AN EFFICIENT ALGORITHM FOR COMPUTING THE ENTROPY OF OUTPUT SEQUENCES FOR BITSHIFT CHANNELS

Stan Baggen and Vladimir B. Balakirsky
Philips Research Laboratories, EIDMA
stan.baggen@philips.com, v.b.balakirsky@tue.nl

We consider data transmission over the so–called "bitshift channel", which models the storage and read–out of data of magnetic and optical recording systems. We consider the case of $(2, \infty)$*–constrained sequences, where a valid input sequence to the channel is a binary sequence having at least two 0's between any 1's. If a "1" is located at position* j *in the input sequence, then it can be received at positions* $j - 1$*,* $j + 1$*, and* j *in the output sequence with probabilities* ε*,* ε*, and* $1 - 2\varepsilon$*, respectively. We obtain tight lower and upper bounds on the entropy of output sequences and on the mutual information between input and output sequences when the input sequence is generated as follows : select bit 1 with probability* p *and encode it by the subblock* 010*; select bit 0 with probability* $1 - p$ *and encode it by the subblock* 0*. An efficient algorithm for computing the bounds is proposed.*

INTRODUCTION

For hard-limiting recording channels such as magnetic and optical recording, the channel can only be in one out of two states. We call these states "low" and "high", respectively. Since the information for such channels is contained in the transitions from "low" to "high" and vice versa, we might as well consider sequences, where a "1" stands for a transition and a "0" for the absence of a transition. Since transitions cannot be arbitrarily close to each other on such channels because of bandwidth constraints, we usually impose a so–called "d–constraint" on the $(0, 1)$–sequences representing a string of transitions, where d refers to the minimum number of 0's between any two 1's [1]. In the sequel, we consider only $d = 2$ constrained sequences, but our approach is easily generalized to any $d \geq 2$.

A major impairment of the recorded signal during read–out is a shift of the recorded transitions with respect to their intended locations during writing. The bitshift channel [2] is a simplified discrete–time model for such shifts of transitions. A binary $d = 2$ constrained sequence $\mathbf{x} = (x_1, x_2, \ldots)$ is a valid input sequence for a bitshift channel if and only if it can be represented as the result of concatenation of subblocks 0 and 010, i.e., if and only if $\mathbf{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots)$, where $\boldsymbol{x}_j \in \{0, 010\}$ for all $j = 1, 2, \ldots$ A binary sequence $\mathbf{y} = (y_1, y_2, \ldots)$ is a possible output sequence of a bitshift channel when $\mathbf{x}$ was transmitted if and only if $\mathbf{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots)$, where $\boldsymbol{x}_j = 0$ implies $\boldsymbol{y}_j = 0$ and $\boldsymbol{x}_j = 010$ implies $\boldsymbol{y}_j \in \{100, 010, 001\}$. We assume that subblocks of the input sequence are independently transmitted over a channel and that $W_{\boldsymbol{x}}(\boldsymbol{y})$ denotes the probability of receiving subblock $\boldsymbol{y}$ when subblock $\boldsymbol{x}$ was sent. Already for such a simple channel model it turns out to be surprisingly difficult to compute information–theoretic quantities. The channel capacity is unknown since the optimizing input distribution is unknown. Even for some given input distribution, only bounds on the mutual information are obtained [2], [3], [4]. Our paper presents bounds on relevant information–theoretic quantities for the case that the binary input is Markovian, i.e., the input in terms of subblocks is memoryless. We first derive the easily computable quantities like the entropy of the input sequences and the conditional entropy of the output sequences given the input sequence. Next we present bounds on the entropy of output sequences that can be easily computed within the desired accuracy; these results also give bounds on the mutual information between input and output sequences.

EASILY COMPUTABLE INFORMATION–THEORETIC QUANTITIES

We assume that subblocks of the input sequence are generated by a Bernoulli source and denote the probability of selecting subblock $\boldsymbol{x}$ at some time instant by $P(\boldsymbol{x})$. Let

$$\begin{aligned} P(\boldsymbol{x}) &\triangleq (1-p)\chi\{\boldsymbol{x} = 0\} + p\chi\{\boldsymbol{x} = 010\}, \\ W_{\boldsymbol{x}}(\boldsymbol{y}) &\triangleq \chi\{\boldsymbol{x} = \boldsymbol{y} = 0\} \\ &\quad + (1-2\varepsilon)\chi\{\boldsymbol{x} = \boldsymbol{y} = 010\} + \varepsilon\chi\{\boldsymbol{x} = 010, \boldsymbol{y} \in \{100, 001\}\}, \end{aligned}$$

where χ stands for the indicator function and the parameters $p \in [0, 1]$, $\varepsilon \in [0, 1/2]$ are given. The entropy of input sequences and the conditional entropy output sequences given a valid input sequence are easily computed. Namely,

$$H_{\text{in}} \triangleq \lim_{t \to \infty} \frac{H(X^t)}{t} = \frac{1}{1+2p} h(p)$$

