November 1, 2002
Submitted to ISIT 2003

Simulation-Based Computation of Information Rates: Upper and Lower Bounds

Dieter Arnold, Aleksandar Kavčić, Hans-Andrea Loeliger,
Pascal O. Vontobel, and Wei Zeng

Abstract

It has recently become feasible to compute information rates of finite-state source/channel models with not too many states. Such methods can also be used to compute upper and lower bounds on the information rate of very general (non-finite-state) channels with memory by means of finite-state approximations. We review these methods and present new reduced-state bounds.

1 Introduction

We consider the problem of computing the information rate

$$I(X;Y) \triangleq \lim_{n\to\infty} \frac{1}{n} I(X_1,\ldots,X_n;Y_1,\ldots,Y_n) \tag{1}$$

between the input process $X=(X_1,X_2,\ldots)$ and the output process $Y=(Y_1,Y_2,\ldots)$ of a time-invariant channel with memory. We will assume that X is Markov or hidden Markov, and we will primarily be interested in the case where the channel input alphabet $\mathcal{X}$ (i.e., the set of possible values of X_k) is finite.

For *finite-state* channels (to be defined in Section 2), a practical method for the computation of (1) was presented independently by Arnold and Loeliger [1], by Sharma and Singh [11], and by Pfister et al. [10]. That method consists essentially of sampling both a long input sequence $x^n \triangleq (x_1,\ldots,x_n)$ and the corresponding output sequence $y^n \triangleq (y_1,\ldots,y_n)$, followed by the computation of $\log p(y^n)$ (and, if necessary, of $\log p(y^n|x^n)$) by means of a forward sum-product recursion on the joint source/channel trellis. We will review this method in Section 2.

Extension of such methods to very general (non-finite state) channels were presented in [2]. These extensions use finite-state approximations of the actual channel. By simulations of the actual source/channel and computations using the finite-state model, both an upper bound and a lower bound on the information rate of the actual channel are obtained. We will review these bounds in Section 3 and give new numerical results.

In Section 4, we propose a new upper bound and a generic new lower bound on the information rate, which complement the bounds of [2].

Related earlier and parallel work includes [6] [12] [13] [5] [7] [14], see [2].

2 Computing $I(X;Y)$ for Finite-State Channels

In this section, we review the method of [1] [11] [10]. We will assume that X, Y, and $S=(S_0,S_1,S_2,\ldots)$ are stochastic processes such that

$$p(x_1,\ldots,x_n,y_1,\ldots,y_n,s_0,\ldots,s_n) = p(s_0)\prod_{k=1}^{n} p(x_k,y_k,s_k|s_{k-1}) \tag{2}$$

D. Arnold and H.-A. Loeliger are with the Dept. of Information Technology and Electrical Engineering, ETH, CH-8092 Zürich, Switzerland. Email: {arnold, loeliger}@isi.ee.ethz.ch.
A. Kavčić and W. Zeng are with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA, 02138. Email: {kavcic, wzeng}@deas.harvard.edu.
P. O. Vontobel is with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801. Email: vontobel@ifp.uiuc.edu. Supported by Grant NSF CCR 99-84515.

for all $n > 0$ and with $p(x_k, y_k, s_k|s_{k-1})$ not depending on k. We will assume that the state S_k takes values in a *finite* set and we will assume that the process S is ergodic; under the stated conditions, a sufficient condition for ergodicity is $p(s_k|s_0) > 0$ for all s_0, s_k for all sufficiently large k.

For the sake of clarity, we will further assume that the channel input alphabet $\mathcal{X}$ is a finite set and that the channel output Y_k takes values in $\mathbb{R}$; none of these assumptions is essential, however. With these assumptions, the left-hand side of (2) should be understood as a probability mass function in x_k and s_k, and as a probability density in y_k.

Under the stated assumptions, the limit (1) exists. Moreover, the sequence $-\frac{1}{n}\log p(X^n)$ converges with probability 1 to the entropy rate $H(X)$, the sequence $-\frac{1}{n}\log p(Y^n)$ converges with probability 1 to the differential entropy rate $h(Y)$, and $-\frac{1}{n}\log p(X^n, Y^n)$ converges with probability 1 to $H(X)+h(Y|X)$, cf. [4].

