RZ 3341 (# 93387) 05/07/01
Computer Science 8 pages

Research Report

On the Information Rate of Binary-Input Channels with Memory

D. Arnold

IBM Research
Zurich Research Laboratory
8803 Rüschlikon
Switzerland
e-mail: arn@zurich.ibm.com

H.-A. Loeliger

ISI, ETH Zurich
Sternwartstr. 7
CH-8092 Zürich, Switzerland
e-mail: loeliger@isi.ee.ethz.ch

IBM Research
Almaden · Austin · Beijing · Delhi · Haifa · T.J. Watson · Tokyo · Zurich

On the Information Rate of Binary-Input Channels with Memory

D. Arnold

IBM Research, Zurich Research Laboratory, 8803 Rüschlikon, Switzerland

H.-A. Loeliger *ISI, ETH Zurich, Sternwartstr. 7, CH-8092, Zürich, Switzerland*

Abstract

The entropy rate of a finite-state hidden Markov model can be estimated by forward sum-product trellis processing (i.e., the forward recursion of the Baum-Welch/BCJR algorithm) of simulated model output data. This can be used to compute information rates of binary-input AWGN channels with memory.

1 Introduction

The magnetic recording channel is often modeled as a binary-input linear channel with memory and additive white Gaussian noise (AWGN). Channels of this type, which are often called partial response (PR) channels, arise also in many communication systems. Despite their ubiquity, no practical algorithm for computing the capacity of such channels is known.

A simpler problem is the computation of the mutual information $I(X;Y)$ between the input X and the output Y of such a channel for the case where X_k, $k \in \mathbb{Z}$, are independent and identically distributed (i.i.d.) random variables with uniform distribution over the (binary) input alphabet. That mutual information will be called the uniform-input information rate.

The problem of computing the uniform-input information rate for binary-input partial response channels was studied by Hirt [1] and by Shamai et al. [2]. Hirt [1] obtains upper and lower bounds from Monte-Carlo simulation using input blocks of finite length n. If m is the channel memory, the difference between the upper bound and the lower bound is at most m/n bits/symbol, and the complexity of this method is proportional to 2^n. The bounds reported in [2] yield less accurate results even for channels with small memory.

Recently, a lower bound on the capacity has been conjectured by Shamai and Laroia [3]. In contrast to Hirt's method, this conjectured lower bound is easy to compute also for channels with large memory. By numerical comparison, it has been observed [3] that the conjectured Shamai-Laroia lower bound (SLLB) on the capacity coincides with the uniform-input information rate for the channels with small memory considered in [1].

In this paper, we present a new practical method to compute uniform-input and nonuniform-input information rates of such channels. Like Hirt's method, our method is simulation-based; however, its complexity is only proportional to 2^m. The new method applies, in fact, to the more general class of channels consisting of a deterministic finite-state channel followed by an additive noise channel, and it extends also to finite-state hidden-Markov input processes. The pivotal observation behind the new method is that the entropy rate of the channel output can be computed by standard forward sum-product trellis processing of simulated channel output.

This paper is structured as follows. The new method is described in Section 2. Numerical results for some standard PR channel models as well as a comparison with Hirt's method are given in Section 3. In Section 4, we go beyond uniform i.i.d. input: for selected channels higher information rates—i.e., improved lower bounds on channel capacity—are obtained with input processes that have memory. Some concluding remarks are given in Section 5.

2 Computing the Information Rates

Consider a channel with binary input $X_i \in \{+1,-1\}$, $i \in \mathbb{Z}$, and output Y_i, $i \in \mathbb{Z}$, given by

$$Y_i = \sum_{k=0}^{m} g_k X_{i-k} + Z_i, \tag{1}$$

where $g_0, g_1, \ldots, g_m$ are fixed real parameters and where the "noise" $Z_i \in \mathbb{R}$ is assumed to be a white (i.e., i.i.d.) process. The channel input process X is assumed to be stationary. We will use the notation $X^n \triangleq (X_1, X_2, \ldots, X_n)$. Entropy rates and information rates are defined in the usual way; e.g., the differential entropy rate $h(Z)$ is defined as $h(Z) \triangleq \lim_{n\to\infty} h(Z^n)/n$. The results of this paper apply also if Z_i, $i \in \mathbb{Z}$, is discrete. In this case, $h(Z)$, $h(Y)$, and