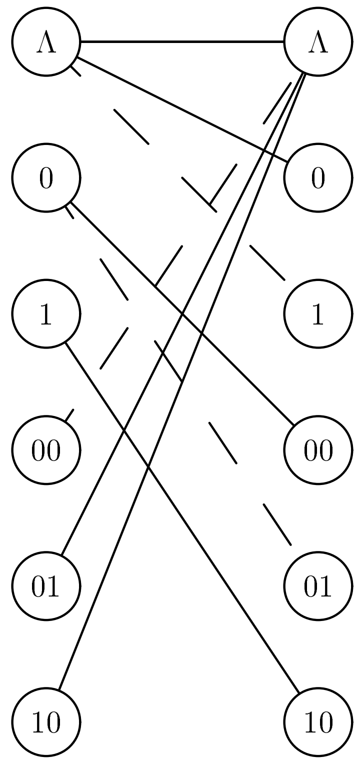
$\pi_s(s';y)$	s $\quad$ s'
$\pi_\Lambda(\Lambda;0) = 1-p$ $\pi_\Lambda(0;0) = p(1-\varepsilon)$ $\pi_\Lambda(1;1) = p\varepsilon$ $\pi_0(00;0) = \varepsilon/(1-\varepsilon)$ $\pi_0(01;1) = 1-\varepsilon/(1-\varepsilon)$ $\pi_1(10;0) = 1$ $\pi_{00}(\Lambda;1) = 1$ $\pi_{01}(\Lambda;0) = 1$ $\pi_{10}(\Lambda;0) = 1$	

Figure 1: Transitions of the machine.

and

$$\lim_{t\to\infty} \frac{H(Y^t|X^t)}{t} = \frac{p}{1+2p}\Big(2\varepsilon + h(2\varepsilon)\Big),$$

where $h(z) \triangleq -z\log z - (1-z)\log(1-z)$, $z \in [0,1]$, stands for the binary entropy function. Note that $2(-\varepsilon\log\varepsilon) - (1-2\varepsilon)\log(1-2\varepsilon) = 2\varepsilon + h(2\varepsilon)$ corresponds to "the jitter entropy". As we are interested in potential limits on the communication over the bitshift channel, we want to know the estimates of the mutual information

$$\frac{1}{t}I(X^t;Y^t) = \frac{1}{t}(H(Y^t) - H(Y^t|X^t)) = \frac{1}{t}(H(X^t) - H(X^t|Y^t)), \quad t = 1,2,\ldots$$

The computation of the entropies $H(Y^t)$, and $H(X^t|Y^t)$, $t = 1,2,\ldots$, is a difficult problem and we develop an algorithm for estimating them in the next section.

All valid input sequences and all possible output sequences can be generated as follows. Introduce a machine having the set of states $\mathcal{S} = \Big\{ \Lambda, 0, 1, 00, 01, 10 \Big\}$ whose transition from the state s to the state s' is associated with the conditional probability $\pi_s(s'; y)$, which is defined as the probability of transition $s \to s'$ and sending $y \in \{0,1\}$ to the output. The probability of generating any fixed sequence of states and corresponding output sequence is defined as the product of conditional probabilities associated with the transitions. All positive conditional probabilities (when $p, \varepsilon > 0$) are given in Figure 1, where transitions with the output symbol 1 are shown by dashed edges. One can easily check that if the machine starts in state Λ and $\varepsilon = 0$, then each $\mathbf{y}_t$ is the prefix of some valid input sequence to the channel, and if $\varepsilon > 0$, then $\mathbf{y}_t$ is the prefix of some possible output sequence. One can also see that

$$\Big(Q_\Lambda, Q_0, Q_1, Q_{00}, Q_{01}, Q_{10} \Big) = \frac{1}{1+2p} \Big(1,\ p(1-\varepsilon),\ p\varepsilon,\ p\varepsilon,\ p(1-2\varepsilon),\ p\varepsilon \Big)$$

is the stationary probability distribution on the states of the machine.