From the above remarks, an obvious algorithm for the numerical computation of $I(X;Y) = h(Y) - h(Y|X)$ is as follows:

1. Sample two "very long" sequences x^n and y^n.
2. Compute $\log p(x^n)$, $\log p(y^n)$, and $\log p(x^n, y^n)$. If $h(Y|X)$ is known analytically, then it suffices to compute $\log p(y^n)$.
3. Conclude with the estimate

$$\hat{I}(X;Y) = \frac{1}{n}\log p(x^n, y^n) - \frac{1}{n}\log p(x^n) - \frac{1}{n}\log p(y^n) \tag{3}$$

or, if $h(Y|X)$ is known analytically, $\hat{I}(X;Y) = -\frac{1}{n}\log p(y^n) - h(Y|X)$.

The computations in Step 2 can be carried out by forward sum-product message passing through the factor graph of (2), as illustrated in Fig. 1. Since the graph represents a trellis, this computation is just the forward sum-product recursion of the BCJR algorithm [3].

Consider, for example, the computation of

$$p(y^n) = \sum_{x^n, s^n} p(x^n, y^n, s^n) \tag{4}$$

with $s^n \triangleq (s_0, s_1, \ldots, s_n)$. By straightforward application of the sum-product algorithm [8], we recursively compute the messages (i.e., state metrics)

$$\mu_{\mathrm{f}}(s_k) = \sum_{x_k, s_{k-1}} \mu_{\mathrm{f}}(s_{k-1})\, p(x_k, y_k, s_k|s_{k-1}) \tag{5}$$

$$= \sum_{x^k, s^{k-1}} p(x^k, y^k, s^k) \tag{6}$$

for $k = 1, 2, 3, \ldots$, as illustrated in Fig. 1. The desired quantity (4) is then obtained as

$$p(y^n) = \sum_{s_n} \mu_{\mathrm{f}}(s_n), \tag{7}$$

the sum of all final state metrics.

In practice, the recursion rule (5) is modified to include a suitable scale factor, cf. [2].

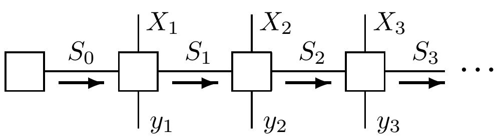

Figure 1: Computation of $p(y^n)$ by message passing through the factor graph of (2).

3 Computing Bounds on $I(X;Y)$ for General Channels

Let $p(x^n, y^n)$ be some ergodic source/channel law. Let $q(y^n|x^n)$ be another ergodic channel and define $q_p(y^n) \triangleq \sum_{x^n} p(x^n)q(y^n|x^n)$. As described in [2], we then have

$$\underline{I}_q(X;Y) \le I(X;Y) \le \overline{I}_q(X;Y) \tag{8}$$

with

$$\overline{I}_q(X;Y) \triangleq \lim_{n\to\infty} \mathrm{E}_{p(\cdot,\cdot)}\left[\frac{1}{n}\log p(Y^n|X^n) - \frac{1}{n}\log q_p(Y^n)\right] \tag{9}$$

and

$$\underline{I}_q(X;Y) \triangleq \lim_{n\to\infty} \mathrm{E}_{p(\cdot,\cdot)}\left[\frac{1}{n}\log q(Y^n|X^n) - \frac{1}{n}\log q_p(Y^n)\right]. \tag{10}$$

Now assume that $p(\cdot|\cdot)$ is some "difficult" (non-finite-state) ergodic channel. As shown in [2], we can compute the bounds $\overline{I}_q(X;Y)$ and $\underline{I}_q(X;Y)$ on the information rate $I(X;Y)$ by the following algorithm:

1. Choose a finite-state source $p(\cdot)$ and an auxiliary finite-state channel $q(\cdot|\cdot)$ so that their concatenation is a finite-state source/channel model as defined in Section 2.
2. Connect the source to the *original* channel $p(\cdot|\cdot)$ and sample two "very long" sequences x^n and y^n.
3. Compute $\log q_p(y^n)$ and, if necessary, $\log p(x^n)$ and $\log q(y^n|x^n)p(x^n)$ by the method described in Section 2.
4. Conclude with the estimates

$$\hat{\overline{I}}_q(X;Y) = -\frac{1}{n}\log q_p(y^n) - h(Y|X) \tag{11}$$

and

$$\hat{\underline{I}}_q(X;Y) = \frac{1}{n}\log q(y^n|x^n)p(x^n) - \frac{1}{n}\log p(x^n) - \frac{1}{n}\log q_p(y^n). \tag{12}$$

Note that the term $h(Y|X)$ in the upper bound (11) refers to the original channel and cannot be computed by means of the auxiliary channel.

