(i \oplus v_k = n \ominus m) = p(v_k = n \ominus m \ominus i) \end{aligned} \tag{6}$$

Thereby, (6) is determined by probabilities $p(v_k = i)$, $i \in \{0, 1, \ldots, M-1\}$, which depend on statistical model used for the channel gain amplitude $a_k$ in (3).

Statistics of process noise $\{\eta_k\}$ is determined by state transition probabilities of channel phase state process (4) as

$$\begin{aligned} q_{i,j} &= p(s_k = j | s_{k-1} = i) \\ &= M \iint_{\Omega(i)\times\Omega(j)} f(\theta_k, \theta_{k-1})\, d\theta_k d\theta_{k-1} \end{aligned} \tag{7}$$

The term $f(\theta_k, \theta_{k-1})$ in (7) is the probability density function of flat-fading channel phase at consecutive time instants $k$ and $k-1$, which depends on statistical model used for the channel gain phase in $\theta_k$ (3).

It is important to notice that the time-varying channel phase state process (5b) introduces channel phase correlation into the signal phase observation process (5a).

## IV. Differential encoding and detection

Differential encoding of $M$-ary sequence $x_k$ is given by

$$x_k = b_k \oplus x_{k-1}, \quad k = 1, 2, 3, \ldots \tag{8}$$

where $b_k$ is $k$th raw information symbol and $x_0$ is the reference symbol.

The differentially encoded sequence $x_k$ (8) is transmitted over the $M$-state, $M$-ary symmetric channel, given by the state-space model (5a), (5b). Differential decoding of the received signal phase $y_k$ is performed as follows:

$$\begin{aligned} d_k &= y_k \ominus y_{k-1} \\ &= [s_{k-1} \oplus \eta_k \oplus b_k \oplus x_{k-1} \oplus v_k] \\ &\quad \ominus [s_{k-1} \oplus x_{k-1} \oplus v_{k-1}] \\ &= b_k \oplus v_k \ominus v_{k-1} \oplus \eta_k = b_k \oplus \varepsilon_k \oplus \eta_k \end{aligned} \tag{9}$$

Although the channel phase process is assumed to be correlated (5b), sequence $d_k$ from (9) is determined by the innovation $\eta_k = s_k - s_{k-1}$ of the channel process (5b), which is essentially i.i.d. (uncorrelated). It enables the adoption of multiple-symbol differential detection for the case of time uncorrelated channels, which exploits the phase distortion correlation from the sequence $\varepsilon_k = v_k - v_{k-1}$, by using a sequence of $N+1$ samples to detect jointly $N$ transmitted symbols.

## V. *Equivalent* FSM CHANNEL

In order to analyze mutual information performance of the differential encoding/detection scheme, we define an *equivalent* FSM channel for a cascade which consists of the differential encoder (8), $M$-state $M$-ary symmetric channel and differential decoder (9). The *equivalent* FSM structure is based on the state-space model (5a), (5b) for $M$-state, $M$-ary symmetric channel and expression (9).

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

$$s_k^{(e)} = c_{i,j}^{(e)} = [v_k = i, v_{k-1} = j],$$

where $i, j \in \{0, 1, \ldots, M-1\}$, with a total number of $M^2$ states. The transition structure of the *equivalent* FSM channel is given by

$$\begin{aligned} q^{(e)}_{(M\cdot i+j),(M\cdot m+n)} &= p(s_k^{(e)} = c_{m,n}^{(e)} | s_{k-1}^{(e)} = c_{i,j}^{(e)}) \\ &= \begin{cases} p(v_k = m) & \text{for } i = n \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{10}$$

where $i, j, m, n \in \{0, 1, \ldots, M-1\}$.

The transition structure (10) (and its memory) is not determined by the original channel phase process correlation, but by the phase noise sequence distribution $p(v_k = m)$.

Additionally, the *equivalent* FSM channel state law

$$\begin{aligned} p_{m,n}^{(e)}(c_{i,j}^{(e)}) &= p(d_k = n | b_k = m, s_k^{(e)} = c_{i,j}^{(e)}) \\ &= p(\eta_k = n \ominus m \ominus i \oplus j) \end{aligned} \tag{11}$$

is determined by the distribution of innovation $\eta_k = s_k - s_{k-1}$ of the phase process of the original channel.

*Theorem 1:* The information capacity of the *equivalent* FSM channel is equal to the information capacity of the original $M$-state, $M$-ary symmetric channel i.e., the differential encoding/detection scheme is information lossless.

