# Uncoded Transmission of Markov Sources over Noisy Channels

Jossy Sayir*
Vienna Telecommunications Research Center (FTW)
*jossy@ftw.at*

The converse to the coding theorem teaches us that reliable communication is not possible at rates above channel capacity. At rates below capacity, it is possible to achieve any positive probability of error by applying block coding to the source output sequence. However, block coding is not necessarily the only way to achieve reliable transmission. For example, if our source is a Markov source whose entropy rate lies below the capacity of the channel, what reliability can be achieved by transmitting the output of the source uncoded over the channel? Are certain Markov sources better than others for uncoded transmission over noisy channels? Before we proceed to seek an answer to these questions, I will start with a short story that is vaguely related to the subject of this paper. The reader should be warned that the story is only partially true.

One sunny afternoon in early 1998, Jim Massey called his assistants into his office at the ETH Zurich, where he had been busy for several days packing his belongings. They were to be shipped to his new home in Copenhagen. His office had been a marvelous treasury of books, notes, pipes and other paraphernalia. It was now mostly empty, but for a large cardboard box in its middle. The box, we were told, contained all the books and notes he had decided to part with. If we found anything in it that we wanted, we were encouraged to help ourselves. I took a few books that I would probably never use. After digging deeper into the rubble, I was puzzled to find a file containing typewritten notes. The pages were yellowing and many of the notes had faded with time. The title, however, was clearly legible. It read: "Stochastic Signals and Information Theory, by James L. Massey, 1963." There were at least 200 pages, adding up to what was clearly a nearly completed book. Excited by this unexpected find, I asked Jim why he had decided to throw away this early masterpiece. In his parting lecture at the ETH, he had announced that he intended to write a book in the coming years. What better way to start writing a book than to take this one up where he had left it 35 years earlier? Jim sighed sadly and answered that the notes were unusable because too much of the text had been erased by the strains of time and pipe smoke. Figure 1 reproduces a paragraph out of Jim's book.

```
  _artley's p__n_er__g work se_m_ _o ___e _ad __r_ littl_ ____ct. He
is _uc_ _ore r_me__ere_ f_r _is electron__ os_illa_o_ _h_n f_r his
mea_ure _f _nform_____. The _n_y trace _f this la__er contri_u_io_
li_s in __e fact tha_ _n_o_mat__n th_ori__s ha__ a__e__ to call
Sha_no_'s uni_ of in_o_m___on ``th_ H___ley'' _hen the _a__ 10 is __ed
for th_ logar_thms. _his _s _ que_tionable __nor sin_e no one _s
l_kel_ t_ use that b_se an_, _ore__e_, _t __ i_approp__ate _ecau_e
_ar_le_ _learly _e_ogn_zed the arbit__riness i___lve_ i_ _he ch_ice of
t_e bas_ used wit_ _n_or_at__n measu_es. S_c _r_nsit gl_ria ___di.
```

Figure 1: A paragraph out of Jim's book

[0]A shorter version of this paper will be presented at the 2nd Asian-European Workshop on "Concepts in Information Theory" in Breisach, Germany in June 2002.

A coding expert would deplore the fact that Jim did not apply block coding to the text of his book when he wrote it. We could use the added redundancy to recover the missing text. But is coding really necessary? Couldn't we use the natural redundancy of the English language to recover the text? Is written English a good code for transmission over the erasure channel of time and pipe smoke? Would it have been preferable if Jim had written his book in German, French, or Japanese?

Uncoded transmission is not the only motivation for considering the properties of Markov sources relating to the transmission over noisy channels. In [1], the use of a modified arithmetic encoder was considered as a joint source and channel encoder for transmission over noisy channels. This encoder is closely related to Han & Hoshi's interval algorithm [2] for random number generation. Taking this concept further, the output of a stationary source can be encoded into a sequence whose joint probability distribution approaches that of a Markov source with any desired parameters. How should these parameters be chosen if the resulting output sequence is to be transmitted over a noisy channel?

The third and final motivation for considering the transmission of Markov source outputs over noisy channels is that we hope to learn something new about convolutional encoders. A convolutional encoder can be modeled as a Markov source, albeit a non-stationary one. Although we will restrict ourselves to stationary Markov sources, our model can be made to approach the model of a convolutional code, while remaining stationary.