4 Reduced-State Bounds

Let $\mathcal{S}'_k$ be a subset of the time-k states. If the sum in the recursion rule (5) is modified to

$$\mu_{\mathrm{f}}(s_k) = \sum_{x_k, s_{k-1}\in\mathcal{S}'_{k-1}} \mu_{\mathrm{f}}(s_{k-1})\, p(x_k, y_k, s_k|s_{k-1}), \tag{13}$$

the sum of the final state metrics will be a lower bound on $p(y^n)$ and the corresponding estimate of $h(Y)$ will be increased. We have proved:

Theorem 1. Omitting states from the computation (5) yields an upper bound on $h(Y)$. □

The sets $\mathcal{S}'_k$ may be chosen arbitrarily. An obvious strategy is to keep only a fixed number of states with the largest metrics.

By a similar argument, one may obtain

Theorem 2. Merging states in the computation (5) yields a lower bound on $h(Y)$. □

So far, however, only the upper bound has proved useful.

The upper bound of Theorem 1 can also be applied to non-finite state channels as follows. Consider, e.g., the autoregressive channel of Fig. 2 and assume that, at time zero, the channel is in some fixed initial state. At time one, there will be two states; at time two, there will be four states, etc. We track all these states according to (5) until there are too many of them, and then we switch to the reduced-state recursion (13).

5 Numerical Examples

We consider binary-input linear intersymbol interference channels with

$$Y_k = \sum_i g_i X_{k-i} + Z_k, \tag{14}$$

with $X_i \in \{+1, -1\}$, and where $Z = (Z_1, Z_2, \ldots)$ is white Gaussian noise with variance σ^2. The fixed channel coefficients $g_i \in \mathbb{R}$, $i \in \mathbb{Z}$, will be specified by their D transform $G(D) \triangleq \sum_i g_i D^i$, and we will assume

$$\sum_i g_i^2 = 1. \tag{15}$$

The signal-to-noise ratio (SNR) in the plots is defined as $1/\sigma^2$ (i.e., the noise power is normalized with respect to the channel input). The source process $X = (X_1, X_2, \ldots)$ will be a sequence of independent and uniformly distributed (i.u.d.) random variables taking values in $\{+1, -1\}$.

Channel 1: Memory 10 FIR filter with $G(D) = \gamma \sum_{i=0}^{10} \frac{1}{1+(i-5)^2} D^i$ where $\gamma \in \mathbb{R}$ is the scale factor required by (15). Fig. 4 shows the following curves. Bottom: The exact information rate, computed as described in Section 2 (with sampled sequences of length $n = 10^6$). Top: The reduced-state upper bound (RSUB) of Section 4, using the 100 "best" (out of 1024) states. Middle: The reduced-state upper bound applied to the equivalent minimum-phase channel.

The trick behind the middle curve in Fig. 4 is as follows. Let

$$G(D) = \beta \prod_i (1 - \zeta_i D). \tag{16}$$

Assuming that $G(D)$ has no zeros on the unit circle, the equivalent minimum-phase filter is

$$G'(D) = \beta \prod_{i:|\zeta_i|<1} (1 - \zeta_i D) \cdot \prod_{i:|\zeta_i|>1} (D - \overline{\zeta_i}), \tag{17}$$

which has all zeros outside the unit circle. It is easy to see that

$$H(D) \triangleq G'(D)/G(D) \tag{18}$$

$$= \frac{\prod_{i:|\zeta_i|>1} (D - \overline{\zeta_i})}{\prod_{i:|\zeta_i|>1} (1 - \zeta_i D)} \tag{19}$$

is an all-pass filter with a stable inverse. Therefore, replacing $G(D)$ by $G(D)H(D) = G'(D)$ does not change the information rate of the channel.

