Implicit channel estimation for ML sequence detection over finite-state Markov communication channels

Zarko B. Krusevac  Rodney A. Kennedy
National ICT Australia and the Department of Information Engineering
Research School of Information Sciences and Engineering
The Australian National University
Canberra ACT 0200, Australia

Predrag B. Rapajic
Medway School of Engineering
University of Greenwich at Medway Pembroke
Chatham Maritime KENT
ME4 4TB United Kingdom
P.Rapajic@gre.ac.uk

Abstract—This paper shows the existence of the optimal training, in terms of achievable mutual information rate, for an output feedback implicit estimator for finite-state Markov communication channels. Implicit (blind) estimation is based on a measure of how modified is the input distribution when filtered by the channel transfer function and it is shown that there is no modification of an input distribution with maximum entropy rate. Input signal entropy rate reduction enables implicit (blind) channel process estimation, but decreases information transmission rate. The optimal input entropy rate (optimal implicit training rate) which achieves the maximum mutual information rate, is found.

I. INTRODUCTION

A. Background and Motivation

Finite state Markov channels (FSMC) have been used extensively in literature to model time-varying communications channels, including Rayleigh fading channels, Ricean fading channels, indoor channels, Nakagami-fading channels and satellite channels ([1]–[4] and references therein). They have also been used for system design and system performance analysis.

Assuming the FSMC transition probabilities are independent of the input, the memory of the FSMC process comes from the dependence of the current channel state on past channel states [1]. Reliable communication over a channel with memory is theoretically possible at any rate below capacity [5]. However, good maximum-likelihood (ML) coding strategies for channel with memory are difficult to determine [5]. First, the length (and therefore the decoding complexity) of such codes would depend on the length of the channel memory. This is apparent from the fact that the coding error exponent for channels with memory depends on the block length $N$, whereas for memoryless channel it is independent of $N$. Second, much less is known about good codes for such channels than for memoryless ones.

The conventional solution is to use channel estimation based methods in combination with the standard ML coding tools for time-invariant channels. Channel estimation algorithms for time-varying channels exploit channel process memory to predict the upcoming channel quality and improve channel process estimation [1]–[3]. Once estimated, time-varying channel process does not interfere with ML coding methods for time-invariant channels.

The traditional approach to channel parameter estimation is to send a training sequence (a pilot signal) before information transmission. We call this explicit channel estimation. However, the training sequence is explicitly known at the receiver and it does not carry any information from the information source. Since regular inclusion of the training sequence reduces information transmission rate [6], channel estimation without explicit receiver training might be preferred. Blind channel estimation [7] is performed by knowing the input distribution and observing only the output of the system. We call it implicit estimation, rather than blind, since it is shown here that information from the input signal structure is implicitly used for the channel estimation.

The decision feedback decoder (DFD) for ML sequence detection for the two-state Gilbert-Elliott channel and a more general class of FSMCs is presented in [3] and [1], respectively. Although in [1], [3] it is shown that if the error propagation is ignored, a system employing the DFD on uniformly symmetric variable noise channel is information-lossless, BER performance analysis in [2] reveals that the DFD occasionally diverges (i.e., loses track of the fading channel) due to bursts of decision errors. Consequently, a known header (training sequence) is required for each interleaving block of data in order to guarantee reliable performance [2]. Thus, the DFD state estimator in [1], [3], is in fact, a decision directed mode of a training based (explicit) FSMC estimator, which is presented here.

If the ML metric function is calculated based only on past channel output observation, the DFD can be reduced back to the output-feedback decoder (OFD) [2]. The state estimator of OFD is an implicit estimator, since it is based on the state distribution conditioned on past outputs alone (no training sequence is used).

Two main issues motivate this paper. First, since no training sequence is included, what sort of information is used by the...
implicit (blind) estimator for the channel estimation, if any? Second, since the channel state process of the FSMC is not completely observable at the receiver side in the presence of channel noise [5], how much training, and in what form, would achieve the true optimum in information transmission over an FSM channel?

B. Contributions

This paper provides an information theoretic analysis of channel estimation based methods for ML sequence detection over time-varying communication channels, modeled as FSMC channels. The decision-feedback explicit channel estimator structure for the FSMC is presented. The explicit estimator uses periodic training sequences in order to avoid bursts of decision errors and to guarantee reliable performance. Additionally, the concept of implicit (blind) FSM channel estimation is introduced by means of channel process estimation, by knowing the input distribution and observing only the output of the system.

