An Entropy Theorem for Computing the Capacity of Weakly (d, k)-Constrained Sequences

Augustus J. E. M. Janssen, *Senior Member, IEEE,* and
Kees A. Schouhamer Immink, *Fellow, IEEE*

***Abstract*—In this correspondence we find an analytic expression for the maximum of the normalized entropy $-\sum_{i\in T} p_i \ln p_i / \sum_{i\in T} ip_i$ where the set T is the disjoint union of sets S_n of positive integers that are assigned probabilities P_n, $\sum_n P_n = 1$. This result is applied to the computation of the capacity of weakly (d, k)-constrained sequences that are allowed to violate the (d, k)-constraint with small probability.**

***Index Terms*—Capacity, constrained code, (d, k) sequence, entropy, magnetic recording, RLL sequence, runlength-limited.**

I. Introduction and Announcement of Results

Let T be a set of positive integers, and assume that T is the disjoint union of a (finite or infinite) number of nonempty sets S_n, $n \in M$. Also assume that there are given numbers $P_n \geq 0$, $n \in M$, with $\sum_n P_n = 1$. We show the following result.

Theorem: The maximum of

$$H := \frac{-\sum\limits_{i\in T} p_i \ln p_i}{\sum\limits_{i\in T} ip_i} \tag{1}$$

under the constraints that $p_i \geq 0$ and

$$\sum_{i\in S_n} p_i = P_n, \qquad n \in M$$

equals z_0, where $z_0 > 0$ is the unique solution z of the equation

$$-\sum_{n\in M} P_n \ln Q_n(z) = -\sum_{n\in M} P_n \ln P_n \tag{2}$$

with $Q_n(z)$ given for $z > 0$

$$Q_n(z) := \sum_{i\in S_n} e^{-iz}, \qquad n \in M. \tag{3}$$

Moreover, the optimal p_i are given by

$$p_i = \frac{P_n}{Q_n(z_0)} e^{-iz_0}, \qquad i \in S_n,\, n \in M \tag{4}$$

and for these p_i we have that

$$\sum_{i\in T} ip_i = \frac{d}{dz}\left[-\sum_{n\in M} P_n \ln Q_n(z)\right](z_0). \tag{5}$$

As an application of this result we consider *weakly constrained* (d, k) sequences [1]. A binary (d, k)-constrained sequence has by definition at least d and at most k "zeros" between consecutive "ones." Such sequences are applied in mass storage devices such as the compact disc (CD) and the DVD. Weakly constrained codes do not strictly work to the rules, as they produce sequences that violate the specified constraints with a given (small) probability, see Section III for an explicit description. It is argued that if the channel is not free of errors, it is pointless to feed the channel with perfectly constrained sequences. Clearly, the extra freedom will result in an increase of the channel capacity. A (d, k)-constrained sequence can be thought to be composed of "phrases" 10^i, $d \leq i \leq k$, where 0^i means a series of i "zeros." In order to compute the channel capacity, i.e., the maximum $z_0/\ln 2$ of the entropy $H/\ln 2$, we define

$$\begin{aligned} T = &\{1, \cdots, d\} \cup \{d+1, \cdots, k+1\} \cup \{k+2, k+3, \cdots\} \\ &=: S_1 \cup S_2 \cup S_3 \end{aligned} \tag{6}$$

where $d = 0, 1, \cdots,$ and $k = d+1, d+2, \cdots$ are given, and we compute the capacity for the case that the probabilities P_1, P_3 assigned to the sets S_1, S_3 are both small. Clearly, the quantities P_1 and P_3 denote the probabilities that phrases are transmitted that are either too short or too long, respectively. We find that the familiar capacities of (d, k)-constrained sequences [2] are approached from above as P_1, $P_3 \to 0$ with an error $A(P_1 \ln P_1 + P_3 \ln P_3)$, where we can evaluate the A explicitly. We obtain a similar result for the case that T is as in (6) with S_1, S_3 merged into a single set $S_1 \cup S_3$.

II. Proof of the Theorem

We present the proof of the theorem for the case that the set T, and consequently the sets M and S_n, $n \in M$, are finite. The case that some of these sets may be infinite gives no particular problems, but complicates the presentation given below somewhat. At the end of this section, we shall indicate some modifications that are needed to have the argument work for this more general case as well.

