Optimizing Information Rate Bounds for Channels with Memory

Parastoo Sadeghi
The Australian National University
Canberra 0200 ACT, Australia
parastoo.sadeghi@anu.edu.au

Pascal O. Vontobel
Hewlett-Packard Laboratories
Palo Alto, CA 94304, USA
pascal.vontobel@ieee.org

Ramtin Shams
The Australian National University
Canberra 0200 ACT, Australia
ramtin.shams@anu.edu.au

***Abstract*— We consider the problem of optimizing information rate upper and lower bounds for communication channels with (possibly large) memory. A recently proposed auxiliary-channel-based technique allows one to efficiently compute upper and lower bounds on the information rate of such channels. Towards tightening these bounds, we propose iterative expectation-maximization (EM) type algorithms to optimize the parameters of the auxiliary finite-state machine channel (FSMC). From a channel coding perspective, optimizing the lower bound is related to increasing the achievable mismatched information rate, i.e. the information rate of a communication system where the maximum-likelihood decoder at the receiver is matched to the auxiliary channel and not to the true channel. We provide explicit solutions for optimizing the upper bound and the difference between the upper and the lower bound and we discuss a method for the optimization of the lower bound for data-controllable channels with memory. We discuss examples of channels with memory, for which application of the developed theory results in noticeably tighter information rate bounds.**

Keywords — Finite-state machine channels, information rate, upper bounds, lower bounds, optimization, mismatch decoding.

I. Introduction

We consider communication channels with memory where the channel input is constrained to a finite alphabet size and where neither the sender nor the receiver has side information about the channel state. Practical examples of channels with memory include the correlated time-varying flat-fading channel (FFC) in wireless communication systems [1] and the partial response channel in magnetic and optical recording, as well as in communications over band-limited channels with inter-symbol interference (ISI) [2]. Although the information rate of such channels is generally formulated [3], the direct computation of the information rate has remained an open problem [1], [3].

Alternative strategies have been proposed in the literature for efficient stochastic and numerical computation of the information rate of finite-state machine channels (FSMCs) [3]–[6]. These techniques work efficiently only for finite-state channels with not too many states and so it is natural to try to come up with efficiently (stochastically) computable upper and lower bounds on the information rate. Such upper and lower bounds were proposed in [3], [7] and were based on the introduction of an auxiliary channel. The lower bound in [3], [7] happens to be a special case of the generalized mutual information (GMI) lower bound for mismatched decoding [8]. The bound signifies achievable information rates when the receiver is equipped with the maximum likelihood (ML) decoding algorithm matched to the auxiliary channel model and hence, usually mismatched to the original channel (over which the actual communication takes place).

The computational complexity of these upper and lower bounds can be controlled by choosing the auxiliary channel to be an FSMC model with not too many states. However, the tightness of the bounds is affected by the number of states and the chosen parameters of the FSMC model. Therefore, for a fixed number of FSMC states, it is desirable to choose the optimum FSMC parameters that give the tightest information rate upper and lower bounds.

The main contribution of this paper is to optimize the parameters of the auxiliary FSMC model in order to tighten information rate bounds introduced in [3], [7]. In our approach we fix the number of FSMC states and optimize other FSMC parameters, i.e. the state transition probabilities and the channel observation law in each state. The proposed approach results in an optimization procedure which is similar in nature to the Baum-Welch algorithm, and therefore also to the expectation-maximization (EM) algorithm.

The paper is structured as follows. In Section II we define the channels and the information rate bounds under consideration. After having shown the general optimization idea in Section III-A, we discuss the optimization of the upper bound in Section III-B and the optimization of the difference of the upper and lower bound in Section III-C. In Section III-D we provide a heuristic iterative optimization technique for the lower bound for data-controllable channels with memory. Finally, in Section IV we apply the developed theory to practical examples of partial response channels and FFCs.

II. Definitions and Information Rate Bounds

A. Definitions

The following definitions are mainly adapted from [9].

1) ***Index Sets and Vectors****:* We will use the index set

$$\mathcal{I}_N \triangleq [-N+1, N] = \{-N+1, \ldots, N\}, \qquad (1)$$

where we assume N to be a positive integer. Note that in all our results we will mainly be interested in the limit $N \to \infty$.