1) We show that the output-feedback FSMC estimator exploits information from the input signal structure. The structure reduces the input signal entropy rate, at the price of decreasing information transmission rate.

2) We show the existence of an optimal implicit (blind) training rate for the output feedback implicit estimator, which achieves the maximum mutual information rate. For the Gilbert-Elliot channel [3] with memory $\mu = 0.9$ and good-to-bad ratio $g/b = 5$, the optimal implicit training rate is 0.014 bit, which means that 1.4% of information should be used for implicit training in order to achieve the maximum mutual information rate.

3) We show that the maximum mutual information rate, assuming optimal implicit (blind) training, is below the channel information capacity, due to noisy time-varying channel estimation, and it becomes a tight lower bound when channel memory approaches its maximum (the channel becomes time invariant). For the time-varying binary symmetric channel [6], the optimal implicit (blind) training achieves 0.885 (88.5%) of the channel information capacity, for channel memory $\mu = 0.99$ and 0.956 (95.6%) for $\mu = 0.999$.

II. CHANNEL MODEL

A. Finite State Markov Channels

For the purpose of this analysis, we assume the channel model belongs to the class of irreducible, aperiodic, stationary Markov chains. The finite channel state space $C = \{c_0, c_1, ..., c_{M-1}\}$ corresponds to $M$ different discrete memoryless channels with common finite input and output alphabets denoted by $X$ and $Y$, respectively [1]. The state transition structure of the FSMC model is uniquely defined by the initial state probabilities vector $P_0$ and the state transition matrix $P$ [4], where

$$P_0 = [P_{0k} = p(S_0 = c_k)] \quad (1a)$$
$$P = [P_{km} = p(S_{n+1} = c_k | S_n = c_m)] \quad (1b)$$

for $k, m = 0, ..., M - 1$ and $S_n \in C$ is the channel state at the time $n$. The conditional input/output probability at time $n$ is determined by $p(y_n | x_n, S_n)$, where $x_n$ and $y_n$ denote the input and output of the FSMC at the time $n$ [1].

B. Uniformly Symmetric Variable-Noise FSMC

A discrete memoryless channel is output symmetric if the rows of the matrix of output input probabilities $M_{i,j} = p(y = j | x = i), j \in Y, i \in X$ are permutations of each other and the columns of $M$ are permutations of each other [1]. A FSMC is uniformly symmetric if every channel state $c_k \in C$ is output symmetric [1].

A FSMC is a variable noise channel if there exists a function $\phi$ such that for $Z_n = \phi(X_n, Y_n)$, $p(Z^n | X^n) = p(Z^n)$, and $Z^n$ is a sufficient statistic for $S^n$ (so $S^n$ is independent of $X^n$ and $Y^n$ given $Z^n$) [1]. $X^n$ and $Y^n$ denote the input and output, respectively, of the FSMC.

Uniformly symmetric, variable noise FSMC channels include channels varying between any finite number of binary symmetric channels (BSCs), as well as quantized additive white noise channels with symmetric PSK inputs and time varying noise statistics or amplitude and phase fading [1]. Thus, uniformly symmetric, variable noise FSMC channels are of particular importance for time-varying communication channel modeling and this paper analysis is mostly related, but not limited, to this class of FSMCs. However, we use the following two FSMC models from this class, to present our results:

1) Gilbert-Elliot Channel: The Gilbert-Elliot channel [3] is a stationary two-state uniformly symmetric variable noise Markov channel. The states are appropriately designed as discrete memoryless BSCs with the crossover probability $p_G$ ($0 \leq p_G \leq 0.5$) for good state and $p_B$ ($0 \leq p_B \leq 0.5$) for bad state, where $p_G < p_B$. The state transition probabilities, to jump from the bad state to the good state and to jump from the good state to the bad state are given by $g$ and $b$, respectively. The channel memory $\mu$ is defined in [3] as

$$\mu = 1 - b - g \quad (2)$$

The Gilbert-Elliot channel is well known as a useful model for time-varying channel analysis [3]. However, due to the relatively high information capacity of memoryless Gilbert-Elliot channels with high good-to-bad ratio, some mutual information performance results for these channels may be misleading.