The plan of the proof is as follows. We fix $x > 0$ in a range $[x_-, x_+]$ to be specified below, and we maximize, using Lagrange's theorem, the quantity

$$\frac{-1}{x} \sum_{i\in T} p_i \ln p_i \tag{7}$$

over $p_i \geq 0$ under the constraints that

$$\sum_{i\in T} ip_i = x, \quad \sum_{i\in S_n} p_i = P_n, \qquad n \in M. \tag{8}$$

The maximum value of (7) thus obtained is maximized over $x \in [x_-, x_+]$ and this yields the maximum H in (1) under the constraints on the p_i in the theorem.

The range of x to be considered in (7) and (8) is equal to $[x_-, x_+]$, where

$$x_- = \sum_{n\in M} P_n \min_{i\in S_n} i \qquad x_+ = \sum_{n\in M} \frac{P_n}{|S_n|} \sum_{i\in S_n} i. \tag{9}$$

To see this, we observe that for any choice of p_i, $i \in T$, satisfying $\sum_{i\in S} p_i = P_n$, we can increase the value of H in (1) by ordering the p_i's per set S_n decreasingly. Indeed, this does not change the values of $-\sum_{i\in T} p_i \ln p_i$ and $\sum_{i\in S_n} p_i$, $n \in M$, while it decreases the value of $\sum_{i\in T} ip_i$. Now x_- corresponds to the case that all mass P_n of S_n is assigned to the minimal element of S_n, $n \in M$, while x_+ corresponds to the case that all elements of S_n are assigned equal masses $|S_n|^{-1} P_n$.

To minimize (7) under the constraint (8), we observe that (7) is a continuous, strictly concave function of the p_i's restricted to the convex set described by (8) and $p_i \geq 0$, $i \in T$. Hence the maximum of (7) under the given constraints exists and is unique. By applying Lagrange's multiplier rule, we easily find that the maximizing p_i are of the form

$$p_i = e^{-(\lambda i + \mu_n)x - 1}, \qquad i \in S_n,\, n \in M \tag{10}$$

with the Lagrange multipliers λ and μ_n, $n \in M$, corresponding to the first constraint and the second constraints in (8), respectively, such that the p_i's in (10) satisfy (8). From what has been said above we have $\lambda \geq 0$.

Manuscript received May 20, 1999; revised November 19, 1999. The material in this correspondence will be presented at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 25–30, 2000.

A. J. E. M. Janssen is with the Philips Research Laboratories, WY 81, 5656 AA Eindhoven, The Netherlands (e-mail: A.J.E.M.Janssen@philips.com).

K. A. S. Immink is with the Institute for Experimental Mathematics, 45326 Essen, Germany (e-mail: immink@exp-math.uni-essen.de).

Communicated by E. Soljanin, Associate Editor for Coding Techniques.
Publisher Item Identifier S 0018-9448(00)02898-4.

It is easy to show that the constraints (8) imply that

$$\mu_n x + 1 = \ln[Q_n(\lambda x)/P_n], \qquad n \in M \tag{11}$$

with Q_n given in (3), and that

$$x = \sum_n P_n \frac{R_n(\lambda x)}{Q_n(\lambda x)} \tag{12}$$

with R_n given for $z \geq 0$ by

$$R_n(z) := \sum_{i \in S_n} ie^{-iz} = -Q_n'(z), \qquad n \in M. \tag{13}$$

We shall now show that for any $x \in (x_-, x_+]$ there is a unique solution $\lambda \in [0, \infty)$ of (12). Indeed, we have for $x > 0$ fixed that

$$\frac{d}{d\lambda}\left[\frac{R_n(\lambda x)}{Q_n(\lambda x)}\right] = x\,\frac{R_n'(\lambda x)Q_n(\lambda x) - R_n(\lambda x)Q_n'(\lambda x)}{Q_n^2(\lambda x)} \tag{14}$$

and for $n \in M$, $z \geq 0$

$$\begin{aligned} &R_n'(z)Q_n(z) - R_n(z)Q_n'(z) \\ &\qquad = -\sum_{i\in S_n} i^2 e^{-iz} \sum_{i \in S_n} e^{-iz} + \left(\sum_{i \in S_n} ie^{-iz}\right)^2 \leq 0 \end{aligned} \tag{15}$$

by the Cauchy–Schwarz inequality with equality if and only if S_n is a singleton. Also

$$\sum_n P_n \frac{R_n(0)}{Q_n(0)} = x_+ \qquad \lim_{\lambda \to \infty} \sum_n P_n \frac{R_n(\lambda x)}{Q_n(\lambda x)} = x_-. \tag{16}$$

Hence, except in the trivial case that all S_n's with $P_n > 0$ are singletons, the right-hand side function in (12) strictly decreases from x_+ at $\lambda = 0$ to x_- at $\lambda = \infty$, as required.