Throughout the paper, we use the following finite windows of the input, output, state, and branch processes:

$$\mathbf{x} \triangleq \mathbf{x}_{-N+1}^{N}, \qquad \mathbf{s} \triangleq \mathbf{s}_{-N+1}^{N}, \qquad \mathbf{b} \triangleq \mathbf{b}_{-N+1}^{N}, \tag{2}$$

$$\mathbf{y} \triangleq \mathbf{y}_{-N+1}^{N}, \qquad \mathbf{s}' \triangleq \mathbf{s}'^{N}_{-N+1}, \qquad \mathbf{b}' \triangleq \mathbf{b}'^{N}_{-N+1}. \tag{3}$$

*2) **Finite-State Machine Sources (FSMS)**:* A time-independent (discrete-time) FSMS has a state sequence $\ldots, S_{-1}, S_0, S_1, \ldots$ and an output sequence $\ldots, X_{-1}, X_0, X_1, \ldots$ where $S_\ell \in \mathcal{S}$ and $X_\ell \in \mathcal{X}$ for all $\ell \in \mathbb{Z}$. We assume that the alphabets $\mathcal{X}$ and $\mathcal{S}$ are finite and that for any $N > 0$ the joint pmf decomposes as

$$P_{\mathbf{S},\mathbf{X}|S_{-N}}(\mathbf{s},\mathbf{x}|s_{-N}) = \prod_{\ell \in \mathcal{I}_N} P_{S_\ell,X_\ell|S_{\ell-1}}(s_\ell, x_\ell|s_{\ell-1}), \tag{4}$$

where we assume $P_{S_\ell,X_\ell|S_{\ell-1}}$ to be independent of ℓ. It is useful to introduce the random variable $B_\ell \in \mathcal{B}$ as $B_\ell \triangleq (S_{\ell-1}, X_\ell, S_\ell)$; then, $P_{B_\ell}(b_\ell)$ represents the probability of choosing branch $b_\ell = \big(s_{\ell-1}(b_\ell), x_\ell(b_\ell), s_\ell(b_\ell)\big)$ at time index ℓ, i.e. the probability to be in state $s_{\ell-1}(b_\ell)$ at time index $\ell-1$, to choose symbol $x_\ell(b_\ell)$ at time index ℓ, and to be in state $s_\ell(b_\ell)$ at time index ℓ. We will only consider sources where there is a one-to-one relationship between $(s_{-N}, \mathbf{x}_{-N+1}^{N})$ and $\mathbf{s}_{-N}^{N}$. We will use the notation

$$\begin{aligned} Q(s_\ell, x_\ell|s_{\ell-1}) &\triangleq P_{S_\ell,X_\ell|S_{\ell-1}}(s_\ell, x_\ell|s_{\ell-1}) \\ &= P_{B_\ell|S_{\ell-1}}\big((s_{\ell-1}, x_\ell, s_\ell)|s_{\ell-1}\big). \end{aligned} \tag{5}$$

*3) **Finite-State Machine Channels (FSMC)**:* A time-independent (discrete-time) FSMC has an input process $\ldots, X_{-1}, X_0, X_1, \ldots$, an output process $\ldots, Y_{-1}, Y_0, Y_1, \ldots$, and a state process $\ldots, S'_{-1}, S'_0, S'_1, \ldots$, where $X_\ell \in \mathcal{X}$, $Y_\ell \in \mathcal{Y}$, and $S'_\ell \in \mathcal{S}'$ for all $\ell \in \mathbb{Z}$. We assume that the alphabets $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{S}'$ are finite and that for any $N > 0$ the joint pmf decomposes as

$$\begin{aligned} &P_{\mathbf{S}',\mathbf{Y}|S'_{-N},\mathbf{X}}(\mathbf{s}', \mathbf{y}|s'_{-N}, \mathbf{x}) \\ &\quad\triangleq \prod_{\ell \in \mathcal{I}_N} P_{S'_\ell,Y_\ell|S'_{\ell-1},X_\ell}(s'_\ell, y_\ell|s'_{\ell-1}, x_\ell) \\ &\quad\triangleq \prod_{\ell \in \mathcal{I}_N} P(s'_\ell|s'_{\ell-1}, x_\ell) \cdot \prod_{\ell \in \mathcal{I}_N} P(y_\ell|s'_{\ell-1}, x_\ell, s'_\ell), \end{aligned} \tag{6}$$

where $P_{S'_\ell,Y_\ell|S'_{\ell-1},X_\ell}$ is independent of ℓ. We denote by $\mathcal{B}'$ the set of legal transitions where for each branch b'_ℓ at time ℓ, there is a branch label $\big(x_\ell; P(s'_\ell|s'_{\ell-1}, x_\ell)\big)$. The "left" (or "previous") state of $b' \in \mathcal{B}'$ is denoted by s'_{p}. We use the notations $W_{S'_\ell,Y_\ell|S'_{\ell-1},X_\ell} \triangleq P_{S'_\ell,Y_\ell|S'_{\ell-1},X_\ell}$, $W(s'_\ell|s'_{\ell-1}, x_\ell) \triangleq P(s'_\ell|s'_{\ell-1}, x_\ell)$, and $W(y_\ell|s'_{\ell-1}, x_\ell, s'_\ell) \triangleq P(y_\ell|s'_{\ell-1}, x_\ell, s'_\ell)$.

In the following, we will make the following important assumptions.