2) Time Varying Binary Symmetric Channel (TV-BSC): The TV-BSC [6] is the simplest non-trivial two-state uniformly symmetric, variable noise Markov channel, consists of channel state $c_0 = S_{NX}$, which is the non-inverting BSC with crossover probability $p$, $0 \leq p \leq 0.5$ and channel state $c_1 = S_X$, which is the inverting BSC with crossover probability $1 - p$. State transition probabilities are given by: $p(S_n = S_{n-1}) = 1 - q$ and $p(S_n \neq S_{n-1}) = q$, $0 \leq q \leq 0.5$. The TV-BSC state process memory is given by [6]

$$\mu = 1 - 2q \quad (3)$$

Although expression (3) looks like a special case of expression (2), the TV-BSC exhibits some unique non-trivial
features of time varying channels [6], which makes results of our analysis more obvious and easier to understand. While the Gilbert-Elliott channel physically models channel amplitude time variations, the TV-BSC models time variations of the channel phase. Additionally, the information capacity of a memoryless TV-BSC is equal to zero [6]. Consequently, any approach which ignores channel process memory results in zero mutual information rate.

III. TIME-VARYING CHANNEL PROCESS ENTROPY AND ESTIMATION

If the channel is time varying, then the channel (state) process is an additional statistical process at work (apart from the information source and noise). For the FSMC, channel state process entropy rate is given by

\[ H_{CP}(S) = \lim_{n \to \infty} \frac{1}{n} H(S^n); \quad S^n = (S_1, S_2, \ldots, S_n). \] (4)

Although the channel state process \( H_{CP}(S) \) is independent of the channel noise, it is not completely observable at the channel output in the presence of channel noise [5]. Thus, we can define the observable channel entropy rate \( H_C \) as

\[ H_C = C_{SI} - C \] (5)

where \( C_{SI} \) is the channel information capacity with channel state information (CSI) assumption. For uniformly symmetric, variable noise FSMCs, under an uniform i.i.d. input distribution assumption, the channel information capacity is given by [1]

\[ C = \log|\gamma| - H(Z) \] (6)

where \( Z = \phi(X_n, Y_n) \) is the FSMC error function [1] and \( H(Z) = \lim_{n \to \infty} H(Z^n)/n \). With CSI knowledge, expression (6) becomes

\[ C_{SI} = \log|\gamma| - H(Z|S) \] (7)

Based on (6) and (7), the observable channel entropy rate \( H_C \) is

\[ H_C = H(Z) - H(Z|S) = H_{CP}(S) - H(S|Z) \] (8)

The channel parameter estimation reduces the channel process entropy rate \( H_{CP}(S) \). However, in the presence of channel noise, \( H(S|Z) > 0 \) and \( H_C(S) < H_{CP}(S) \), i.e., the observable channel process entropy rate is smaller than the actual channel process entropy rate. This opens the issue of how much training would achieve the true optimum information transmission over an FSM channel.

IV. EXPlicit CHANNEL ESTIMATION AND DECISION-FEEDBACK ESTIMATOR FOR FSMC

A. Explicit channel estimation

Explicit channel estimation is performed by sending a training sequence, before information transmission, to perform initial parameter estimation. Once the parameters are estimated sufficiently well, a switch is made to the decision directed mode. The training sequence is explicitly known at the receiver and it does not carry any information from the information source. Consequently, the regular inclusion of a training sequence in each packet carries its own training sequence to perturb the estimator operations in decision directed mode. The training sequence for time-varying channel estimation reduces the information transmission rate [6]. As an illustration, in a GSM system each packet carries its own training sequence 26 bits long, representing 20% overhead in each packet.

B. Decision-feedback explicit channel estimator for FSMC

The decision-feedback decoder (DFD) for FSMCs, presented in [1], [3], is an ML sequence decoder based on FSMC state estimation. The channel state estimator exploits the following recursive formula for the state distribution conditioned on past input/output pairs \( (\pi_n(k) = p(S_n = s_k|x^{n-1}, y^{n-1})) \)

\[ \pi_{n+1} = \frac{\pi_n D(x_n, y_n) P}{\pi_n D(x_n, y_n) I} \] (9)

where \( \pi_n = (\pi_n(0), \ldots, \pi_n(M-1)) \), \( D(x_n, y_n) \) is a diagonal \( M \times M \) matrix, with \( k \)th diagonal term \( p(y_n|x_n) \). \( I = [1, \ldots, 1]^T \) is a M-dimensional vector and \( P \) is the matrix of the channel state transition probabilities (1b).

