Titre: Title:	Analysis of the sequential decoding metric by Markov Chains
Auteurs: Authors:	David Haccoun
Date:	1977
Type:	Rapport / Report
Référence: Citation:	Haccoun, D. (1977). Analysis of the sequential decoding metric by Markov Chains. (Technical Report n° EP-R-77-48). https://publications.polymtl.ca/6239/

DÉPARTEMENT DE GÉNIE ÉLECTRIQUE
SECTION AUTOMATIQUE, COMMUNICATION, INFORMATIQUE

Rapport Technique EP77-R-48

Classification: Library of Congress no......

ANALYSIS OF THE SEQUENTIAL DECODING METRIC BY MARKOV CHAINS

par

David Haccoun, Ph.D.
Professeur Agrégé

Ecole Polytechnique de Montréal

Campus de l'Université
de Montréal
Case postale 6079
Succursale 'A'
Montréal, Québec
H3C 3A7

ANALYSIS OF THE SEQUENTIAL DECODING METRIC BY MARKOV CHAINS*

DAVID HACCOUN Ph.D.

Professeur Agrégé
Département de Génie Electrique
Ecole Polytechnique de Montréal

RAPPORT TECHNIQUE No. EP 77-R-48

ECOLE POLYTECHNIQUE DE MONTREAL
Décembre 1977

* This research was supported in part by the National Research Council of Canada, grant No. A-9336.

ANALYSIS OF THE SEQUENTIAL DECODING METRIC BY MARKOV CHAINS

by

DAVID HACCOUN

ABSTRACT

A new approach to the analysis of the factors responsible for the computational variability of sequential decoding is presented. The analysis uses Massey's Markov Chain model of the difference between a node on the correct path and its smallest succeeding value. In this model the states of the chain correspond to the possible metric dip values includind zero, and it is shown that the average separation between breakout nodes is equal to the reciprocal of the stationary probability of state zero. The analysis is extended to multiple-path sequential decoding yielding a closed-form expression of the improvement over the single-path sequential decoding of the average separation between non-search nodes. Finally by regarding the probability distribution of the correct path metric dip, Pr (Dip < H) as the probability that starting from state zero the systems returns to that state without visiting any state equal or larger than H, and exact expression of this distribution is derived. All the expressions are given in terms of the set of the correct path branch metric values and their probability assignments. They are easily computable, directly applicable to particular cases, and as shown by specific examples the theoretical results are in excellent agreement with experimental measurements.

I. INTRODUCTION

Sequential decoding is one of the most powerful decoding techniques for convolutionally encoded messages transmitted over a discrete memoryless channel. It is a suboptimum tree search procedure which attemps to find the most likely data sequence or path $\underline{U} = (u_1, u_2, u_3, \ldots)$ through the encoded tree in which the branches are assigned likelihood or branch metric values $\{\gamma\}$. For a tree of length L branches the objective of the decoder is to find the path $\underline{U}$ for which the total accumulated metric

$$\Gamma_L^{(\underline{U})} = \sum_{i=1}^{L} \gamma_i^{(\underline{U})}$$

is the largest [1]. The central idea of sequential decoding is to explore the tree one branch at a time without searching the entire tree. This exploration is performed along the path that appears to be the most likely, that is whose metric is the largest among those examined. With this procedure the decoder must occasionnaly go back in the tree and reverse an earlier decison, but in order to minimize this backing up and extension of unlikely paths, the metric is biased in such a way that on the average it increases along the correct path and decreases along all incorrect paths.

The average number of computations per decoded bit is typically very small, but the main problem with sequential decoding is the computational variability which is asymptotically Pareto distributed [2]. This variability can be reduced by having the decoder explore simultaneously the M, M > 1 most likely paths (instead of the single most likely path), and exploit the trellis structure of the code to eliminate merging redundant paths [3].

A particularly important quantity with sequential decoding is the difference Δ_k between the cumulative metric at a node on the correct path and its smallest succeeding value,

$$\Delta_k = \Gamma_k - \min_{j \geq k} \Gamma_j \tag{1}$$

If for a node $\underline{U}_k$ on the correct path $\underline{U}$ we have

$$\Gamma_k \leq \Gamma_j \quad , \qquad j > k \tag{2}$$

then $\Delta_k = 0$ and this node is called "breakout" [4]. If inequality (2) is strict, then $\underline{U}_k$ is a "strict" breakout node. Breakout nodes are called "nonsearch" since they are decoded by a single computation. However, since all the incorrect paths that the decoder will ever explore emerge from nonbreakout nodes on the correct path, as Δ_k increases, the number of computations necessary to decode $\underline{U}_k$ increases exponentially. Nonbreakout nodes are called "search" nodes. Fig. (1) shows a typical segment of the correct path metric.

Clearly, decoding proceeds smoothly when the decoder moves along consecutive breakout nodes on the correct path. The decoding effort increases and becomes variable only when, because of the noise in the channel, nonbreakout nodes are encountered on the correct path. Each of these nonbreakout nodes becomes the root node of a potential subtree of incorrect paths that must be explored as long as their metric value does not drop below that of the next breakout node. Therefore the decoding effort increases as both the correct path metric drops and the number of consecutive nonbreakout nodes increases. Prime determinants for the computational behaviour of a sequential decoder are thus the distribution of the correct path metric dips and the separation (in number of branches) between consecutive breakout nodes.

In this paper a new approach to the analysis of these two factors responsible for the computational variability of sequential decoding is presented. The approach uses a Markov Chain model of the metric differences Δ_k introduced by Massey et al [5] . In this chain, the states are the possible metric dip values, and decoding is regarded as executing a random walk with a reflecting barrier at the origin. The Markov Chain model is reviewed in section II and several properties of the chain are interpreted with respect to sequential decoding. The Markov Chain model is then used in sections III and IV to derive closed-form expressions for the average separation between nonsearch nodes for the single-path and the multiple-path sequential decoders, and to determine an exact expression for the cumulative distribution of the correct path metric dips. These closed-form expressions are especially interesting because they are easily computable on a computer and also because they are given in terms of the actual set of channel transition probabilities and the actual set of branch metrics used by the decoder. Finally the accuracy of the model was verified by the excellent agreement obtained between the theoretical results and actual measurements performed on computer simulations.

II. MARKOV CHAIN MODEL OF THE CORRECT PATH METRIC DIPS

In this section we present Massey's Markov Chain model [5] for the correct path metric dips and interpret some of its properties with respect to sequential decoding.