- We only consider indecomposable FSMCs [10, Ch. 4.6].
- All expression involving probabilities (and therefore also quantities like P, Q, W, etc.) will be implicitly conditioned on $(S_{-N}, S'_{-N}) = (s_{-N}, s'_{-N})$ for some fixed states s_{-N} and s'_{-N}. (For notational conciseness we chose not to show this conditioning explicitly in the expressions to come.) Because of the assumed indecomposability of the FSMCs, this conditioning is irrelevant in the limit $N \to \infty$.[1]
- Similarly, when talking about the auxiliary channels, we will implicitly condition on $\hat{S}'_{-N} = \hat{s}'_{-n}$ and on $\tilde{S}'_{-N} = \tilde{s}'_{-n}$ for some fixed states $\hat{s}'_{-n}$ and $\tilde{s}'_{-n}$, respectively.

B. Information Rate Bounds

*1) **Information Rate Upper Bound [7]**:* Assume that the original channel with channel law $W(\mathbf{y}|\mathbf{b})$ is connected to an FSMS. Moreover, assume that an auxiliary FSMC with channel law $\hat{W}(\mathbf{y}|\mathbf{b})$ (and with the same input alphabet $\mathcal{X}$ and the same output alphabet $\mathcal{Y}$ as the original channel) is fed by the same FSMS. Let the auxiliary FSMC states and branches be denoted by $\hat{s}'$ and $\hat{b}'$, respectively. Then upper bound on the information rate of the original channel is given as

$$\overline{I}^{(N)}(\hat{W}) \triangleq \frac{1}{2N} \sum_{\mathbf{b},\mathbf{y}} Q(\mathbf{b}) W(\mathbf{y}|\mathbf{b}) \log\left(\frac{W(\mathbf{y}|\mathbf{b})}{\hat{R}(\mathbf{y})}\right), \tag{7}$$

where the asymptotic version is $\overline{I}(\hat{W}) \triangleq \lim_{N\to\infty} \overline{I}^{(N)}(\hat{W})$ and where $\hat{R}(\mathbf{y}) = \sum_{\mathbf{b}} Q(\mathbf{b}) \hat{W}(\mathbf{y}|\mathbf{b})$ is induced by the fixed FSMS distribution Q and the auxiliary channel law $\hat{W}$. It is noted that for computing the upper bound an analytical or numerical evaluation method for

$$H^{(N)}(W|Q) \triangleq -\frac{1}{2N} \sum_{\mathbf{b},\mathbf{y}} Q(\mathbf{b}) W(\mathbf{y}|\mathbf{b}) \log\big(W(\mathbf{y}|\mathbf{b})\big) \tag{8}$$

is required.[2]

*2) **Information Rate Lower Bound [7]**:* Using the same setup as in Section II-B.1, the lower bound on the information rate of the original channel is

$$\underline{I}^{(N)}(\hat{W}) \triangleq \frac{1}{2N} \sum_{\mathbf{b},\mathbf{y}} Q(\mathbf{b}) W(\mathbf{y}|\mathbf{b}) \log\left(\frac{\hat{W}(\mathbf{y}|\mathbf{b})}{\hat{R}(\mathbf{y})}\right), \tag{9}$$

where the asymptotic version is $\underline{I}(\hat{W}) \triangleq \lim_{N\to\infty} \underline{I}^{(N)}(\hat{W})$. We note that the lower bound is invariant to the multiplication of $\hat{W}(\mathbf{y}|\mathbf{b})$ by a positive function of $\mathbf{y}$.

*3) **Difference between the Upper and the Lower Bound**:* Using the definition of bounds in Sections II-B.1 and II-B.2, the difference between the upper and the lower bound is

$$\Delta^{(N)}(\hat{W}) = \frac{1}{2N} \sum_{\mathbf{b},\mathbf{y}} Q(\mathbf{b}) W(\mathbf{y}|\mathbf{b}) \log\left(\frac{W(\mathbf{y}|\mathbf{b})}{\hat{W}(\mathbf{y}|\mathbf{b})}\right), \tag{10}$$

where the asymptotic version is $\Delta(\hat{W}) \triangleq \lim_{N\to\infty} \Delta^{(N)}(\hat{W})$.

III. Optimizing Information Rate Bounds

A. Optimization Approach

It turns out that the direct optimization of these information rate bounds is in general intractable. Therefore, we propose to

[1] Note that the treatment of the initial state here slightly differs from the approach taken in [9], see especially [9, Def. 35 and the paragraphs following it]. However, in the limit $N \to \infty$, this is irrelevant.

[2] Alternatively, a lower bound on $H^{(N)}(W|Q)$ can be used to obtain an upper bound on $\overline{I}^{(N)}(\hat{W})$.