$h(Y|X)$ are replaced by $H(Z)$, $H(Y)$, and $H(Y|X)$ respectively. We consider the problem of computing

$$I(X;Y) = h(Y) - h(Y|X) \tag{2}$$

(or an estimate thereof), for the case that X is a process with memory at most m, i.e., $p(x_k|x_{k-1}, x_{k-2}, \ldots) = p(x_k|x_{k-1}, \ldots, x_{k-m})$, for all $k \in \mathbb{Z}$. Since $h(Y|X) = h(Z)$, the problem reduces to the computation of $h(Y)$ (or an estimate thereof).

As is well known, for any given block length n and any given channel output $y^n = (y_1, y_2, \ldots, y_n)$, the probability $p(y^n)$ can be computed by the forward recursion of the Baum-Welch/BCJR algorithm [4, 5], which operates on the trellis of the channel. For this computation, each trellis branch b is assigned the "metric" $\mu(b) = p(b|\,\mathrm{lst}(b))p(y|b)$, where $\mathrm{lst}(b)$ is the starting (left-hand) state of b, and each of the initial (leftmost) states of the trellis is assigned the "metric" $\mu(s) = p(s)$ according to the stationary state distribution. The trellis is then processed from left (initial states) to right (final states), computing state metrics $\mu(s)$ according to the rule

$$\mu(s) = \sum_{b:\,\mathrm{rst}(b)=s} \mu(\mathrm{lst}(b))\mu(b), \tag{3}$$

where $\mathrm{lst}(b)$ and $\mathrm{rst}(b)$ denote the left-hand (starting) and right-hand (ending) state, respectively, of branch b. Then the sum of $\mu(s)$ over all final (rightmost) trellis states s equals $p(y^n)$.

An estimate of $h(Y^n) = -E[\log(p(y^n))]$ is thus obtained by the following algorithm. Simulate the channel N times, each time starting with the stationary state distribution and simulating n channel inputs x^n and outputs y^n. For each such simulation, compute $p(y^n)$ as described above. Let ρ_k be the resulting $p(y^n)$ of the k-th simulation. Then $-\frac{1}{N}\sum_{k=1}^{N}\log(\rho_k)$ is an estimate of $h(Y^n)$ that converges (with probability 1) to $h(Y^n)$ for $N \to \infty$.

The above algorithm for the computation of $h(Y^n)$ was formulated mainly to give some intuition. We now describe the algorithm that will actually be used. For a stationary ergodic finite-state hidden-Markov process Y (and thus certainly for models as in (1)), the Shannon-McMillan-Breimann theorem [6, Section 15.7] applies also if Y is continuous [7]. Hence,

$$-\frac{1}{n}\log(p(Y^n)) \to h(Y) \tag{4}$$

with probability one.

An estimate of $h(Y) = \lim_{n\to\infty} h(Y^n)$ is thus obtained by a single long simulation of y^n, $n = 1, 2, 3, \ldots$, and the corresponding single forward sum-product recursion, which yields $p(y^n)$, $n = 1, 2, 3, \ldots$, as the sum of the time-n state metrics. Due to (4), the sequence of estimates $-\frac{1}{n}\log(p(y^n))$ converges to $h(Y)$.

For large n, the state metrics $\mu(s)$ computed according to (3) quickly tend to zero. In practice, the recursion rule (3) is therefore changed to

$$\mu(s) = \sum_{b:\,\mathrm{rst}(b)=s} \lambda_k \mu(\mathrm{lst}(b))\mu(b), \tag{5}$$

where λ_k is a scale factor for the k-th trellis section. If λ_k is chosen such that, for each time k, the sum of the time-k state metrics equals 1, then

$$\frac{1}{n}\sum_{k=1}^{n}\log(\lambda_k) = -\frac{1}{n}\log(p(y^n)). \tag{6}$$

The left-hand side of (6) thus converges to $h(Y)$.