Minimum-phase polynomials concentrate the signal energy into the leading tap weights [9], which makes the reduced-state bound tighter.

Channel 2: First order IIR filter as in Fig. 2 with $G(D) = \gamma/(1 - \alpha D) = \gamma(1 + \alpha D + \alpha^2 D^2 + \ldots)$, where $\gamma \in \mathbb{R}$ is the scale factor required by (15).

Fig. 5 shows the following curves. Rightmost: The (indistinguishable) upper and lower bounds (AUB and ALB) of Section 3, computed using the finite-state model of Fig. 3 with 512 states, with an optimized uniform quantizer, and with optimized σ'. Very close to the left: The reduced-state upper bound (RSUB) of Section 4 using only 4 (!) states. Leftmost: The memoryless binary-input (BPSK) channel.

Fig. 6 shows information rates vs. the number of trellis states used in the computation (for $\sigma^2 = 1$). Top and bottom: the upper and lower bounds of Section 3 (AUB and ALB). Middle: the reduced-state upper bound (RSUB).

Channel 3: IIR filter of order 6 with

$$G(D) = \gamma/(1.0000 + 0.3642 \cdot D + 0.0842 \cdot D^2 + 0.2316 \cdot D^3 - 0.2842 \cdot D^4 + 0.2084 \cdot D^5 + 0.2000 \cdot D^6).$$

Fig. 7 shows the following curves. Leftmost: BPSK. Middle: Reduced-state upper bound using only 2 (!) states. Rightmost: Reduced-state upper bound using 128 states.

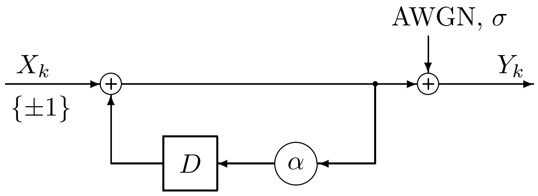

Figure 2: IIR filter channel.

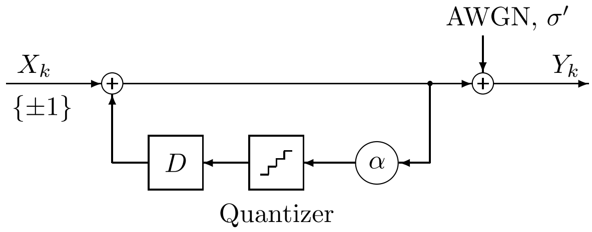

Figure 3: A quantized version of Fig. 2.

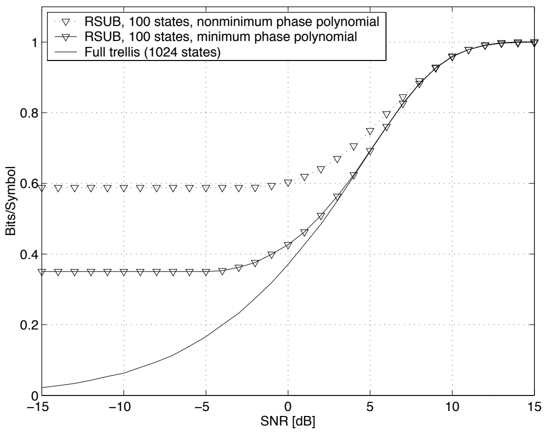

Figure 4: Memory 10 FIR filter.

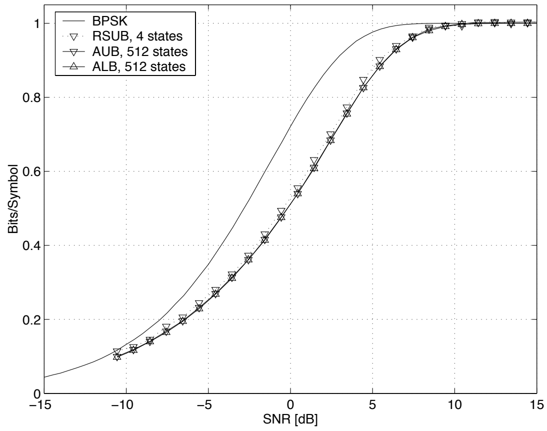

Figure 5: Bounds for Fig. 2 vs. SNR.

