Channel Estimation and Differential Detection for MPSK Signaling in Time-Varying Flat-Fading Communication Channels

Zarko B. Krusevac
National ICT Australia and
University of New South Wales
Sydney, NSW 2052 Australia
Email: zarko.k@student.unsw.edu.au

Predrag B. Rapajic
School of Elect. Eng. and Telecomm.
University of New South Wales
Sydney, NSW 2052 Australia
Email: p.rapajic@unsw.edu.au

Parastoo Sadeghi
National ICT Australia and
University of New South Wales
Sydney, NSW 2052 Australia
Email: parastoo@student.unsw.edu.au

***Abstract*— This paper addresses the issue of capacity achieving reliable information transfer over time-varying flat-fading communication channels. We use M-state, M-ary symmetric finite-state Markov channels (FSMC) to model time variations of flat-fading channels. We analyze the conventional (separate) approach to channel parameter estimation and data detection by using decision-feedback and output-feedback FSMC estimators. Our analysis includes the metric function update analysis for the decision-feedback estimator and the mutual information rate penalty caused by the input signal entropy reduction, for the output-feedback estimation. Then, we consider the implementation of differential detection instead of channel estimation. We show that differential detection transfers the memory of the channel process into a latent form, which does not interfere with the operation of standard ML coding for memoryless channels. Furthermore, we show that multiple-symbol differential detection practically achieves the channel information capacity with observation times only on the order of a few additional symbol intervals.**

I. Introduction

1) Background and Motivation: Good maximum-likelihood (ML) coding strategies for time-varying communication channels are difficult to determine and decoder complexity grows exponentially with the memory length [1]. This is apparent from the fact that the coding error exponent for channels with memory depends on the block length N [2], whereas for time-invariant (memoryless) channels it is independent of N. Furthermore, much less is known about good codes for time-varying channels than for time-invariant (memoryless) ones [3].

The conventional solution is to fragment and disperse the memory of time-varying channels by using interleaver/deinterleaver [1], [3]. However, this cascaded channel has a lower inherent Shannon capacity then the original one, since standard ML decoding algorithms ignore the latent channel process memory in the interleaved channel [3]. Most other ML decoding strategies use channel estimation prior to ML decoding. The decision-feedback estimator (DFE) and output-feedback estimator (OFE) for the Gilbert-Elliott channel and more general finite-state Markov channels (FSMC) are presented in [3], [1] and [4], respectively. The low complexity of the proposed decoders is based on the Markov structure of the channel. However, the DFE requires a known header (training sequence) for each interleaved block of data to guarantee reliable performance [4]. Furthermore, the OFE uses channel input entropy rate reduction for (implicit) channel estimation (Section III).

All channel estimation algorithms use channel resources for training, explicitly (inclusion of a training sequence) or implicitly (channel input entropy rate reduction) [5]. Furthermore, the time-varying channel parameters are not completely observable at the channel output in the presence of channel noise [2]. Acknowledging that, separate channel estimation followed by ML detection may not be optimal in achieving the channel information capacity.

2) Contributions: In this paper, we analyze channel estimation approach and differential detection approach for ML decoding for flat-fading time-varying communication channels. We use the M-state, M-ary symmetric Markov channel to model time variations of flat-fading channels [6]. The main contributions of this paper are summarized as follows:

- In Section III, we use DFE [3], [1] and OFE [4] to analyze the conventional (separate) approach to channel parameters estimation and data detection, including the metric function update analysis for the DFE and mutual information performance penalty caused by the input signal entropy reduction for the OFE.

- In Section IV, we use differential detection instead of channel parameter estimation. We show that the differential detection transfers the memory of the channel process into a latent form, which does not interfere with the operation of standard ML coding for memoryless channels. Furthermore, we show that multiple-symbol differential detection practically achieves the channel information capacity with observation times only on the order of a few additional symbol intervals.

