RZ 3349 (# 93395) 06/18/01
Electrical Engineering 5 pages

Research Report

An Upper Bound on the Capacity of Channels with Memory and Constraint Input

Pascal O. Vontobel[1] and Dieter M. Arnold[2]

[1]Laboratory for Signal and Information Processing
ETH Zurich
Sternwartstr. 7
CH-8092 Zurich, Switzerland
vontobel@isi.ee.ethz.ch

[2]IBM Research
Zurich Research Laboratory
8803 Rüschlikon
Switzerland
arn@zurich.ibm.com

An Upper Bound on the Capacity of Channels with Memory and Constraint Input

Pascal O. Vontobel[1] and Dieter M. Arnold[2]

[1] *Laboratory for Signal and Information Processing, ETH Zurich, Sternwartstr. 7, CH-8092 Zurich, Switzerland*

[2] *IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland*

Abstract

A method for computing upper bounds on capacity for a class of time-invariant indecomposable finite state channels is presented. It extends a result for finite-input memoryless channels from [1]. Numerical results are provided for selected channels having memory and constraint (binary) input.

[1] This work was supported by Grant TH-16./99-3.

1 Introduction

In many communication systems the channel is modeled as a binary-input linear channel with finite-length memory and additive white Gaussian noise (AWGN). Channels of this type arise, for instance, in magnetic recording where the actual channel is equalized to a so-called partial response (PR) target polynomial [2]. Despite their widespread use, the capacity of such channels is unknown. In particular, the capacity is unknown even for channels with memory $M = 1$.

A lower bound on the capacity has been conjectured by Shamai and Laroia [3]. By numerical comparison, it has been observed [3] that for channels with small memory the conjectured Shamai-Laroia lower bound (SLLB) on the capacity coincides with the information rate of the channel where the input is a 0.5-Bernoulli sequence over the binary input alphabet [4]. In [7], it is demonstrated that this so-called symmetric information rate can be computed exactly with standard forward sum-product trellis processing of simulated channel output data. Assuming the inputs to be generated by an optimized Markov source, non-symmetric information rates are obtained as well [7, 8].

In this paper we derive an upper bound on the channel capacity whose calculation complexity depends on the memory length M of the channel and on a parameter L which will be introduced.

2 Channel Model and Notation

Time-invariant indecomposable finite state channels (FSC) are defined in Ch. 4.6 of [6]. We consider a class of FSCs which can be represented as follows by a trellis with N states. Let s_k be the state after time k, x_k be the input at time k, u_k be the branch used at time k, $\gamma(u_k)$ be the associated output symbol at time k, and, finally, let y_k be the noisy output at time k with the following relations: $s_k = \alpha(s_{k-1}, x_k)$, $u_k = \beta(s_{k-1}, x_k)$ and $y_k = \gamma(u_k) + z_k$, where $\{z_k\}$ is an additive white noise (AWN) process (see upper right part of Fig. 1). The presented method works for this general scenario but here we will focus on time-invariant channels with memory that are specified by a Kronecker delta response $g(D)$ of order M, whose input is constraint to be binary, i.e. $x_k = -1$ or $x_k = +1$, and the noise is AWGN. The resulting trellis has $N = 2^M$ states and given the input sequence $X(D)$, the output can then be written as $Y(D) = g(D)X(D) + Z(D)$ (see upper left part of Fig. 1), where D is the delay operator. Of course, any channel with memory M can also be represented by a trellis having $2^{M'}$ states where $M' \geq M$. As typical examples we will use $g_{\mathrm{Ch1}}(D) = (1 - D)/\sqrt{2}$

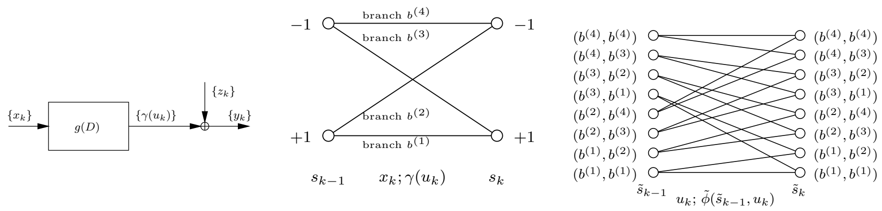

Figure 1: Channel model (upper left part), section of a channel trellis with $M' = 1$ (upper right part), section of the new trellis with $M' = 1$ and $L = 2$ (lower part).

and $g_{\text{Ch2}}(D) = (1 - D/2)/\sqrt{5/4}$: the normalization has been chosen such that the sum of the squared Kronecker delta response is one.