In the next section, we will derive some information-theoretic measures that can be used to analyze the transmission of stationary sources over noisy channels. In the following section, an algorithm to estimate one of these measures in a simulation environment is used to obtain experimental data on the transmission of Markov source outputs over noisy channels. In the last section, the binary erasure channel is considered as a special case. For this channel, the information-theoretic measures of interest can be derived analytically and bounded, providing a partial explanation of the properties observed in our experiment.

# 1 Equivocation and Equivocation Rate

Let the input and output sequences of a noisy channel be designated as shown in Figure 2. How

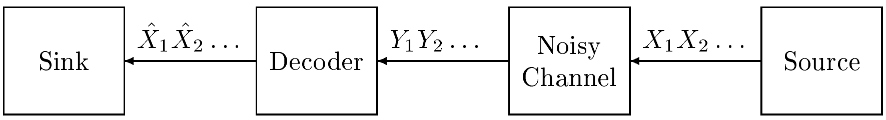


Figure 2: Uncoded Transmission of a Markov Source

are we to assess whether a given source is suitable for a given noisy channel? We choose to assess sources based on the *equivocation* $\mathcal{E}_N$, defined as the entropy of a block of $N$ channel input symbols given the corresponding $N$ output symbols, i.e., $\mathcal{E}_N = H(X_1 \ldots X_N | Y_1 \ldots Y_N)$. Some may object that a better criterion for our assessment would be the probability of error of a maximum likelihood decoder applied to the channel output sequence. We claim that the equivocation is a better criterion, because it is related not only to the probability of error, but also to the expected number of guesses of a list decoder. Figure 3 shows the region of possible equivocation – probability of error points, bounded on one side by Fano's inequality and on the other side by a bound given in [3]. The figure implies that the equivocation tends to zero if and only if the probability of error tends to zero.

We consider only discrete stationary sources and discrete memoryless channels without feedback. For those, we can write

$$\begin{aligned}
\mathcal{E}_N &= H(X_1 \ldots X_N) + H(Y_1 \ldots Y_N | X_1 \ldots X_N) - H(Y_1 \ldots Y_N) \\
&= H(X_1 \ldots X_N) - H(Y_1 \ldots Y_N) + H(Y_1 | X_1 \ldots X_N) + H(Y_2 | X_1 \ldots X_N Y_1) + \ldots
\end{aligned}$$

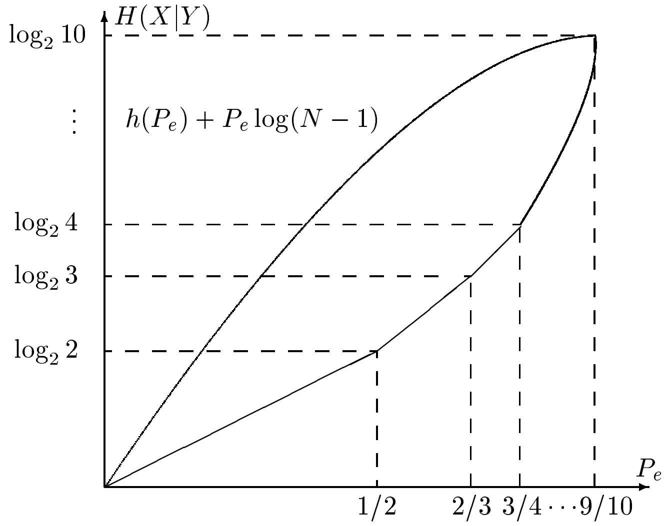


Figure 3: Fano's inequality and its counterpart for a decimal alphabet

$$= \quad H(X_1 \ldots X_N) - H(Y_1 \ldots Y_N) + NH(Y_1|X_1). \tag{1}$$

A necessary and sufficient condition for a source to achieve arbitrarily reliable communication over a noisy channel is for the equivocation $\mathcal{E}_N$ to vanish as $N$ goes to infinity.