3 Uniform-Input Information Rates

The algorithm described in Section 2 was used to compute the uniform-input information rate for the channels listed in Table 1. The channel input alphabet is the set $\{+1, -1\}$. The channel tap coefficients $g_0, g_1, \ldots, g_m$ are normalized so that $\sum_{k=0}^{m} g_k^2 = 1$. The input energy is $E_\mathrm{s} = 1$ and the noise is AWGN with variance $N_\mathrm{o}/2$.

Table 1: Impulse responses of selected channels.

Channel name	Normalized impulse response
DICODE	$g(D) = (1 - D)/\sqrt{2}$
EPR4	$g(D) = (1 + D - D^2 - D^3)/2$
E2PR4	$g(D) = (1 + 2D - 2D^3 - D^4)/\sqrt{10}$
CH6	$g(D) = 0.19 + 0.35D + 0.46D^2 + 0.5D^3 +$ $0.46D^4 + 0.35D^5 + 0.19D^6$

The resulting uniform-input information rates are shown in Fig. 1. Note that, for the memoryless channel $g(D)$=1, which is also shown in Fig. 1, the uniform-input information rate coincides with the channel capacity. The following observations can be made:

- For a given signal to noise ratio (SNR), the uniform-input information rate decreases with increasing memory.
- At a rate of 0.9 bits/symbol, the uniform-input information rate of the DICODE channel is 0.8 dB away from the capacity of the memoryless channel. The E2PR4 channel is another dB off. This is in agreement with results reported in [1, 3].

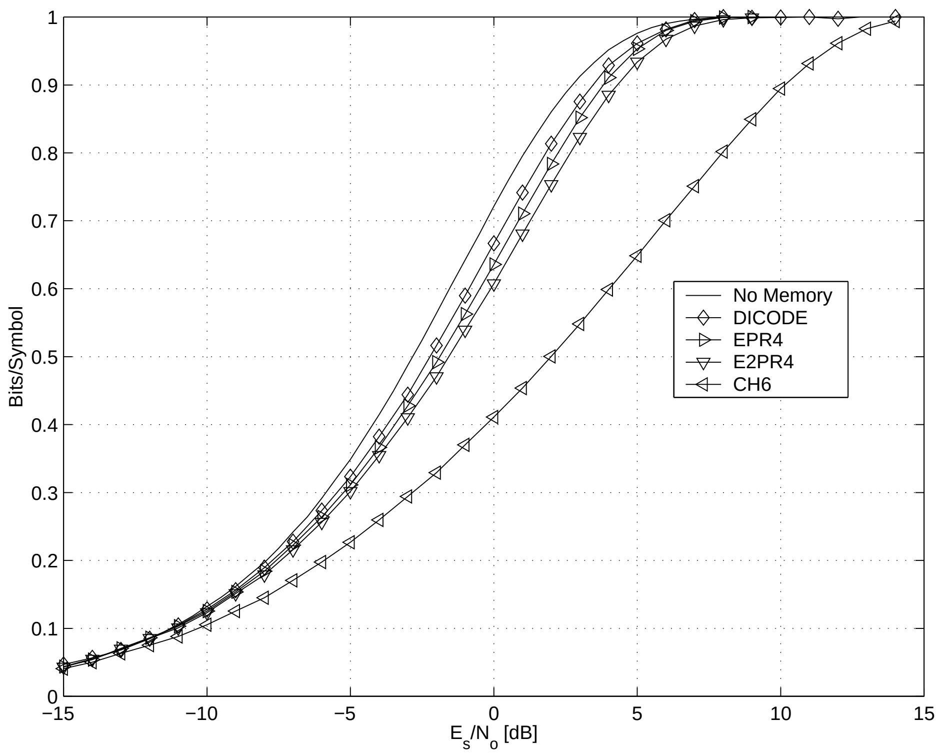

Figure 1: Uniform-input information rates of selected channels.

The convergence behavior of the algorithm is illustrated by Fig. 2. The uniform-input information rate for the DICODE channel at 0 dB was computed 10+100+1000 times, each time by a simulation run of 10^6 symbols and with a new random seed. For each blocklength n, Fig. 2 shows the minimum and the maximum computed estimate of the information rate among the first 10, the next 100, and the remaining 1000 simulation runs.