Consider some node $\underline{U}_k$ on the correct path $\underline{U}$, and let $\{\gamma_i\}$ be the set of correct branch metric values. A fundamental quantity closely related to the decoding of $\underline{U}_k$ is the metric difference given by Eq(1), repeated below

$$\Delta_k = \Gamma_k - \mathrm{Min}(\Gamma_k, \Gamma_{k+1}, \Gamma_{k+2}, \ldots) \tag{3}$$

Naturally,

$$\Delta_k \geq 0 \quad , \quad k = 1,2,\ldots \tag{4}$$

where the equality holds only for breakout nodes.

Since Γ_k is the cumulative metric over the first k branches, then

$$\Gamma_k = \sum_{i=1}^{k} \gamma_i \tag{5}$$

and Eq (3) becomes

$$\Delta_k = - \mathrm{Min}\,\{0, \gamma_{k+1}, (\gamma_{k+1} + \gamma_{k+2}), \ldots\} \tag{6}$$

which can be expressed recursively by

$$\Delta_k = \begin{cases} (\Delta_{k+1} - \gamma_{k+1}) & , \quad \text{if } (\Delta_{k+1} - \gamma_{k+1}) \geq 0 \\ 0 & \quad \text{otherwise} \end{cases} \tag{7}$$

or

$$\Delta_k = \text{Max}\ \{0,\ (\Delta_{k+1} - \gamma_{k+1})\} \tag{8}$$

Since all the correct path branch metrics γ are statistically independent random variables with the same distribution, then the non-negative random variables $\Delta_{k+1}, \Delta_k, \Delta_{k-1}, \ldots$, induce a queuing process. This process is equivalent to a Markov Chain where the states are the possible values of Δ_k, and where the 1-step transition probabilities are the probabilities associated with the γ's. If J is the largest negative branch metric value and Q the largest positive branch value, then

$$\begin{aligned} p_\ell &= P(\gamma = -\ell)\ , \quad 0 \leq \ell \leq J \\ q_\ell &= P(\gamma = \ell)\ , \quad 0 < \ell \leq Q \end{aligned} \tag{9}$$

This set of probabilities is readily obtained from the channel transition probabilities and the definition of the metric. The states Δ_k are integers since the branch metric values are always rounded to integers in practical sequential decoders.

If a node $\underline{U}_k$ has some value $\Delta_k = j$, then state j of the chain is occupied. The next state to be occupied, say state ℓ, will be determined by the 1-step transition from state j, according to the set of probabilities of Eq.(9), and $\Delta_{k-1} = \ell$. For example, in Fig. (2), the set of the possible branch metrics is { 2, 1, 0, -3, -7} and the corresponding probabilities are $\{q_2, q_1, p_0, p_3, p_7\}$. As shown in Fig. (2), for this particular example all states to the right of state 2 exhibit the same transition pattern.

A breakout node of the correct path locates the decoder at state 0 whereas a nonbreakout node will locate it at the corresponding nonzero state. Clearly decoding the correct path is equivalent to executing an interger-valued random walk where a visit to any nonzero state initiates a search for the correct path. Since from state zero the decoder could either remain in that state or go into a search, state zero acts as a reflecting barrier at the origin of the random walk. For a soft-quantized channel there are in general Q, Q > 1 transitions towards the zero-state and J possible transitions away from the zero state, corresponding respectively to the possible Q positive and J negative branch metrics values. The form of the probability transition matrix for such a random walk with Q = 2 and J = 4 is given below

	0	1	2	3	4	5	6	7	8	9	...
0	$(p_0+q_1+q_2)$	p_1	p_2	p_3	p_4	0	0	0	0	0	..
1	(q_1+q_2)	p_0	p_1	p_2	p_3	p_4	0	0	0	0	..
2	q_2	q_1	p_0	p_1	p_2	p_3	p_4	0	0	0	..
3	0	q_2	q_1	p_0	p_1	p_2	p_3	p_4	0	0	..
4	0	0	q_2	q_1	p_0	p_1	p_2	p_3	p_4	0	..
5	0	0	0	q_2	q_1	p_0	p_1	p_2	p_3	p_4	..
6	0	0	0	0	q_2	q_1	p_0	p_1	p_2	p_3	..
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	..

This Markov chain model for Δ_k was first proposed by Massey et al [5] who considered the particular case in which only a single transition of unit length is permitted toward the origin state, i.e., Q = 1, $P(\gamma_k = +1) = q$, but any finite number J of transitions away from the origin are permitted.

Regardless of the chain, when a node $\underline{U}_k$ on the correct path has a positive metric difference, $\Delta_k > 0$, there is a search associated with this node, and clearly the correct path metric must go through a dip at least equal to Δ_k before reaching the next breakout node. We can therefore associate the Δ's with the correct path metric dips. Of particular importance with this representation is the stationary probability distribution of the different states or dip values, and the average separation between consecutive breakout nodes.

Properties of the Markov Chain

We now interpret some well known properties of Markov Chains [6] with respect to sequential decoding.

Property 1:

The average branch separation between breakout nodes on the correct path is equal to the mean recurrence time of the origin state.

Suppose state j can be reached from state i in n steps. Let T_{ij} be the waiting time (in number of transitions required) for the first entrance to state j from the initial state i. Then

$$f_{ij}^{(n)} = P(T_{ij} = n) \tag{10}$$

is the probability to reach state j for the first time from state i in exactly n steps. The probability of eventually reaching state j in a finite number of steps from state i is

$$f_{ij} = P(T_{ij} < \infty) = \sum_{n=1}^{\infty} f_{ij}^{(n)} \quad , \forall\, i,j \tag{11}$$

In particular T_{ii} is the return time to state i (or recurrence time of i) and if $f_{ii} = 1$, then state i is said to be recurrent or persistent. If $f_{ii} < 1$, state i is said to be transient.

For a recurrent state i, the mean recurrence time (or mean return time) is defined as

$$\mu_i = \sum_{n=1}^{\infty} n\, f_{ii}^{(n)} \tag{12}$$

If $\mu_i = \infty$, state i is called a null state, and if $\mu_i < \infty$ it is called a nonnull state.

For sequential decoding the probability that the first return to state 0 (a breakout node on the correct path) takes n steps is

$$P(T_{00} = n) = f_{00}^{(n)} \tag{13}$$

Hence the average number of steps to return to state 0, or equivalently the average separation d_o (in number of branches) between consecutive breakout nodes of the correct path is the mean recurrence time of state 0,

$$d_o = \mu_o = \sum_{n=1}^{\infty} n \, f_{oo}^{(n)} \tag{14}$$

The average separation d_o is finite only if state 0 is recurrent nonnull. We know that for sequential decoding, the only incorrect paths of interest emerge from nonbreakout nodes on the correct path. Therefore in general both the average value and the variability of the computational effort increase with the average separation between consecutive breakout nodes. A decoding algorithm which effectively reduces the average separation d_o will also reduce the variability of the decoding effort, and if $d_o = 1$ this algorithm will never get into a search for the correct path.