Proof: The $M$-state $M$-ary symmetric channel is uniformly symmetric, variable noise channel [11] and assuming an input distributions $p(X)$ that is uniform i.i.d., the channel information capacity is given by [11]

$$C = \log_2 M - \lim_{N\to\infty} \frac{1}{N+1} H(Z^{N+1}) \tag{12}$$

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

$$C^{(e)} = \log_2 M - \lim_{N\to\infty} \frac{1}{N} H({Z^{(e)}}^N) \tag{13}$$

for an input distribution $p(B)$, that is uniform i.i.d. $Z$ in (12) and $Z^{(e)}$ in (13) are error functions [11], for the original and *equivalent* channel, respectively and $i, j \in \{0, 1, \ldots, M-1\}$. $H(\cdot)$ is the entropy function [12].

For a stationary stochastic process $Z_N$ [12]

$$\lim_{N\to\infty} \frac{1}{N} H(Z^N) = \lim_{N\to\infty} H(Z_N | Z^{N-1}) \tag{14}$$

By the chain rule [12],

$$(Z^N) = \sum_{i=1}^{N} H(Z_i | Z^{i-1}) \tag{15}$$

and

$$\lim_{N\to\infty} H(Z_N | Z^{N-1}) = \lim_{N\to\infty} \left(H(Z^N) - H(Z^{N-1})\right) \tag{16}$$

Combining (14) and (16), expression (12) becomes

$$C = \log_2 M - \lim_{N\to\infty} \left(H(Z^{N+1}) - H(Z^N)\right) \tag{17}$$

Similarly, (13) can be expressed

$$C^{(e)} = \log_2 M - \lim_{N\to\infty} \left(H({Z^{(e)}}^N) - H({Z^{(e)}}^{N-1})\right) \tag{18}$$

Lemma 1 in Appendix I proves that

$$H(Z^{N+1}) = \log_2 M + H({Z^{(e)}}^N). \tag{19}$$

Consequently,

$$H(Z^{N+1}) - H(Z^N) = H({Z^{(e)}}^N) - H({Z^{(e)}}^{N-1}) \tag{20}$$

Thus, $C = C^{(e)}$. $\square$

Due to existence of the symbol $x_0$, serving as a reference for the differential codec, $N$-symbol observation interval for the $M$-state, $M$-ary symmetric channel implicitly assumes a $N+1$-symbol observation interval for differential detection and the *equivalent* channel. However, the reference symbol $x_0$ represents a negligible amount of information for the actual information transfer.

The above result means there is a potential fundamental advantage of using multiple-symbol differential detection over coherent detection for time-correlated, time-varying communication channels in the presence of channel noise. Although explicit or implicit (blind) channel estimation methods for coherent detection exploit the time-varying channel process correlation (memory) to improve channel estimation, coherent detection may not be optimal (in terms of achievable mutual information rate) in the presence of channel noise [13]. The reason is that the time varying channel process is not completely observable in the presence of channel noise [14].

## VI. SIMULATION ANALYSIS

Here we provide performance analysis of the maximum mutual information rate versus the average received SNR per bit, $\gamma_b$, for $N$-symbol differential detection of BPSK and QPSK. Probabilities $p(v_k = i)$ in (5b) (which determines the channel state law (6)), the probability density function $f(\theta_k, \theta_{k-1})$ in (7) (which determines the state transition probabilities (7)) and $\gamma_b = 2\sigma^2 \mathcal{E}_b / N_0$ are calculated assuming Clarke's model [15] for the fading channel gain (3) with normalized fading power $2\sigma^2 = 1$.

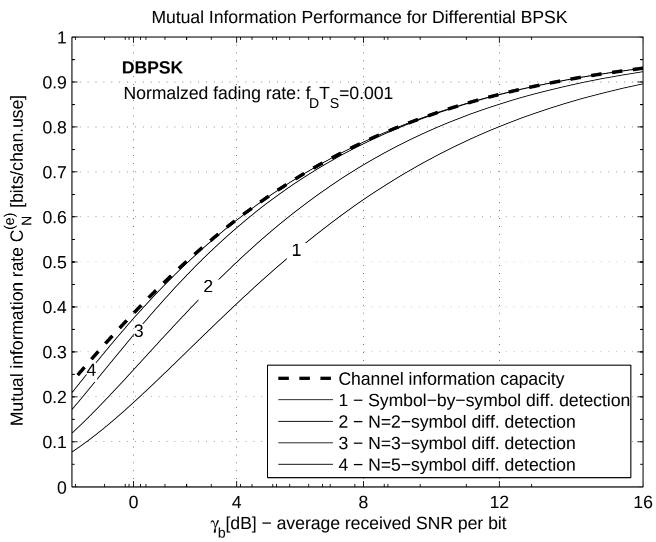


Fig. 1. Mutual information rate $C_N^{(e)}$ (21) for symbol-by-symbol and $N$-symbol differential BPSK