There are the following "trivial" upper bounds for the capacity of such channels. Firstly, it is upper bounded by the logarithm of the size of the input alphabet; in the case of binary inputs this is 1 bit/channel use. Secondly, it is upper bounded by the capacity of the same channel with arbitrary continuous input of the same variance; its solution is given by the usual water-filling argument [5]

3 Derivation of the Upper Bound

First we consider finite-input continuous-output memoryless channels. Let $W(y|x)$ be the probability density of receiving y when sending x. For any probability density $R(.)$ over the output alphabet, the capacity is upper bounded by $C \leq \max_x D\big(W(\cdot|x)||R(\cdot)\big)$ (this is a consequence of, for example, the comment on p. 142 of [1], or as a generalization of Problem 4.17 on p. 524f. in [6]).

Following Ch. 4.6 of [6], the capacity of an indecomposable finite state channel is given by

$$\begin{aligned} C &= \lim_{n\to\infty} \sup_{P_{\mathbf{x}_1^n}} \min_{s_0} \frac{1}{n} I(\mathbf{X}_1^n; \mathbf{Y}_1^n | S_0 = s_0) \\ &= \lim_{n\to\infty} \sup_{P_{\mathbf{x}_1^n}} \max_{s_0} \frac{1}{n} I(\mathbf{X}_1^n; \mathbf{Y}_1^n | S_0 = s_0). \end{aligned} \tag{1}$$

Interchanging the "sup" and the "max" in Eq. (1) and using the above bounding technique, the capacity of a channel with memory represented by a trellis with $2^{M'}$ states is therefore upper bounded by

$$C \leq \lim_{n\to\infty} \max_{\mathbf{u}_1^n} \frac{1}{n} D\big(W_1^n(\cdot|\mathbf{u}_1^n)||R_1^n(\cdot)\big),$$

where the maximization is over all possible branch sequences $\mathbf{u}_1^n$ of length n starting at time zero. We will call the maximizing branch sequence the "worst-case" branch sequence. $R_1^n(.)$ is any probability density over the output sequence $\mathbf{y}_1^n$ of length n. $W_1^n(\mathbf{y}_1^n|\mathbf{u}_1^n)$ is the probability density of receiving $\mathbf{y}_1^n$ when going through the branch sequence $\mathbf{u}_1^n$. Because we have AWN, we can write $W_1^n(\mathbf{y}_1^n|\mathbf{u}_1^n) = \prod_{k=1}^n W(y_k|u_k)$.

In the sequel we use the abbreviations $R(\mathbf{y}_\ell^{\ell'}) = R_{\mathbf{Y}_\ell^{\ell'}}(\mathbf{y}_\ell^{\ell'})$ for a probability density and $R(\mathbf{y}_\ell^{\ell'}|\mathbf{y}_m^{m'}) = R_{\mathbf{Y}_\ell^{\ell'}|\mathbf{Y}_m^{m'}}(\mathbf{y}_\ell^{\ell'}|\mathbf{y}_m^{m'})$ for a conditional probability density. If $\ell<1$ and $\ell'<1$ then $\mathbf{y}_\ell^{\ell'}$ is void; if the lower index ℓ of $\mathbf{y}_\ell^{\ell'}$ is smaller than one, we set $\mathbf{y}_\ell^{\ell'} = \mathbf{y}_1^{\ell'}$. Similar relations hold for $\mathbf{u}_\ell^{\ell'}$. Our approach is to model the output as the following product with parameter $L \geq 1$:

$$\begin{aligned} R(\mathbf{y}) &= R(y_1) \cdot R(y_2|y_1) \cdots R(y_n|\mathbf{y}_{n-L}^{n-1}) \\ &= \prod_{k=1}^n R(y_k|\mathbf{y}_{k-L}^{k-1}). \end{aligned}$$

Then

$$\begin{aligned}
&\frac{1}{n} D\big(W_1^n(\cdot|\mathbf{u}_1^n)||R_1^n(\cdot)\big) \\
&= \frac{1}{n} \mathrm{E}_{W_1^n(\cdot|\mathbf{u}_1^n)}\left[\log\left(\frac{\prod_{k=1}^n W(y_k|u_k)}{\prod_{k=1}^n R(y_k|\mathbf{y}_{k-L}^{k-1})}\right)\right] \\
&= \frac{1}{n}\sum_{k=1}^n \mathrm{E}_{W_1^n(\cdot|\mathbf{u}_1^n)}\left[\log\left(\frac{W(y_k|u_k)}{R(y_k|\mathbf{y}_{k-L}^{k-1})}\right)\right] \\
&= \frac{1}{n}\sum_{k=1}^n \mathrm{E}_{W_{k-L}^k(\cdot|\mathbf{u}_{k-L}^k)}\left[\log\left(\frac{W(y_k|u_k)}{R(y_k|\mathbf{y}_{k-L}^{k-1})}\right)\right] \\
&= \frac{1}{n}\sum_{k=1}^n \phi(\mathbf{u}_{k-L}^k) = \frac{1}{n}\sum_{k=1}^n \tilde{\phi}(\tilde{s}_{k-1}, u_k),
\end{aligned}$$