Property 2: If $R < R_{comp}$ for the sequential decoder, then state 0 is recurrent.

R_{comp} being the computational cutoff rate of sequential decoding [2] if $R < R_{comp}$ the average number of computations to decode one branch on the correct path is finite. Therefore the decoder always returns to a breakout node or zero state of the chain. Hence $f_{oo} = 1$ and state 0 is recurrent.

Property 3: The Markov chain is irreducible

There is no absorbing state in our Markov Chain, every state can be reached from every other state and hence the chain is irreducible. Since in any irreducible Markov Chain all state are of the same class (recurrent a transient), if state 0 is recurrent then all other states are recurrent.

III. AVERAGE SEPARATION BETWEEN BREAKOUT NODES

We have shown that the average separation between breakout nodes is equal to the mean recurrence time of state zero. Now, if a chain has a stationary probability distribution $\{v_j\}$ defined by

$$v_j = \sum_{i=0}^{\infty} v_i P_{ij}, \quad j = 0, 1, 2, \ldots \tag{15}$$

then the unconditional distribution of occupancy of the states becomes independent of time: the process is in statistical equilibrium. The stationary is unique when it exists and the v_j are given by

$$v_j = \frac{1}{\mu_j} \tag{16}$$

where μ_j is the mean recurrence time of state j.

If such a distribution exists for the state of the chain, then

$$P(\Delta_k = \ell) = v_\ell, \quad \ell \geq 0 \tag{17}$$

Of special interest is the stationary probability v_0 of occupying state 0, since it represents the relative frequency of breakout nodes on the correct path which a sequential decoder will decode by a single computation. From Eq (14) and (16) it follows that the average separation between breakout nodes is given by

$$d_0 = \frac{1}{v_0} \tag{18}$$

Since an irreducible Markov Chain has a stationary distribution if and only if all states are recurrent nonnull, then it follows from property 2 and 3, that if $R < R_{comp}$, the stationary distribution for our chain exists. The determination of the $\{v_j\}$ according to Eq (15) is in general quite cumbersome. However, for the simple case where there are any finite number J of transitions away from the origin state, but where only a single transition of unit length towards the origin is permitted, Massey et al [5] give a simple recursive technique to determine the stationary probabilities v_i, given they exist. For such a chain, the average drift toward the origin is

$$z = q - \sum_{i=0}^{J} i\, p_i \tag{19}$$

and assuming $z > 0$, v_0 is then given by

$$v_0 = \frac{z}{q} \tag{20}$$

and hence

$$d_0 = \frac{1}{v_0} = \frac{q}{z} \tag{21}$$

Unfortunately for the general case where there are Q, $Q > 1$ positive branch metric values, this "drift balancing" technique cannot be used [5]. However it can be applied for the binary symmetric channel when all metric values are normalized so that the positive branch metric has unit value.

For example consider sequential decoding for rate 1/2 codes over a binary symmetric channel with a cross over probability $\varepsilon = 0.0330$. The rounded off branch metrics are (1, -4, -9) and the corresponding probabilities are {0.9351 , 0.0638 , 0.0011 }. From Eq(20) the stationary probability v_0 is then equal to 0.7165 and the calculated average separation between breakout nodes is equal to 1.3956. Actual measurement of d_0 on a computer simulation of a path of length 200,000 branches using identical branch metrics and identical transition probabilities has given a value of 1.3979 which is in excellent agreement with the theoretical value.

The stationary probability v_0 represents the proportion of the nodes on the correct path that will be decoded by a single computation. Hence a proportion $(1-v_0)$ of the nodes on the correct path are nonbreakout and are thus responsible for the variability of the computational effort of sequential decoding. Eq. (21) is intuitively agreeable since as v_0 increases, d_0 and the proportion of nonbreakout nodes both decrease, decreasing also the variability of the decoding effort. Clearly a procedure which reduces the effective d_0 will consequently reduce the variability of the computation [3] . This reduction of d_0 is investigated next.

MULTIPLE PATH SEQUENTIAL DECODING

Consider a sequential decoder that extends simultaneously the M, M > 1 most likely paths instead of the single most likely path. As a consequence of the multiple-path extension it has been shown [3], [8] that the small dips on the correct path metric are simply ignored by the decoder in the sense that the decoder does not get in a search for the correct path. For these small dips the nonbreakout nodes behave like breakout nodes. They become "pseudo" breakout or, just like the true breakout nodes, "nonsearch" nodes, and consequently, compared to sequential decoding the average separation between the nonsearch nodes is therefore reduced.

Suppose all dips smaller than some value H are ignored by the multiple-path decoder. Since all nonsearch nodes have a state smaller than H, if the decoder starts from state zero and returns to it without ever visiting any state at or beyond H, all states in the path correspond to nonsearch nodes. If ${}_{H}d_{o}$ is the new average separation between the nonsearch nodes, then the ratio

$$\frac{d_o}{{}_{H}d_o}$$

may be taken as a measure of the improvement afforded by the multiple-path decoder over the single-path sequential decoder. Using the Markov Chain model we now proceed to determine an exact expression of this improvement in terms of the set of the branch metric values and their associated probabilities.

Let $\mathcal{H}$ be the set of states smaller than state H, and let the first return probability to state zero in n steps without ever leaving $\mathcal{H}$ be defined as

$$ {}_{\mathcal{H}}f_{oo}^{(n)} = P_r\left[\Delta_n = 0,\ \Delta_j \neq 0,\ \Delta_j \in \mathcal{H},\ \ j = 1,2,\ldots n-1 \mid \Delta_o = 0\right] \tag{23}$$

Likewise define the first entrance probability to state 0 in n steps without leaving H as

$$ {}_{H}f_{io}^{(n)} = P_r\left[\Delta_n = 0,\ \Delta_j \neq 0,\ \Delta_j \ \varepsilon\ H,\ j = 1,2,\ldots n-1 \mid \Delta_o = i\right] \tag{24} $$

The events of returning the state 0 can be classified into 2 disjoint sets depending on whether all states visited "en route" belong to H or not. Therefore writing ${}^{c}f_{oo}^{(n)}$ as the complement to ${}_{H}f_{oo}^{(n)}$, we have

$$ f_{oo}^{(n)} = {}_{H}f_{oo}^{(n)} + {}^{c}f_{oo}^{(n)} \tag{25} $$

where $f_{oo}^{(n)}$ is the first return probability to state 0.