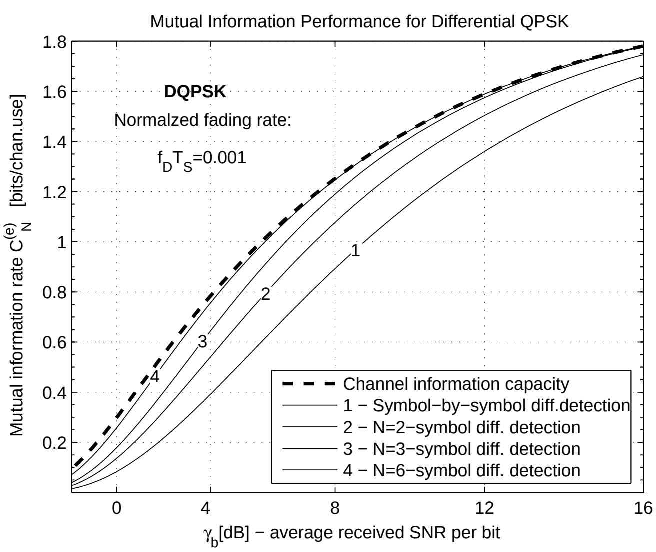


Fig. 2. Mutual information rate $C_N^{(e)}$ (21) for symbol-by-symbol and $N$-symbol differential QPSK

The maximum mutual information rate $C_N^{(e)}$ , for $N$-symbol, multiple-symbol differential MPSK is calculated as

$$C_N^{(e)} = \log_2 M - \frac{1}{N} H({Z^{(e)}}^N) \tag{21}$$

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 [16].

Fig. 1 and Fig. 2 depict the mutual information rate $C_N^{(e)}$ (21) over average received SNR per bit, $\gamma_b$, for symbol-by-symbol and $N = 2, 3, 4$-symbol differential BPSK (M=2) and QPSK (M=4), respectively, for normalized fading rate $f_D T_s = 0.001$. It is shown that practically, the channel information capacity is approached with observation times only on the order of a few symbol intervals.

## APPENDIX I

*Lemma 1:*

$$H(Z^{N+1}) = H({Z^{(e)}}^N) + \log_2 M \tag{22}$$

Proof: Due to constellation and channel state symmetry of the $M$-state $M$-ary symmetric channel one can denote $p_m \triangleq p_{i,j}$ and $q_m \triangleq q_{i,j}$ for $|j - i| = m$. By using expressions (11) and (10), the distribution of the error function sequence ${Z^{(e)}}^N$ of the *equivalent* channel, after $N$ time instants can be expressed

$$\begin{aligned} &p({Z^{(e)}}^N) \\ &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} p({Z^{(e)}}^N | s_0^{(e)} = c_{i,j}^{(e)}) p(s_0^{(e)} = c_{i,j}^{(e)}) \\ &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \Big[ \sum_{k=0}^{M-1} \sum_{\ell=0}^{M-1} p({Z^{(e)}}^N, s_N^{(e)} = c_{k,\ell}^{(e)} | s_0^{(e)} = c_{i,j}^{(e)}) \Big] p_i \cdot p_j \\ &= \sum_{i=0}^{M-1} p_i \Big[ \sum_{k=0}^{M-1} \Big[ \sum_{\ell=0}^{M-1} p({Z^{(e)}}^N, s_N^{(e)} = c_{k,\ell}^{(e)} | s_0^{(e)} = c_{i,0}^{(e)}) \Big] \Big] \end{aligned} \tag{23}$$

