

An Interpretation of Identification Entropy

Rudolf Ahlswede and Ning Cai, *Senior Member, IEEE*

Abstract—After Ahlswede introduced identification for source coding he discovered identification entropy and demonstrated that it plays a role analogously to classical entropy in Shannon's noiseless source coding. We give now even more insight into this functional interpreting its two factors.

Index Terms—Identification entropy, operational justification, source coding for identification.

I. INTRODUCTION

A. Terminology

Identification in source coding started in [3]. Then identification entropy was discovered and its operational significance in noiseless source coding was demonstrated in [4].

Familiarity with that paper is helpful, but not necessary here. As far as possible we use its notation.

Differences come from the fact that we use now a q -ary coding alphabet $\mathcal{X} = \{0, 1, \dots, q-1\}$, whereas earlier only the case $q = 2$ was considered and it was remarked only that all results generalize to arbitrary q . In particular the identification entropy, abbreviated as ID-entropy, for the source (\mathcal{U}, P, U) has the form

$$H_{I,q}(P) = \frac{q}{q-1} \left(1 - \sum_{u \in \mathcal{U}} P_u^2 \right). \quad (1.1)$$

Shannon (in 1948) has shown that a source (\mathcal{U}, P, U) with output U satisfying $\text{Prob}(U = u) = P_u$, can be encoded in a prefix code $\mathcal{C} = \{c_u : u \in \mathcal{U}\} \subset \{0, 1, \dots, q-1\}^*$ such that for the q -ary entropy

$$H_q(P) = \sum_{u \in \mathcal{U}} -P_u \log_q P_u \leq \sum_{u \in \mathcal{U}} P_u \|c_u\| \leq H_q(P) + 1$$

where $\|c_u\|$ is the length of c_u .

We use a prefix code, abbreviated as PC, \mathcal{C} for another purpose, namely, noiseless identification, that is, every user who wants to know whether a v ($v \in \mathcal{U}$) of his interest is the actual source output or cannot consider the random variable (RV) C with $C = c_u = (c_{u_1}, \dots, c_{u_{\|c_u\|}})$ if $U = u$ and check whether $C = (C_1, C_2, \dots)$ coincides with c_v in the first, second, etc., letter and stop when the first different letter occurs or when $C = c_u$. Let $L_{\mathcal{C}}(P, u)$ be the expected number of checkings, if code \mathcal{C} is used.

Related quantities are

$$L_{\mathcal{C}}(P) = \max_{v \in \mathcal{U}} L_{\mathcal{C}}(P, v) \quad (1.2)$$

that is, the expected number of checkings for a person in the worst case, if code \mathcal{C} is used

$$L(P) = \min_{\mathcal{C}} L_{\mathcal{C}}(P), \quad (1.3)$$

the expected number of checkings in the worst case for a best code, and finally, if v 's are chosen by an RV V independent of U and defined by $\text{Prob}(V = v) = Q_v$ for $v \in \mathcal{V} = \mathcal{U}$, we consider

$$L_{\mathcal{C}}(P, Q) = \sum_{v \in \mathcal{U}} Q_v L_{\mathcal{C}}(P, v) \quad (1.4)$$

the average number of expected checkings, if code \mathcal{C} is used, and also

$$L(P, Q) = \min_{\mathcal{C}} L_{\mathcal{C}}(P, Q) \quad (1.5)$$

the average number of expected checkings for a best code.

A natural special case is the mean number of expected checkings

$$\bar{L}_{\mathcal{C}}(P) = \sum_{u=1}^N \frac{1}{N} L_{\mathcal{C}}(P, u), \quad \text{if } \mathcal{U} = [N] \quad (1.6)$$

which equals $L_{\mathcal{C}}(P, Q)$ for $Q = (\frac{1}{N}, \dots, \frac{1}{N})$, and

$$\bar{L}(P) = \min_{\mathcal{C}} \bar{L}_{\mathcal{C}}(P). \quad (1.7)$$

Another special case of some "intuitive appeal" is the case $Q = P$. Here we write

$$L(P, P) = \min_{\mathcal{C}} L_{\mathcal{C}}(P, P). \quad (1.8)$$

It is known that Huffman codes minimize the expected code length for PC.

This is not always the case for $L(P)$ and the other quantities in identification.

In this correspondence an important incentive comes from Theorem 4 of [1].

For $P^N = (2^{-\ell_1}, \dots, 2^{-\ell_N})$, that is with 2-powers as probabilities $L(P^N, P^N) = H_I(P^N)$. Here the assumption means that there is a **complete** prefix code (i.e., equality holds in Kraft's equality).

B. A Terminology Involving Proper Common Prefixes

The quantity $L_{\mathcal{C}}(P, Q)$ is defined below also for the case of not necessarily independent U and V . It is conveniently described in a terminology involving proper common prefixes. For an encoding $c : \mathcal{U} \rightarrow \mathcal{X}^*$, we define for two words $w, w' \in \mathcal{X}^*$ $cp(w, w')$ as the number of proper common prefixes including the empty word, which equals the length of the maximal proper common prefix plus 1.

For example, $cp(11\ 000) = 1$, $cp(0110\ 0100) = 3$, and $cp(1001\ 1000) = 4$ (since the proper common prefixes are $\emptyset, 01, 100$).

Now with encoding c for PC \mathcal{C} and RVs U and V , $cp(c_U, c_V)$ measures the time steps it takes to decide whether U and V are equal, that is, the checking time or waiting time, which we denote by

$$W_{\mathcal{C}}(U, V) = cp(c_U, c_V). \quad (1.9)$$

Clearly, we can write the expected waiting time as

$$\mathbb{E}W_{\mathcal{C}}(U, V) = \mathbb{E}cp(c_U, c_V). \quad (1.10)$$

It is readily verified that for independent U, V , that is, $\text{Pr}(U = u, V = v) = P_u Q_v$

$$\mathbb{E}W_{\mathcal{C}}(U, V) = L_{\mathcal{C}}(P, Q) = \mathbb{E}cp(c_U, c_V). \quad (1.11)$$

We give now another description for $\mathbb{E}W_{\mathcal{C}}(U, V)$. For a word $w \in \mathcal{X}^*$ and a code \mathcal{C} define as subset of \mathcal{U}

$$\mathcal{U}(\mathcal{C}, w) = \{u \in \mathcal{U} : c_u \text{ has proper prefix } w\} \quad (1.12)$$

Manuscript received April 13, 2005; revised April 4, 2006.
Both authors are with the Fakultät für Mathematik, University of Bielefeld, D-33501 Bielefeld, Germany (e-mail: hollmann@math.uni-bielefeld.de; cai@mathematik.uni-bielefeld.de).

Communicated by S. A. Savari, Associate Editor for Source Coding.