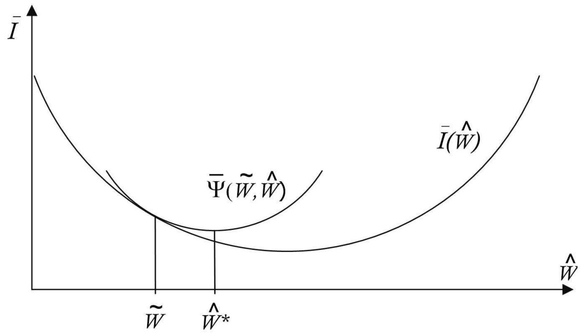

Fig. 1. Iterative minimization of the upper bound via a surrogate function.

use an iterative approach. In the case of the optimization of the upper bound, the underlying idea of such an approach is as follows (see also Fig. 1):

- Assume that at the current iteration we have found an auxiliary FSMC model with channel law $\tilde{W}$ and with the corresponding information rate upper bound $\overline{I}(\tilde{W})$.[3]
- At the next iteration we would like to find a "better" auxiliary channel model with channel law $\hat{W}^*$ which results in a tighter upper bound. To this end, we introduce a surrogate function $\overline{\Psi}(\tilde{W},\hat{W})$ which locally approximates $\overline{I}(\hat{W})$ around $\hat{W} = \tilde{W}$: we require that the surrogate function assume the same value at $\hat{W} = \tilde{W}$ as $\overline{I}(\hat{W})$ does, i.e. $\overline{\Psi}(\tilde{W},\tilde{W}) = \overline{I}(\tilde{W})$, and that $\overline{\Psi}(\tilde{W},\hat{W})$ be never below $\overline{I}(\hat{W})$, i.e., $\overline{\Psi}(\tilde{W},\hat{W}) \geq \overline{I}(\hat{W})$ for all $\hat{W}$.
- Let us assume for the moment that we can find such a surrogate function that can easily be minimized and let us call $\hat{W}^*$ the channel law that achieves the minimum of $\overline{\Psi}(\tilde{W},\hat{W})$ over $\hat{W}$ (one such function is given in Section III-B).
- With this we can get a new channel law $\hat{W}^*$ with $\overline{I}(\hat{W}^*) \leq \overline{I}(\tilde{W})$. Updating $\tilde{W}$ to be equal to $\hat{W}^*$, we then repeat the above procedure until some termination criterion is met.

It is important to note that unlike the idealistic situation depicted in Fig. 1, $\overline{I}(\hat{W})$ is in general not a unimodal function of $\hat{W}$ and therefore $\overline{I}(\hat{W})$ can have multiple local minima.

B. Optimizing the Upper Bound

Let

$$\overline{\Psi}^{(N)}(\tilde{W},\hat{W}) \triangleq \overline{I}^{(N)}(\hat{W}) + \frac{1}{2N}\sum_{\mathbf{y}} R(\mathbf{y})\, D_{\hat{\mathbf{b}}'}\left(\tilde{P}(\hat{\mathbf{b}}'|\mathbf{y}) \,\middle\|\, \hat{P}(\hat{\mathbf{b}}'|\mathbf{y})\right), \tag{11}$$

be the surrogate function for the upper bound $\overline{I}^{(N)}(\hat{W})$ and let its asymptotic version be $\overline{\Psi}(\tilde{W},\hat{W}) \triangleq \lim_{N\to\infty} \overline{\Psi}^{(N)}(\tilde{W},\hat{W})$. Due to the well-known properties of relative entropy, the desired properties (as discussed in Section III-A) of $\overline{\Psi}(\tilde{W},\hat{W})$ with respect to $\overline{I}^{(N)}(\hat{W})$ follow immediately.

[3] In order to simplify notation we use $\tilde{W}$ instead of the more precise $\tilde{\tilde{W}}$.

Lemma 1 (Minimizing $\overline{\Psi}$) *Let $\tilde{W}$ be given. Define*

$$\tilde{T}_1^{(N)}(\hat{b}') \triangleq \frac{1}{2N}\sum_{\ell\in\mathcal{I}_N}\sum_{\mathbf{y}} R(\mathbf{y}) \sum_{\hat{b}'_\ell} \tilde{P}(\hat{b}'_\ell|\mathbf{y}) \left[\hat{b}'_\ell = \hat{b}'\right], \tag{12}$$

$$\tilde{T}_2^{(N)}(\hat{b}',y) \triangleq \frac{1}{2N}\sum_{\ell\in\mathcal{I}_N}\sum_{\mathbf{y}} R(\mathbf{y}) \cdot \sum_{\hat{b}'_\ell} \tilde{P}(\hat{b}'_\ell|\mathbf{y}) \left[\hat{b}'_\ell = \hat{b}'\right][y_\ell = y], \tag{13}$$

with its asymptotic versions $\tilde{T}_1(\hat{b}') \triangleq \lim_{N\to\infty} \tilde{T}_1^{(N)}(\hat{b}')$ and $\tilde{T}_2(\hat{b}',y) \triangleq \lim_{N\to\infty} \tilde{T}_2^{(N)}(\hat{b}',y)$. (Here, $[\hat{b}'_\ell = \hat{b}']$ is defined to equal one if $\hat{b}'_\ell = \hat{b}'$ and to equal zero otherwise, etc.)