The basic idea of introducing the machine can be explained as follows. There can be many ways of partitioning a given output sequence $\mathbf{y}$ in subbblocks, since a concatenation of elements of $\{0, 100, 010, 001\}$ is not a uniquely decodable code. Each partitioning is implemented in such a way that we move a pointer from left to the right by 1 position at each time instant and insert a special symbol "," as soon as the current subblock is complete. The state of our machine is the vector that is observed between the last "," and the current position of the output sequence; if a "," is put at the current position, then we assume that the state is Λ. This procedure is illustrated below by 3 possible ways of partitioning of the prefix vector $\mathbf{y}_4 = 0100$ of some valid output sequence, where $\tilde{\mathbf{y}}_4$ denotes a vector obtained by inserting symbols "," in the vector $\mathbf{y}_4$ and $\mathbf{s}_4$ is the corresponding vector of states :

$\tilde{\mathbf{y}}_4 =$	0,	1	0	0,	0	1	0,	0,	0	1	0,	0
$\mathbf{s}_4 =$	Λ	1	10	Λ	0	01	Λ	Λ	0	01	Λ	0

All possible states that can be observed form the set $\mathcal{S}$. In the general case, generating of possible output sequences is represented as passing through some path in the trellis constructed as a result of repetition of the section of the transition

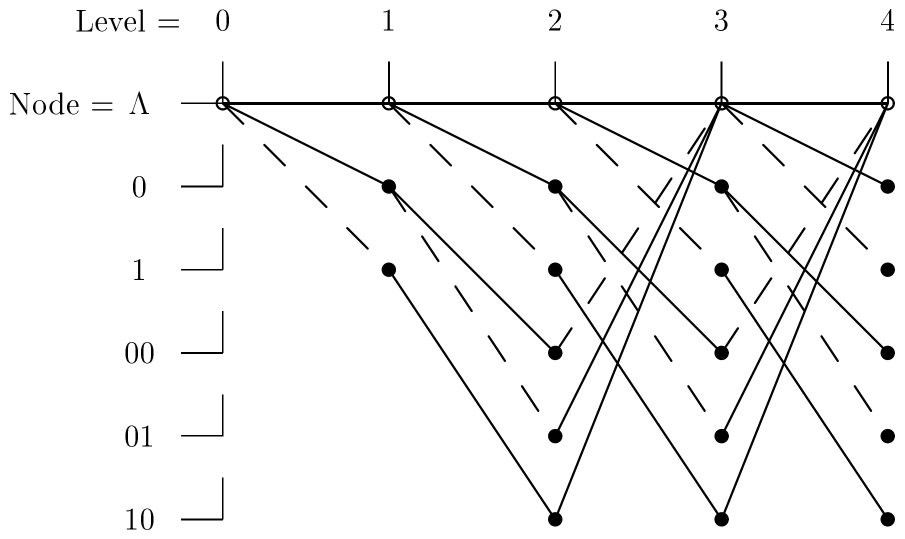


Figure 2: Fragment of the trellis for generating all valid input sequences and all possible output sequences of a bitshift channel.

diagram in Figure 1. The first 4 sections of the trellis and paths corresponding to the output vector 0100 are given in Figures 2, 3, respectively.

Proposition.

$$\begin{aligned}\lim_{t\to\infty}\frac{H(X^t|Y^t)}{t} &= \lim_{t\to\infty}\frac{1}{t}\sum_{\tau=1}^{t} H(X_\tau|X^{\tau-1},Y^t)\\ &= Q_\Lambda \lim_{\ell\to\infty}\sum_{\mathbf{y}_\ell\in\{0,1\}^\ell}\Theta(\mathbf{y}_\ell)H(S_1|S_0=\Lambda,\mathbf{y}_\ell), \qquad (1)\end{aligned}$$

where

$$\Theta(\mathbf{y}_\ell) \triangleq \sum_{s_1,\ldots,s_\ell\in\mathcal{S}}\prod_{j=1}^{\ell}\pi_{s_{j-1}}(s_j;y_j), \quad s_0=\Lambda,$$

is the probability of observing the prefix $\mathbf{y}_\ell$ *of a valid sequence at the output of the bitshift channel.*

If one computes the sum on the right–hand side of (1) taken on all $\mathbf{y}_{\ell_{\max}}$ for some $\ell_{\max}<\infty$, then it would require the inspection of $2^{\ell_{\max}}$ output vectors, and the algorithm can be implemented only for sufficiently small $\ell_{\max}$. Furthermore, we can only obtain an upper bound on $\lim_{t\to\infty} H(X^t|Y^t)/t$ in this way.