Denoting the unique solution of λ of (12) by $\lambda(x)$ for $x \in (x_-, x_+]$, we find from (11) and (12) for the maximum value of (7) under the constraints (8) that

$$H(x) = \lambda(x)x - \frac{1}{x}\sum_n P_n \ln P_n + \frac{1}{x}\sum_n P_n \ln Q_n(\lambda(x)x). \tag{17}$$

To maximize $H(x)$ over $x \in (x_-, x_+]$ we differentiate $H(x)$ with respect to x, and we get using (13)

$$\begin{aligned} H'(x) &= (\lambda(x)x)' + \frac{1}{x^2}\sum_n P_n \ln P_n \\ &\quad + \frac{-1}{x^2}\sum_n P_n \ln Q_n(\lambda(x)x) \\ &\quad - \frac{1}{x}(\lambda(x)x)' \sum_n P_n \frac{R_n(\lambda(x)x)}{Q_n(\lambda(x)x)}. \end{aligned} \tag{18}$$

By (12) and the definition of $\lambda(x)$ it thus follows that

$$H'(x) = \frac{1}{x^2}\left(\sum_n P_n \ln P_n - \sum_n P_n \ln Q_n(\lambda(x)x)\right). \tag{19}$$

We shall next show that there is a unique $x_0 \in (x_-, x_+)$ such that

$$H'(x_0) = 0; \quad H'(x) > 0,\ x < x_0; \quad H'(x) < 0,\ x > x_0. \tag{20}$$

We first observe that $\lambda(x)x$ decreases from ∞ to 0 as x increases from x_- to x_+. Indeed, from (12) and the definition of $\lambda(x)$ we have

$$\begin{aligned} 1 &= \sum_n P_n \frac{d}{dx}\left[\frac{R_n(\lambda(x)x)}{Q_n(\lambda(x)x)}\right] \\ &= (\lambda(x)x)' \sum_n P_n \left(\frac{d}{dz}\frac{R_n(z)}{Q_n(z)}\right) \ (z = \lambda(x)x). \end{aligned} \tag{21}$$

As in (14) and (15) we have that $(R_n(z)/Q(z))' \leq 0$, with equality signs for all n only in trivial cases, whence $(\lambda(x)x)' < 0$. Also, it is easy to see that

$$\lim_{x \uparrow x_+} \lambda(x)x = 0 \qquad \lim_{x \downarrow x_-} \lambda(x)x = \infty \tag{22}$$

as required. Hence, except in trivial cases, we have that $\sum_n P_n \ln Q_n(\lambda(x)x)$ increases from ∞ to

$$\sum_n P_n \ln|S_n| \geq 0 > \sum_n P_n \ln P_n$$

as x increases from x_- to x_+. Therefore, there is a unique x_0 such that (20) holds.

Evidently, $H(x)$ assumes its maximum at the x_0 of the previous paragraph, and we have at this x_0 from (17) that

$$H(x_0) = \lambda(x_0)x_0 =: z_0. \tag{23}$$

Thus we see that the maximum of $H(x)$ over $x \in (x_-, x_+]$ equals z_0, where z_0 is the unique solution $z = \lambda(x)x$ of the equation

$$-\sum_n P_n \ln Q_n(z) = -\sum_n P_n \ln P_n. \tag{24}$$

This proves (2) of the theorem. From (12) and (13) and the definition of $\lambda(x)$ we get the formula (5), and the explicit expression for the p_i in (4) follows from (10). This completes the proof of the theorem.

We now briefly comment on the required modifications to have the argument of the proof also work for the case that some of the sets S_n are infinite. Now $x_+ = \infty$, and we must consider $x \in (x_-, x_+)$. Also, for $x \in (x_-, \infty)$ fixed, the right-hand side of (12) strictly decreases in λ from ∞ at $\lambda = 0$ to x_- at $\lambda = \infty$, whence there is a unique solution $\lambda = \lambda(x)$ of (12). Finally, the maximization of $H(x)$ over $x \in (x_-, \infty)$ can be done in a similar way as in the case of finite T.