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 (10) and $\sum_{j=0}^{M-1} p_j = 1$. Furthermore, applying probability balancing, one can find

$$p(s_0^{(e)} = c_{i,j}^{(e)}) = p_i \cdot p_j. \tag{24}$$

However, expression $\sum_{\ell=0}^{M-1} p({Z^{(e)}}^N, s_N^{(e)} = c_{k,\ell}^{(e)} | s_0^{(e)} = c_{i,0}^{(e)})$ in (23), can be calculated by using backward recursion. The recursion starts

$$\begin{aligned} &\sum_{\ell=0}^{M-1} p({Z^{(e)}}^N, s_N^{(e)} = c_{k,\ell} | s_0^{(e)} = c_{i,0}) \\ &= \sum_{\ell=0}^{M-1} p(Z_N^{(e)}, {Z^{(e)}}^{N-1}, s_N^{(e)} = c_{k,\ell}^{(e)} | s_0^{(e)} = c_{i,0}^{(e)}) \\ &= \sum_{\ell=0}^{M-1} \sum_{n=0}^{M-1} \sum_{t=0}^{M-1} p({Z^{(e)}}^{N-1}, s_{N-1}^{(e)} = c_{n,t}^{(e)} | s_0^{(e)} = c_{i,0}^{(e)}) \\ &\quad \cdot p(Z_N^{(e)} | s_N^{(e)} = c_{k,\ell}^{(e)}) \cdot p(c_{k,\ell}^{(e)} | c_{n,t}^{(e)}) \\ &= p_k \cdot \sum_{\ell=0}^{M-1} q_{(Z^{(e)} | c_{k,\ell}^{(e)})} \cdot \sum_{t=0}^{M-1} p({Z^{(e)}}^{N-1}, s_{N-1}^{(e)} = c_{\ell,t}^{(e)} | s_0^{(e)} = c_{i,0}^{(e)}) \end{aligned} \tag{25}$$

and ends by

$$\begin{aligned}
&\sum_{\ell=0}^{M-1} p(Z_1^{(e)}, s_1^{(e)} = c_{k,\ell}^{(e)} | s_0^{(e)} = c_{i,j}^{(e)}) \\
&= \sum_{\ell=0}^{M-1} p(Z_1^{(e)} | s_1^{(e)} = c_{k,\ell}^{(e)}) p(s_1^{(e)} = c_{k,\ell}^{(e)} | s_0^{(e)} = c_{i,j}^{(e)}) \\
&= p(Z_1^{(e)} | s_1^{(e)} = c_{k,i}^{(e)}) p(s_1^{(e)} = c_{k,i}^{(e)} | s_0^{(e)} = c_{i,j}^{(e)}) \\
&= p_k \cdot q_{(Z^{(e)}|c_{k,i}^{(e)})} \qquad (26)
\end{aligned}$$

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 (11).

Combining expression (23) with (25) and (26), one can write

$$\begin{aligned}
p({Z^{(e)}}^N) =& \sum_{i=0}^{M-1} p_i \sum_{k=0}^{M-1} p_k \sum_{\ell=0}^{M-1} q_{(Z^{(e)}|c_{k,\ell}^{(e)})} \cdot p_\ell \sum_{t=0}^{M-1} q_{(Z^{(e)}|c_{\ell,t}^{(e)})} \\
& \cdot p_t \cdot \ldots \cdot \sum_{v=0}^{M-1} q_{(Z^{(e)}|c_{m,v}^{(e)})} \cdot p_v \cdot q_{(Z^{(e)}|c_{v,i}^{(e)})} \\
=& \sum_{k=0}^{M-1} p_k \sum_{\ell=0}^{M-1} q_{(Z^{(e)}|c_{k,\ell}^{(e)})} \cdot p_\ell \sum_{t=0}^{M-1} q_{(Z^{(e)}|c_{\ell,t}^{(e)})} \qquad (27) \\
& \cdot p_t \cdot \ldots \cdot \sum_{v=0}^{M-1} q_{(Z^{(e)}|c_{m,v}^{(e)})} \cdot p_v \sum_{i=0}^{M-1} p_i \cdot q_{(Z^{(e)}|c_{v,i}^{(e)})}
\end{aligned}$$