Now suppose there are N true breakout nodes over a very long path, whose average separation between breakout nodes is d_o and average separation between nonsearch nodes is ${}_{H}d_o$. We can write

$$ d_oN = {}_{H}d_oN\left[{}_{H}f_{oo}^{(1)} + \sum_{n=2}^{\infty} n\ {}_{H}f_{oo}^{(n)} + \sum_{n=1}^{\infty} {}^{c}f_{oo}^{(n)}\right] \tag{26} $$

where the terms in the brackets represent respectively, the proportion of consecutive breakout nodes, consecutive "pseudo" breakout nodes, and search nodes. Using Eq(25), Eq(26) becomes

$$ d_o = {}_{H}d_o\left[\sum_{n=1}^{\infty}(n-1)\ {}_{H}f_{oo}^{(n)} + \sum_{n=1}^{\infty} f_{oo}^{(n)}\right] \tag{27} $$

Since state 0 is recurrent, $\sum_{n=1}^{\infty} f_{oo}^{(n)} = 1$ and we finally obtain

$$ \frac{d_o}{{}_{H}d_o} = 1 + \sum_{n=1}^{\infty}(n-1)\ {}_{H}f_{oo}^{(n)} \tag{28} $$

or

$$ \frac{d_o - {}_{H}d_o}{{}_{H}d_o} = \sum_{n=1}^{\infty}(n-1)\ {}_{H}f_{oo}^{(n)} \tag{29} $$

Before proceeding to evaluate the infinite sum, we can check the validity of Eq(28) by making $H = 1$ (sequential decoding) and $H = \infty$ (Viterbi decoding*)

(i) Sequential decoding. Since $H = 1$, clearly ${}_H f_{oo}^{(n)} = 0$ for $n > 1$, and Eq(28) yields $d_o = {}_1 d_o$ as required.

(ii) Viterbi decoding. All nodes are nonsearch nodes. Therefore,

$$ {}_H f_{oo}^{(n)} = f_{oo}^{(n)} \tag{30}$$

and Eq (28) becomes

$$ \frac{d_o}{{}_H d_o} = 1 + \sum_{n=1}^{\infty} n f_{oo}^{(n)} - \sum_{n=1}^{\infty} f_{oo}^{(n)} \tag{31}$$

Using Eq (14) we obtain

$$ \frac{d_o}{{}_H d_o} = 1 + d_o - 1 = d_o $$

that is

$$ {}_H d_o = 1 $$

As expected, for Viterbi decoding the average separation ${}_H d_o$ between the nonsearch nodes (that is all the nodes) on the correct path is equal to 1. Having verified the validity of Eq(28) for the two extreme cases, we now show that the infinite sum of Eq(28) can be evaluated for any H from the set of transition probabilities of the Markov chain of the correct path.

Direct evaluation of the righthand side of Eq(29) is rather cumbersome and we shall compute separately $\sum_{n=1}^{\infty} {}_H f_{oo}^{(n)}$ and $\sum_{n=1}^{\infty} n \, {}_H f_{oo}^{(n)}$.

---

* If the multiple-path decoder operates on a trellis and exploits the path remergers, it can be shown that if $M = 2^m$ where m is the memory of the convolutional code, then the decoder is essentially a Viterbi decoder [3], [8] Consequently all nodes on the correct path are nonsearch.

From the definitions of the first return probability Eq(23), and first entrance probability Eq(24) we can write

$$ {}_H f_{oo}^{(n+1)} = \sum_{j=1}^{H-1} P_{oj} \; {}_H f_{jo}^{(n)} \tag{32}$$

and

$$ {}_H f_{jo}^{(n+1)} = \sum_{i=1}^{H-1} P_{ji} \; {}_H f_{io}^{(n)}, \quad j = 1,2,\ldots H-1 \tag{33}$$

where the $\{P_{ij}\}$ are the 1-step transition probabilities of the Markov Chain.

The form of Eq (32) and (33) suggests the following matrix representation: define the vector

$$ \underline{F}_o = \begin{bmatrix} {}_H f_{oo}^{(1)} \\ {}_H f_{10}^{(1)} \\ {}_H f_{20}^{(1)} \\ \vdots \\ {}_H f_{H-1,0}^{(1)} \end{bmatrix} = \begin{bmatrix} P_{oo} \\ P_{10} \\ P_{20} \\ \vdots \\ P_{H-1,0} \end{bmatrix} \tag{34}$$

and let $[A]$ be the H x H transition probability matrix of the chain where the 1rst column is set equal to zero

$$ A = \begin{bmatrix} 0 & p_1 & p_2 & \cdots & p_{H-1} \\ 0 & p_o & p_1 & \cdots & p_{H-2} \\ 0 & q_1 & p_o & & p_{H-3} \\ 0 & q_2 & q_1 & \ddots & p_{H-4} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & q_{H-2} & \cdots & q_1 & p_o \end{bmatrix} \tag{35}$$

then from Eqs (32) and (33) we obtain

$$\underline{F}_1 = \begin{bmatrix} {}_Hf_{00}^{(2)} \\ {}_Hf_{10}^{(2)} \\ \vdots \\ {}_Hf_{H-1,0}^{(2)} \end{bmatrix} = [A]\,\underline{F}_o \tag{36}$$

Likewise we obtain

$$\underline{F}_2 = \begin{bmatrix} {}_Hf_{00}^{(3)} \\ {}_Hf_{10}^{(3)} \\ \vdots \\ f_{H-1,0}^{(3)} \end{bmatrix} = [A]\ \underline{F}_1 = [A]^2\ \underline{F}_o \tag{37}$$

and in general

$$\underline{F}_n = [A]^n\,\underline{F}_o \tag{38}$$

Let the vector $\underline{F}_n$ be the (n+1)th column of a matrix $[B]$ having H rows and an infinite number of columns

$$[B] = [\underline{F}_o \vdots \underline{F}_1 \vdots \underline{F}_2 \vdots \cdots \vdots \underline{F}_n \vdots \cdots] \tag{39}$$

then

$$[B] = [\underline{F}_o \vdots [A]\,\underline{F}_o \vdots [A]^2\ \underline{F}_o \vdots \cdots \vdots [A]^n\ \underline{F}_o \vdots \cdots] \tag{40}$$

The elements b_{1i} of the first row of $[B]$ are the ${}_Hf_{00}^{(n)}$ of interest.