Digital Object Identifier 10.1109/TIT.2006.879972

and its indicator function $1_{\mathcal{U}(\mathcal{C}, w)}$. Now

$$\begin{aligned} \mathbb{E} cp(c_U, c_V) &= \sum_{u, v \in \mathcal{U}} \Pr(U = u, V = v) cp(c_u, c_v) \\ &= \sum_{u, v \in \mathcal{U}} \Pr(U = u, V = v) \sum_w 1_{\mathcal{U}(\mathcal{C}, w)}(u) 1_{\mathcal{U}(\mathcal{C}, w)}(v) \\ &= \sum_w \Pr(U \in \mathcal{U}(\mathcal{C}, w), V \in \mathcal{U}(\mathcal{C}, w)) \end{aligned}$$

and by (1.11)

$$\mathbb{E} W_{\mathcal{C}}(U, V) = \sum_w \Pr(U \in \mathcal{U}(\mathcal{C}, w), V \in \mathcal{U}(\mathcal{C}, w)). \quad (1.13)$$

C. Matrix Notation

Next we look at the double infinite matrix

$$\Lambda = (cp(w, w'))_{w \in \mathcal{X}^*, w' \in \mathcal{X}^*} \quad (1.14)$$

and its minor $\Lambda^{(L)}$ labeled by sequences in $\mathcal{X}^{\leq L}$.

Henceforth we assume that U and V are independent and have distributions P and Q . We can then use (1.11) For a prefix code \mathcal{C} P induces, the distribution $P_{\mathcal{C}}$ and Q induces the distribution $Q_{\mathcal{C}}$, when for $u, v \in \mathcal{U}$

$$P_{\mathcal{C}}(c_u) = P_u, Q_{\mathcal{C}}(c_v) = Q_v \quad (1.15)$$

and

$$P_{\mathcal{C}}(x) = Q_{\mathcal{C}}(x) = 0, \quad \text{for } x \in \mathcal{X}^* \setminus \mathcal{C}. \quad (1.16)$$

Viewing both $P_{\mathcal{C}}$ and $Q_{\mathcal{C}}$ as row vectors, for the corresponding column vector $Q_{\mathcal{C}}^T$ (1.11) can be written in the form

$$L_{\mathcal{C}}(P, Q) = P_{\mathcal{C}} \Lambda Q_{\mathcal{C}}^T. \quad (1.17)$$

It is clear from (1.10) that a noncomplete prefix code, that is one for which the Kraft sum is smaller than 1, can be improved for identification by shortening a suitable codeword. Hence, **an optimal ID source code is necessarily complete**. In such a code

$$\max_{u \in \mathcal{U}} \|c_u\| \leq |\mathcal{U}| - 1 \quad (1.18)$$

and one can replace Λ by its submatrix $\Lambda^{(L)}$ for $L = |\mathcal{U}| - 1$. This implies

$$L_{\mathcal{C}}(P, Q) = P_{\mathcal{C}}^{(L)} \Lambda^{(L)} (Q_{\mathcal{C}}^{(L)})^T, \quad (1.20)$$

where $P_{\mathcal{C}}^{(L)}$ and $Q_{\mathcal{C}}^{(L)}$ are row vectors obtained by deleting the components $y \notin \mathcal{X}^{\leq L}$.

Sometimes, the expressions (1.17) or (1.19) are more convenient for the investigation of $L_{\mathcal{C}}(P, Q)$. For example, it is easy to see that Λ and, therefore, also $\Lambda^{(L)}$ are positive semidefinite. Indeed, let Δ (resp., $\Delta^{(L)}$) be a matrix whose rows are labeled by sequences in \mathcal{X}^* (resp., $\mathcal{X}^{\leq L}$) and whose columns are labeled by sequences in \mathcal{X}^* (resp., $\mathcal{X}^{\leq L-1} \cup \{\text{empty sequence}\}$) such that its (x, y) -entry is

$$\delta_y^*(x) = \begin{cases} 1, & \text{if } y \text{ is a proper prefix of } x \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\Delta \Delta^T = \Lambda \quad \text{and} \quad \Delta^{(L)} (\Delta^{(L)})^T = \Lambda^{(L)} \quad (1.21)$$

and hence Λ and $\Lambda^{(L)}$ are positive semidefinite. Therefore, by (1.19), $L_{\mathcal{C}}(P, P)$ is (\mathcal{U}) -convex in P .

Furthermore, for sources (\mathcal{U}, P) with $|\mathcal{U}| = 2^k$ and the block code $\mathcal{C} = \{0, 1\}^k$, the uniform distribution on \mathcal{U} achieves $\min_P L_{\mathcal{C}}(P, P)$.¹

Another interesting observation on (1.20) is that as the w th component of $P_{\mathcal{C}}^{(L)} \Delta^{(L)}$ (resp., $Q_{\mathcal{C}}^{(L)} \Delta^{(L)}$) is $P(\mathcal{U}(\mathcal{C}, w))$ (resp., $Q(\mathcal{U}(\mathcal{C}, w))$), application of the Cauchy-Schwarz inequality to (1.20) yields

$$\begin{aligned} & \left[P_{\mathcal{C}}^{(L)} \Lambda^{(L)} (Q_{\mathcal{C}}^{(L)})^T \right]^2 \\ & \leq \left[P_{\mathcal{C}}^{(L)} \Lambda^{(L)} (P_{\mathcal{C}}^{(L)})^T \right] \cdot \left[Q_{\mathcal{C}}^{(L)} \Lambda^{(L)} (Q_{\mathcal{C}}^{(L)})^T \right] \quad (1.22) \end{aligned}$$

and equality holds iff for all w

$$P(\mathcal{U}(\mathcal{C}, w)) = Q(\mathcal{U}(\mathcal{C}, w)).$$

We state this in equivalent form as follows.

Lemma 1:

$$L_{\mathcal{C}}(P, Q)^2 \leq L_{\mathcal{C}}(P, P) L_{\mathcal{C}}(Q, Q) \quad (1.23)$$

and equality holds iff for all w

$$P(\mathcal{U}(\mathcal{C}, w)) = Q(\mathcal{U}(\mathcal{C}, w)),$$

which implies $L_{\mathcal{C}}(P, Q) = L_{\mathcal{C}}(P, P) = L_{\mathcal{C}}(Q, Q)$.

This suggests to introduce

$$\mu_{\mathcal{C}}(P, Q) = \frac{L_{\mathcal{C}}(P, Q)^2}{L_{\mathcal{C}}(P, P) L_{\mathcal{C}}(Q, Q)} \leq 1$$

as a measure of similarity of sources P and Q with respect to the code \mathcal{C} .

Intuitively, we feel that for a good code for source P and Q as user distribution P and Q should be very dissimilar, because then the user waits less time until he knows that the output of U is not what he wants.

This idea will be used later for code construction. Actually, it is clear even in the general case where U and V are not necessarily independent.

To simplify the discussion, we assume here that the alphabet \mathcal{X} is binary, i.e., $q = 2$.

Then the first bit of a codeword partitions the source \mathcal{U} into two parts $\bar{\mathcal{U}}(i_1)$; $i_1 = 0, 1$; where $\bar{\mathcal{U}}(i_1) = \{u \in \mathcal{U} : c_{u1} = i_1\}$. By (1.13), to minimize $\mathbb{E} W_{\mathcal{C}}(U, V)$ one has to choose a partition such that $\Pr(U \in \bar{\mathcal{U}}(i_1), V \in \bar{\mathcal{U}}(i_1))$'s are small simultaneously for $i_1 = 0, 1$. To construct a good code one can continue this line: partition $\bar{\mathcal{U}}(i_1)$ to $\bar{\mathcal{U}}(i_1, i_2)$'s such that

$$\Pr(U \in \bar{\mathcal{U}}(i_1, i_2), V \in \bar{\mathcal{U}}(i_1, i_2) \mid U \in \bar{\mathcal{U}}(i_1), V \in \bar{\mathcal{U}}(i_1))$$

are as small as possible for $i_1, i_2 = 0, 1$, and so on.