Furthermore, for the M-state, M-ary symmetric channel

$$\begin{aligned}
&\sum_{i=0}^{M-1} p(z_{N+1} = m, Z^N | s_0 = i) \\
&= \sum_{i=0}^{M-1} \sum_{k=0}^{M-1} p(z_{N+1} = m, Z^N, s_N = k | s_0 = i) \\
&= \sum_{i=0}^{M-1} \sum_{k=0}^{M-1} p(Z^N, s_N = k | s_0 = i) \cdot p(z_{N+1} = m | s_N = k) \\
&= \sum_{i=0}^{M-1} \sum_{k=0}^{M-1} p(Z^N, s_N = k | s_0 = i) \\
&\cdot \sum_{\ell=0}^{M-1} p(z_{N+1} = m | s_{N+1} = \ell) \cdot p(s_{N+1} = \ell | s_N = k) \\
&= \sum_{i=0}^{M-1} \sum_{k=0}^{M-1} p(Z^N, s_N = k | s_0 = i) \sum_{\ell=0}^{M-1} p_{|m-\ell|} q_{|\ell-k|} \\
&= \sum_{\ell=0}^{M-1} p_{|m-\ell|} \sum_{k=0}^{M-1} q_{|\ell-k|} \big[ \sum_{i=0}^{M-1} p(Z^N, s_N = k | s_0 = i) \big] \qquad (28)
\end{aligned}$$

However, expression $\sum_{i=0}^{M-1} p(Z^N, s_N = k | s_0 = i)$ in (28), can be calculated by using backward recursion. It stars

$$\begin{aligned}
&\sum_{i=0}^{M-1} p(Z^N, s_N = k | s_0 = i) = \sum_{i=0}^{M-1} p(Z_N, Z^{N-1}, s_N = k | s_0 = i) \\
&= \sum_{i=0}^{M-1} \sum_{n=0}^{M-1} p(Z^{N-1}, s_{N-1} = n | s_0 = i) \\
&\qquad \cdot p(Z_N | s_N = k) \cdot p(s_N = k | s_{N-1} = n) \\
&= p_{(Z,k)} \cdot \sum_{n=0}^{M-1} q_{|k-n|} \sum_{i=0}^{M-1} p(Z^{N-1}, s_{N-1} = n | s_0 = i) \qquad (29)
\end{aligned}$$

and ends by

$$\begin{aligned}
&\sum_{i=0}^{M-1} p(Z_1, s_1 = k | s_0 = i) = \sum_{i=0}^{M-1} p(Z_1 | s_1 = k) p(s_1 = k | s_0 = i) \\
&= p_{(Z,k)} \underbrace{\sum_{i=0}^{M-1} q_{|k-i|}}_{=1} = p_{(Z,k)} \qquad (30)
\end{aligned}$$

where $p_{(Z,k)} = p(Z|s = k)$ is the channel law at the state $s = k$, defined by (6). Combining expression (28) with (29) and (30), one can get

$$\begin{aligned}
&\sum_{i=0}^{M-1} p(z_{N+1} = m, Z^N | s_0 = i) \\
&= \sum_{\ell=0}^{M-1} p_{|m-\ell|} \sum_{k=0}^{M-1} q_{|\ell-k|} \cdot p_{(Z,k)} \cdot \ldots \cdot \sum_{v=0}^{M-1} q_{|r-v|} \cdot p_{(Z,v)} \qquad (31)
\end{aligned}$$

for any $m \in \{0, 1, \ldots, M-1\}$.

However, due to constellation symmetry, (27) and (31) are the same combinations of the same $p_i$ and $q_j$ multiplications. It leads to the following equality

$$\begin{aligned}
&\sum_{Z^N} \big[ \sum_{i=0}^{M-1} p(z_{N+1} = m, Z^N | s_0 = i) \\
&\qquad \cdot \log_2 \big( \sum_{i=0}^{M-1} p(z_{N+1} = m, Z^N | s_0 = i) \big) \big] \\
&\qquad = \sum_{{Z^{(e)}}^N} \big[ p({Z^{(e)}}^N) \log_2 p({Z^{(e)}}^N) \big] \qquad (32)
\end{aligned}$$

for any $m \in \{0, 1, \ldots, M-1\}$.