III. Application to Weakly (d, k)-Constrained Sequences

We shall now apply our theorem to the computation of the capacity of weakly (d, k)-constrained sequences, these being allowed to violate the (d, k)-constraint with (small) probability. Accordingly, we let d, k be two nonnegative integers, $k > d$ (with k possibly ∞), and we consider the set $T = \{1, 2, \cdots\}$ partitioned as

$$\begin{aligned} T &= \{1, \cdots, d\} \cup \{d+1, \cdots, k+1\} \cup \{k+2, k+3, \cdots\} \\ &= S_1 \cup S_2 \cup S_3 \end{aligned} \tag{25}$$

where the sets S_n, $n = 1, 2, 3$ are assigned probabilities $P_n \geq 0$ with $P_1 + P_2 + P_3 = 1$. For this kind of application it is customary to consider the normalized entropy

$$\overline{H} = \frac{-\sum_i p_i \log_2 p_i}{\sum_i ip_i} = \frac{1}{\ln 2} H \tag{26}$$

with H of (1).

We compute for $z > 0$

$$Q_1(z) = \sum_{i=1}^{d} e^{-iz} = \frac{1 - e^{-dz}}{1 - e^{-z}}\, e^{-z} \tag{27}$$

$$Q_2(z) = \sum_{i=d+1}^{k+1} e^{-iz} = \frac{e^{-dz} - e^{-(k+1)z}}{1 - e^{-z}}\, e^{-z} \tag{28}$$

$$Q_3(z) = \sum_{i=k+2}^{\infty} e^{-iz} = \frac{e^{-(k+1)z}}{1 - e^{-z}}\, e^{-z}. \tag{29}$$

By the theorem, given P_1, P_3, the maximum value $C(d, k; P_1, P_3)$ of $\overline{H}$ under the given constraints is equal to $z_0(P_1, P_3)/\ln 2$, where $z_0 = z_0(P_1, P_3)$ is the unique solution z of

$$\begin{aligned}-P_1 \ln Q_1(z) - P_2 \ln Q_2(z) - P_3 \ln Q_3(z) \\ = -P_1 \ln P_1 - P_2 \ln P_2 - P_3 \ln P_3 \quad (30)\end{aligned}$$

with $P_2 = 1 - P_1 - P_3$.

We are particularly interested in the behavior of $C(d, k; P_1, P_3)$ as a function of P_1, P_3 small. We first observe that for $P_1 = P_3 = 0$, $P_2 = 1$, (30) reduces to

$$Q_2(z) = \frac{e^{-dz} - e^{-(k+1)z}}{1 - e^{-z}} e^{-z} = 1 \quad (31)$$

i.e., with $y = e^z$ to

$$y^{k+2} - y^{k+1} - y^{k-d+1} + 1 = 0. \quad (32)$$

This is the familiar equation associated with perfectly (d, k)-constrained sequences for which the capacity $C(d, k)$ is given by $\log_2 y_{00} = z_{00}/\ln 2$, where z_{00} is the unique positive solution of (31) and y_{00} is $\exp(z_{00})$. Since the $Q_n(z)$ are smooth functions of $z > 0$, there holds for z close to z_{00}

$$Q_n(z) = Q_n(z_{00}) + (z - z_{00})Q_n'(z_{00}) + O((z - z_{00})^2). \quad (33)$$

From (30) it follows from some elementary considerations that

$$z_0(P_1, P_3) = z_{00} + O(P_1 \ln P_1 + P_3 \ln P_3).$$

For small P_1, P_3 we thus get that $z_0(P_1, P_3)$ satisfies

$$\begin{aligned}&\ln Q_2(z_0(P_1, P_3)) \\ &\quad = P_1 \ln P_1 + P_3 \ln P_3 - (P_1 + P_3) - P_1 \ln Q_1(z_{00}) \\ &\qquad - P_3 \ln Q_3(z_{00}) + O((P_1 + P_3)(P_1 \ln P_1 + P_3 \ln P_3)). \quad (34)\end{aligned}$$