We next compare the method of Section 2 with that used by Hirt [1]. The latter method computes the quantities I_n and $\hat{I}_n$, which are defined as follows:

$$I_n \triangleq I(X^n; Y^n | X_0, X_{-1}, \ldots, X_{1-m})/n, \tag{7}$$

and

$$\hat{I}_n \triangleq I(X^n; Y^{n+m} | X_0, \ldots, X_{1-m}, X_{n+1}, \ldots, X_{n+m})/n, \tag{8}$$

both for uniform i.i.d. binary input X. By standard arguments,

$$I_n \leq \hat{I}_n \tag{9}$$

and

$$\lim_{n\to\infty} I_n = \lim_{n\to\infty} \hat{I}_n = I(X;Y). \tag{10}$$

The difference between I_n and $\hat{I}_n$ is at most m/n. For i.i.d. inputs I_n is a lower bound on the information rate $I(X;Y)$, i.e. $I_n \leq I(X;Y)$ for $n \geq 1$.

Hirt computes I_n and $\hat{I}_n$ by numerical integration based on Monte-Carlo simulation. Note that this could also be done by the algorithm for the computation of $h(Y^n)$ described in Section 2.

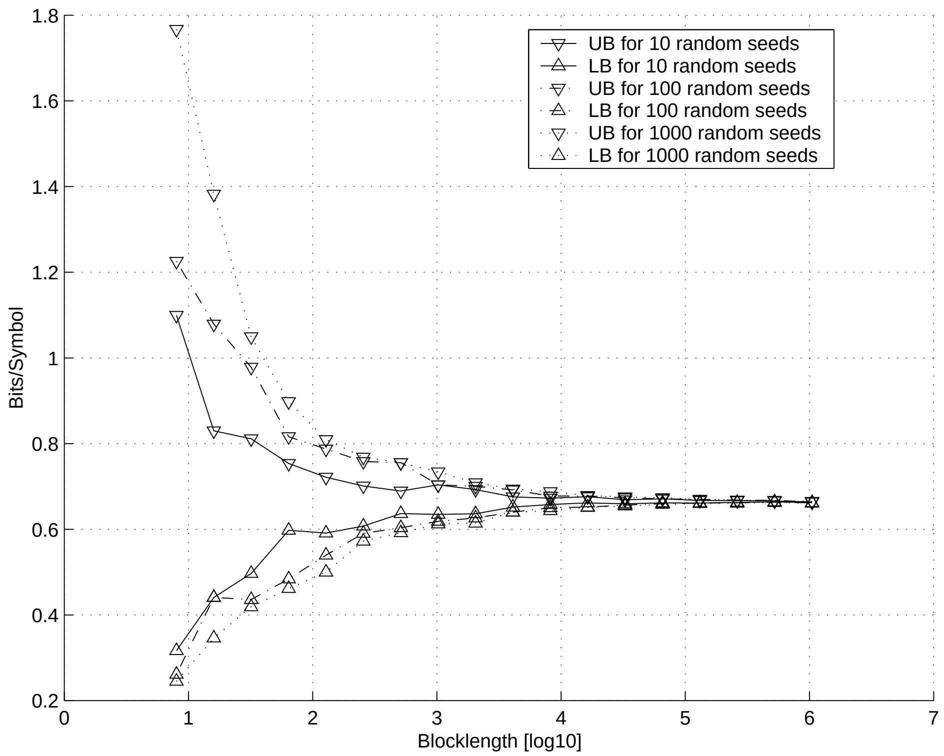

Figure 2: Convergence behaviour of the new algorithm.

Numerical results for Hirt's method are shown in Figures 3 and 4 for the DICODE channel and the channel CH6, respectively, at -8 dB, 0 dB, and 5 dB. The plots show I_n (there denoted by "Lower bound") and $\hat{I}_n$ ("Upper bound") as a function of n. For comparison, the plots also show the exact uniform-input information rate computed by the algorithm of Section 2 (denoted by FSPA) as well as the Shamai-Laroia conjectured lower bound on capacity (SLLB) [3]. The latter is computed by evaluating a single 1-dimensional integral. In the computed examples, it is very close to the uniform-input information rate.