The $\hat{W}$ that minimizes $\overline{\Psi}(\tilde{W},\hat{W})$ is given by

$$\hat{W}^*(\hat{s}'|\hat{s}'_{\mathrm{p}},x) \propto \tilde{T}_1(\hat{b}'), \tag{14}$$

$$\hat{W}^*(y|\hat{b}') \propto \tilde{T}_2(\hat{b}',y), \tag{15}$$

where $\hat{b}' = (\hat{s}'_{\mathrm{p}},x,\hat{s}')$ and the proportionality constants are chosen such that the constraints

$$\sum_{\hat{s}'} \hat{W}(\hat{s}'|\hat{s}'_{\mathrm{p}},x) = 1 \quad \text{(for all } (\hat{s}'_{\mathrm{p}},x)\text{)}, \tag{16}$$

$$\sum_{y} \hat{W}(y|\hat{b}') = 1 \quad \text{(for all } \hat{b}'\text{)} \tag{17}$$

are fulfilled. It is reminded that $\hat{s}'_{\mathrm{p}}$ denotes the left state of $\hat{b}'$.

The proof is provided in Appendix I. We observe that the update equations in Lemma 1 look very much like the update equations for the Baum-Welch algorithm, except that we are also averaging over $\mathbf{y}$.[4]

C. Optimizing the Difference between the Bounds

Let

$$\Psi_\Delta^{(N)}(\tilde{W},\hat{W}) \triangleq \Delta^{(N)}(\hat{W}) + \frac{1}{2N}\sum_{\mathbf{b},\mathbf{y}} Q(\mathbf{b})W(\mathbf{y}|\mathbf{b}) \cdot D_{\hat{\mathbf{b}}'}\left(\tilde{P}(\hat{\mathbf{b}}'|\mathbf{b},\mathbf{y}) \,\middle\|\, \hat{P}(\hat{\mathbf{b}}'|\mathbf{b},\mathbf{y})\right), \tag{18}$$

be the surrogate function for the difference between the upper and the lower bound $\Delta^{(N)}(\hat{W})$ and let its asymptotic version be $\Psi_\Delta(\tilde{W},\hat{W}) \triangleq \lim_{N\to\infty} \Psi_\Delta^{(N)}(\tilde{W},\hat{W})$. Due to the well-known properties of relative entropy, the desired properties (as discussed in Section III-A) of $\Psi_\Delta(\tilde{W},\hat{W})$ with respect to $\Delta^{(N)}(\hat{W})$ follow immediately.

Lemma 2 (Minimizing Ψ_Δ) *Let $\tilde{W}$ be given. Define*

$$\tilde{T}_3^{(N)}(\hat{b}') \triangleq \frac{1}{2N}\sum_{\ell\in\mathcal{I}_N}\sum_{\mathbf{b},\mathbf{y}} Q(\mathbf{b})W(\mathbf{y}|\mathbf{b}) \cdot \sum_{\hat{b}'_\ell} \tilde{P}(\hat{b}'_\ell|\mathbf{b},\mathbf{y}) \left[\hat{b}'_\ell = \hat{b}'\right],$$

$$\tilde{T}_4^{(N)}(\hat{b}',y) \triangleq \frac{1}{2N}\sum_{\ell\in\mathcal{I}_N}\sum_{\mathbf{b},\mathbf{y}} Q(\mathbf{b})W(\mathbf{y}|\mathbf{b}) \cdot \sum_{\hat{b}'_\ell} \tilde{P}(\hat{b}'_\ell|\mathbf{b},\mathbf{y}) \left[\hat{b}'_\ell = \hat{b}'\right] \cdot [y_\ell = y],$$

with its asymptotic versions $\tilde{T}_3(\hat{b}') \triangleq \lim_{N\to\infty} \tilde{T}_3^{(N)}(\hat{b}')$ and $\tilde{T}_4(\hat{b}',y) \triangleq \lim_{N\to\infty} \tilde{T}_4^{(N)}(\hat{b}',y)$. The $\hat{W}$ that minimizes $\Psi_\Delta(\tilde{W},\hat{W})$ is given by

$$\hat{W}^*(\hat{s}'|\hat{s}'_{\mathrm{p}},x) \propto \tilde{T}_3(\hat{b}'), \tag{19}$$

$$\hat{W}^*(y|\hat{b}') \propto \tilde{T}_4(\hat{b}',y). \tag{20}$$

[4] The Baum-Welch algorithm is an early instance of the EM algorithm. The EM theory was later generalized in 1977 by Dempster, Laird, and Rubin.