Hence, using that $Q_2(z_{00}) = 1$

$$Q_2(z_0(P_1, P_3)) = Q_2(z_{00}) + \Delta + \epsilon \quad (35)$$

where

$$\begin{aligned}\Delta = P_1 \ln P_1 + P_3 \ln P_3 - (P_1 + P_3) \\ - P_1 \ln Q_1(z_{00}) - P_3 \ln Q_3(z_{00}) \quad (36)\end{aligned}$$

and here and in the sequel ϵ denotes an O-term as in the third line of (34). Therefore,

$$z_0(P_1, P_3) - z_{00} = \frac{\Delta}{Q_2'(z_{00})} + \epsilon \quad (37)$$

and it follows that

$$C(d, k; P_1, P_3) - C(d, k) = \frac{z_0(P_1, P_3) - z_{00}}{\ln 2} = \frac{\overline{\Delta}}{Q_2'(z_{00})} + \epsilon \quad (38)$$

with

$$\begin{aligned}\overline{\Delta} = \frac{\Delta}{\log_2 2} = P_1 \log_2 P_1 + P_3 \log_2 P_3 - P_1 \log_2 Q_1(z_{00}) \\ - P_3 \log_2 Q_3(z_{00}) - \frac{P_1 + P_3}{\ln 2}. \quad (39)\end{aligned}$$

Thus the difference $C(d, k; P_1, P_3) - C(d, k)$ consists of a linear combination of terms $P_1 \log_2 P_1$, $P_3 \log_2 P_3$, P_1, P_3, and an ϵ-error as $P_1 \downarrow 0$, $P_3 \downarrow 0$.

We next present two examples. The first example is merely meant to check that the theorem yields results that are in agreement with what one can also obtain by more elementary means. The second example is relevant for storage practice.

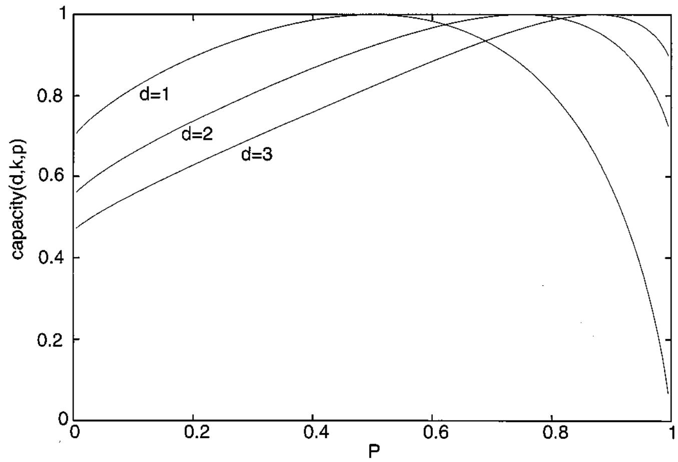

Fig. 1. The capacity $C(d, \infty; P_1)$ of weakly d-constrained sequences as a function of the probability P_1 that the sequence violates the given d constraint.

Example 1: Take $k = \infty$ so that the terms with index 3 disappear altogether. We have now

$$Q_1(z) = \frac{e^{dz} - 1}{e^{(d+1)z} - e^{dz}} \qquad Q_2(z) = \frac{1}{e^{(d+1)z} - e^{dz}}. \quad (40)$$

Equation (30) becomes

$$\begin{aligned}-P_1 \ln Q_1(z) - (1 - P_1) \ln Q_2(z) \\ = -P_1 \ln P_1 - (1 - P_1) \ln(1 - P_1) \quad (41)\end{aligned}$$

with solution $z_0(P_1)$ for $0 \le P_1 \le 1$. Observe that

$$\begin{aligned}C(d, \infty; P_1 = 0) &= \frac{z_0(0)}{\ln 2} = C(d, \infty) \\ C(d, \infty; P_1 = 1) &= \frac{z_0(1)}{\ln 2} = C(0, d - 1). \quad (42)\end{aligned}$$

In Fig. 1 we have plotted $C(d, \infty; P_1)$ as a function of $P_1 \in (0, 1)$ for $d = 1, 2, 3$. It is seen that $C(d, \infty; P_1)$ has maximum unity, and we shall show that this maximum occurs at $P_1 = 1 - 2^{-d}$. Indeed, using (40) and (41) we get for $z = z_0(P_1)$

$$\begin{aligned}\ln e^{dz} + \ln(e^z - 1) - P_1 \ln\left(e^{dz} - 1\right) \\ = -P_1 \ln P_1 - (1 - P_1) \ln(1 - P_1). \quad (43)\end{aligned}$$

Differentiating implicitly with respect to P_1 and setting

$$(dz_0(P_1))/dP_1 = 0$$

we easily obtain

$$P_1 = 1 - e^{-dz_0(P_1)}. \quad (44)$$

Substituting this P_1 back into (43) we then exactly obtain $z_0(P_1) = \ln 2$, as required. The above result can also be understood by noting that if the capacity is unity, the distribution is given by $p_i = 2^{-i}, i \ge 1$ [3], so that the maximum occurs at $P_1 = 1 - 2^{-d}$.