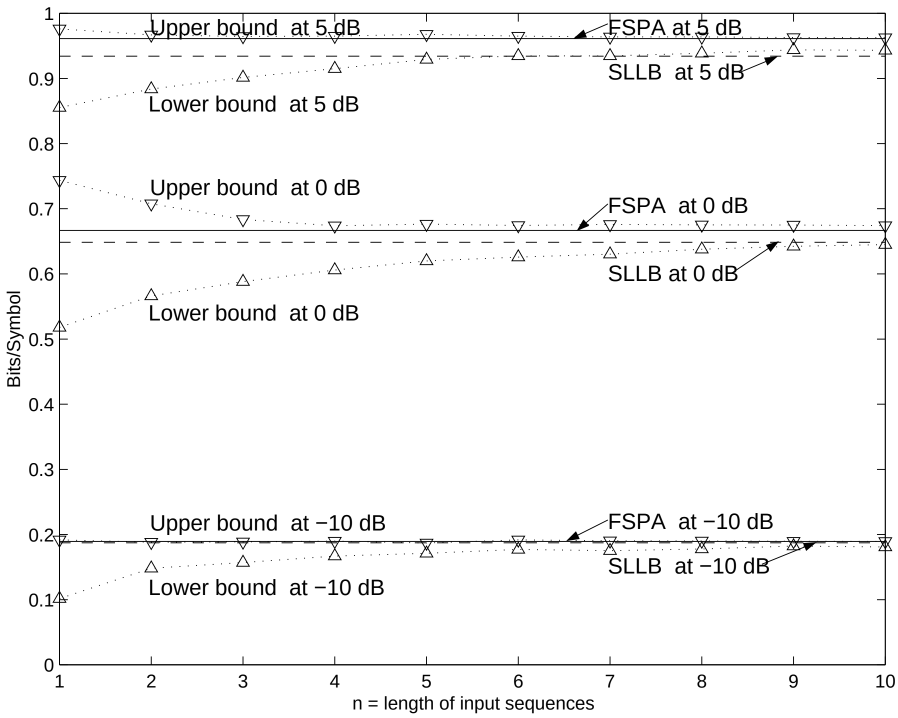

Figure 3: Comparison of uniform-input information rates obtained by different methods for the DICODE-channel.

4 Nonuniform-Input Information Rates

We now go beyond uniform i.i.d. input and consider stationary input distributions of the form $p(x_k|x_{k-1}, x_{k-2}, \ldots) = p(x_k|x_{k-1}, \ldots, x_{k-m})$. Clearly, the method of Section 2 can still be used to compute the information rate $I(X;Y)$. If necessary, the channel memory parameter m may be artificially increased to allow for input processes with larger memory.

In the sequel, we focus on the DICODE channel, which is equivalent to the PR4-channel (with $g(D) = (1 - D^2)/\sqrt{2}$) that is widely used in magnetic recording. We computed the maximum information rate of this channel for input processes with memory 1 and 2 (i.e., 2 states and 4 states, respectively). The results are shown in Fig. 5. The branching probabilities of the Markov model were numerically optimized for maximum $h(Y)$. It is easy to see that, for SNR $\rightarrow \infty$, the uniform i.i.d. process is optimal. For low SNR, however, the optimal input distribution tries to produce more output symbols with high energy.

Also shown in Fig. 5 are the capacity both of the memoryless AWGN channel and of the DICODE channel for nonbinary power-limited (Gaussian) input. The plot shows that, at low