where we have introduced $\tilde{s}_{k-1} = \mathbf{u}_{k-L}^{k-1}$. We also used the definition

$$\begin{aligned}
&\phi(\mathbf{u}_{k-L}^k) = \tilde{\phi}(\tilde{s}_{k-1}, u_k) \\
&= \mathrm{E}_{W_{k-L}^k(\cdot|\mathbf{u}_{k-L}^k)}\left[\log\left(\frac{W(y_k|u_k)}{R(y_k|\mathbf{y}_{k-L}^{k-1})}\right)\right] \\
&= -H(Z_k) - \mathrm{E}_{W_{k-L}^k(\cdot|\mathbf{u}_{k-L}^k)}\big[\log R(y_k|\mathbf{y}_{k-L}^{k-1})\big].
\end{aligned}$$

To specify $R(y_k|\mathbf{y}_{k-L}^{k-1})$ we introduce the probability density $\bar{R}(\mathbf{y}_{k-L}^k)$. With $\bar{R}(\mathbf{y}_{k-L}^{k-1}) = \sum_{y_k} \bar{R}(\mathbf{y}_{k-L}^k)$, we can define $R(y_k|\mathbf{y}_{k-L}^{k-1}) = \bar{R}(\mathbf{y}_{k-L}^k)/\bar{R}(\mathbf{y}_{k-L}^{k-1})$ such that

$$\begin{aligned}
&\tilde{\phi}(\tilde{s}_{k-1}, u_k) \\
&\quad = -H(Z_k) - \mathrm{E}_{W_{k-L}^k(\cdot|\mathbf{u}_{k-L}^k)}\big[\log \bar{R}(\mathbf{y}_{k-L}^k)\big] \\
&\quad\quad + \mathrm{E}_{W_{k-L}^k(\cdot|\mathbf{u}_{k-L}^k)}\big[\log \bar{R}(\mathbf{y}_{k-L}^{k-1})\big]. \qquad (2)
\end{aligned}$$

We choose $\bar{R}_{k-L}^k(.)$ such that $\bar{R}_{k-L}^k(\mathbf{y}_1^{L+1}) = \bar{R}_1^{L+1}(\mathbf{y}_1^{L+1})$ for all $\mathbf{y}_1^{L+1}$. Thus, because of the time-invariance of the channel we have $\phi(\mathbf{u}_{k-L}^k) = \phi(\mathbf{u}_1^{L+1}) = \tilde{\phi}(\tilde{s}_{k-1}, u_k) = \tilde{\phi}(\tilde{s}_L, u_{L+1})$ for all $L+1 \leq k \leq n$ independently of the current value of n. Hence we can build a new trellis with $2^{M'} \cdot 2^L$ states with state $\tilde{s}_k$ after time k, and a branch between $\tilde{s}_{k-1} = \mathbf{u}_{k-L}^{k-1}$ and $\tilde{s}_k = \mathbf{u}_{k-L+1}^k$ with metric $\tilde{\phi}(\tilde{s}_{k-1}, u_k)$ (see lower part of Fig. 1). Note that the trellis sections become identical after L time steps at the latest. For a specific $\mathbf{u}_1^n$, $D(W_1^n(\cdot|\mathbf{u}_1^n)||R_1^n(\cdot))$ equals the cumulative metric along the corresponding path in the new trellis. In the calculation of $\tilde{\phi}(\tilde{s}_{k-1}, u_k)$ we approximate the expectation values in Eq. (2) by stochastic averaging. For $\bar{R}(\mathbf{y}_1^{L+1})$ we may take any multi-dimensional density we want: we take the output density of the $L+1$ trellis sections of the channel model with $2^{M'}$ states and optimized branching probabilities as gotten in [7, 8]. The values of $\bar{R}_1^{L+1}(.)$ can be efficiently calculated using a modified version of the sum-product algorithm.