As to (38) and (39), we let $y_{00} = \exp(z_{00})$, so that y_{00} is the unique solution $y > 1$ of

$$y^{d+1} - y^d = 1 \quad (45)$$

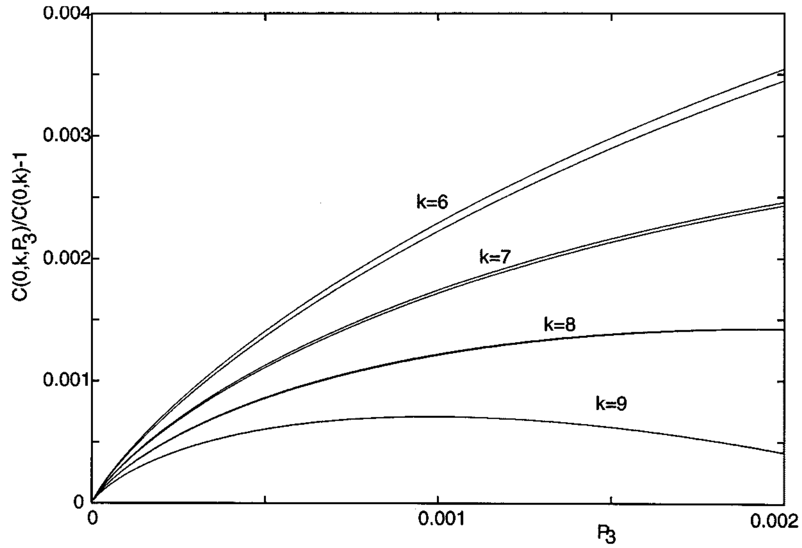

Fig. 2. The relative capacity gain $C(0,\, k;\, P_3)/C(0,\, k) - 1$ of weakly k-constrained sequences as a function of the probability P_3 that the sequence violates the given k-constraint. The upper curves are computed with full accuracy, while the lower curves are computed with approximation (55).

and we compute

$$Q_1(z_{00}) = y_{00}^d - 1 \qquad Q_2'(z_{00}) = -(d + 1 + y_{00}^d). \tag{46}$$

Therefore, there holds

$$C(d,\, \infty;\, P_1) - C(d,\, \infty) = -\frac{P_1(\log_2\, P_1 - \log_2(y_{00}^d - 1) - 1/\ln 2)}{d + 1 + y_{00}^d} \tag{47}$$

and, indeed, Fig. 1 shows a $-P_1\, \log_2\, P_1$-behavior of $C(d,\, \infty;\, P_1) - C(d,\, \infty)$ near $P_1 = 0$.

Example 2: We now consider the case that $d = 0$ and $k \geq 6$, so that

$$T = \{1,\, \cdots,\, k+1\} \cup \{k+2,\, k+3,\, \cdots\} = S_2 \cup S_3 \tag{48}$$

where the sets S_2 and S_3 are assigned probabilities P_2 and P_3, $P_2 + P_3 = 1$, with P_3 small. In fact, this is Example 1 with d replaced by $k+1$, S_1 replaced by S_2, S_2 replaced by S_3, and P_1 replaced by $P_2 = 1 - P_3$. Hence we consider Fig. 1 at the far right-hand side of the P_1-axis. Accordingly, we have

$$Q_2(w) = \frac{e^{(k+1)w} - 1}{e^{(k+2)w} - e^{(k+1)w}} \qquad Q_3(w) = \frac{1}{e^{(k+2)w} - e^{(k+1)w}} \tag{49}$$

where we have written w rather than z to avoid confusion with the z in Example 1. Equation (30) becomes

$$-(1 - P_3)\, \ln\, Q_2(w) - P_3\, \ln\, Q_3(w) = -(1 - P_3)\, \ln(1 - P_3) - P_3\, \ln\, P_3 \tag{50}$$

the solution w of which we denote by $w_0(P_3)$. For the corresponding capacity at $P_3 = 0$ we have

$$C(0,\, k;\, P_3 = 0) = \frac{w_0(0)}{\ln\, 2} = C(0,\, k). \tag{51}$$

Denoting $w_0(0) = w_{00}$ and $x_{00} = \exp{(w_{00})}$ we have that x_{00} is the unique solution $x > 1$ of the equation

$$x^{k+2} - 2x^{k+1} + 1 = 0, \qquad \text{i.e., } x = 2 - \frac{1}{x^{k+1}}. \tag{52}$$