Consequently

$$\sum_{n=1}^{\infty} {}_Hf_{00}^{(n)} = \sum_{i=1}^{\infty} b_{1i} \tag{41}$$

Factoring the common vector $\underline{F}_o$, Eq(40) becomes

$$[B] = \left[(I \; \vdots \; [A] \; \vdots [A]^2 \vdots \; \ldots \; \vdots \; [A]^n \; \vdots \; \ldots\ldots) \; \underline{F}_o \right] \tag{42}$$

The matrix power series $I + [A] + [A]^2 + \, .. + [A]^n$ .. will converge to $(I - [A])^{-1}$ if the eigenvalues of $[A]$ are less than 1. Since $[A]$ is the truncation of a stochastic matrix, the sum converges. Letting

$$\underline{X} = \left[I - [A] \right]^{-1} \underline{F}_o \tag{43}$$

it then follows that the sum of the elements of the first row of $[B]$ is the first element x_1 of $\underline{X}$. Therefore we obtain the intermediate result

$$\sum_{n=1}^{\infty} {}_H f_{oo}^{(n)} = x_1 \tag{44}$$

Likewise each element of the sum $\sum_{n=1}^{\infty} n \;\; f_{oo}^{(n)}$ is an element of the first row of the infinite matrix

$$[B]^* = \left[\underline{F}_o \; \vdots \; 2 \; [A] \; \underline{F}_o \; \vdots \; 3 \; [A]^2 \; \underline{F}_o \vdots \; \ldots \; \vdots (n+1) [A]^n \; \underline{F}_o \; \vdots \; \ldots \right] \tag{45}$$

$$= \left[(I \; \vdots \; 2 \; [A] \vdots \; 3 \; [A]^2 \; \vdots \; \ldots \; \vdots \; (n+1) \; [A]^n \; \vdots \; \ldots) \; \underline{F}_o \right] \tag{46}$$

The sum of the columns converges to the vector

$$\underline{Y} = ([I - [A]]^{-1})^2 \; \underline{F}_o \tag{47}$$

and hence

$$\sum_{n=1}^{\infty} n \;\; f_{oo}^{(n)} = y_1 \tag{48}$$

where y_1 is the first element of the vector $\underline{Y}$. Finally combining Eq (44) and (48) we have the result

$$\frac{d_o}{{}_H d_o} = 1 + y_1 - x_1 \tag{49}$$

or

$$\frac{d_o - {}_H d_o}{{}_H d_o} = y_1 - x_1 \tag{50}$$

Observing that $\underline{F}_o$ and $[A]$ are known, and that H is finite the closed form expressions (49) or (50) can easily be evaluated on a computer. This computation has been carried out for a rate 1/2 convolutional code operating at E_b/N_o = 3.0dB and 3.5dB over a discrete memoryless Gaussian channel using a 3 bit quantization. The set of transition probabilities $\{q_i\}$ and $\{p_i\}$ of the associated Markov Chain are given in Table 1 and Table 2. Results of the computation of Eq(49) are plotted in Fig. (3) The curve shows that the ratio $d_o/{}_H d_o$ seems to level off for $H > 120$, which indicates that ${}_H d_o$ must be close to 1 for those values of H; that is d_o is slightly larger than 1.51 for E_b/N_o = 3.5dB and slightly larger than 1.64 for E_b/N_o = 3.0dB.

In order to verify experimentally these results, ${}_H d_o$ was measured for $1 \leq H \leq 160$ on a computer simulation of the correct path metric performed under strictly identical conditions. Based on simulation runs of a correct path metric of length 200,000 branches, d_o was found to be equal to 1.6745 for E_b/N_o = 3.0dB and 1.5179 for E_b/N_o = 3.5dB. As for the ratio $d_o/{}_H d_o$ the experimental values are so close to the theoretical results that hardly any difference can be observed on the curve of Fig.(3). Table 3 gives the experimental and theoretical values and demonstrates how closely (within 1%) the theory fits to the measurements.

IV CUMULATIVE DISTRIBUTION OF THE CORRECT PATH METRIC DIPS

Defining a dip on the correct path as the largest metric difference between two consecutive nonadjacent breakout nodes (see Fig. 4a), we shall use again the Markov Chain model and the previous results to derive an exact expression of its cumulative distribution

$$P_H = Pr\ (\text{Dip} < H) \tag{51}$$

Following the same approach as in section III, we let $\mathcal{H}$ be the set of states smaller than state H. Then, clearly, as illustrated in Fig (4b), P_H is the probability that starting from state 0, the system returns to that state without ever leaving the set $\mathcal{H}$. Recalling that ${}_{\mathcal{H}}f_{oo}^{(n)}$ is the first return probability to state 0 in n steps without ever leaving $\mathcal{H}$ (see Eq (24)) , it follows that Eq(51) can be written as

$$P_H = \sum_{n=1}^{\infty} {}_{\mathcal{H}}f_{oo}^{(n)} \tag{52}$$

This expression was encountered in section III and is given by Eq.(44). This closed form expression of the cumulative distribution of the correct path metric dips is, again, easily evaluated on a computer.

Using the set of transition probabilities given in Table 1 for $E_b/N_o = 3.5$dB computation of P_H has been carried out and plotted in Fig.(5) as the complement $(1 - P_H) = Pr(Dip \geq H)$. Results of the corresponding measurement obtained from a computer simulation on a path of 2×10^5 branches are also plotted for comparison.

The distributions are clearly exponentially decreasing with H and again, the measurements fit remarquably well with the theoretical results. Although the exponential behaviour was expected, The Markov Chain approach to the analysis of the metric dips is superior to the traditional bounding technique which considers a random ensemble of codes, and gives for P_H a loose upper bound rather than an equality [7]. On top of yielding easily computable expressions, the Markov Chain technique presents the additional advantage of taking into account the particular channel quantization scheme and the actual set of integer branch metrics used, making it directly applicable to any particular case.

V. CONCLUSION

The analysis of the correct path metric of sequential decoding presented in this paper departs from the traditional random coding technique over ensemble of codes and emphasizes the importance of the Markov Chain model. Interpreting this model with sequential decoding we have derived closed form expressions of the average separation between nonsearch nodes for both the single and multiple-path sequential decoding, and an exact expression of the cumulative distribution of the correct path metric dips. These new expressions are easy to evaluate, and being expressed in terms of the set of the correct path branch metric values and their probability assignments they are directly applicable to specific cases. The excellent agreement obtained between theoretical and experimental results demonstrates the accuracy of the model. Finally the Markov Chain approach seems to be applicable to other important problems of sequential decoding such as buffer overflow [5] and computational effort.

REFERENCES

[1] Fano, R.M., "A heuristic discussion of probabilistic decoding", IEEE Transactions on Information Theory, Vol. IT-9, pp. 64-73, April 1963.

[2] Jacobs, I.M., and Berlekamp, E.R., "A lower bound to the distribution of computation for sequential decoding", IEEE Transactions on Information Theory, Vol, IT-13, pp. 167-174, April 1967.