When U and V are independent the requirement for a good code is that the difference between $P(\bar{\mathcal{U}}(i_1, \dots, i_k))$ and $Q(\bar{\mathcal{U}}(i_1, \dots, i_k))$ is large.

We call this the LOCAL UNBALANCE PRINCIPLE in contrast to the GLOBAL BALANCE PRINCIPLE below.

Another extremal case is that U and V are equal with probability one and in this case one may never use the unbalance principle. However, in this case the identification for the source makes no sense: The user knows that his output definitely comes! But still we can investigate the problem by assuming that with high probability $U = V$. More specifically, we consider the limit of $\mathbb{E} W_{\mathcal{C}}(U_k, V_k)$ for a sequence of random variables $(U_k, V_k)_{k=1}^{\infty}$ such that U_k converges to V_k

¹A proof is given in the forthcoming Ph.D. dissertation "L-identification for sources" written by C. Heup at the Department of Mathematics of the University of Bielefeld, Bielefeld, Germany.

in probability. Then it follows from Proposition 1 that $\mathbb{E} W_C(U_k, V_k)$ converges to the average length of codewords, the **classical object in source coding**! In this sense, identification for sources is a generalization of source coding (data compression).

One of the discoveries of [4] is that ID-entropy is a lower bound to $L_C(P, P)$. In Section II we repeat the original proof and we give in Section III another proof of this fact via two basic tools, Lemmas 3 and 4 for $L_C(P^n, P^n)$, where P^n is the distribution of a memoryless source. It provides a clear information-theoretical meaning of the two factors $\frac{2}{q-1}$ and $(1 - \sum_{u \in \mathcal{U}} P_u^2)$ of ID-entropy. Next we consider in Section IV sufficient and necessary conditions for a prefix code \mathcal{C} to achieve the ID-entropy lower bound for $L_C(P, P)$. *Quite surprisingly, it turns out that the ID-entropy bound for ID-time is achieved by a variable-length code iff the Shannon entropy bound for the average length of codewords is achieved by the same code* (Theorem 2).

Finally, we end the correspondence in Section V with a global balance principle to find good codes (Theorem 3).

II. AN OPERATIONAL JUSTIFICATION OF ID-ENTROPY AS LOWER BOUND FOR $L_C(P, P)$

Recall from the Introduction that for $q = 2$

$$H_I(P) = 2 \left(1 - \sum_{u=1}^N P_u^2 \right), \quad \text{for } P = (P_1 \dots P_N).$$

We repeat the first main result for $L(P, P)$ from [4].

Central in our derivation is a proof by induction based on a **decomposition formula for trees**.

Starting from the root a binary tree \mathbb{T} goes via 0 to the subtree \mathbb{T}_0 and via 1 to the subtree \mathbb{T}_1 with sets of leaves \mathcal{U}_0 and \mathcal{U}_1 , respectively. A code \mathcal{C} for (\mathcal{U}, P) can be viewed as a tree \mathbb{T} , where \mathcal{U}_i corresponds to the set of codewords \mathcal{C}_i , $\mathcal{U}_0 \cup \mathcal{U}_1 = \mathcal{U}$.

The leaves are labeled so that $\mathcal{U}_0 = \{1, 2, \dots, N_0\}$ and $\mathcal{U}_1 = \{N_0 + 1, \dots, N_0 + N_1\}$, $N_0 + N_1 = N$. Using probabilities

$$Q_i = \sum_{u \in \mathcal{U}_i} P_u, \quad i = 0, 1$$

we can give the decomposition in the following.

Lemma 2: [4] For a code \mathcal{C} for (\mathcal{U}, P^N)

$$\begin{aligned} L_C((P_1, \dots, P_N), (P_1, \dots, P_N)) \\ = 1 + L_{C_0} \left(\left(\frac{P_1}{Q_0}, \dots, \frac{P_{N_0}}{Q_0} \right), \left(\frac{P_1}{Q_0}, \dots, \frac{P_{N_0}}{Q_0} \right) \right) Q_0^2 \\ + L_{C_1} \left(\left(\frac{P_{N_0+1}}{Q_1}, \dots, \frac{P_{N_0+N_1}}{Q_1} \right), \left(\frac{P_{N_0+1}}{Q_1}, \dots, \frac{P_{N_0+N_1}}{Q_1} \right) \right) Q_1^2. \end{aligned}$$

This readily yields the following result.

Theorem 1: [4] For every source (\mathcal{U}, P^N)

$$L(P^N) \geq L(P^N, P^N) \geq H_I(P^N).$$

Proof: We proceed by induction on N . The base case $N = 2$ can be established as follows. For $N = 2$ and any \mathcal{C} $L_C(P^2, P^2) \geq P_1 + P_2 = 1$, but

$$\begin{aligned} H_I(P^2) &= 2(1 - P_1^2 - (1 - P_1)^2) \\ &= 2(2P_1 - 2P_1^2) = 4P_1(1 - P_1) \leq 1. \end{aligned}$$

For the induction step, for any code \mathcal{C} use the decomposition formula in Lemma 2 and, of course, the desired inequality for N_0 and N_1 as induction hypothesis

$$\begin{aligned} L_C((P_1, \dots, P_N), (P_1, \dots, P_N)) \\ \geq 1 + 2 \left(1 - \sum_{u \in \mathcal{U}_0} \left(\frac{P_u}{Q_0} \right)^2 \right) Q_0^2 + 2 \left(1 - \sum_{u \in \mathcal{U}_1} \left(\frac{P_u}{Q_1} \right)^2 \right) Q_1^2 \\ \geq H_I(Q) + Q_0^2 H_I(P^{(0)}) + Q_1^2 H_I(P^{(1)}) = H_I(P^N) \end{aligned}$$

where $Q = (Q_0, Q_1)$, $1 \geq H_I(Q)$, $P^{(i)} = (\frac{P_u}{Q_i})_{u \in \mathcal{U}_i}$, and the grouping identity is used for the equality. This holds for every \mathcal{C} and therefore also for $\min_C L_C(P^N)$. \square

The approach readily extends also to the q -ary case.