For the following consideration we neglect the initial L trellis sections in the new trellis which are not relevant for the calculation of the upper bound. In the remaining trellis, all sections are identical and independent of the current value of n. Finding the "worst-case" sequence $\mathbf{u}_1^n$ for $n \to \infty$ is now equivalent to finding the semi-infinite path $\mathbf{u}_1^\infty$ in this new trellis with the largest metric or equivalently with the largest (time-) average metric per trellis section. We argue that if the worst-case sequence is unique, it must be periodic with a period length not larger than $2^{M'+L}$. This is because in the new trellis all sections are identical, such that if we are in a specific state at any time step, always the same branch is taken to the next time step. As there are $2^{M'+L}$ states, we must return to the same state after at most $2^{M'+L}$ time steps. To determine this largest (time-) average metric per trellis section we can use a modified Viterbi (max-sum) algorithm. If there are several paths having the same largest (time-) average metric, there is also a periodic one among them and as we are only interested in the (time-) average metric per trellis section and not the path itself we can proceed by searching for this periodic "worst-case" sequence.

4 Results

Upper bounds on the capacity of the $g_{\mathrm{Ch1}}(D)$ and $g_{\mathrm{Ch2}}(D)$ channels with $M = 1$ can be seen in Fig. 2; we used $M' = 3$ and $L = 6$. We also plotted the capacity (water-filling upper bound) for the same channels with arbitrary continuous input of the same variance; the lower bound (optimized information rate) was obtained by the method presented in [7, 8] with a memory three Markov process at the input. For large SNR the bound is not tight in the case of $g_{\mathrm{Ch1}}(D)$ (here with overshoot $\propto 1/(L+1)$) because the trellis is not observable. This means that even without noise, the channel state cannot always be determined by seeing only a finite window of the output sequence (because the all-zero sequence can be generated in two different ways).

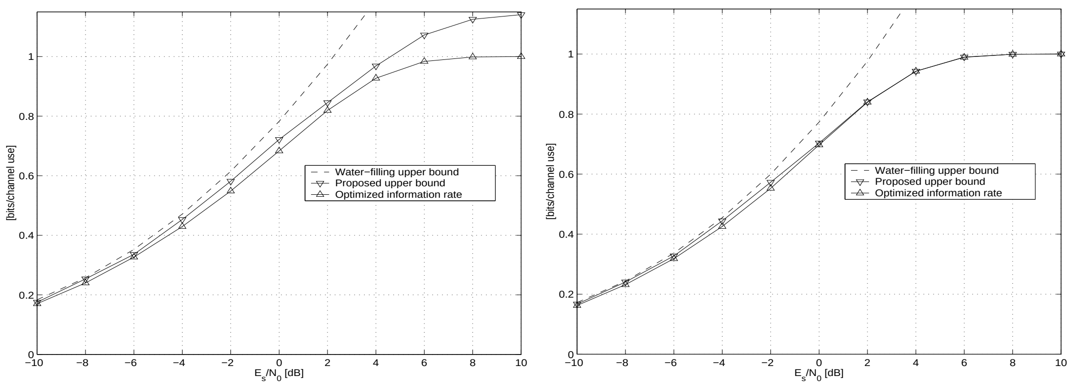

Figure 2: Capacity bounds for the $g_{\mathrm{Ch1}}(D) = (1 - D)/\sqrt{2}$ channel (upper part) and the $g_{\mathrm{Ch2}}(D) = (1 - D/2)/\sqrt{5/4}$ channel (lower part) with $M' = 3$ and $L = 6$ (see also text).

References

[1] I. Csiszár and J. Körner, *Information Theory.* Budapest: Akadémiai Kiadó (Publishing House of the Hungarian Academy of Sciences), 1981. Coding theorems for discrete memoryless systems.

[2] R. D. Cideciyan, F. Dolivo, R. Hermann, W. Hirt, and W. Scott, "A PRML System for Digital Magnetic Recording," *IEEE J. Sel. Areas Comm.*, vol. 10, pp.38–56, Jan. 1992.

[3] S. Shamai and R. Laroia, "The Intersymbol Interference Channel: Lower Bounds on Capacity and Channel Precoding Loss," *IEEE Trans. on Inform. Theory*, vol. 42, pp. 1388–1404, Sept. 1996.

[4] W. Hirt, *Capacity and Information Rates of Discrete-Time Channels with Memory.* PhD thesis, Swiss Federal Institute of Technology, ETH Zurich, October 1998.

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

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

[7] D. Arnold, H.-A. Loeliger, "On the Information Rate of Binary-Input Channels with Memory," accepted for publication at ICC 2001, Helsinki. Available as IBM research report RZ3341 under `http://domino.watson.ibm.com/library/ Cyberdig.nsf/home`.

[8] A. Kavčić, "Capacity of Markov Sources over Noisy Channels," submitted to GLOBECOM 2001.