Since \( \pi_n \) is a sufficient statistic for the current output given all past inputs and outputs [1], the system, composed of a block interleaver, FSMC, DFD and deinterleaver, reduces the FSMC to a discrete memoryless channel and a conventional ML sequence decoder can be implemented [1]. The ML decoder metric given the sufficient statistic \( \pi_n \), which is updated is at each \( n \), is \( m(x^n, y^n) = \sum_{j=1}^{n} m(x_j, y_j) \), where the metric update \( m(x_j, y_j) \) is given by [1]

\[ m(x_j, y_j) = -\log(\sum_k p(y_j|x_j, \pi_j) \pi_j(k)) \] (10)

In Fig. 1 we propose a training-based (explicit) decision-feedback explicit channel estimator. Fig. 1(a) shows the system model employing the proposed structure. The explicit state estimator, Fig. 1(b), periodically includes known training sequence (switch position (1)) in order to keep track of time-varying CSI ML metric function, which assumes perfect knowledge of channel state information (CSI). Between headers the estimator operates in decision directed mode (switch...
position (2)), which uses feedback from the decoder output, as proposed in [1], [3].

Fig. 2 shows the ML metric function update \( m(x_j, y_j) \), (10), for the explicit decision feedback estimator for the Gilbert-Elliott channel. In training mode, the estimator keeps track of channel time variations, but no information is transferred.

V. CONCEPT OF IMPLICIT (BLIND) CHANNEL ESTIMATION

Here we introduce the concept of implicit (blind) FSM channel estimation, which assumes the identification of the channel process \( S \) by knowing the input distribution and observing only the output \( \{y_n\} \) of the system. Implicit estimation is based on a measure of how modified is the input distribution when filtered by the FSMC transfer function.

Since no training sequence is included, the net information rate delivered to channel is equal to the input sequence entropy rate. However, our analysis in the next Section reveals that the implicit channel estimation is not possible with an input distribution which produces the maximum input entropy rate and achieves the channel information capacity. By choosing an input distribution, which reduces input entropy rate, one can introduce structure into the input signal process. This structure enables implicit channel state process estimation, but reduces information transmission rate below the channel information capacity. Thus, implicit (blind) estimation implicitly uses information “hidden” in the transmitted signal structure and consequently, we use the term implicit rather than blind estimation.

VI. OUTPUT-FEEDBACK IMPLICIT FSMC ESTIMATOR

Fig. 3(a) shows the system model for output-feedback decoder for ML sequence detection [2], employing output-feedback estimator structure, Fig. 3(b). The implementation of the output-feedback estimator is based on a similar recursive formula to (9), for the state distribution conditioned on past outputs alone, \( \rho_n(l) = p(S_n = c_l | y^{n-1}) \), under the assumption of independent channel inputs [1]

\[
\rho_{n+1} = \rho_n B(y_n | P) \frac{B(y_n)}{B(y_n)}
\]

(11)

where \( \rho_n = (\rho_n(0), ..., \rho_n(M-1)) \) and \( B(y_n) \) is a diagonal \( MxM \) with \( k \)th diagonal term \( p(y_n | S_n = c_k) \). The initial conditions for \( \rho_n \) are given by (1a), i.e.,

\[
\rho_0 = P_0 = (p(S_0 = c_0), ..., p(S_0 = c_{M-1}))
\]

(12)

Since the channel estimation is performed by observing only the output \( \{y_n\} \) of the system, Fig. 3(b), the output-feedback estimator is implicit. The ML metric function is based on statistic \( \rho_n \), and given by \( m(x^n, y^n) = \sum_{j=1}^{n} m(x_j, y_j) \), with the ML metric update

\[
m(x_j, y_j) = -\log \left( \sum_k p(y_j | x_j, S_k = c_j) \rho_j(k) \right)
\]

(13)

Lemma 1: For an uniformly symmetric, variable noise FSMC and an uniform i.i.d. input distribution, the recursive formula (11) converges toward the stationary state probability vector \( \tau = (p(c_0), ..., p(c_{M-1})) \) which is obtained by solving the eigenvector equation \( \lambda \tau = \rho \tau \), where \( \lambda \) is given by (1b).