The proportionality constants are chosen such that the constraints in (16) *and* (17) *are fulfilled.*

Since the proof of Lemma 2 is similar in structure to the proof of Lemma 1, we have omitted the details for brevity.

D. Optimizing the Lower Bound

So far, we have not been able to find a suitable surrogate function for the lower bound that is provably never above the function $\underline{I}^{(N)}(\hat{W})$. However, using a certain surrogate function (that possibly violates the requirements in Section III-A), one can come up with the following update algorithm for the auxiliary FSMC channel law for tightening the information rate lower bound for data-controllable channels with memory. The results in Section IV show an acceptable algorithm performance. Let $\tilde{W}$ be the given auxiliary channel law at current iteration. The new auxiliary channel law $\hat{W}^*$ is then given by

$$\hat{W}^*(y|\hat{b}') = \gamma(y) \frac{\sum\limits_{\hat{b}} V(b', \hat{b}'|y)}{\tilde{T}_5(\hat{b}', y)}, \tag{21}$$

where $\gamma(y)$ is an arbitrary positive function of y.[5] Here, $V(b, \hat{b}'|y)$ and $\tilde{T}_5^{(N)}(\hat{b}', y)$ are defined to be

$$V(b', \hat{b}'|y) \triangleq \frac{Q_{b', \hat{b}'} W(y|b')}{R(y)}, \tag{22}$$

$$\tilde{T}_5^{(N)}(\hat{b}', y) \triangleq \frac{1}{R(y)} \frac{1}{2N} \sum_{\ell \in \mathcal{I}_N} \sum_{\mathbf{y}} R(\mathbf{y}) \sum_{\hat{b}'_\ell} \frac{\tilde{P}(\hat{b}'_\ell|\mathbf{y})}{\tilde{W}(y_\ell|\hat{b}'_\ell)} \left[\hat{b}'_\ell = \hat{b}'\right] [y_\ell = y], \tag{23}$$

with the asymptotic version $\tilde{T}_5(\hat{b}', y) \triangleq \lim_{N\to\infty} \tilde{T}_5^{(N)}(\hat{b}', y)$, and $Q_{b',\hat{b}'}$ is defined as

$$Q_{b', \hat{b}'} \triangleq \begin{cases} Q(\hat{b}') \left[\hat{b}' = b'\right] & (\text{if } \mathcal{B}' = \hat{\mathcal{B}}') \\ Q(b') \left[\hat{b}' = \text{corresponding part of } b'\right] & (\text{if } \mathcal{B}' \supset \hat{\mathcal{B}}') \\ Q(\hat{b}') \left[b' = \text{corresponding part of } \hat{b}'\right] & (\text{if } \mathcal{B}' \subset \hat{\mathcal{B}}') \end{cases} \tag{24}$$

Let us conclude this section by remarking that for auxiliary FSMCs with not too many states, $\tilde{T}_1, \ldots, \tilde{T}_5$ are efficiently stochastically computable quantities.

IV. Numerical Results

Fig. 2 shows application of the upper and the lower bound optimization for a partial response channel. The original channel has a memory order 3. The channel model is

$$y_\ell = \sum_{i=0}^{3} h_i x_{\ell-i} + n_\ell, \tag{25}$$

where x, y, n, h are the channel input, channel output, additive white Gaussian noise, and channel coefficient, respectively. In our analysis we used $[h_0, h_1, h_2, h_3] = [+0.5, +0.5, -0.5, -0.5]$, which is also known as the EPR4 channel. We assume a binary, independent and uniformly distributed (i.u.d) input with $\mathcal{X} = \{-1, +1\}$. The output was quantized with partition points at $[-15, -2.5{:}0.5{:}{+2.5}, +15]$.

We looked at three different classes of auxiliary channels, the first having the same trellis as a partial response channel of memory order 1, the second having the same trellis as a partial response channel of memory order 2, and the third having the same trellis as a partial response channel of memory order 3. For the first two classes we used two different initialization methods:

- In the first method, the initial guess for the auxiliary channel law is derived from the truncated version of the original channel as $[\hat{h}_0, \hat{h}_1] = [+0.5, +0.5]$ and $[\hat{h}_0, \hat{h}_1, \hat{h}_2] = [+0.5, +0.5, -0.5]$, respectively.
- In the second initialization method, the initial guess for the auxiliary channel is obtained by optimizing the difference between the upper and the lower bound according to Section III-C.

For the third class we randomly initialized the auxiliary channel for the upper bound and used the mean of channel law across all branches in the original EPR4 channel as initialization for the lower bound; an empirical proof of the effectiveness of the proposed algorithms can be seen by the fact that the upper and lower bounds coincide. Also, the lower bound converges to the true value with only one iteration.