III. AN ALTERNATIVE PROOF OF THE ID-ENTROPY LOWER BOUND FOR $L_C(P, P)$

First, we establish Lemma 3 below, which holds for the more general case $\mathbb{E} W_C(U, V)$. Let $((U^n, V^n))_{n=1}^\infty$ be a discrete memoryless correlated source with a generic pair of variables (U, V) . Again, U^n serves as (random) source and V^n serves as random user. For a given code \mathcal{C} for (U, V) let \mathcal{C}^n be the code obtained by encoding the components of sequence $u^n \in \mathcal{U}^n$ iteratively. That is, for all $u^n \in \mathcal{U}^n$

$$c_{u^n}^n = (c_{u_1}, c_{u_2}, \dots, c_{u_n}). \quad (3.1)$$

Lemma 3:

$$\mathbb{E} W_{\mathcal{C}^n}(U^n, V^n) = \mathbb{E} W_C(U, V) \left(1 + \sum_{t=1}^{n-1} \Pr(U^t = V^t) \right) \quad (3.2)$$

and therefore,

$$\lim_{n \rightarrow \infty} \mathbb{E} W_{\mathcal{C}^n}(U^n, V^n) = \frac{\mathbb{E} W_C(U, V)}{1 - \Pr(U = V)}. \quad (3.3)$$

Proof: Since

$$\Pr(U^n = V^n) = \prod_{t=1}^n \Pr(U_t = V_t) = \Pr^n(U = V)$$

(3.3) follows from (3.2) immediately by the summation formula for geometric series.

To show (3.2), we define first for all $t \geq 2$ random variables

$$Z_t = \begin{cases} 0, & \text{if } U^{t-1} \neq V^{t-1} \\ 1, & \text{otherwise} \end{cases} \quad (3.4)$$

and for $t = 1$ we let Z_1 be a constant for convenience of notation. Further, we let W_t be the waiting time for the random user V^n in the t th block.

Conditional on $Z_t = 1$ it is defined like $W_C(U, V)$ in (1.9) and conditional on $Z_t = 0$ obviously $\Pr(W_t = 0 | Z_t = 0) = 1$, because the random user has made his decision before the t 's step. Moreover, by the definition of \mathcal{C}^n

$$\mathbb{E}[W_t | Z_t = 1] = \mathbb{E} W_C(U, V) \quad (3.5)$$

and, consequently,

$$\begin{aligned} \mathbb{E}[\mathbb{E}(W_t | Z_t)] \\ = \begin{cases} \Pr(U^{t-1} = V^{t-1}) \mathbb{E} W_C(U, V), & \text{for } t = 2, 3, \dots, n \\ \mathbb{E} W_C(U, V), & \text{for } t = 1 \end{cases} \end{aligned} \quad (3.6)$$

where (3.6) holds in case $t = 1$, because the random user has to wait for the first outcome. Therefore it follows that

$$\begin{aligned}\mathbb{E}W_{\mathcal{C}^n}(U^n, V^n) &= \mathbb{E}W^n = \sum_{t=1}^n \mathbb{E}W_t = \sum_{t=1}^n \mathbb{E}[\mathbb{E}(W_t | Z_t)] \\ &= \mathbb{E}W_{\mathcal{C}}(U, V) + \sum_{t=1}^{n-1} \Pr(U^t, V^t) \mathbb{E}W_{\mathcal{C}}(U, V)\end{aligned}$$

as we wanted to show.

Next we consider the case where U and V are independent and identically distributed with distribution P so that

$$\Pr(U^n = u^n, V^n = v^n) = \prod_{t=1}^n P_{u_t} \cdot P_{v_t}. \quad (3.7)$$

More specifically, we are looking for a lower bound on $L_{\mathcal{C}}(P^n, P^n)$ for **all** prefix codes \mathcal{C} over \mathcal{U}^n .

Lemma 4: For all $\varepsilon > 0$ there exists an $\eta > 0$ such that for sufficiently large n and all positive integers

$$L_n = \lfloor n(H(P) - \varepsilon)(\log q)^{-1} \rfloor \quad (3.8)$$

for all prefix codes \mathcal{C} over \mathcal{U}^n

$$L_{\mathcal{C}}(P^n, P^n) > (1 - 2^{-n\eta}) \sum_{t=0}^{L_n-1} q^{-t}. \quad (3.9)$$

Proof: For given $\varepsilon > 0$ we choose $\delta > 0$ such that for a $\tau > 0$ and sufficiently large n for familiar sets $T_{P,\delta}^n$ of typical sequences

$$P^n(T_{P,\delta}^n) > 1 - 2^{-n\tau}$$

and for all $u^n \in T_{P,\delta}^n$

$$P(u^n) < 2^{-n(H(P) - \frac{\varepsilon}{2})}.$$

Since for a prefix code \mathcal{C}

$$|\{u^n \in \mathcal{U}^n : \|c_{U^n}\| \leq L_n\}| \leq q^{L_n} \quad (3.10)$$

$$\begin{aligned}\Pr(\|c_{U^n}\| \leq L_n) &= \Pr(\|c_{V^n}\| \leq L_n) \\ &\leq \Pr(V^n \notin T_{P,\delta}^n) + \Pr(V^n \in T_{P,\delta}^n, \|c_{V^n}\| \leq L_n) \\ &< 2^{-n\tau} + |\{u^n : \|c_{u^n}\| \leq L_n\}| \cdot 2^{-n(H(P) - \frac{\varepsilon}{2})} \\ &\leq 2^{-n\tau} + q^{L_n} 2^{-n(H(P) - \frac{\varepsilon}{2})}.\end{aligned} \quad (3.11)$$

However, (3.8) implies that

$$q^{L_n} \leq 2^{n(H(P) - \varepsilon)}.$$

This together with (3.11) yields

$$\Pr(\|c_{U^n}\| \leq L_n) < 2^{-n\tau} + 2^{-n\frac{\varepsilon}{2}} < 2^{-n\delta} \quad (3.12)$$

for $\delta \triangleq \min(\frac{\tau}{2}, \frac{\varepsilon}{4})$.

Next, for the distribution P and the code \mathcal{C} over \mathcal{U}^n we construct a related source (\tilde{U}, \tilde{P}) and a code $\tilde{\mathcal{C}}$ over $\tilde{\mathcal{U}}$ as follows.

The new set $\tilde{\mathcal{U}}$ contains $\{u^n \in \mathcal{U}^n : \|c_{u^n}\| \leq L_n\}$ and for its elements $\tilde{P}(u^n) = P^n(u^n)$ and the new \sim -coding is $\tilde{c}_{u^n} = c_{u^n}$.

Now we define the additional elements in $\tilde{\mathcal{U}}$ with its \tilde{P} and $\tilde{\mathcal{C}}$.