The formulas (38) and (39) yield for $P_3 \to 0$

$$C(0,\, k;\, P_3) - C(0,\, k) = \frac{P_3\, \log_2\, P_3 - P_3\, \log_2\, Q_3(w_{00}) - P_3/\ln 2}{Q_2'(w_{00})} + \epsilon. \tag{53}$$

We compute, using (52) repeatedly

$$Q_3(w_{00}) = \frac{1}{x_{00}^{k+1} - 1}$$
$$Q_2'(w_{00}) = -2 + \frac{k}{x_{00}^{k+1} - 1}. \tag{54}$$

Finally, when $k \geq 6$ we have (see the second formula in (52)) that x_{00} is close to 2, where $Q_3(w_{00}) \approx x_{00}^{-k-1}$, $Q_2'(w_{00}) \approx -2$. This yields the approximation as $P_3 \to 0$

$$C(0,\, k;\, P_3) - C(0,\, k) \approx \frac{1}{2}(-P_3\, \log_2\, P_3 - (k+1)P_3\, \log_2\, x_{00} + P_3/\ln 2). \tag{55}$$

In Fig. 2 we have plotted the relative capacity gain

$$\frac{C(0,\, k;\, P_3) - C(0,\, k)}{C(0,\, k)} \tag{56}$$

for $k = 6,\, 7,\, 8,\, 9$ and $P_3 \in (0,\, 0.002]$. It is seen that (56) exhibits the expected $\frac{1}{2}P_3\, \log_2\, P_3$ behavior for P_3 very near to zero, but that the linear terms at the right-hand side of (55) dominate the $P_3\, \log_2\, P_3$ term from $P_3 = 2^{-k-1}$ onwards (see end of Example 1). As we can see in Fig. 2, the approximation given in (55) is quite accurate, especially for larger values of k.

We finally consider the case that, with d and k as before, the set $T = \{1,\, 2,\, \cdots\}$ is partitioned as

$$T = \{1,\, \cdots,\, d,\, k+2,\, k+3,\, \cdots\} \cup \{d+1,\, \cdots,\, k+1\} = S_1 \cup S_2 \tag{57}$$

so that sets S_1, S_3 in (25) are merged into one set S_1, with probabilities P_1, P_2 assigned to S_1, S_2. The analysis for this case proceeds along the same lines as for the partitioning of T as in (25). In particular, we have now

$$Q_1(z) = \frac{1 - e^{-dz} + e^{-(k+1)z}}{1 - e^{-z}} e^{-z}$$
$$Q_2(z) = \frac{e^{-dz} - e^{-(k+1)z}}{1 - e^{-z}} e^{-z} \tag{58}$$

(the same Q_2 as in (28)), and $C(d, k; P_1 = 0) = C(d, k)$. Also, z_{00} is the same as before, and for the behavior of $C(d, k; P_1)$ as $P_1 \downarrow 0$ we now find

$$C(d, k; P_1) - C(d, k) = \frac{P_1(\log_2 P_1 - \log_2 Q_1(z_{00}) - 1/\ln 2)}{Q_2'(z_{00})} + O(P_1^2 \log_2 P_1). \tag{59}$$

IV. CONCLUSIONS

We have presented an analytic expression for the maximum of the normalized entropy $-\sum_{i\in T} p_i \ln p_i / \sum_{i \in T} ip_i$ under the condition that T is the disjoint union of sets S_n of positive integers that are assigned probabilities P_n, $\sum_n P_n = 1$. This result has been applied to compute the capacity of weakly (d, k)-constrained sequences that are allowed to violate the (d, k)-constraint with a given (small) probability.

REFERENCES

[1] K. A. S. Immink, "Weakly constrained codes," *Electron. Lett.*, vol. 33, no. 23, pp. 1943–1944, Nov. 1997.

[2] ——, *Codes for Mass Data Storage Systems*. Amsterdam, The Netherlands: Shannon Foundation Publishers, 1999.

[3] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, July 1948.