We define the normalized equivocation $\overline{\mathcal{E}}_N \stackrel{\text{def}}{=} \frac{1}{N}\mathcal{E}_N$ and the *equivocation rate* $\overline{\mathcal{E}}_\infty \stackrel{\text{def}}{=} \lim_{N\to\infty} \overline{\mathcal{E}}_N$. For the equivocation rate, we can develop (1) to obtain[1]

$$\overline{\mathcal{E}}_\infty = H_\infty(X) - I(X_1; Y_1) + H(Y_1) - H_\infty(Y). \tag{2}$$

A necessary (but not sufficient) condition for arbitrarily reliable transmission over a noisy channel is for the equivocation rate $\overline{\mathcal{E}}_\infty$ to be equal to zero. Setting $\overline{\mathcal{E}}_\infty$ to zero yields

$$I(X_1; Y_1) - H_\infty(X) = H(Y_1) - H_\infty(Y) \tag{3}$$

Equation 3 has interesting implications:

- $H_\infty(X)$ is the information rate of the sequence at the input of the channel. It is equivalent to the rate $R$, commonly used when coding is applied;
- the right-hand side of (3) must necessarily be positive. Therefore, (3) cannot be satisfied when the left-hand side is negative, i.e., when $H_\infty(X) = R > I(X_1; Y_1)$. In particular, when $H_\infty(X) = R > \max_{P_x} I(X_1; Y_1) = C$, arbitrarily reliable communication can never be achieved, which is in line with the converse to the coding theorem;
- when $H_\infty(X) = R = C$, then the source must have a capacity-achieving marginal distribution such that $I(X_1; Y_1) = C$. We obtain the condition $H(Y_1) = H_\infty(Y)$, which is equivalent to demanding that $Y_1, Y_2, \ldots$ be a sequence of *independent and identically distributed* random variables.

## 2 Measuring the Equivocation Rate

Equation 2 can be used to estimate the equivocation rate. $H_\infty(X)$ is easy to determine when the source model is known. When no model is known, it can be estimated using universal source coding.

[1] $H_\infty(X)$ denotes the entropy rate of the sequence $X_1, X_2, \ldots$, defined as $H_\infty(X) \stackrel{\text{def}}{=} \lim_{N\to\infty} \frac{1}{N} H(X_1 \ldots X_N)$.

$H(Y_1)$ and $I(X_1;Y_1)$ depend only on the marginal distribution of the source and are thus easy to determine. The difficulty in estimating the equivocation rate lies in estimating $H_\infty(Y)$. For Markov sources, [4] suggests the use of a simplified forward-backward algorithm to estimate the entropy rate at the output of the channel. The algorithm works by transmitting a source sequence over the channel and computing the probability of the channel output sequence. When the sequence length tends to inifinity, by the asymptotic equi-partition property, the negative logarithm of its probability will tend towards the entropy rate with probability one. We implemented this algorithm and used it to estimate the equivocation rate for various Markov sources and various channels.

The Markov sources in our experiment were binary finite-memory unifilar sources. Their state can be thought of as a binary $\mu$-tuple corresponding the last $\mu$ symbols emitted by the source, where $\mu$ is called the *memory* of the source. Every state has at most two transitions leading out of it, corresponding to a zero or a one being emitted. The next state is constructed by left-shifting the old state and adding the most recent output symbol to its right (as its least significant digit.) Let $T$ be the transition matrix of the Markov source, whose $j$-th column contains the vector of output probabilities from the $j$-th state of the source. The stationary state distribution $P_S$ of the Markov source can be determined by solving the matrix equation $P_S = TP_S$. The solution is the right eigenvector of $T$ corresponding to the eigenvalue 1. In our experiment, we used only Markov sources with doubly stochastic transition matrices, i.e., where the sum of the transition probabilities entering any state is equal to one (the sum of the transition probabilities leaving a state are equal to one by definition, as they form a conditional probability distribution.) It is easy to see that for a doubly stochastic matrix, the all-one vector is a right eigenvector corresponding to the eigenvalue 1. Therefore, the stationary state distribution $P_S$ of a Markov source with a doubly stochastic transition matrix is the uniform distribution, where $P_S(s) = 1/2^\mu$ for each state $s$.