We partition $\{u^n \in \mathcal{U}^n : \|c_{u^n}\| > L_n\}$ into subsets S_j ($1 \leq j \leq J$) according to the L_n th prefix and use letter g_j to represent S_j and put the set $\tilde{\mathcal{U}} = \{g_j : 1 \leq j \leq J\}$ into $\tilde{\mathcal{U}}$ so that

$$\tilde{\mathcal{U}} = \{u^n \in \mathcal{U}^n : \|c_{u^n}\| \leq L_n\} \cup \tilde{\mathcal{U}}.$$

Then we define $\tilde{P}(g_j) = \sum_{u^n \in S_j} P(u^n)$ and let \tilde{c}_{g_j} be the common L_n th prefix of the c_{u^n} 's for the u^n 's in S_j . That is, we consider all u^n sharing the same L_n th prefix in c_{u^n} as a single element. Obviously

$$L_{\mathcal{C}}(P^n, P^n) \geq L_{\tilde{\mathcal{C}}}(\tilde{P}, \tilde{P}). \quad (3.13)$$

Finally, let \tilde{U}_n and \tilde{V}_n be random variables for the new source and new random user with distribution \tilde{P} and let Z be a random variable such that

$$Z = \begin{cases} 0, & \text{if both } \|c_{U^n}\| \text{ and } \|c_{V^n}\| \text{ are larger than } L_n \\ 1, & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned}L_{\tilde{\mathcal{C}}}(\tilde{P}, \tilde{P}) &= \mathbb{E}W = \mathbb{E}(W | Z) \geq \Pr(Z = 0) \mathbb{E}(W | Z = 0) \\ &= \Pr(\|c_{U^n}\| \geq L_n) \Pr(\|c_{V^n}\| \geq L_n) \cdot L_{\tilde{\mathcal{C}}}(\tilde{P}, \tilde{P})\end{aligned} \quad (3.14)$$

where W is the random waiting time, \tilde{P} is the common conditional distribution of \tilde{U}_n given $\tilde{U}_n \in \tilde{\mathcal{U}}$, and \tilde{V}_n given $\tilde{V}_n \in \tilde{\mathcal{U}}$, i.e.,

$$\tilde{P}(g_j) = \frac{\tilde{P}(g_j)}{\tilde{P}(\tilde{\mathcal{U}})}, \quad \text{for } g_j \in \tilde{\mathcal{U}}$$

and $\tilde{\mathcal{C}}$ is the restriction of \mathcal{C} to $\tilde{\mathcal{U}}$.

Notice that $\tilde{\mathcal{C}}$ is a block code of length L_n . In order to bound $L_{\tilde{\mathcal{C}}}(\tilde{P}, \tilde{P})$ we extend $\tilde{\mathcal{U}}$ to a set of cardinality q^{L_n} in the case of necessity and assign zero probabilities and a codeword of length L_n not in $\tilde{\mathcal{C}}$. This little modification obviously does not change the value of $L_{\tilde{\mathcal{C}}}(\tilde{P}, \tilde{P})$. Thus, if we denote the uniform distribution over the extended set $\tilde{\mathcal{U}}$ by \bar{P} , we have

$$L_{\tilde{\mathcal{C}}}(\tilde{P}, \tilde{P}) \geq L_{\tilde{\mathcal{C}}}(\bar{P}, \bar{P}) \quad (3.15)$$

where $\tilde{\mathcal{C}}$ is a bijective block code $\tilde{\mathcal{U}} \rightarrow \mathcal{X}^{L_n}$.

It is clear that $\mathcal{U}(\tilde{\mathcal{C}}, \omega) \neq \emptyset$ iff the length of ω is smaller than $L_n - 1$ and

$$\mathcal{U}(\tilde{\mathcal{C}}, \omega) = \mathcal{X}^{L_n-1}, \quad \text{if } \|\omega\| = \ell \leq L_n - 1.$$

Then it follows from (3.13) that

$$L_{\tilde{\mathcal{C}}}(\bar{P}, \bar{P}) = \sum_{t=0}^{L_n-1} q^t [q^{L_n-t} \cdot q^{-L_n}]^2 = \sum_{t=0}^{L_n-1} q^{-t}. \quad (3.16)$$

Finally, we combine (3.12), (3.13), (3.14), (3.15) and (3.16) and Lemma 4 follows.

An immediate consequence is the following corollary.

Corollary 1:

$$\lim_{n \rightarrow \infty} L(P^n, P^n) \geq \sum_{t=0}^{\infty} q^{-t} = \frac{q}{q-1}. \quad (3.17)$$

Furthermore, for independent and identically distributed random variables U, V with distribution P we have

$$\Pr(U = V) = \sum_{u \in \mathcal{U}} P_u^2$$

and from (3.3) and (3.17) follows the ID-entropy bound.

Corollary 2: (See [4, Theorem 2])

$$L_C(P, P) \geq \frac{q}{q-1} \left(1 - \sum_{u \in \mathcal{U}} P_u^2 \right). \quad (3.18)$$

This derivation provides a clear information-theoretical meaning to the two factors in ID-entropy: $\frac{q}{q-1}$ is a universal lower bound on the ID-waiting time for a discrete memoryless source with an independent user having the same distribution P . $\frac{1}{1 - \sum_{u \in \mathcal{U}} P_u^2}$ is the cost paid for coding the source componentwise and leaving time for the random user in the following sense.

Let us imagine the following procedure.

At a unit of time, the random source U^n outputs a symbol U_t and the random user V^n , who wants to know whether $U^n = V^n$, checks whether U_t coincides with his own symbol V_t . He will end if not. Then the waiting time for him is ℓ with probability

$$\begin{aligned} \Pr(U^{\ell-1} = V^{\ell-1}) \Pr(U_\ell \neq V_\ell) \\ = \Pr(U = V)^{\ell-1} (1 - \Pr(U = V)), \quad \text{for } \ell \leq n. \end{aligned}$$

Letting $n \rightarrow \infty$ we obtain a geometric distribution.

The expected waiting time is

$$\begin{aligned} \mathbb{E}W &= \sum_{\ell=0}^{\infty} \ell \Pr(U = V)^{\ell-1} (1 - \Pr(U = V)) \\ &= \sum_{\ell=0}^{\infty} (\ell+1) \Pr(U = V)^{\ell} - \sum_{\ell=0}^{\infty} \Pr(U = V)^{\ell} \\ &= \sum_{\ell=0}^{\infty} \Pr(U = V)^{\ell} = \frac{1}{1 - \Pr(U = V)} \end{aligned} \quad (3.19)$$

which equals $\frac{1}{1 - \sum_u P_u^2}$ in the case of independent and identically distributed random variables.

(Actually (3.2) holds for all stationary sources and we choose a memoryless source for simplicity.) In general, (3.3) has the form

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}W_{C^n}(U^n, V^n) \\ = \mathbb{E}W_C(U, V) \cdot \lim_{n \rightarrow \infty} \left(1 + \sum_{t=1}^{n-1} \Pr(U^t = V^t) \right). \end{aligned} \quad (3.20)$$

By monotonicity, the limit at the right-hand side and therefore also at the left-hand side exists and equals a positive finite or infinite value.

When it is finite, one may replace $\Pr(U = V)^{t-1}$, $\Pr(U = V)$, and $\Pr(U = V)^t$ in the first lines of (3.19) by $\Pr(U^{t-1} = V^{t-1})$, $\Pr(U_t = V_t | U^{t-1} = V^{t-1})$, and $\Pr(U^t = V^t)$, respectively, and obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(1 + \sum_{t=1}^{n-1} \Pr(U^t = V^t) \right) \\ = \sum_{t=0}^{\infty} t \Pr(U^{t-1} = V^{t-1}) \cdot \Pr(U_t \neq V_t | U^{t-1} = V^{t-1}) \\ = \mathbb{E}L, \end{aligned} \quad (3.21)$$

the expectation of random leaving time L for a stationary source.