II. Channel Model

In flat-fading channels, the received signal r_k at discrete time index k is defined in the observation equation as [7]

$$r_k = g_k u_k + n_k = a_k e^{j\theta_k} u_k + n_k, \qquad (1)$$

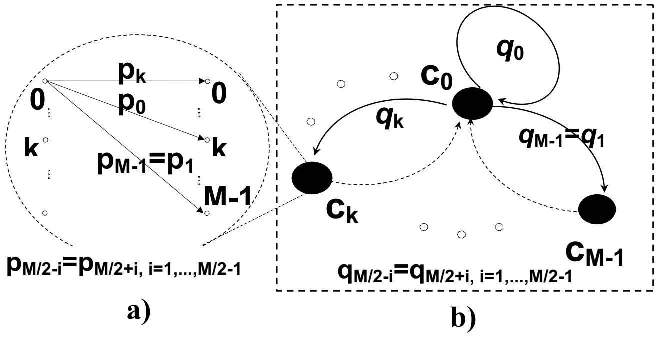

Fig. 1. **Time-varying M-ary symmetric channel model** - FSM modeling of time-varying flat-fading channels: a) Channel state $c_k, \quad k = 0, 1...M-1$ b) Channel transition structure

where $u_k \triangleq \sqrt{\mathcal{E}_s} e^{j\frac{2\pi x_k}{M}}$ is the transmitted MPSK signal [7] with energy per symbol $\mathcal{E}_s$, x_k is the information symbol that takes M possible integer values from 0 to $M-1$. n_k in (1) is complex-valued AWGN, and $g_k = a_k e^{j\theta_k}$ is complex-valued Gaussian flat-fading gain. In general, the actual realization of complex-valued Gaussian flat-fading gain g_k in (1) is unknown to the receiver *a priori* and varies with time.

1) FSMC modeling of time-varying flat-fading channel phase ambiguity: Since the most fundamental channel phase ambiguity in MPSK signaling is whether the flat-fading channel phase causes an MPSK symbol to be rotated to other symbols, flat-fading channel phase is partitioned into M equiprobable, non-overlapping intervals. Each partition corresponds to an FSMC state c_i, $i \in \{0, \cdots, M-1\}$, (Fig. 1) as

$$\theta_k \quad \in \quad [\frac{2\pi i - \pi}{M}, \frac{2\pi i + \pi}{M}) \Leftrightarrow S_k = c_i = i \qquad (2)$$

where θ_k and S_k are channel phase and the corresponding channel state, at the time instant $k = 0, 1...$, respectively. Referring to (1) and (2), it is observed that whenever flat-fading channel phase θ_k is in the i^{th} interval, the channel state is $S_k = c_i$ and an average additional phase shift of $\frac{2\pi i}{M}$ is introduced to the transmitted MPSK signal u_k, as if the information symbol $x_k + i$ was used for modulation [6].

State transition probabilities of flat-fading phase states (Fig. 1-b) are obtained using (2) as

$$\begin{aligned} q_{j\ominus i} &= \Pr(S_k = c_j = j | S_{k-1} = c_i = i) \qquad (3) \\ &= M \int_{\frac{2\pi i - \pi}{M}}^{\frac{2\pi i + \pi}{M}} \int_{\frac{2\pi j - \pi}{M}}^{\frac{2\pi j + \pi}{M}} f(\theta_k, \theta_{k-1}) d\theta_k d\theta_{k-1}, \end{aligned}$$

where $\ominus$ is modulo-M operator, $f(\theta_k, \theta_{k-1})$ is the probability density function of flat-fading channel phase at two consecutive time intervals k and $k-1$, in the following form:

$$f(\theta_k, \theta_{k-1}) = \frac{1-\rho_1^2}{4\pi^2} \frac{\sqrt{(1-\delta^2)} + \delta(\pi - \cos^{-1}(\delta))}{\sqrt{(1-\delta^2)^3}}, \qquad (4)$$

where $\delta \triangleq \rho_1 \cos(\theta_k - \theta_{k-1})$ and $\rho_1 = \sigma^2 J_0(2\pi f_D T_s)$ is the autocorrelation function of the fading process. σ^2 is the fading variance per dimension, $J_0(\cdot)$ is the zero-order Bessel function of the first kind, $f_D T_s$ is the normalized fading rate, f_D is the Doppler frequency and T_s is the symbol period.

Time variations of flat-fading channel phase are modeled in the state equation as

$$S_k = S_{k-1} \oplus \eta_k, \qquad (5)$$

where $\oplus$ is modulo-M operator, $\{\eta_k\}$ is an M-ary i.i.d process, where $Pr(\eta_k = j \ominus i)$ is defined by (3).