Proof: We assume an uniformly symmetric variable noise FSMC with the capacity achieving uniform i.i.d. input distribution [1]. Since each \( c_k \) is output symmetric and the marginal \( p(x^n) \) is uniform, then \( p(y_n | S_n = c_k) \) is also uniform [1], i.e., \( p(y_n | S_n = c_k) = \frac{1}{|\mathcal{Y}|} \), \( |\mathcal{Y}| \) is the output alphabet. Hence, for the uniformly symmetric variable-noise FSMC, with the uniform i.i.d. input distribution, the output-feedback estimator statistic (11) becomes

\[
\rho_{n+1}(l) = \frac{\sum_{j=1}^{K} p(y_n | S_n = c_j) p(S_n = c_j | y^{n-1}) P_{j\ell}}{\sum_{k=1}^{K} p(y_n | S_n = c_k) p(S_n = c_k | y^{n-1})}
\]

\[
= \frac{\left( \frac{1}{|\mathcal{Y}|} \sum_{j=1}^{K} \rho_n(j) P_{j\ell} \right)}{\left( \frac{1}{|\mathcal{Y}|} \sum_{k=1}^{K} \rho_n(k) \right)}
\]

\[
= \sum_{j=1}^{K} \rho_n(j) P_{j\ell}
\]

(14)
Thus, the recursive formula (11) for \( \rho_n \), converges toward the solution of the eigenvector equation \( P^T \rho = \rho \), which is the stationary state probability vector \( \tau = (p(c_0), \ldots, p(c_{M-1})) \).

**Lemma 2:** For an uniformly symmetric variable noise FSMC and an input distribution that is uniform i.i.d., \( \mathcal{H}(\rho^n) = \mathcal{H}^{NM}(S^n) \), i.e. the statistic \( \rho_n \) of the output-feedback estimator does not reduce the channel process entropy rate (\( \rho_n \) does not estimate the channel state process).

**Proof:** Follows directly from Lemma 1 and initial conditions (12), as

\[
p(S_n = c_t) = \sum_{j=1}^{K} p(S_{n-1} = c_j) P_{jt} = \sum_{j=1}^{K} \rho_n(j) P_{jt}
\]

\[
= \rho_n(t) = p(S_n = c_t | y^{n-1})
\]

\[
\Rightarrow \mathcal{H}(\rho^n) = \frac{1}{n} \sum_{i=1}^{n} H(\rho_i) = \frac{1}{n} \sum_{i=1}^{n} H(S_i) = \mathcal{H}^{NM}(S^n)
\]

Fig.4 shows the entropy rate \( \mathcal{H}(\rho^n) \) for binary signaling over a Gilbert-Elliott channel. For an input distribution that is uniform i.i.d. \( p(x_j = 0) = 0.5 \) in Fig.4, the implicit estimator does not estimate the channel process and \( \mathcal{H}(\rho^n) = \mathcal{H}^{NM}(S^n) \), for any channel memory \( \mu \), given by (2). An “unbalanced” input distribution, \( p(x_j = 0) \neq 0.5 \), reduces \( \mathcal{H}(\rho^n) \), thus enabling channel estimation. More “unbalanced” input distributions \( p(x_j = 0) \) closer to 1) provide more reduced \( \mathcal{H}(\rho^n) \) and thus better channel estimation. Additionally, if the channel variation is slower (\( \mu = 0.99 \) in Fig.4), then \( \mathcal{H}(\rho^n) \) is reduced compared to the faster channel (\( \mu = 0.9 \)), for any input distribution \( p(x) \neq 0.5 \).

Fig.5 shows the tracking ability of the output feedback implicit estimator for binary signaling over the Gilbert-Elliott channel. For the uniform i.i.d. input distribution \( p(x_j = 0) = 0.5 \), the statistic \( \rho_s \) does not estimate the channel and the ML metric update (13) does not keep track of the CSI ML metric update. However, “unbalanced” input distributions \( p(x_j = 0) \neq 0.5 \) do enable tracking of the CSI ML metric.