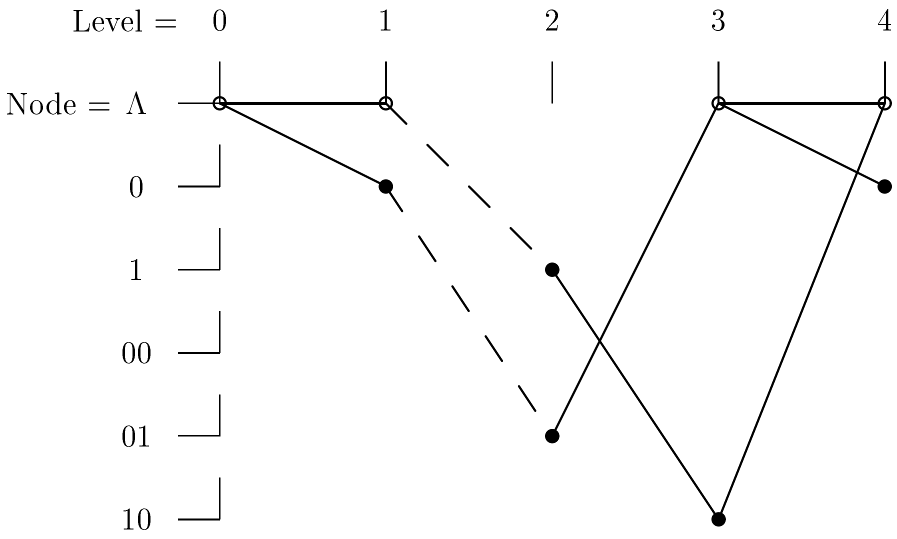


Figure 3: Paths of the trellis corresponding to the output vector $\mathbf{y}_4 = 0100$.

Theorem. *Suppose that the integer $\ell_{\max}$ and the sets $\mathcal{Y}^\ell \subseteq \{0,1\}^\ell$, $\ell = \overline{1, \ell_{\max}}$, are fixed in such a way that their union is a complete prefix subset of the set of sequences that can be generated by the machine in Figure 1, which starts in state Λ, i.e., for all $t \geq \ell_{\max}$ and $\mathbf{y}_t \in \{0,1\}^t$ with $\Theta(\mathbf{y}_t) > 0$, there exists a unique $\ell \in \{1, \ldots, \ell_{\max}\}$ such that $\mathbf{y}_\ell \in \mathcal{Y}^\ell$ is a prefix of $\mathbf{y}_t$ and*

$$\sum_{\ell=1}^{\ell_{\max}} \sum_{\mathbf{y}_\ell \in \mathcal{Y}^\ell} \Theta(\mathbf{y}_\ell) = 1.$$

Then

$$Q_\Lambda \sum_{\ell=1}^{\ell_{\max}} \sum_{\mathbf{y}_\ell \in \mathcal{Y}^\ell} \Theta(\mathbf{y}_\ell) H^{(\mathrm{L})}(\mathbf{y}_\ell) \leq \lim_{t \to \infty} \frac{H(X^t | Y^t)}{t} \leq Q_\Lambda \sum_{\ell=1}^{\ell_{\max}} \sum_{\mathbf{y}_\ell \in \mathcal{Y}^\ell} \Theta(\mathbf{y}_\ell) H^{(\mathrm{U})}(\mathbf{y}_\ell),$$

where

$$\begin{aligned} H^{(\mathrm{U})}(\mathbf{y}_\ell) &\triangleq H(S_1 | S_0 = \Lambda, \mathbf{y}_\ell), \\ H^{(\mathrm{L})}(\mathbf{y}_\ell) &\triangleq H^{(\mathrm{U})}(\mathbf{y}_\ell) - I(S_1; S_\ell | S_0 = \Lambda, \mathbf{y}_\ell) = H(S_1 | S_0 = \Lambda, S_\ell, \mathbf{y}_\ell). \end{aligned}$$

Table 1: The values obtained from the bounds on the entropy of output sequences for $p = 1/2$, $\varepsilon = 0.1$, where $K_{\ell_{\max}} = \sum_{\ell=1}^{\ell_{\max}} |\mathcal{Y}^\ell|$ and $H^{(\mathrm{L})}_{\ell_{\max}}, H^{(\mathrm{U})}_{\ell_{\max}}$ are values of the lower and upper bounds on $\lim_{t\to\infty} H(Y^t)/t$.