Thus (3.20) is rewritten as

$$\lim_{n \rightarrow \infty} \mathbb{E}W_{C^n}(U^n, V^n) = \mathbb{E}W_C(U, V) \mathbb{E}L. \quad (3.22)$$

Now the information-theoretical meaning of (3.22) is quite clear. One encodes a source $(U^n, V^n)_{n=1}^{\infty}$ with alphabet \mathcal{U} component by component by a variable length code \mathcal{C} . The first term at the right-hand side of (3.22) is the expected waiting time in a block and the second term is the expected waiting time for different U_t and V_t .

IV. SUFFICIENT AND NECESSARY CONDITIONS FOR A PREFIX CODE \mathcal{C} TO ACHIEVE THE ID-ENTROPY LOWER BOUND OF $L_C(P, P)$

Quite surprisingly, the ID-entropy bound to ID-waiting time is achieved by a variable-length code iff the Shannon entropy bound to the average lengths of codewords is achieved by the same code.

For the proof, we use a simple consequence of the Cauchy-Schwarz inequality, which states for two sequences of real numbers (a_1, a_2, \dots, a_k) and (b_1, b_2, \dots, b_k) that

$$\left(\sum_{i=1}^k a_i b_i \right)^2 \leq \left(\sum_{i=1}^k a_i^2 \right) \left(\sum_{i=1}^k b_i^2 \right) \quad (4.1)$$

with equality iff for some constant, say γ , $a_i = \gamma b_i$ for all i or $b_i = \gamma a_i$ for all i .

Choosing $b_i = 1$ for all i one has

$$\left(\sum_{i=1}^k a_i \right)^2 \leq k \sum_{i=1}^k a_i^2 \quad (4.2)$$

with equality iff $a_1 = a_2 = \dots = a_k$.

Theorem 2: Let \mathcal{C} be a prefix code. Then the following statements are equivalent.

- i) $\sum_{u \in \mathcal{U}} P_u \|c_u\| = H(P)$.
- ii) For all $\omega \in \mathcal{X}^*$ with $\mathcal{U}(\mathcal{C}, \omega) \neq \emptyset$

$$P(\mathcal{U}(\mathcal{C}, \omega)) = q^{-\|\omega\|} \quad (4.3)$$

and for all $u, u' \in \mathcal{U}$ such that $\|c_u\| = \|c_{u'}\|$ and such that c_u and $c_{u'}$ share the same prefix of length $\|c_u\| - 1$ implies

$$P_u = P_{u'}. \quad (4.4)$$

iii)

$$L_C(P, P) = \frac{q}{q-1} \left(1 - \sum_{u \in \mathcal{U}} P_u^2 \right). \quad (4.5)$$

Proof: It is well known that i) is equivalent to

i') For all $u \in \mathcal{U}$

$$\|c_u\| = -[\log q]^{-1} \log P_u \text{ or } P_u = q^{-\|c_u\|}. \quad (4.6)$$

Notice that for i) the code \mathcal{C} is necessarily complete. We shall show that

$$\text{i') } \Rightarrow \text{ii) } \Rightarrow \text{iii) } \Rightarrow \text{i').}$$

Ad i') \Rightarrow ii): For all ω with $\mathcal{U}(\mathcal{C}, \omega) \neq \emptyset$, the code \mathcal{C}_ω obtained by deleting the common prefix ω from all the codewords c_u , $u \in \mathcal{U}(\mathcal{C}, \omega)$, is a complete code on $\mathcal{U}(\mathcal{C}, \omega)$, because \mathcal{C} is a complete code. That is,

$$\sum_{u \in \mathcal{U}(\mathcal{C}, \omega)} q^{-[\|c_u\| - \|\omega\|]} = 1$$

and, consequently, by (4.6)

$$\begin{aligned} P(\mathcal{U}(\mathcal{C}, \omega)) &= \sum_{u \in \mathcal{U}(\mathcal{C}, \omega)} P_u = \sum_{u \in \mathcal{U}(\mathcal{C}, \omega)} q^{-\|c_u\|} \\ &= q^{-\|\omega\|} \sum_{u \in \mathcal{U}(\mathcal{C}, \omega)} q^{(\|c_u\| - \|\omega\|)} = q^{-\|\omega\|}. \end{aligned}$$

Ad ii) \Rightarrow iii): Suppose (4.3) holds for all ω and we prove iii) by induction on $\ell_{\max}(C) = \max_{u \in \mathcal{U}} |c_u|$.

In case $\ell_{\max}(C) = 1$, both sides of (4.5) are one. Assume iii) holds for all codes C' with $\ell_{\max}(C') \leq L-1$ and let $\ell_{\max}(C) = L$. Let $\mathcal{U}_1(C)$ and $\mathcal{U}_{(\alpha)}(C)$ be as in the proof of (1.11) and let $C_{(\alpha)}$ be the prefix code for the source with alphabet $\mathcal{U}_{(\alpha)}(C)$ and distribution $P_{(\alpha)}$ such that for all $u \in \mathcal{U}_{(\alpha)}(C)$ and $\mathcal{X}' = \{c_u : u \in \mathcal{U}_1(C)\}$

$$P_{(\alpha)}(u) = P^{-1}(\mathcal{U}_{(\alpha)}(C))P_u.$$

Then (4.3) and (4.4) imply that ii) holds for all $C_{(\alpha)}$, $\alpha \in \mathcal{U}_1(C)$, and for all $\beta \in \mathcal{U}_1(C)$

$$P_\beta = |\mathcal{U}_1(C)|^{-1} P(\mathcal{U}_1(C)). \quad (4.7)$$

Next, we apply (4.3) to all ω with $\mathcal{U}(C, \omega)$ and $\|\omega\| = 1$ and obtain

$$\Pr(U \notin \mathcal{U}_1(C)) = (q - |\mathcal{U}_1(C)|)q^{-1} \quad (4.8)$$

which with (4.7) yields for all $\beta \in \mathcal{U}_1(C)$

$$P_\beta = q^{-1}. \quad (4.9)$$

Moreover, by the induction hypothesis for all $C_{(\alpha)}$ and $P_{(\alpha)}$, $\alpha \in \mathcal{U}_1(C)$

$$L_{C_{(\alpha)}}(P_{(\alpha)}, P_{(\alpha)}) = \frac{q}{q-1} \left(1 - q^2 \sum_{u \in \mathcal{U}_{(\alpha)}(C)} P_u^2 \right) \quad (4.10)$$

as by (4.3)

$$P(\mathcal{U}_{(\alpha)}(C)) = q^{-1} \quad (4.11)$$

for all $\alpha \in \mathcal{X}^\Delta = \mathcal{X} \setminus \{c_u : u \in \mathcal{U}_1(C)\}$ (say).