As noted earlier, higher lower bounds mean higher achievable information rates using a mismatched receiver that is equipped with the ML decoding for the auxiliary channel with 2 (respectively, 4 or 8 states) whereas actual communication takes place over the EPR4 channel with 8 states.

Fig. 3 shows the application of the upper bound optimization and a non-optimized lower bound for a non-finite-state FFC. See [11] for more details on the FSMC modeling of FFCs. The normalized fading rate is $f_D T = 0.1$ and binary i.u.d input is assumed. For such an input, a closed-form expression for $H^{(N)}(W|Q)$ in (8) is available. The SNR is 0 dB. We have optimized the upper bound using an FSMC model for the fading channel with 8 phase states, 2 amplitude states, and memory order 1. The initial FSMC state transition probability and channel law were obtained by assuming equiprobable fading channel phase and amplitude partitioning. It is observed that by applying the optimization algorithm we are able to tighten the upper bound from 0.42 to 0.39 bits/ch use. Also in the figure, we have shown the well-known upper bound with perfect channel state information (CSI) assumption. It is observed that the new optimized upper bound is much tighter than the CSI upper bound. Finally, we have shown a (non-optimized) lower bound on FFC information rate using an FSMC model with 8 phase states, 1 amplitude state, and memory order 2.

V. Conclusions

In this paper we optimized the parameters of auxiliary FSMC models in [3], [7] to tighten information rate upper and lower bounds for general channels with memory. We provided explicit solutions for optimizing the upper bound and

[5] As noted earlier, the lower bound is invariant to the multiplication of $\hat{W}(\mathbf{y}|\mathbf{b})$ by a positive function of $\mathbf{y}$.

the difference between the upper and the lower bound and a heuristic method for the optimization of the lower bound for data-controllable channels. We confirmed theory by providing numerical results and showed that optimization of the bounds results in noticeably tighter upper and lower bounds, compared to the case where auxiliary FSMC parameters are chosen heuristically. The optimized channel law for the lower bound also provides a tighter GMI lower bound for the mismatched ML decoding at the receiver using the auxiliary FSMC model. Optimization of the lower bound for non-controllable channels such as FFCs and a better understanding of the global properties of $\overline{I}(\hat{W})$, $\underline{I}(\hat{W})$, and $\Delta(\hat{W})$ is proposed for future research.

APPENDIX I

In this Appendix we prove Lemma 1. The expression that we would like to minimize is

$$\begin{aligned}
&\overline{\Psi}^{(N)}(\tilde{W}, \hat{W}) \qquad (26)\\
&= \overline{I}^{(N)}(\hat{W}) + \frac{1}{2N}\sum_{\mathbf{y}} R(\mathbf{y}) \sum_{\hat{\mathbf{b}}'} \tilde{P}(\hat{\mathbf{b}}'|\mathbf{y}) \log\left(\frac{\tilde{P}(\hat{\mathbf{b}}'|\mathbf{y})}{\hat{P}(\hat{\mathbf{b}}'|\mathbf{y})}\right)\\
&= c_1^{(N)}(W, \tilde{W}) - \frac{1}{2N}\sum_{\mathbf{y}} R(\mathbf{y}) \log\left(\hat{R}(\mathbf{y})\right)\\
&\qquad - \frac{1}{2N}\sum_{\mathbf{y}} R(\mathbf{y}) \sum_{\hat{\mathbf{b}}'} \tilde{P}(\hat{\mathbf{b}}'|\mathbf{y}) \log\left(\hat{P}(\hat{\mathbf{b}}'|\mathbf{y})\right)\\
&= c_1^{(N)}(W, \tilde{W}) - \frac{1}{2N}\sum_{\mathbf{y}} R(\mathbf{y}) \sum_{\hat{\mathbf{b}}'} \tilde{P}(\hat{\mathbf{b}}'|\mathbf{y}) \log\left(\hat{P}(\hat{\mathbf{b}}', \mathbf{y})\right),
\end{aligned}$$

where $c_1^{(N)}(W, \tilde{W})$ is a function of W and $\tilde{W}$ only. Using

$$\hat{P}(\hat{\mathbf{b}}', \mathbf{y}) = Q(\mathbf{x}) \cdot \left(\prod_{\ell \in \mathcal{I}_N} \hat{W}(\hat{s}'_\ell | \hat{s}'_{\ell-1}, x_\ell)\right) \cdot \left(\prod_{\ell \in \mathcal{I}_N} \hat{W}(y_\ell | \hat{b}'_\ell)\right)$$