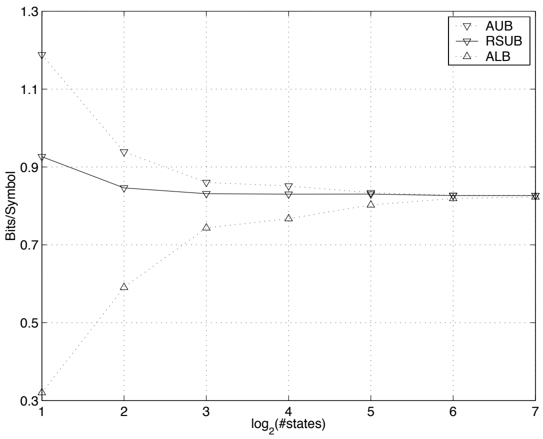

Figure 6: Bounds for Fig. 2 vs. # states.

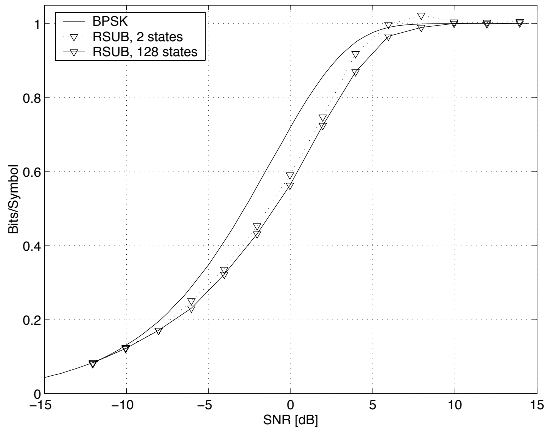

Figure 7: Order 6 IIR filter: upper bounds.

6 Conclusions

It has recently become feasible to compute information rates of finite-state source/channel models with not too many states. By new extensions of such methods, we can compute upper and lower bounds on the information rate of very general non-finite state channels. Bounds from channel approximations and bounds from reduced-state trellis computations can be combined in several ways.

References

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

[2] D. Arnold, H.-A. Loeliger, and P. Vontobel, "Computation of information rates from finite-state source/channel models," *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing,* (Allerton House, Monticello, Illinois), October 2 – October 4, 2002, to appear. Available from `http://www.isi.ee.ethz.ch/~loeliger/`.

[3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Information Theory,* vol. 20, pp. 284–287, March 1974.

[4] A. Barron, "The strong ergodic theorem for densitities: generalized Shannon-McMillan-Breiman theorem," *Annals of Prob.,* vol. 13, no. 4, pp. 1292–1303, 1995.

[5] A. J. Goldsmith and P. P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Information Theory,* vol. 42, pp. 868–886, May 1996.

[6] W. Hirt, *Capacity and Information Rates of Discrete-Time Channels with Memory.* ETH-Diss no. 8671, ETH Zurich, 1988.

[7] A. Kavčić, "On the capacity of Markov sources over noisy channels," *Proc. 2001 IEEE Globecom,* San Antonio, TX, pp. 2997–3001, Nov. 25–29, 2001.

[8] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Information Theory,* vol. 47, pp. 498–519, Feb. 2001.

[9] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing,* 2nd ed. Prentice Hall, 1999.

[10] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite-state ISI channels," *Proc. 2001 IEEE Globecom,* San Antonio, TX, pp. 2992–2996, Nov. 25–29, 2001.

[11] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels", *Proc. 2001 IEEE Int. Symp. Information Theory,* Washington, DC, USA, June 24–29, 2001, p. 283.

[12] Sh. Shamai, L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Information Theory,* vol. 37, pp. 1527–1539, Nov. 1991.

[13] Sh. Shamai and R. Laroia, "The intersymbol interference channel: lower bounds on capacity and channel precoding loss," *IEEE Trans. Information Theory,* vol. 42, pp. 1388–1404, Sept. 1996.

[14] P. O. Vontobel and D. Arnold, "An upper bound on the capacity of channels with memory and constraint input," *Proc. 2001 IEEE Information Theory Workshop,* Cairns, Australia, Sept. 2–7, 2001, pp. 147–149.