Finally, like in the proof of (1.11), we have

$$\begin{aligned} L_C(P, P) &= 1 + \sum_{\alpha \in \mathcal{X}^\Delta} P^2(\mathcal{U}_{(\alpha)}(C)) L_{C_{(\alpha)}}(P_{(\alpha)}, P_{(\alpha)}) \\ &= 1 + \sum_{\alpha \in \mathcal{X}^\Delta} \frac{(1}{q(q-1)} \left[1 - q^2 \sum_{u \in \mathcal{U}_{(\alpha)}(C)} P_u^2 \right] \\ &= 1 + \frac{|\mathcal{X}^\Delta|}{q(q-1)} - \frac{q}{q-1} \sum_{u \notin \mathcal{U}_1(C)} P_u^2 \\ &= 1 + \frac{q - |\mathcal{U}_1(C)|}{q(q-1)} - \frac{q}{q-1} \sum_{u \in \mathcal{U}} P_u^2 \\ &\quad + \frac{q}{q-1} |\mathcal{U}_1(C)| q^{-2} \\ &= \frac{q}{q-1} \left(1 - \sum_{u \in \mathcal{U}} P_u^2 \right), \quad \text{that is (4.5)} \end{aligned}$$

where the second equality holds by (4.10), the third equality holds, because $\{\mathcal{U}_1(C), \mathcal{U}_{(\alpha)}(C), \alpha \in \mathcal{X}'\}$ is a partition of \mathcal{U} , and the fourth equality follows from (4.9) and the definition of \mathcal{X}^Δ .

Ad iii) \Rightarrow i') : Again we proceed by induction on the maximum length of codewords.

Suppose first that for a code C $\ell_{\max}(C) = 1$. Then $L_C(P, P) = 1$ and $|\mathcal{U}| \leq q$. Applying (4.2) to the ID-entropy we get

$$\frac{q}{q-1} \left(1 - \sum_{u \in \mathcal{U}} P_u^2 \right) \leq \frac{q}{q-1} (1 - |\mathcal{U}|^{-1})$$

with equality iff P is the uniform distribution. On the other hand, since $|\mathcal{U}| \leq q$, $\frac{q}{q-1} (1 - |\mathcal{U}|^{-1}) \leq \frac{q}{q-1} (1 - \frac{1}{q}) = 1$, and the equality holds iff $|\mathcal{U}| = q$. Then (4.5) holds iff P is uniform and $|\mathcal{U}| = q$, i.e., (4.6).

Assume now that the implication iii) \Rightarrow i') holds for all codes with maximum lengths $\leq L-1$ and that C is a prefix code of maximum length $\ell_{\max}(C) = L$.

Without loss of generality, we can assume that C is complete, because otherwise we can add "dummy" symbols with 0 probability to \mathcal{U} and assign to them suitable codewords so that the Kraft sum equals 1, but this does not change equality (4.5).

Having completeness, we can assume that for $(ak) \leq q^{L-1}$ there are kq symbols $u(i, j)$ ($1 \leq i \leq k, 0 \leq j \leq q-1$) in \mathcal{U} with $\|c_{u(i, j)}\| = L$ and such that $c_{u(i, 0)}, c_{u(i, 1)}, \dots, c_{u(i, q-1)}$ share a prefix ω_i of length $L-1$ for $i = 1, 2, \dots, k$.

Let $u(1), \dots, u(k)$ be k "new symbols" not in the original \mathcal{U} and consider

$$\mathcal{U}' = [\mathcal{U}\{u(i, j) : 1 \leq i \leq k, 0 \leq j \leq q-1\}] \cup \{u(i) : 1 \leq i \leq k\}$$

and the probability distribution P' defined by

$$P'_{u'} = \begin{cases} P_{u'}, & \text{if } u' \in \mathcal{U} \cap \mathcal{U}' \\ \sum_{j=0}^{q-1} P_{u(i, j)}, & \text{if } u' = u(i) \text{ for some } i. \end{cases} \quad (4.12)$$

Next we define a prefix code C' for the source (\mathcal{U}', P') by using C as follows:

$$c'_{u'} = \begin{cases} c_{u'}, & \text{if } u' \in \mathcal{U} \cap \mathcal{U}' \\ \omega_i, & \text{if } u' = u(i) \text{ for some } i. \end{cases} \quad (4.13)$$

Then for $u' \in \mathcal{U} \cap \mathcal{U}'$, $\|c'_{u'}\| = \|c_{u'}\|$ and $\|c'_{u(1)}\| = \|c'_{u(2)}\| = \dots = \|c'_{u(k)}\| = L-1$.

Therefore, by induction hypothesis

$$L_{C'}(P', P') \geq \frac{q}{q-1} \left(1 - \sum_{u' \in \mathcal{U}'} P_{u'}^2 \right) \quad (4.14)$$

and equality holds iff $P_u = q^{-\|c_u\|}$ for $u \in \mathcal{U} \cap \mathcal{U}'$ and

$$\sum_{j=0}^{q-1} P_{u(i, j)} = P'_{u(i)} = q^{-(L-1)}, \quad \text{for } i = 1, 2, \dots, k.$$

Furthermore, it follows from (4.2) and the definition of $L_C(P, P)$ and $L_{C'}(P', P')$ that

$$\begin{aligned} L_C(P, P) &= L_{C'}(P', P') + \sum_{i=1}^k \left(\sum_{j=0}^{q-1} P_{u(i, j)} \right)^2 \\ &= L_{C'}(P', P') + \sum_{i=1}^k P_{u(i)}^2 \\ &\geq \frac{q}{q-1} \left(1 - \sum_{u' \in \mathcal{U}'} P_{u'}^2 \right) + \sum_{i=1}^k P_{u(i)}^2 \\ &= \frac{q}{q-1} \left(1 - \sum_{u \in \mathcal{U} \cap \mathcal{U}'} P_u^2 \right) + \sum_{i=1}^k \left(1 - \frac{q}{q-1} \right) P_{u(i)}^2 \\ &= \frac{q}{q-1} \left[1 - \sum_{u \in \mathcal{U} \cap \mathcal{U}'} P_u^2 - \sum_{i=1}^k q^{-1} \left(\sum_{j=0}^{q-1} P_{u(i, j)} \right)^2 \right] \\ &\geq \frac{q}{q-1} \left[1 - \sum_{u \in \mathcal{U}} P_u^2 \right]. \end{aligned} \quad (4.15)$$

By (4.13), the first inequality holds iff $P_u = q^{-\|c_u\|}$ for $u \in \mathcal{U} \cap \mathcal{U}'$ and $\sum_{j=0}^{q-1} P_{u(i, j)} = q^{-(L-1)}$ for $i = 1, 2, \dots, k$; it follows from (4.2) that the last inequality holds and with equality iff

$$P_{u(i, 0)} = P_{u(i, 1)} = \dots = P_{u(i, q-1)}, \quad \text{for } i = 1, 2, \dots, k.$$

In order to have

$$L_C(P, P) = \frac{q}{q-1} \left[1 - \sum_{u \in \mathcal{U}} P_u^2 \right]$$

the two inequalities in (4.15) must be equalities. However, this is equivalent with (4.6), i.e., i').