2) Output observation process and channel state modeling: The channel output observation process can be modeled in discrete M-ary format as

$$y_k = S_k \oplus x_k \oplus v_k, \qquad (6)$$

where v_k is the symbol detection error sequence in discrete M-ary format, when flat-fading channel phase is known to the receiver. It is important to notice that the channel phase process memory is incorporated in (6) through the channel phase state process in (5).

Based on (2) and (6), channel states c_i, $i = 0, 1, ...M-1$ can be modeled as M-ary symmetric channels (Fig. 1-a), with crossover probabilities:

$$\begin{aligned} p_{n\ominus m} &= Pr(y_k = n | x_k = m) = Pr(c_i \oplus v_k = n \ominus m) \\ &= Pr(v_k = n \ominus m \ominus i); \; n, m = 0, ..., M-1 \quad (7) \end{aligned}$$

Probabilities $Pr(v_k = i)$, $i = 0, 1, 2, ...M-1$, are calculated assuming memoryless Rayleigh-distributed fading amplitudes a_k and AWG channel noise [7].

III. Time-varying channel process entropy and estimation

A. Channel Process Entropy

For time-varying channels, the channel (state) process is an additional statistical process in work, apart from the information source and noise. Channel (state) process entropy rate of the M-state, M-ary symmetric FSMC is given by:

$$\mathcal{H}_{CP}(S) = \lim_{n\to\infty} \frac{1}{n} H(S^n); \; S^n = (S_1, S_2, ..S_n) \qquad (8)$$

where $H(S^n)$ is the entropy of channel state process S^n after n time instants, $S_n \in \{c_0, c_1, ...c_{M-1}\}$.

The M-state, M-ary symmetric channel is variable-noise, uniformly-symmetric FSMC [1] and under uniform i.i.d. input distribution assumption, the channel information capacity can be calculated as [1]:

$$C = \log|Y| - \mathcal{H}(Z) \qquad (9)$$

where $\mathcal{H}(Z) = \lim_{n\to\infty} \frac{H(Z^n)}{n}$ and $Z = \phi(X, Y)$ is the FSMC error function [1]. With the channel state information (CSI) knowledge, expression (9) becomes:

$$C^{SI} = \log|Y| - \mathcal{H}(Z|S) \qquad (10)$$

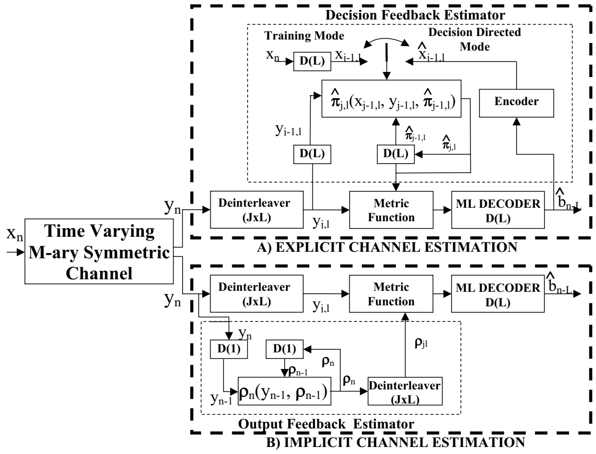

Fig. 2. **Channel estimation of M-state, M-ary symmetric channel**: a) Decision-feedback estimator for explicit channel estimation b) Output-feedback estimator for implicit channel estimation

Although $\mathcal{H}_{CP}(S)$ is independent of the channel noise, it is not completely observable at the channel output in the presence of channel noise [2]. Thus, if we define the observable channel entropy rate as $\mathcal{H}_C = C^{SI} - C$ then, based on (9) and (11), $\mathcal{H}_C$ becomes:

$$\mathcal{H}_C(S,Z) = \mathcal{H}(Z) - \mathcal{H}(Z|S) = \mathcal{H}_{CP}(S) - \mathcal{H}(S|Z) \quad (11)$$

Since $\mathcal{H}(S|Z) \geq 0$, it is concluded that $\mathcal{H}_{CP}(S) \geq \mathcal{H}_C(S,Z)$. Furthermore, since $\mathcal{H}(S|Z)$ is an indecomposable function of channel process and channel noise process, then the conventional (separate) approach to the channel parameter estimation and data detection may not optimal in terms of achievable mutual information rate.