We identify our Markov sources with a "transition vector" $t = [t_0, t_1, \ldots, t_{2^\mu - 1}]$ that specifies the probability of emitting a zero in each state. The transition matrix $T$ is fully determined by the vector $t$. The condition that $T$ be doubly stochastic can be translated in a condition for the vector $t$ that

$$t_k = 1 - t_{2^{\mu-1}+k} \text{ , for } k = 0 \ldots 2^{\mu-1} - 1,$$

where the states are numbered so that the $k$-th state is the binary representation of the number $k$. In other words, only the first half of the vector $t$ can be chosen at will. The second half must be equal to one minus the first half. Figure 4 shows the Markov source for $t = [.8, .4, .2, .6]$. The entropy rate of a Markov source is computed by averaging the output entropies of the states,

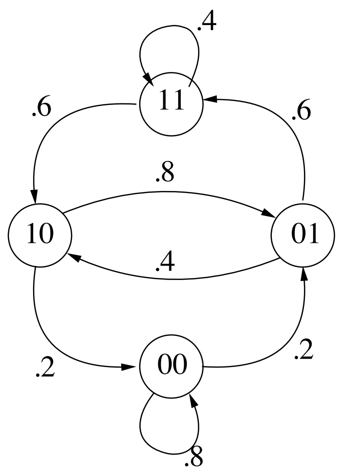


Figure 4: A Markov source with $t = [.8, .4, .2, .6]$

weighted by their stationary probabilities. Since our stationary probability is uniform, the entropy rate $H_\infty(X)$ of the source becomes

$$H_\infty(X) = \frac{1}{2^\mu} \sum_{i=0}^{2^\mu} h(t_i) = \frac{1}{2^{\mu-1}} \sum_{i=0}^{2^{\mu-1}} h(t_i),$$

where $h(.)$ is the binary entropy function. We call $r = [h(t_0), \ldots, h(t_{2^{\mu-1}-1})]$ the *rate vector* of the source.

We measured the equivocation rate for a great number of sources, transmitted over binary symmetric and binary erasure channels. All the sources measured had entropy rate 1/2. Space constraints prevent us from including the graphs of our measurements. The experimental observations resulting from our measurements can be summarized as follows:

- for a four-state source, the rate vector being $r = [r_0, 1 - r_0]$ to achieve an entropy rate of 1/2, the equivocation rate decreases steadily when $r_0$ approaches 0 or 1. The equivocation is maximized for $r_0 = 1/2$;
- we simulated $2^\mu$-state sources where the components of the rate vector were uniformly distributed between 0 and 1, so that their average is 1/2. The equivocation for this type of sources does not decrease visibly with growing memory $\mu$, beyond $\mu = 2$;
- the equivocation of $2^\mu$-state sources decreases drastically when the elements of the rate vector are chosen close to 0 and 1, so that their average remains 1/2. For this type of sources, the equivocation also decreases with growing memory $\mu$.

The results of our experiment can be interpreted as follows: it appears that general Markov sources with "randomly chosen" rate vectors are badly suited for transmission over noisy channels. Indeed, the equivocation for such sources is not improved by an increase in the order of the source, although the complexity of a maximum likelihood decoder increases exponentially with that order. Markov sources suitable for transmission over noisy channels appear to be those that have many "deterministic" states (states with rates equal or close to 0), compensated by many "random" states (states with rates equal or close to 1) to keep the average rate equal to 1/2. Since a rate 1/2 convolutional encoder can be seen as a Markov source where every odd code digit is produced by a random state and every even code digit is produced by a deterministic state, this means that good Markov sources for transmission over noisy channels appear to have output statistics similar to the output statistics of a convolutional encoder.

An experiment to model English, French and German text as Markov sources to determine which language is more suitable for transmission over erasure channels has not been completed at the time of writing. Readers are invited to place bets on their favorite language. If the paper is accepted, I will try to complete the experiment by the time the final paper is due.

# 3 Block Equivocation for the Binary Erasure Channel

For the binary erasure channel with erasure probability $\delta$, there is an analytical expression for the entropy rate at the output of the channel. Equation 1 simplifies to

$$\begin{aligned} \mathcal{E}_N \;=\; & H(X_1 \ldots X_N) - (1-\delta)^N H(X_1 \ldots X_N) - \\ & \delta(1-\delta)^{N-1}[H(\neg X_1) + H(\neg X_2) + \ldots + H(\neg X_N)] - \\ & \delta^2(1-\delta)^{N-2}[H(\neg X_1 X_2) + H(\neg X_1 X_3) + \ldots + H(\neg X_{N-1} X_N)] - \\ & \ldots - \\ & \delta^{N-2}(1-\delta)^2[H(X_1 X_2) + H(X_1 X_3) + \ldots + H(X_{N-1} X_N)] - \\ & \delta^{N-1}(1-\delta)[H(X_1) + H(X_2) + \ldots + H(X_N)], \end{aligned} \tag{4}$$

where $H(\neg X_i)$ is taken to mean the entropy of the block $X_1 \dots X_N$ with the $i$-th letter missing, $H(\neg X_i X_j)$ means the entropy of the block with the $i$-th and the $j$-th letter missing, etc.