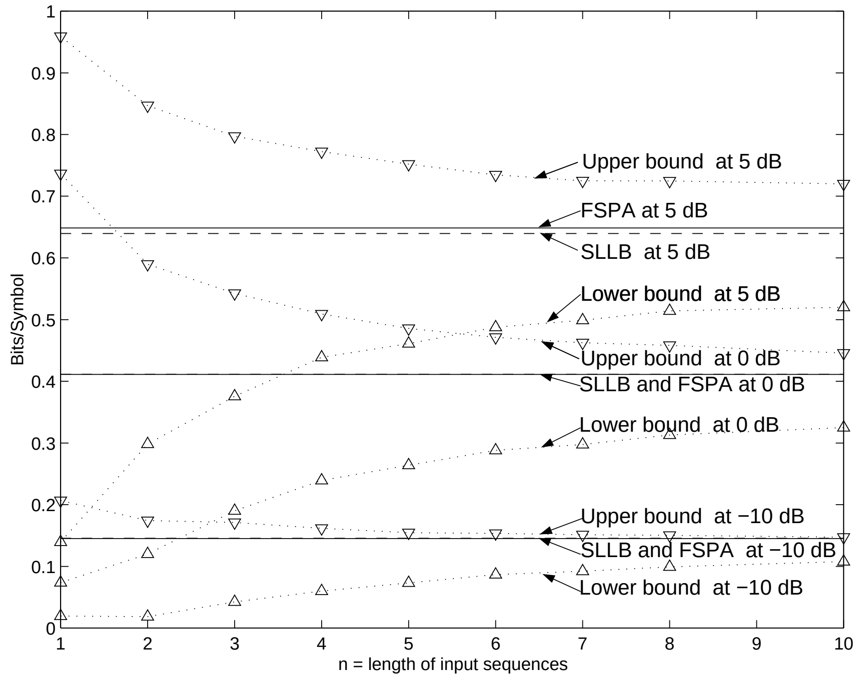

Figure 4: Comparison of uniform-input information rates obtained by different methods for the CH6-channel.

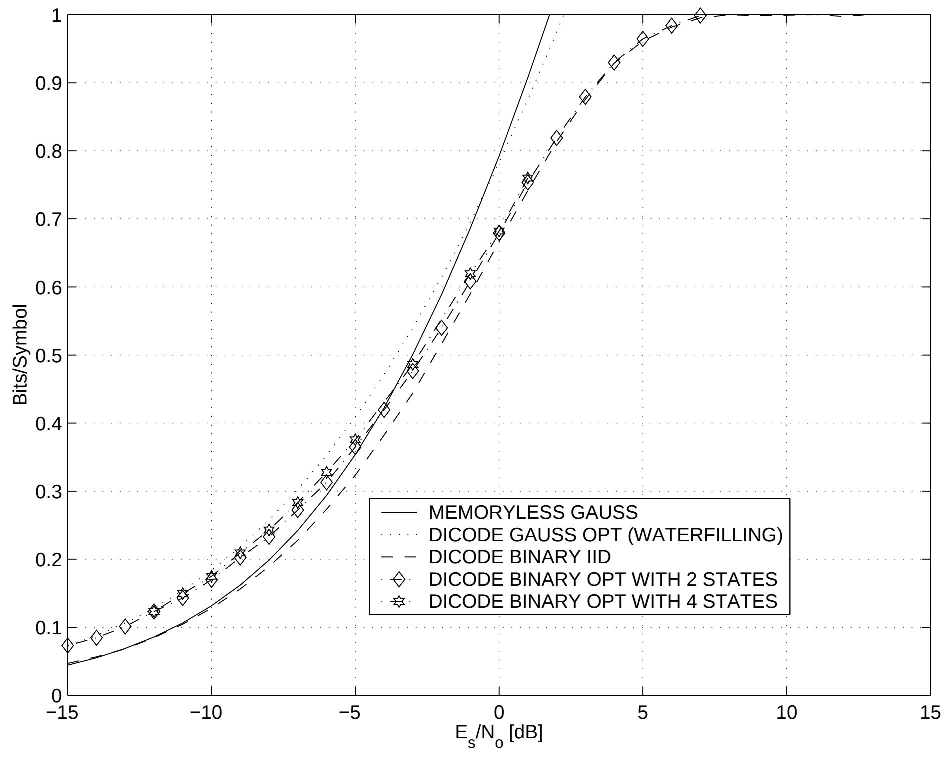

Figure 5: Maximum information rates of the DICODE channel for input processes with memory 0, 1, and 2.