B. Explicit Channel Estimation and Decision-Feedback Estimator

The traditional approach to channel parameter estimation is to send a training sequence (a pilot signal) before information transmission. When the channel is time-varying, the training sequence is periodically used to perform channel parameter estimation, and once the parameters are estimated sufficiently well, a switch is made to the decision-directed mode. We call this method explicit training, since the training sequence is explicitly known at the receiver. Since the training sequence does not carry any information from the information source, a periodic inclusion of explicit training reduces information transmission rate.

The DFE for FSMC's is shown in Fig. 2-a [3], [1] and is based on the following recursive formula for the state distribution conditioned on past input/output pairs ($\pi_n(k) = p(S_n = c_k|x^{n-1}, y^{n-1})$) [1]:

$$\pi_{n+1} = \frac{\pi_n D(x_n, y_n) Q}{\rho_n D(x_n, y_n)\mathbf{I}} \quad (12)$$

where $\pi_n = (\pi_n(0), ..., \pi_n(M-1))$, $D(x_n, y_n)$ is a diagonal MxM matrix, with kth diagonal term $p_k(y_n|x_n)$ defined by (7). $\mathbf{I} = [1, ..., 1]^T$ is an M-dimensional identity vector and Q is the matrix of the channel transition probabilities in (3).

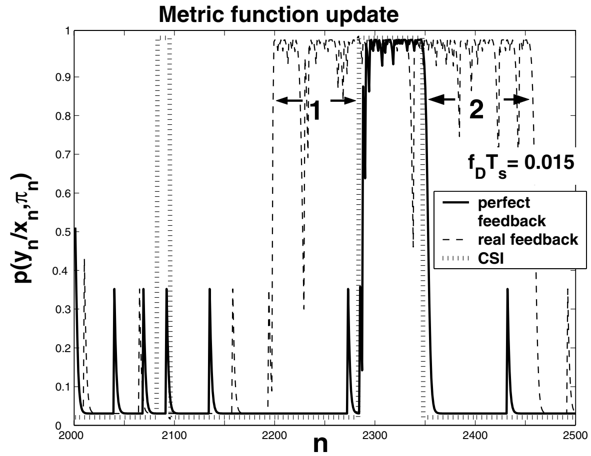

Fig. 3. **Metric function update for the decision-feedback estimator for BPSK over the time-varying M=2-symmetric channel:** CSI assumption, perfect feedback assumption (training mode) and imperfect feedback assumption (decision-directed mode); $f_D T_s$ is the normalized fading rate

Since π_n is a sufficient statistic for the current output given all past inputs and outputs [1], the cascade of a block interleaver, FSMC, DFE and deinterleaver reduce the FSMC to a discrete memoryless channel and a conventional ML decoder can be implemented [1]. The ML decoder metric given the sufficient statistic π_n is defined as $m(x^n, y^n) = \sum_{j=1}^{n} m(x_j, y_j)$ and is updated at each n [1]. The metric update $m(x_j, y_j)$ is given by [1]

$$\begin{aligned} m(x_j, y_j) &= -log[p(y_j|x_j, \pi_j)] \\ &= -log[\sum_k (p(y_j|x_j, \pi_j(k))\pi_j(k)] \end{aligned} \quad (13)$$

Although in [1], [3] it is shown that if the error propagation is ignored, a system employing the DFE on uniformly-symmetric variable-noise channel is information-lossless, the BER performance analysis in [4] reveals that the channel state estimator in the DFE occasionally diverges (i.e., loses track of the fading channel) due to burst of decision errors. Furthermore, Fig. 3 shows that the metric function update $m(x_j, y_j)$ (13), in the decision-directed mode, occasionally loses the track of CSI metric function (markers 1 and 2 in Fig. 3). Consequently, a known header (training sequence) is required for each interleaved block of data in order to guarantee reliable performance [4]. Thus, the DFE from [1], [3] is an explicit channel estimator, which needs a periodic training sequence to guarantee reliable performance. The inclusion of the periodic known header reduces information transmission rate below the channel capacity [5].