![ML metric function update](image)

With asymptotically large interleaving depth \( J \) each channel becomes memoryless (by the analogy with \( \pi \)-output channel in [11]). Consequently, the mutual information rate (16) becomes

\[
\mathcal{I}(Y; \rho; X) = \lim_{J \to \infty} \frac{1}{J} \mathcal{I}(Y^J, \rho^J; X^J)
\]

**VII. MUTUAL INFORMATION PERFORMANCE ANALYSIS OF THE OUTPUT-FEEDBACK ESTIMATOR**

The cascade in Fig.3(a), which includes interleaver, FSM channel, deinterleaver and output-feedback estimator is equivalent to a set of \( J \) parallel \((y, \rho)\)-output channels, where \( J \) is the interleaving depth. To simplify the notation, we will use \( y_{jt} \triangleq y_n \) to explicitly denote that \( y_n \) is in the \( j \)th row and \( j \)th column of the \( J \times L \) sized deinterleaver. Similarly, \( \rho_{jt} \triangleq \rho_n \) and \( x_{jt} \triangleq x_n \) denote, respectively, the state estimate and interleaver input corresponding to \( y_{jt} \).

For a fixed input distribution, the average mutual information rate (per channel use) is calculated as

\[
\mathcal{I}_J = \frac{1}{J} \mathcal{I}(Y^J, \rho^J; X^J) = \frac{1}{J} \sum_{n=1}^{J} H(Y_n|\rho_n) - H(Y_n|X_n, \rho_n)
\]

\[
= \mathcal{H}(Y^J|\rho^J) - \mathcal{H}(Y^J|X^J, \rho^J)
\]

With asymptotically large interleaving depth \( J \) each channel becomes memoryless (by the analogy with \( \pi \)-output channel in [11]). Consequently, the mutual information rate (16) becomes

\[
\mathcal{I}(Y; \rho; X) = \lim_{J \to \infty} \frac{1}{J} \mathcal{I}(Y^J, \rho^J; X^J)
\]
A. Mutual information rate for uniform i.i.d. input distribution

Assuming an input distribution that is uniform i.i.d., based on Lemma 1, the mutual information rate (17) becomes

\[ I(Y, \rho; X)_{|\rho_{u,i.d.}(X)} = \lim_{n \to \infty} \sum_{n=1}^{J} H(Y_n|\rho_n) - H(Y_n|X_n, \rho_n) = \log |\mathbb{Y}| - \lim_{J \to \infty} H(Y^{J}|X^{J}) = C^{NM} \]  

(18)

since \( X^J \) and \( Y^J \) are memoryless (interleaved) sequences, assuming a large interleaving depth \( J \). \( C^{NM} \) in (18) is the information capacity of the equivalent memoryless (interleaved) FSMC.

Although uniform i.i.d. input distribution achieves the information capacity of uniformly symmetric, variable noise FSMC [1], the system model in Fig. 3(a), implementing output feedback implicit estimator, is reduced back to a memoryless channel. It has a (significantly) lower inherent Shannon capacity \( C^{NM} \) than the original one. As an illustration, for the time-varying binary symmetric channel, \( C^{NM} = 0 \) [6].

B. Optimal implicit training rate

In order to increase the entropy rate \( H(Y^J|X^J, \rho^J) \) in (16), by estimating the channel process based on the \( \rho^J \) statistic. The entropy rate \( H(Y^J|X^J, \rho^J) \) can be reduced by reducing the input signal entropy rate \( H(X) \), as it is shown in Fig. 6, for the TV-BSC.

However, it is known [1, Lemmas 5.1 and 5.2] that \( H(Y^J|\rho^J) \) achieves its maximum, which is \( 1 \log |\mathbb{Y}| \), for an input distribution that is uniform i.i.d. By reducing the input entropy rate, the entropy rate \( H(Y^J|\rho^J) \) is reduced as well, decreasing the information transmission rate (17), Fig. 6.

Thus, the mutual information rate \( I_f(Y^J, \rho^J, X^J) \), given by (16), is a tradeoff between better channel estimation \( H(Y^J|X^J, \rho^J) \) reduction, which provides more reliable communication, and information transmission rate reduction (due to \( H(Y^J|\rho^J) \) reduction), both based on input signal entropy rate \( H(X) \) reduction. Fig. 6 shows the mutual information rate \( I_f(Y^J, \rho^J, X^J) \) and entropy rates \( H(Y^J|\rho^J) \) and \( H(Y^J|X^J, \rho^J) \) for output-feedback implicit estimator for binary i.i.d. signaling over the TV-BSC.

As a result of above tradeoffs, there is an optimal input distribution \( \rho_{opt}(X) \), with the entropy rate \( H_{opt}(X) \), which achieves the maximum mutual information rate \( I_{max}(Y^J, \rho^J, X^J) \). The optimal implicit training rate can be expressed as follows

\[ R_{opt} = H_0 - H_{opt} \]

(19)

where \( H_0 \) is the entropy rate for the uniform i.i.d input distribution.