We now wish to simplify (4) for the case when the source sequence $X_1 \dots X_N$ is the output of a finite-memory Markov source. The term $H(X_1 \dots X_N)$ simplifies easily to $H(X_1 \dots X_\mu) + (N-\mu)H_\infty(X)$. For Markov sources whose marginal distribution is uniform, this simplifies to $\mu + (N-\mu)H_\infty(X)$. Since we are interested in the asymptotic behavior of the equivocation for large block sizes, we can safely assume that $N >> \mu$. The difficulty lies in evaluating entropies of blocks with gaps, i.e., symbols missing as in $H(\neg X_7 X_8 X_9)$. We can make two extreme assumptions:

1. after a gap of any length, we have lost all knowledge of the source state. Therefore, we need to apply the marginal state distribution to the first $\mu$ symbols following a gap.

2. after a gap of any length, we still have exact knowledge of the source state. Therefore, the entropy of the symbols following the gap is the entropy rate $H_\infty$ of the source.

The truth lies between these two extremes: for short gaps, we still have some knowledge of the source state following the gap, whereas for long gaps, we loose that knowledge for most sources.

Using these extreme assumptions, it is possible to obtain a lower and an upper bound on the equivocation. This requires a lot of juggling with combinatorial problems and we have not been able to solve it for the general case. For the special case of a binary erasure channel with $\delta = 1/2$, the upper and lower bounds are

$$-\frac{N(2^\mu - 1) + \mu - 2^{\mu+1}}{2^{\mu+1}} + H_\infty(X)\frac{N(2^{\mu+1} - 1) + \mu - 2^{\mu+1}}{2^{\mu+1}} \le \mathcal{E}_N \le \frac{NH_\infty(X)}{2} + \frac{\mu}{2^N}(1 - H_\infty) \quad (5)$$

The capacity of this channel is $1/2$. Note that the lower bound is positive for $H_\infty(X) \ge (2^\mu - 1)/(2^{\mu+1} - 1)$ which lies slightly below the capacity of the channel and only approaches capacity as the memory $\mu$ goes to infinity. In other words, it is *not possible* to achieve arbitrarily reliable communications at rates up to capacity for this channel using *any finite memory system*, even if the block length $N$ is allowed to grow to infinity.

Neither the lower nor the upper bound in (5) are very satisfactory in terms of giving useful indications for the performance of real sources: the lower bound is zero for most of the interesting region and the upper bound is roughly half the block length times the rate. However, understanding how they were obtained gives us an insight as to why we observed what we did in our experiment of the previous section. If a source is quick to approach its stationary distribution, its performance will be closer to the lower bound, since it is fairer to assume that the state information will be lost after a gap. If, however, a source is *slow* to approach its stationary distribution, its performance will be closer to the upper bound, since it is fairer to assume that the state information is kept throughout a gap. Although intuition would have it otherwise, it appears to be good for a source to "forget" its state quickly as it progresses through an erasure gap.

# References


[1] J. Sayir, "Arithmetic Coding for Noisy Channels," *IEEE Information Theory Workshop*, Kruger Park, South Africa, June 10-25 1999.

[2] T. S. Han & M. Hoshi, "Interval Algorithm for Random Number Generation," *IEEE Trans. on Inform. Theory*, **IT-43**, pp. 599-611, March 1997.

[3] M. Feder & N. Merhav, "Relations Between Entropy and Error Probability," *IEEE Trans. on Inform. Theory*, **IT-40**, pp. 259-266, January 1994.

[4] D. Arnold & H.-A. Loeliger, "On the Information Rate of Binary-Input Channels with Memory," *IEEE Conference on Communications,* Helsinki, Finnland, 2001.