C. Implicit Channel Estimation and Output-Feedback Estimator

Since the inclusion of the training sequence reduces the information transmission rate, channel estimation without resort to training is preferred [8]. The decision-directed mode of the DFE is just a special case of the so-called "blind" algorithms for channel parameter estimation, where the stream of

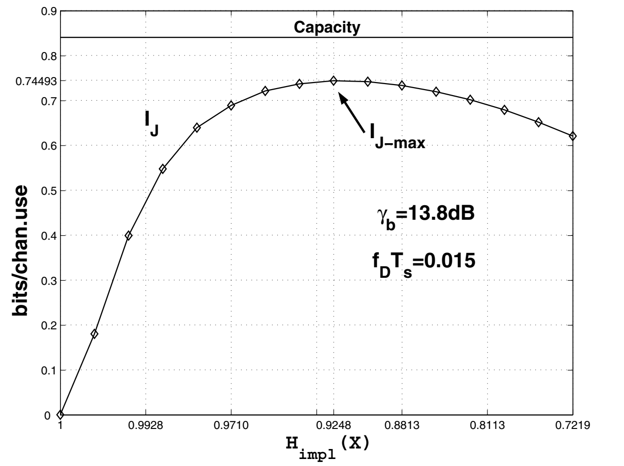

Fig. 4. **Mutual information rate for output-feedback estimator for BPSK over the time-varying M=2-symmetric channel**; $f_D T_s$is the normalized fading rate

information symbols without the presence of explicit training symbols is used for information transmission and the channel parameter estimation [8].

However, a proper initialization and convergence of the blind algorithm is based on information (input) signal entropy rate reduction [5] and consequently, information transmission rate reduction. Thus, we rather use the term implicit training than blind estimation, since it exploits implicit training resources hidden in the transmitted signal structure [5].

The implementation of the OFE [4] is shown in Fig. 2-b and is based on the following recursive formula for the state distribution conditioned on past outputs alone, $\rho_{n+1} = p(S_n = c_l|y^{n-1})$, under the assumption of independent channel inputs [1]

$$\rho_{n+1} = \frac{\rho_n B(y_n) P}{\rho_n B(y_n) \mathbf{I}} \tag{14}$$

where $\rho_n = (\rho_n(0), ..., \rho_n(M-1))$, $B(y_n)$ is a diagonal MxM matrix, with kth diagonal term $p(y_n|S_n = c_k)$. Since no feedback decisions are used by the estimator, no perfect feedback has to be assumed (Fig. 2-b), [4].

The system from Fig. 2-b, up to the output of the state estimator, is equivalent to a set of J parallel ρ-output channels, and each channel is memoryless with asymptotically large interleaving depth J (π-output channels for DFE are explained in [1]). Thus, for a fixed input distribution, the mutual information rate is calculated as:

$$I_J = \frac{1}{J} I(Y^J, \rho^J; X^J) = \sum_{j=1}^{J} H(Y_j|\rho_j) - H(Y_j|\rho_j, X_j) \tag{15}$$

The capacity of uniformly-symmetric variable-noise channel is achieved with an input distribution that is uniform and i.i.d. [1]. However, one can show that, under uniform and i.i.d input distribution assumption, recursion (14) converges to the stationary state probability vector π_0 which can be obtained by solving the eigenvector equation $Q^T \pi_0 = \pi_0$. Thus, with

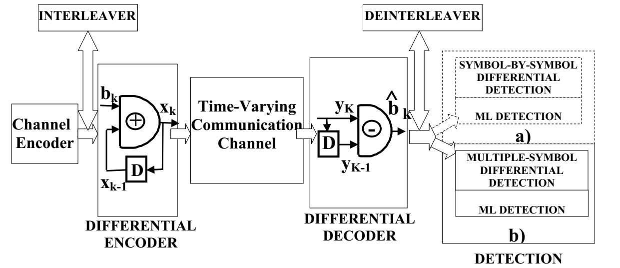

Fig. 5. **System model:** Differential encoding/detection of MPSK signaling over time-varying communication channels: a) Symbol-by-symbol differential detection b) Multiple-symbol differential detection

the uniform i.i.d channel input, the OFE decoder metric does not reduce the channel process entropy rate. In order to estimate the channel state process based on ρ_n estimation, one has to reduce the input process entropy rate $\mathcal{H}_{impl}(X)$ (to 'unbalance' the uniform i.i.d. input distribution). In such a way, the OFE is an implicit channel estimator.