$\ell_{\max}$	γ	$H^{(\mathrm{L})}_{\ell_{\max}}$	$H^{(\mathrm{U})}_{\ell_{\max}}$	$K_{\ell_{\max}}$
16	10^{-2}	.601215	.601545	131
	10^{-3}	.601256	.601323	191
	10^{-4}	.601258	.601310	223
24	10^{-2}	.601231	.601535	414
	10^{-3}	.601282	.601303	991
	10^{-4}	.601285	.601288	1358
32	10^{-2}	.601232	.601535	959
	10^{-3}	.601282	.601302	4233
	10^{-4}	.601285	.601286	6945

The sets $\mathcal{Y}^\ell$, $\ell = \overline{1, \ell_{\max}}$, are recursively chosen as follows. Fix a parameter $\gamma \in (0,1)$ and set

$$\begin{aligned} \mathcal{Y}^\ell &= \Big\{ \mathbf{y}_\ell : \mathbf{y}_\ell \text{ has no prefix in } \mathcal{Y}^1, \ldots, \mathcal{Y}^{\ell-1};\ H^{(\mathrm{L})}(\mathbf{y}_\ell) \geq (1-\gamma) H^{(\mathrm{U})}(\mathbf{y}_\ell) \Big\}, \\ \mathcal{Y}^{\ell_{\max}} &= \Big\{ \mathbf{y}_{\ell_{\max}} : \mathbf{y}_{\ell_{\max}} \text{ has no prefix in } \mathcal{Y}^1, \ldots, \mathcal{Y}^{\ell_{\max}-1} \Big\}, \end{aligned}$$

where $\ell = \overline{1, \ell_{\max} - 1}$.

The numerical illustration in Table 1 shows that this approach can be very efficient. The values of the functions

$$\begin{aligned} H_{\mathrm{out}} &\triangleq \lim_{t\to\infty} \frac{H(Y^t)}{t}, \\ I_{\mathrm{in;out}} &\triangleq \lim_{t\to\infty} \frac{I(X^t; Y^t)}{t} \end{aligned}$$

for $p \in \{0.317, 1/2\}$ are given in Figure 4. Notice that $p = 0.317\ldots$ maximizes the entropy of input sequences to the channel, $h(p)/(1+2p)$.

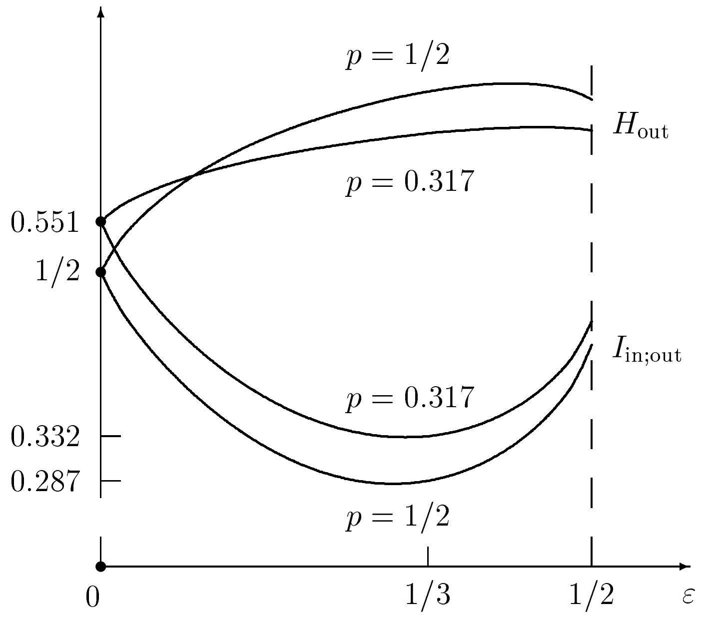


Figure 4: The values of $I_{\mathrm{in;out}}$ and H_{out} as functions of ε for $p \in \{0.317, 1/2\}$.

REFERENCES

[1] P. H. Siegel, "Recording codes for digital magnetic storage," *IEEE Trans. Magnetics*, vol. 21, no. 5, pp.1344–1349, 1985.

[2] S. Shamai (Shitz) and E. Zehavi, "Bounds on the capacity of the bit–shift magnetic recording channel," *IEEE Trans. Inform. Theory*, vol. 37, no. 3, pp. 863–872, 1991.

[3] C. P. M. J. Baggen, *An Information Theoretic Approach to Timing Jitter.* Ph.D. Thesis, UCSD, 1993.

[4] C. P. M. J. Baggen and J. K. Wolf, "On the influence of timing jitter on the capacity of hard–limiting channels," *Proc. URSI Int. Symp. on Signals, Systems and Electronics (ISSSE'92)*, Paris, France, pp. 799–804, 1992.