Consequently,

$$\sum_{Z^{N+1}} \Big[ \sum_{i=0}^{M-1} p(Z^{N+1}|s_0=i) \cdot \log_2 \Big( \sum_{i=0}^{M-1} p(Z^{N+1}|s_0=i) \Big) \Big]$$
$$= \sum_{m=0}^{M-1} \sum_{Z^N} \Big[ \sum_{i=0}^{M-1} p(z_{N+1}=m, Z^N|s_0=i)$$
$$\cdot \log_2 \Big( \sum_{i=0}^{M-1} p(z_{N+1}=m, Z^N|s_0=i) \Big) \Big]$$
$$= \sum_{m=0}^{M-1} \sum_{{Z^{(e)}}^N} \Big[ p({Z^{(e)}}^N) \log_2 p({Z^{(e)}}^N) \Big]$$
$$= -\sum_{m=0}^{M-1} H({Z^{(e)}}^N) = -M \cdot H({Z^{(e)}}^N) \quad (33)$$

Finally, since the initial channel state probability $p(s_0 = i) = 1/M$ for any $m \in \{0, 1, \ldots, M-1\}$, we have:

$$H(Z^{N+1}) = -\sum_{Z^{N+1}} p(Z^{N+1}) \cdot \log_2(p(Z^{N+1}))$$
$$= -\sum_{Z^{N+1}} \Big[ \frac{1}{M} \sum_{i=0}^{M-1} p(Z^{N+1}|s_0=i) \log_2 \Big( \frac{1}{M} \sum_{i=0}^{M-1} p(Z^{N+1}|s_0=i) \Big) \Big]$$
$$= -\frac{1}{M} \Big[ \log_2(\frac{1}{M}) \cdot \underbrace{\sum_{i=0}^{M-1} \underbrace{\sum_{Z^{N+1}} p(Z^{N+1}|s_0=i)}_{=1}}_{=M}$$
$$\underbrace{\sum_{Z^{N+1}} \sum_{i=0}^{M-1} p(Z^{N+1}|s_0=i) \cdot \log_2 \Big( \sum_{i=0}^{M-1} p(Z^{N+1}|s_0=i) \Big) \Big]}_{-M \cdot H({Z^{(e)}}^N) \text{ by (33)}}$$
$$= \log_2(M) + H({Z^{(e)}}^N) \qquad \square$$

## References


[1] J. G. Proakis, *Digital Communications*, 4th ed. New York: Mc Graw Hill, 2000.

[2] D. Divsalar and M. K. Simon, "Multiple-symbol differential detection of MPSK," *IEEE Transactions on Communications*, vol. 38, no. 3, pp. 300–308, Mar. 1990.

[3] K. M. Mackenthun, Jr., "A fast algorithm for multiple-symbol differential detection of MPSK," *IEEE Transactions on Communications*, vol. 42, no. 2/3/4, feb/mar/apr 1994.

[4] D. Divsalar and M. K. Simon, "Maximum-likelihood differential detection of uncoded and trellis coded amplitude phase modulation over AWGN and fading channels - metric and performance," *IEEE Transactions on Communications*, vol. 42, no. 1, pp. 76–89, Jan. 1994.

[5] P. Kaplan and S. Shamai, "On the achievable information rates of DPSK," *IEE Proc.-1*, vol. 139, no. 3, June 1992.

[6] P. Ho and D. Fung, "Error performance of multiple-symbol differential detection of PSK signals transmitted over correlated Rayleigh fading channels," *IEEE Transactions on Communications*, vol. 40, no. 10, pp. 1566–1569, Oct. 1992.

[7] M. K. Simon and M. S. Alouini, "Multiple-symbol differential detection with diversity reception," *IEEE Transactions on Communications*, vol. 49, no. 8, pp. 300–308, Aug. 1990.

[8] L. Lampe, R. Fischer, and S. Calabro, "Differential encoding strategies for transmission over fading channels," *Electronics Letters*, vol. 35, no. 3, Feb. 1999.

[9] C. Komninakis and R. D. Wesel, "Joint iterative channel estimation and decoding in flat correlated Rayleigh fading," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1706–1717, Sept. 2001.

[10] P. Sadeghi and P. B. Rapajic, "Capacity performance analysis of coherent detection in correlated fading channels using finite-state Markov models," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, LA, USA, Sept. 2004, pp. 2158–2162.

[11] A. J. Goldsmith and P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Transactions on Information Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. New York: Wiley, 1991.

[13] Z. B. Krusevac, P. B. Rapajic, and R. A. Kennedy, "Optimal implicit channel state estimation for finite-state Markov communication channels," in *Proc. IEEE Int. Symp. on Inform. Theory (ISIT'06)*, Seattle, Washington, USA, July 2006, pp. 2657–2661.

[14] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[15] R. H. Clarke, "A statistical theory of mobile-radio reception," *Bell Syst. Tech. J.*, vol. 47, no. 6, pp. 957–1000, 1968.

[16] L. R. Rabiner, "A tutorial on hidden Markov models and selected application in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.