Fig. 4 shows the mutual information rate (15) for the OFE for BPSK signaling over time-varying M=2-ary symmetric channel. $\gamma_b = 2\sigma^2 \frac{E_b}{N_0}$ from Fig. 4 is the average received SNR per bit, E_b is the average transmitted energy per bit and σ^2 is the fading variance per dimension of complex valued flat-fading gain g_k. The mutual information rate in Fig. 4 is a tradeoff between the input process entropy rate reduction and channel process estimation. Furthermore, for a fixed normalized fading rate $f_D T_s$, one can find the optimal input distribution which achieves the maximal mutual information rate. This maximum is still below the channel capacity, Fig. 4.

IV. Differential coding/detection

The system model for MPSK signaling over time-varying communication channels (Fig. 5) uses either symbol-by-symbol or multiple-symbol differential detection, instead of channel estimation. The differential encoding process starts with an arbitrarily first reference symbol $x_0 \in \{0, .., M-1\}$. Then, the differentially encoded sequence x_k is given by:

$$x_k = b_k \oplus x_{k-1}, \;\; k = 1, 2, 3..., \;\; x_0 \text{ -reference symbol} \tag{16}$$

where b_k is k^{th} information symbol.

By using (5), (6) and (16), the differential detection scheme from Fig. 5 can be quantitatively described in the following form:

$$\hat{b}_k = y_k \ominus y_{k-1} = b_k \oplus v_k \ominus v_{k-1} \oplus \eta_k = b_k \oplus \varepsilon_k \tag{17}$$

where $\varepsilon_k = v_k \ominus v_{k-1} \oplus \eta_k$.

Thus, the differential detection (17) transfers the memory of the channel state process S, from the output observation (6), into a latent form (η_k in (5) is an i.i.d. process), enabling the adoption of standard coding/ML detection schemes for memoryless channels. The memory of sequence $\varepsilon_k = v_k \ominus v_{k-1} \oplus \eta_k$, created by the differential detection, still exists, independently of the time-varying channel process. Multiple-symbol differential detection [9] uses this memory to improve the performance of differential detection.

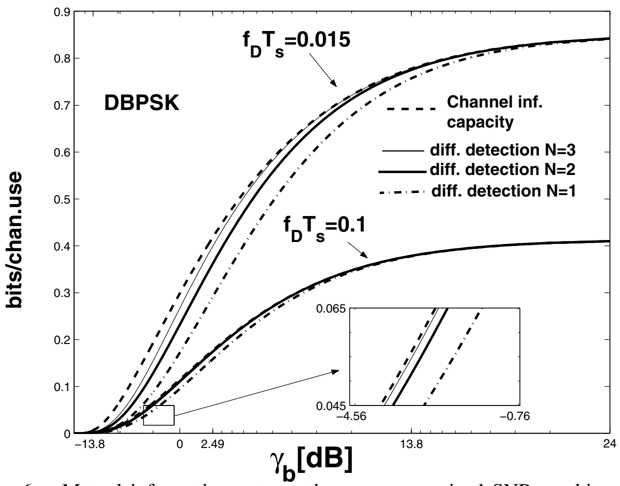

Fig. 6. Mutual information rate vs. the average received SNR per bit, γ_b, for N-symbol differential detection of BPSK ($N = 1, 2, 3$);

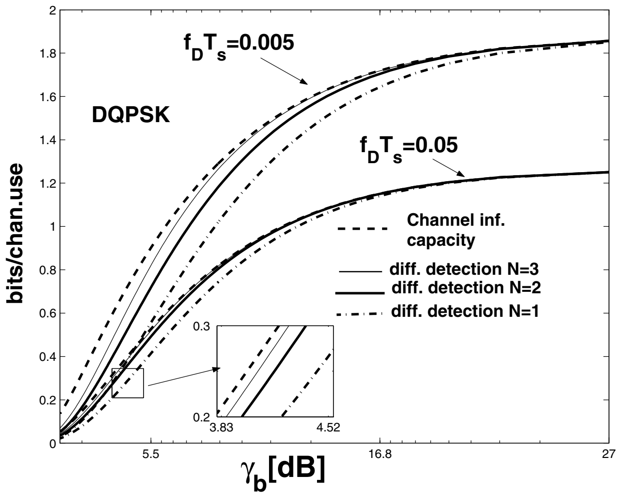

Fig. 7. Mutual information rate vs. the average received SNR per bit, γ_b, for N-symbol differential detection of QPSK ($N = 1, 2, 3$);