SNR, the optimized Markov models achieve noticeably higher information rates than uniform i.i.d. processes. These information rates exceed even the capacity of the memoryless Gaussian channel. Similar results were obtained for the CH6 channel by systematically optimizing the branching transition probabilities. At rate one half, a memory 6 input process improves the lower bound on capacity of the CH6 channel by 2 dB (see Fig. 6).

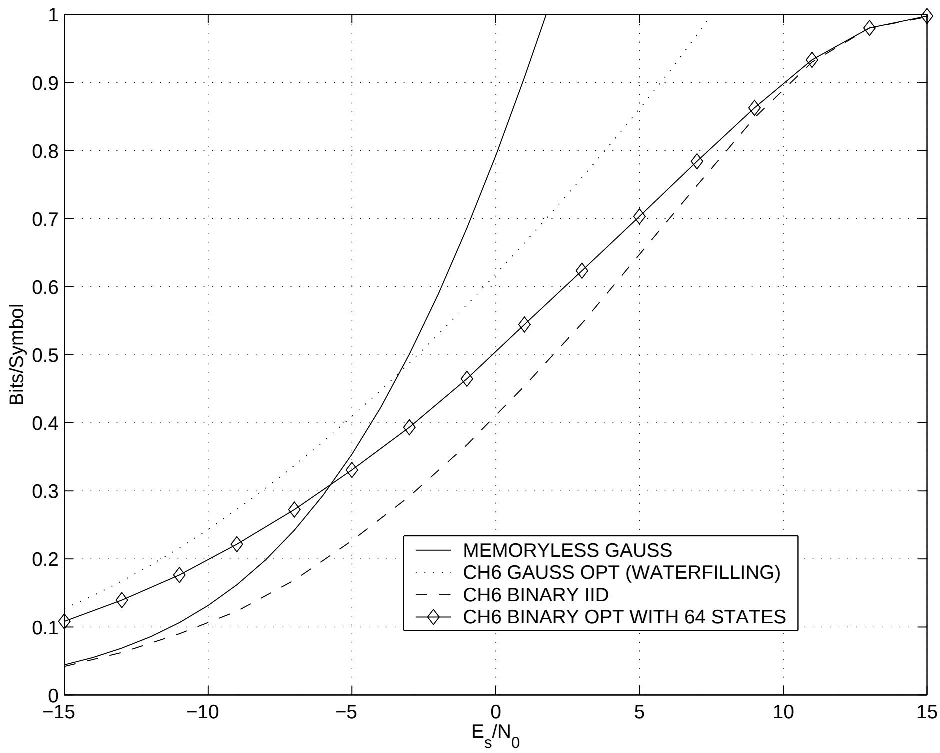

Figure 6: Maximum information rate of the CH6 channel for an input process with memory 6.

5 Summary and Conclusions

It has been pointed out that uniform-input and nonuniform-input information rates for discrete-time binary-input channels with memory can be computed by standard forward sum-product processing of simulated channel output. The complexity of this method is proportional to 2^m, where m is the channel memory; with todays computing power, the method is easily applicable to channels with memory up to about 15. The method extends to general finite-state channel input processes. For the DICODE and the CH6 channel, maximum information rates for input processes with memory were computed. By increasing the memory, the capacity of these channels can be closely approximated.

References

[1] W. Hirt, "Capacity and Information Rates of Discrete-Time Channels with Memory," *Ph.D. thesis no. 8671*, ETH Zurich, 1988.

[2] Sh. Shamai, L.H. Ozarow, A.D. Wyner, "Information Rates for a Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs," *IEEE Trans. Inform. Theory*, vol. 37, no. 6, pp. 1527-1539, Nov. 1991.

[3] Sh. Shamai, R. Laroia, "The Intersymbol Interference Channel: Lower Bounds on Capacity and Channel precoding Loss," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1388-1404, Sept. 1996.

[4] L.E. Baum, T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Stat.*, vol. 37, pp. 1554-1563, 1966.

[5] L.R. Bahl, J. Cocke, F. Jelinek, J. Raviv, "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Trans. Inform. Theory*, vol. 20, pp. 284-287, March, 1974.

[6] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[7] B.G. Leroux, "Maximum-likelihood estimation for hidden Markov models," *Stochastic Processes and their Applications*, vol. 40, pp. 127-143, 1992.