[3] Haccoun, D., and Ferguson, M., "Generalized stack algorithms for decoding convolutional codes", IEEE Transactions on Information Theory, Vol. IT-21, pp. 638-651, November 1975.

[4] Gallager, R.G., "Information theory and reliable communication" chap. 6, John Wiley, 1968.

[5] Massey, J.L, Sain, M.K., and Geist, J.M., "Certain infinite Markov Chains and sequential decoding", Discrete Mathematics, Vol. 3, pp. 163-175, September 1972.

[6] Feller, W., "An introduction to probability theory and its applications", Vol. 1, chap XV, John Wiley, 1968.

[7] Forney, G.D. Jr, "Convolutional codes III. Sequential decoding", Information and Control, Vol. 25, pp. 267-297, July 1974.

[8] Haccoun, D., "Multiple-path stack algorithms for decoding convolutional codes", Ph.D. dissertation, Dept of Electrical Engineering, McGill University, June 1974, and Technical Report EP 74-R-47, Ecole Polytechnique de Montréal, Oct. 1974.

ACKNOWLEDGEMENT

The author wishes to thank Mr Alain Janelle, Research Associate in the Department of Electrical Engineering for his discussions and helpful suggestions about this work. Moreover Mr Janelle performed the measurements on the computer simulation of the sequential decoder.

TABLE CAPTIONS

TABLE 1 : Markov Chain transition probabilities for a 3-bit quantized channel and $E_b/N_o = 3.5$dB.

TABLE 2 : Markov Chain transition probabilities for a 3-bit quantized channel and $E_b/N_o = 3.0$dB.

TABLE 3 : Comparison of the theoretical and experimental values of $d_o/_H d_o$.

FIGURE CAPTIONS

FIGURE 1: Segment of the correct path metric

FIGURE 2: Example of a Markov Chain with $Q = 2$, $J = 7$

FIGURE 3: Calculated values of $d_o/_H d_o$

FIGURE 4: (a) Definition of a dip of the correct
(b) Markov Chain representation

FIGURE 5: Experimental and calculated distribution $P_H = Pr$ (DIP $\geq$ H)

TABLE 1

Markov Chain transition probabilities for a 3-bit quantized channel and E_b/N_o = 3.5dB

q_8 =	0.477151	p_{32} =	.001469
q_7 =	0.207368	p_{35} =	.006750
q_6 =	0.022530	p_{36} =	.001467
q_3 =	0.127200	p_{40} =	.000900
q_2 =	0.027640	p_{46} =	.000276
		p_{48} =	4.32×10^{-4}
		p_{60} =	1.877×10^{-3}
p_2 =	.008477	p_{61} =	4.08×10^{-4}
p_5 =	.061074	p_{62} =	1.62×10^{-4}
p_6 =	.013271	p_{65} =	2.50×10^{-4}
p_{10} =	.008141	p_{73} =	1.2×10^{-4}
p_{18} =	.001954	p_{78} =	2.4×10^{-5}
p_{19} =	.022951	p_{87} =	4.5×10^{-5}
p_{20} =	.004987	p_{103} =	1.32×10^{-5}
p_{24} =	.003059	p_{128} =	2.0×10^{-6}

TABLE 2

Markov Chain transition probabilities for a 3-bit quantized channel and E_b/N_o = 3.0dB

q_8 = .4364452	p_{30} = .002025
q_7 = .2101586	$p_{32} = 8.062\times10^{-3}$
q_6 = .0252991	$p_{33} = 1.941\times10^{-3}$
q_3 = .1343121	$p_{37} = 1.2405\times10^{-3}$
q_2 = .032372	$p_{42} = 3.965\times10^{-4}$
	$p_{45} = 6.206\times10^{-4}$
	$p_{56} = 2.3558\times10^{-3}$
p_2 = .0103333	$p_{61} = 3.625\times10^{-4}$
p_5 = .0671917	$p_{69} = 1.813\times10^{-4}$
p_6 = .016177	$p_{72} = 3.72\times10^{-5}$
p_{10} = .010338	$p_{81} = 7.1\times10^{-5}$
p_{17} = .026309	$p_{96} = 2.18\times10^{-5}$
p_{18} = .008920	$p_{120} = 3.2\times10^{-6}$
p_{22} = .004048	

TABLE 3

Comparison of the theoretical and experimental values of $d_o/_H d_o$

	E_b/N_o = 3.0 dB			E_b/N_o = 3.5 dB		
	Experiment (200,000 branches)		Theory	Experiment (200,000 branches)		Theory
H	$_H do$	$do/_H do$	$do/_H do$	$_H do$	$do/_H do$	$do/_H do$
1	1.6745	1.0000	-	1.5179	1.0000	-
3	1.6594	1.0091	1.0087	1.5065	1.0075	1.0073
4	1.6594	1.0091	1.0087	1.5065	1.0075	1.0073
6	1.5596	1.0737	1.0731	1.4206	1.0685	1.0672
8	1.5316	1.0933	1.0928	1.3986	1.0852	1.0834
10	1.5285	1.0955	1.0953	1.3964	1.0870	1.0851
12	1.4836	1.1287	1.1286	1.3647	1.1122	1.1126
15	1.4736	1.1336	1.1333	1.3581	1.1177	1.1158
20	1.3582	1.2329	1.2340	1.2785	1.1872	1.1879
25	1.3054	1.2828	1.2833	1.2292	1.2349	1.2378
30	1.2827	1.3055	1.3080	1.2111	1.2533	1.2587
35	1.2134	1.3800	1.3800	1.1933	1.2720	1.2775
40	1.1792	1.4200	1.4215	1.1374	1.3345	1.3367
45	1.1588	1.4451	-	1.1143	1.3621	-
50	1.1336	1.4771	1.4762	1.0987	1.3816	1.3851
60	1.0887	1.5381	1.5305	1.0773	1.4089	1.4156
70	1.0610	1.5782	1.5702	1.0438	1.4542	1.4586
80	1.0419	1.6072	1.5960	1.0293	1.4746	1.4803
90	1.0281	1.6287	1.6138	1.0178	1.4913	1.4942
100	1.019	1.643	1.626	1.013	1.498	1.503
110	-	-	1.633	-	-	1.509
120	1.008	1.661	1.638	1.006	1.509	1.513
130	-	-	1.641	-	-	1.516
140	1.003	1.669	-	1.002	1.515	-
160	1.001	1.672	-	1.001	1.517	-