The information capacity of the differential MPSK system form Fig. 5 is given by [2]:

$$C_{DMPSK} = \lim_{N\to\infty} \max_{P(b^N)} \frac{1}{N} I(\hat{b}^N, b^N) \tag{18}$$

where $I(\hat{b}^N, b^N)$ denotes the mutual information between input process b^N and output process $\hat{b}^N$ after N symbol intervals and $P(b^N)$ denotes the set of all input distributions of b^N.

Fig. 6 and Fig. 7 depict the maximal mutual information rate versus the average received SNR per bit, γ_b, for the differential detection of BPSK and QPSK, respectively. Two different normalized fading rate are considered, $f_DT_s = 0.015$ and $f_DT_s = 0.1$ for BPSK and $f_DT_s = 0.05$ and $f_DT_s = 0.005$ for QPSK (all correspond to medium-speed fading channels). N-symbol differential detection with $N+1$-symbol observation interval ($N = 1, 2, 3$) is considered. Maximum achievable mutual information rates are calculated as $N = 1$, $N = 2$ and $N = 3$ approximations of (18). The channel information capacity is calculated independently of the differential encoding/detection process.

1. Symbol-by-symbol detection of differential BPSK ($N = 1$ from Fig. 6), achieves a good maximal mutual information performance, especially at high γ_b values and higher fading rate ($f_DT_s = 0.1$ from Fig. 6). Some performance degradations are noticeable at low γ_b values and lower fading rate ($f_DT_s = 0.015$, form Fig. 6).

2. Mutual information performance of the differential BPSK is further improved by implementing multiple-symbol differential detection technique ($N = 2$ and $N = 3$ in Fig. 6). Practically, the channel information capacity is approached with observation times only on the order of a few additional symbol intervals.

3. Since the amount of the performance penalty of interleaved (symbol-by-symbol) differential MPSK compared to that of coherent MPSK increases with the number of phases M [9], results in Fig. 7 show some additional performance degradation of symbol-by-symbol differential detection of QPSK in comparison with differential BPSK. However, multiple-symbol differential detection with a few additional symbol intervals practically approaches the channel information capacity, Fig. 7.

V. Conclusions

This paper addresses the issue of capacity achieving reliable information transfer over time-varying flat-fading communication channels, modeled as M-state, M-ary symmetric Markov channels. Since it is not possible to completely isolate the channel process entropy from the channel output observation in the presence of channel noise, the conventional (separate) approach to the channel parameter estimation and data detection may not be optimal in terms of achievable mutual information rate. However, if multiple-symbol differential detection is used, instead of channel estimation, the channel information capacity is approached with observation times only on the order of a few additional symbol intervals.

References

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

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

[3] M. Mushkin and I. Bar-David, "Capacity and coding for Gilbert-Elliott channels," *IEEE Trans. Inform. Theory*, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.

[4] L. Li and A. J. Goldsmith, "Low-complexity mamximum-likelihood detection of coded signals sent over finite-state Markov channels," *IEEE Trans. Commun.*, vol. 50, no. 4, pp. 524–531, Apr. 2002.

[5] Z. Krusevac and P. B. Rapajic, "On channel uncertainty modeling - an information theoretic approach," in *Proc. IEEE Glob. Telecomm. Conf. (Globecom)*, Dallas, USA, Nov. 2004, pp. 410–414.

[6] 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.

[7] J. G. Proakis, *Digital Communications*, forth ed. New York: Mc Graw Hill, 2000.

[8] R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: a review," *Proc. IEEE*, vol. 86, no. 10, pp. 1927–1950, Oct. 1998.

[9] 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 Trans. Commun.*, vol. 42, no. 1, pp. 76–89, Jan. 1994.