The optimal implicit training rate (19) is the optimal amount of information per bit which has to be used for implicit channel estimation, in order to achieve the maximum mutual information rate.

Fig. 7 shows the optimal implicit estimation rate for output-feedback estimator for binary signaling over the Gilbert-Elliott channel. For the channel memory \( \mu = 0.9 \) (given by (2)) and good-to-bad ratio \( g/b = 5 \), the optimal implicit training rate is 0.014 bit, which means that 1.4% of information should be used for implicit training in order to achieve the maximum mutual information rate.

Fig. 8 depicts optimal implicit estimation for a range of values of channel process memory \( \mu \), for binary signaling over the Gilbert-Elliott channel. As the channel becomes slower (\( \mu \) approaches to 1), \( H_{opt} \) approaches \( H_0 = 1 \) which correspond to the uniform i.i.d. input distribution. This means that less training is needed for optimal performance if the channel variation is slower. For \( \mu = 1 \), the channel stays forever in the initial state (becomes time-invariant) and \( I_{max}(Y, \rho; X) = C^{CSI} \) with \( H_{opt} = H_0 = 1 \) and \( R_{opt} = 0 \) for \( \rho_{opt}(X) \) = uniform i.i.d., which means that no channel resources have to be used for training. However, as the channel becomes very fast (\( \mu \) approaches to 0), performance gain of using channel process memory becomes small and less training should be used for optimal performance in terms of mutual information rate. For \( \mu = 0 \), the channel process is memoryless and
no channel estimation can improve the performance. Consequently, $I_{\text{max}}(Y, \rho, X) = C^M$ with $H_{\text{opt}} = H_0 = 1$ for $p_{\text{opt}}(x) = \text{uniform i.i.d.}$.

The capacity penalty of the output feedback decoder, due to noisy channel estimation, can be expressed as follows

$$C - I_j(Y_j, \rho_j, X^j) = C - I(Y, \rho, X)_{p_{\text{opt}}(X)}$$

$$= \left[ \log |\mathcal{Y}| - \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} H(Y_n|X_n)_{p_{\text{opt}}(X)} \right]$$

$$- \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \left( H(Y_n|X_n, \rho_n)_{p_{\text{opt}}(X)} - H(Y_n|X_n, \rho_n)_{p_{\text{opt}}(X)} \right)$$

where

$$C = \log |\mathcal{Y}| - \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} H(Y_j|X_j, \rho_j)$$

is the information capacity of the uniformly symmetric, variable noise FSMC [1] and $\pi_n$ is the state distribution conditioned on past input/output pairs, given by (9).

It is shown in [1] that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} H(Y_n|X_n)_{\text{opt}} = \log |\mathcal{Y}|$$

for $p_{\text{opt}}(X)$. Additionally [8],

$$H(Y_n|X_n, \rho_n) = H(Y_n|X_n, Y^{n-1}) \geq H(Y_n|X_n, Y^{n-1}, X^{n-1}) = H(Y_n|X_n, \pi_n)$$

with equality if and only if $Y_n$ and $X^{n-1}$ are independent. However, $Y_n$ and $X^{n-1}$ are independent if and only if the channel is time invariant or memoryless. It means that implicit output feedback channel estimation in noisy environment achieves channel information capacity if and only if the channel is time invariant or memoryless, and with an input distribution that is uniform i.i.d. (Fig. 8) shows that capacity penalty due to noisy channel estimation vanishes for $\mu = 1$, since the Gilbert-Elliott channel becomes time-invariant and $I_{\text{max}}(Y, \rho, X) = C^C BSC$. Otherwise a capacity penalty due to noisy channel estimation (20) is unavoidable.

Fig. 9 corroborates the above analysis for binary signaling over the TV-BSC channel. The maximum mutual information rate $I_{\text{max}}$, assuming optimal implicit training, is below the channel information capacity, due to noisy time-varying channel estimation. For the TV-BSC with memory $\mu = 0.99$ ($\mu$ is given by (3)), the optimal implicit training achieves 0.885 (87.5%) of the channel information capacity $C_{TV-BSC}$. For $\mu = 0.999$, the optimal implicit training achieves 0.978 (97.8%) of the channel information capacity $C_{TV-BSC}$. Thus, implicit channel estimation performs better if the channel variation is slower.

REFERENCES