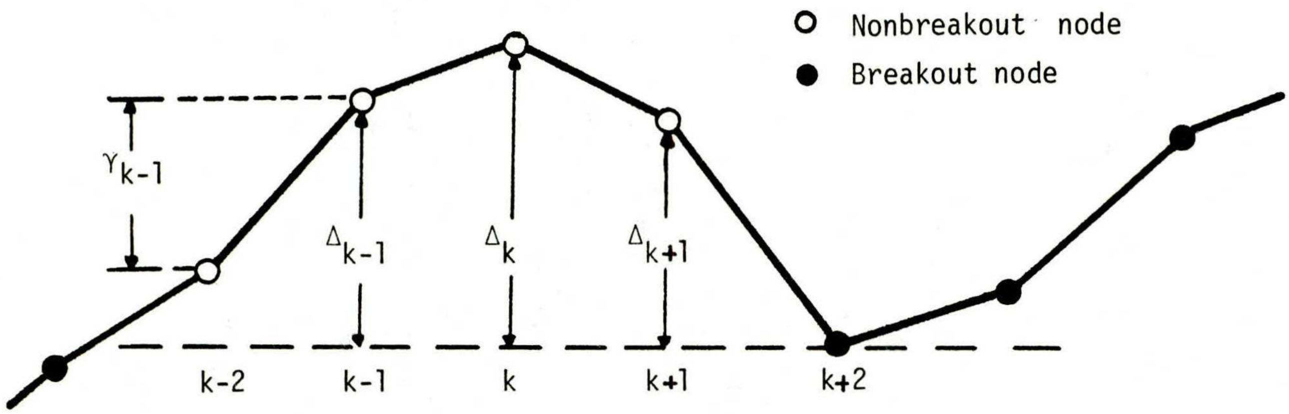


FIGURE 1: Segment of the correct path metric

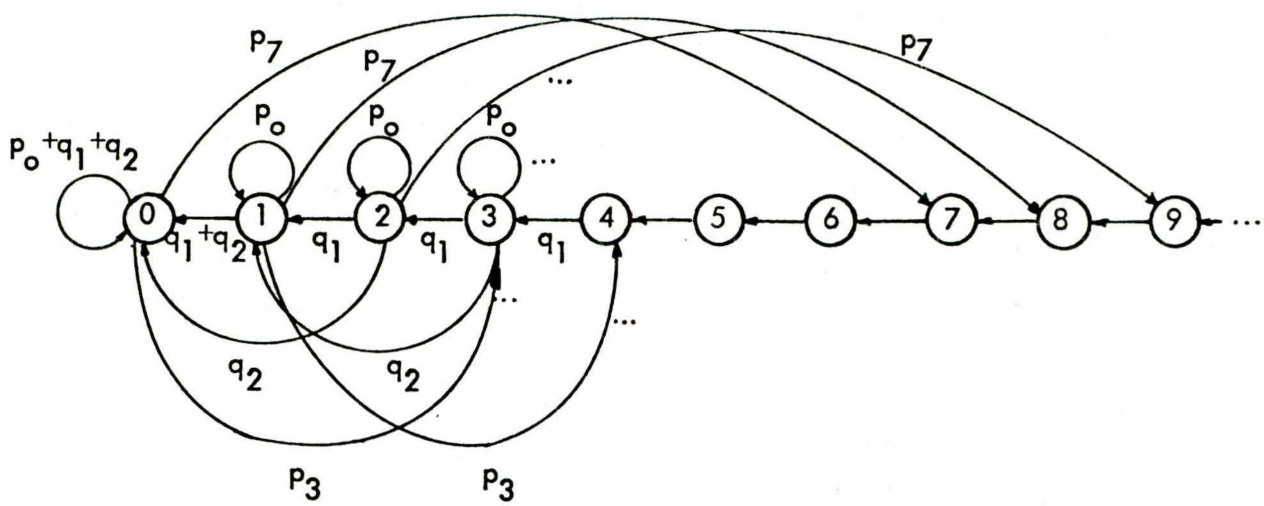
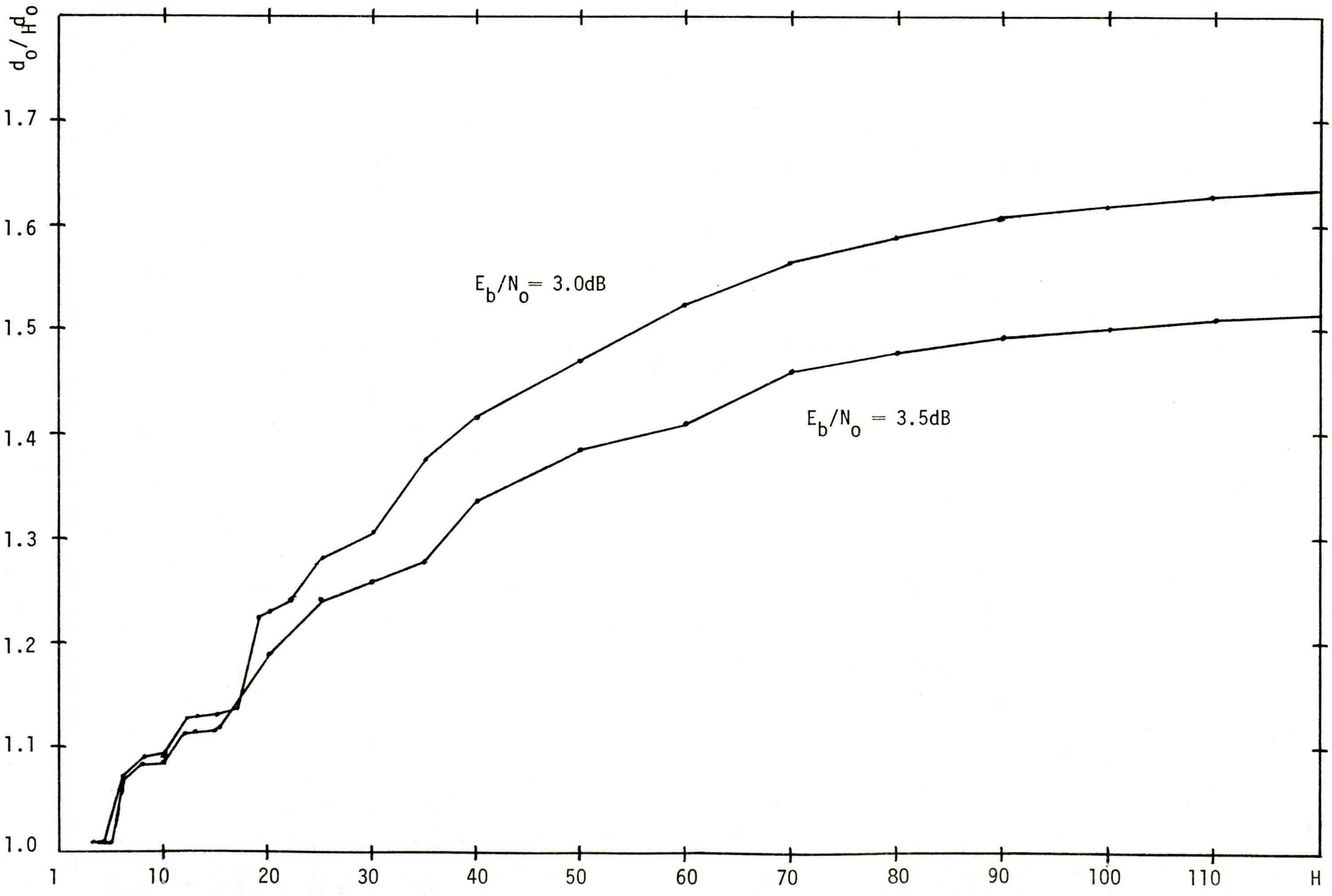