simplifies the surrogate function further to

$$\begin{aligned}
&\overline{\Psi}^{(N)}(\tilde{W}, \hat{W}) = c_2^{(N)}(W, \tilde{W}) \qquad (27)\\
&- \sum_{\hat{b}'} \log\left(\hat{W}(\hat{s}'|\hat{s}'_{\mathrm{p}}, x)\right) \underbrace{\frac{1}{2N}\sum_{\ell \in \mathcal{I}_N}\sum_{\mathbf{y}} R(\mathbf{y}) \sum_{\hat{\mathbf{b}}'} \tilde{P}(\hat{\mathbf{b}}'|\mathbf{y}) \left[\hat{b}'_\ell = \hat{b}'\right]}_{\triangleq \tilde{T}_1^{(N)}(\hat{b}')}\\
&- \sum_{\hat{b}'}\sum_{y} \log\left(\hat{W}(y|\hat{b}')\right) \underbrace{\frac{1}{2N}\sum_{\ell \in \mathcal{I}_N}\sum_{\mathbf{y}} R(\mathbf{y}) \sum_{\hat{\mathbf{b}}'} \tilde{P}(\hat{\mathbf{b}}'|\mathbf{y}) \left[\hat{b}'_\ell = \hat{b}'\right] [y_\ell = y]}_{\triangleq \tilde{T}_2^{(N)}(\hat{b}', y)},
\end{aligned}$$

where $c_2^{(N)}(W, \tilde{W})$ is a function of W and $\tilde{W}$ only. Applying the Lagrange method for minimizing $\overline{\Psi}^{(N)}(\tilde{W}, \hat{W})$ under the constraints (16)-(17) gives Lemma 1.

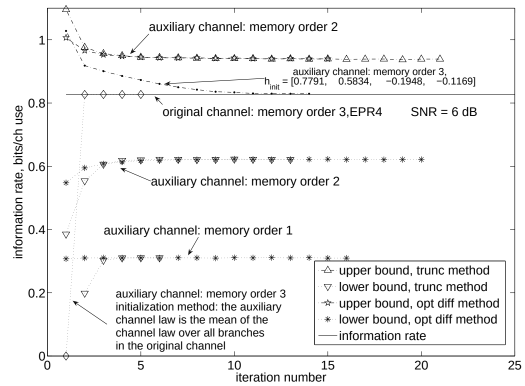

Fig. 2. Optimizing the upper and the lower bounds for the EPR4 channel. (The SNR is defined as in [3].)

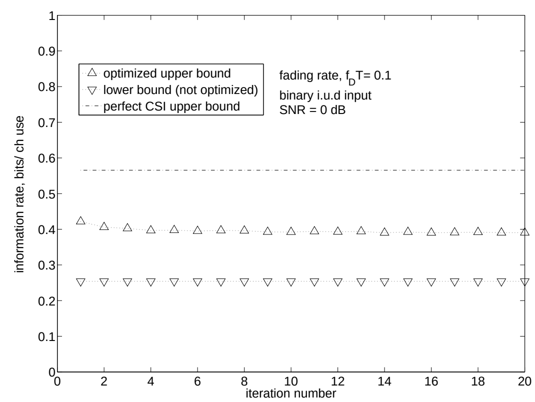

Fig. 3. Optimizing the upper bound for a flat-fading channel from 0.42 to 0.39 bits/ch use, which is much tighter than the upper bound with perfect channel state information. For reference, a lower bound is also shown.

REFERENCES

[1] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[2] J. G. Proakis, *Digital Communications*, 4th ed. New York: McGraw-Hill, 2000.

[3] D. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. on Inform. Theory*, vol. IT–52, no. 8, pp. 3498–3508, Aug. 2006.

[4] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. 2001 IEEE Int. Conf. on Communications*, Helsinki, Finland, June 11–14 2001, pp. 2692–2695.

[5] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," in *Proc. IEEE Intern. Symp. on Inform. Theory*, Washington, D.C., June 24–29 2001, p. 283.

[6] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite-state ISI channels," in *Proc. IEEE GLOBECOM*, San Antonio, TX, Nov. 2001, pp. 2992–2996.

[7] D. Arnold, H.-A. Loeliger, and P. O. Vontobel, "Computation of information rates from finite state/source channel models," in *Proc. 40th Allerton Conf. on Communications, Control, and Computing*, Allerton House, Monticello, Illinois, USA, October 2–4 2002.

[8] A. Ganti, A. Lapidoth, and İ. E. Telatar, "Mismatched decoding revisited: general alphabets, channels with memory, and the wide-band limit," *IEEE Trans. on Inform. Theory*, vol. IT–46, no. 7, pp. 2315–2328, 2000.

[9] P. O. Vontobel, A. Kavčić, D. Arnold, and H.-A. Loeliger, "A generalization of the Blahut-Arimoto algorithm for finite-state channels," *accepted (after revisions) for IEEE Trans. Inform. Theory, available online under* `http://www.hpl.hp.com/personal/Pascal_Vontobel`, Nov. 2005.

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

[11] P. Sadeghi and P. Rapajic, "Capacity analysis for finite-state Markov mapping of flat-fading channels," *IEEE Trans. Commun.*, vol. 53, no. 5, pp. 833–840, May 2005.