Time-Varying Encoders for Constrained Systems: An Approach to Limiting Error Propagation

Jonathan J. Ashley, *Member, IEEE*, and Brian H. Marcus, *Fellow, IEEE*

***Abstract*—Time-varying encoders for constrained systems are introduced. The approach generalizes the state-splitting (ACH) algorithm in a way that yields encoders consisting of multiple phases, with encoding proceeding cyclically from one phase to the next. The framework is useful for design of high-rate codes with reduced decoder error propagation and reduced complexity.**

***Index Terms*—Constrained systems, finite-state encoder, input-restricted channel, PRML, sliding-block decoder, state splitting.**

I. INTRODUCTION

Constrained coding is a special kind of channel coding in which unconstrained user sequences are encoded into sequences that are required to satisfy certain hard constraints such as runlength limits [10], [14], [13].

Manuscript received December 29, 1998.

J. J. Ashley is with Infineon Technologies, Santa Cruz, CA 95060 USA (e-mail: Jonathan.Ashley@infineon.com).

B. H. Marcus is with IBM Almaden Research Center, San Jose, CA 95120 USA (e-mail: marcus@almaden.ibm.com).

communicated by R. M. Roth, Associate Editor for Coding theory.

Publisher Item Identifier S 0018-9448(00)03101-1.

In a finite-state encoder, arbitrary user data sequences are encoded to constrained data sequences via a finite-state machine. The encoder is said to have *rate* $p : q$ if at each step of the encoding process, one p-tuple of user data is encoded to one q-tuple of constrained data in such a way that the concatenation of the encoded q-tuples obeys the given constraint. For the purposes of limiting decoder error propagation, decoding is usually implemented via a sliding-block decoder. One method of constructing finite-state encoders is the *state-splitting algorithm* (also called the ACH algorithm) [1], [14]. The main purpose of our correspondence is to show how to adapt the state-splitting algorithm to the time-varying setting and to use it as an outline for constructing high-rate codes with limited error propagation.

The state-splitting algorithm begins with a representation of the desired constraint and iteratively constructs a sequence of graphs, ultimately arriving at a graph that can be used as an encoder. The last step in the construction is the *data-to-codeword assignment*, where input labels (or tags) are assigned to the edges of this graph. At each state this amounts to a 1-1 assignment between the set of all binary p-tuples and 2^p of the outgoing edges. The choice of assignment can significantly affect the complexity and performance of the code.

When p is relatively small, one can usually find a reasonably good assignment simply by *ad hoc* experimentation. But when p is relatively large, there are far too many possible assignments and a poor choice could lead to a very costly implemenatation. Indeed, for large p, the data-to-codeword assignment becomes more of the heart of the code construction problem. Recently, the data recording industry has been moving toward detection schemes that can function well at very high code rates such as $16 : 17,\ 24 : 25$, and $32 : 33$. Most such codes that can be found in the literature have been designed by clever *ad hoc* procedures.

The design of such a code may appear to be somewhat mind-boggling at first. After all, a rate $24 : 25$ code would involve some assignment of all 2^{24} binary strings of length 24. In much of the previous work, this difficulty is overcome by "breaking down" the coding problem into smaller subproblems: say, by partitioning the coordinates into a few smaller groups, designing an encoding strategy on each of these groups and then putting together the resulting encoded strings in a way that satisfies the constraint; see, for example, [17], [12], [5], [4], [15]. One can think of the constraint and the encoders as changing from each one of these groups to another. In connection with this, we mention that periodically time-varying constraints have been introduced in [3] and [6].

Inspired by this work as well as our own independent efforts to design low-complexity codes, we show in this correspondence how to adapt the state-splitting algorithm to the setting of periodically time-varying encoders. For instance, we could construct a rate $16 : 19$ code by employing two different finite-state encoders in alternating phases: a rate $8 : 9$ phase and a rate $8 : 10$ phase; this can potentially yield a low-complexity encoder. But there is another advantage. For a generic block-decodable $16 : 19$ code, an isolated channel error may corrupt one 16-bit block, equivalently two user bytes. If such a code were constructed as a two-phase rate $8 : 9/8 : 10$ block-decodable code (i.e., each 9-bit block and 10-bit block decodes directly to a user byte), then such an error could corrupt only one user byte. If the latter were not possible, one might try to construct a two-phase $8 : 9/8 : 10$ sliding-block-decodable code with sliding window consisting of two blocks (a 9-bit block followed by a 10-bit block and *vice versa*). In this case, an isolated channel error could corrupt two user bytes. However, a 2-bit channel error (such as a bit shift) could corrupt only three user bytes, whereas in a generic block-decodable $16 : 19$ code it might corrupt four user bytes.