FIGURE 2: Example of a Markov Chain with $Q = 2$, $J = 7$

FIGURE 3: Calculated values of $d_o/{}_Hd_o$



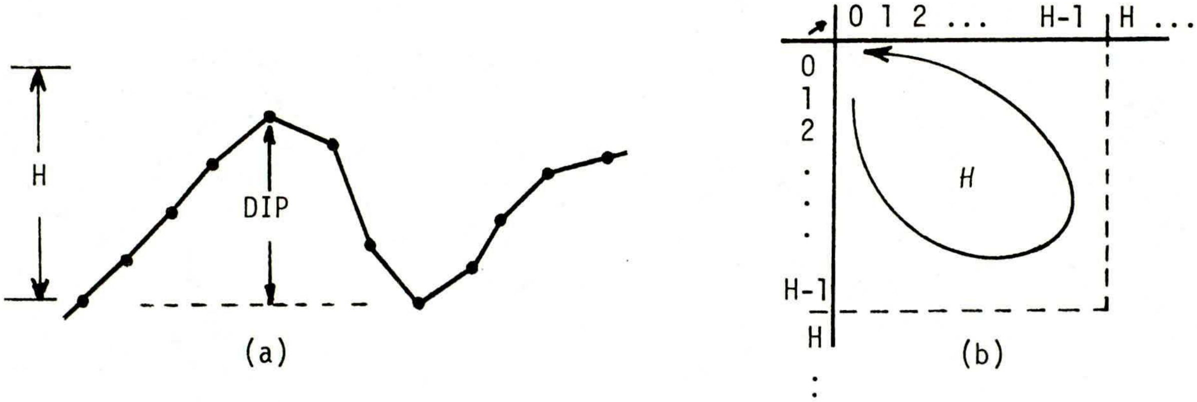


Figure 4: (a) Definition of a dip of the correct
(b) Markov Chain representation

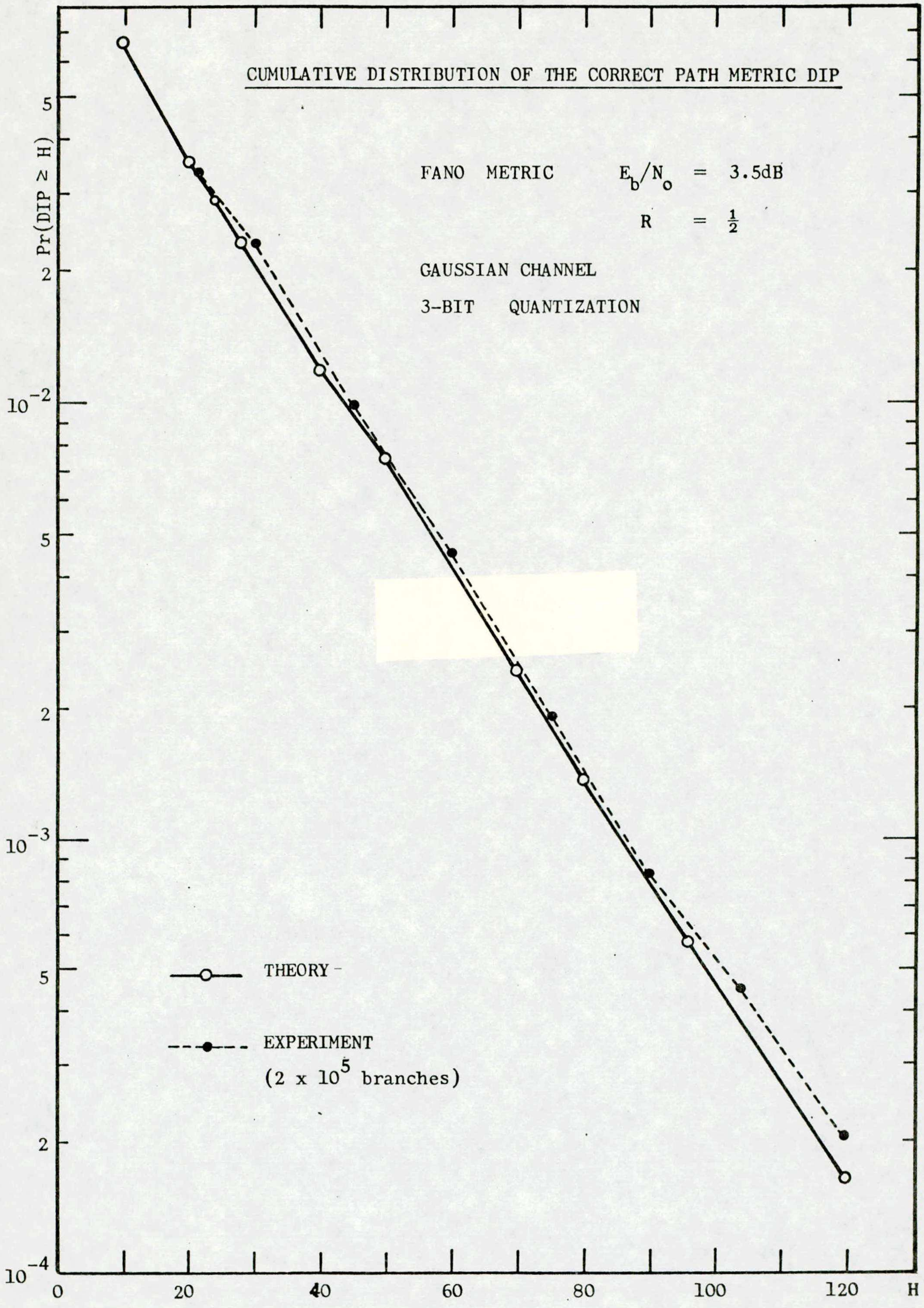


Figure 5: Experimental and calculated distribution $P_H = Pr\ (Dip \geq H)$