V. A GLOBAL BALANCE PRINCIPLE TO FIND GOOD CODES

In case U and V are independent and identically distributed, there is no gain in using the local unbalance principle (LUP). But in this case, Corollary 1 and (4.2) provide a way to find a good code. We first rewrite Corollary 1 as

$$\mathbb{E}W_C(U, V) = \sum_n \sum_{\omega \in \mathcal{X}^n} \Pr(U \in \mathcal{U}(\mathcal{C}, \omega), V \in \mathcal{U}(\mathcal{C}, \omega)).$$

By the assumptions on U and V with their distribution P

$$L_C(P, P) = \sum_n \sum_{\omega \in \mathcal{X}^n} P^2(\mathcal{U}(\mathcal{C}, \omega)). \quad (5.1)$$

Notice that in case

$$P_{n,C} \triangleq \sum_{\omega \in \mathcal{X}^n} P(\mathcal{U}(\mathcal{C}, \omega))$$

is a constant $\sum_{\omega \in \mathcal{X}^n} P^2(\mathcal{U}(\mathcal{C}, \omega))$ is minimized by choosing the $P(\mathcal{U}(\mathcal{C}, \omega))$'s uniformly. This gives us a global balance principle (GBT) for finding good codes.

We shall see the roles of both, the LUP and the GBP in the proof of the following coding theorem for discrete memoryless sources (DMSs).

Theorem 3: For a DMS $(U^n, V^n)_{n=1}^\infty$ with generic distribution $P_{UV} = PQ$, i.e., the generic random variables U and V are independent and $P_U = P, P_V = Q$

$$\lim_{n \rightarrow \infty} L(P^n, Q^n) = \begin{cases} 1, & \text{if } P \neq Q \\ \frac{q}{q-1}, & \text{if } P = Q. \end{cases} \quad (5.2)$$

Proof: Trivially, $L_C(P, Q) \geq 1$ and by Corollary 2, $\frac{q}{q-1}$ is a lower bound to $\lim_{n \rightarrow \infty} L(P^n, P^n)$. Hence, we only have to construct codes to achieve asymptotically the bounds in (5.2).

Case $P \neq Q$: We choose a $\delta > 0$ so that for sufficiently large n

$$T_{P,\delta}^n \cap T_{Q,\delta}^n = \emptyset \quad (5.3)$$

and for a $\theta > 0$

$$P(T_{P,\delta}^n) > 1 - 2^{-n\theta} \quad \text{and} \quad Q(T_{Q,\delta}^n) > 1 - 2^{-n\theta}. \quad (5.4)$$

Partition \mathcal{U}^n into two parts \mathcal{U}_0 and \mathcal{U}_1 such that $\mathcal{U}_0 \supset T_{P,\delta}^n$ and $\mathcal{U}_1 \supset T_{Q,\delta}^n$.

To simplify matters, we assume $q = 2$. This does not loose generality since enlarging the alphabet cannot make things worse.

Let $\ell_i = \lceil \log |\mathcal{U}_i| \rceil$ and $\psi_i : \mathcal{U}_i \rightarrow 2^{[\ell_i]}$ for $i = 1, 2$. Then we define a code \mathcal{C} by $c_{u^n} = (i, \psi_i(u^n))$ if $u^n \in \mathcal{U}_i$ and show that $L_C(P^n, Q^n)$ is arbitrarily close to one if n is sufficiently large. Actually, it immediately follows from Proposition 1

$$\begin{aligned} L_C(P^n, Q^n) &= \sum_{u^n, u'^n \in \mathcal{U}^n} P^n(c_{u^n}) Q^n(c_{u'^n}) cp(c_{u^n}, c_{u'^n}) \\ &= \sum_{u^n \in \mathcal{U}_0} \sum_{u'^n \in \mathcal{U}_0} P^n(c_{u^n}) Q^n(c_{u'^n}) cp(c_{u^n}, c_{u'^n}) \\ &\quad + \sum_{u^n \in \mathcal{U}_0} \sum_{u'^n \in \mathcal{U}_1} P^n(c_{u^n}) Q^n(c_{u'^n}) cp(c_{u^n}, c_{u'^n}) \\ &\quad + \sum_{u^n \in \mathcal{U}_1} \sum_{u'^n \in \mathcal{U}_0} P^n(c_{u^n}) Q^n(c_{u'^n}) cp(c_{u^n}, c_{u'^n}) \\ &\quad + \sum_{u^n \in \mathcal{U}_1} \sum_{u'^n \in \mathcal{U}_1} P^n(c_{u^n}) Q^n(c_{u'^n}) cp(c_{u^n}, c_{u'^n}) \\ &< \ell_0 \sum_{u^n \in \mathcal{U}_0} P^n(c_{u^n}) \sum_{u'^n \in \mathcal{U}_0} Q^n(c_{u'^n}) \end{aligned}$$

$$\begin{aligned} &+ \sum_{u^n \in \mathcal{U}_0} P^n(c_{u^n}) \sum_{u'^n \in \mathcal{U}_1} Q^n(c_{u'^n}) \\ &+ \sum_{u^n \in \mathcal{U}_1} P^n(c_{u^n}) \sum_{u'^n \in \mathcal{U}_0} Q^n(c_{u'^n}) \\ &+ \ell_1 \sum_{u^n \in \mathcal{U}_1} P^n(c_{u^n}) \sum_{u'^n \in \mathcal{U}_1} Q^n(c_{u'^n}) \\ &\leq \left[\sum_{u^n \in \mathcal{U}_0} P^n(c_{u^n}) \sum_{u'^n \in \mathcal{U}_1} Q^n(c_{u'^n}) + \sum_{u^n \in \mathcal{U}_1} P^n(c_{u^n}) \right. \\ &\quad \times \left. \sum_{u'^n \in \mathcal{U}_0} Q^n(c_{u'^n}) \right] + \lceil n \log |\mathcal{U}| \rceil \left[\sum_{u^n \in \mathcal{U}_0} (c_{u^n}) \right. \\ &\quad \times \left. \sum_{u'^n \in \mathcal{U}_0} Q^n(c_{u'^n}) + \sum_{u^n \in \mathcal{U}_1} P^n(c_{u^n}) \sum_{u'^n \in \mathcal{U}_0} Q^n(c_{u'^n}) \right] \\ &\leq 1 + \lceil n \log |\mathcal{U}| \rceil \left[\sum_{u'^n \in \mathcal{U}_0} Q^n(c_{u'^n}) + \sum_{u^n \in \mathcal{U}_1} P^n(c_{u^n}) \right] \end{aligned}$$

and therefore,

$$L_C(P^n, Q^n) < 1 + \lceil n \log |\mathcal{U}| \rceil 2^{-n\theta+1} \rightarrow 1, \quad \text{as } n \rightarrow \infty \quad (5.5)$$

where the second inequality holds because

$$\ell_i = \lceil \log |\mathcal{U}_i| \rceil \leq \lceil \log |\mathcal{U}| \rceil, \quad \text{for } i = 0, 1$$

and the last inequality follows from (5.4).

Case $P = Q$: Now we let $P = Q$. For $0 < \alpha < H(P)$ let $\mathcal{P}_n(> \alpha)$ be the set of n -types (n -empirical distributions) \hat{P} on \mathcal{U} with $|\mathcal{T}_{\hat{P}}^n| > 2^{n\alpha}$. Then there is a positive θ such that the empirical distribution of the output U^n (resp., V^n) is in $\mathcal{P}_n(> \alpha)$ with probability larger than $1 - 2^{-n\theta}$. Next we choose an integer ℓ_n such that for

$$\beta \triangleq \frac{1}{4} \min(\theta, \alpha), \quad 2^{\frac{3}{2}\beta} < q^{\ell_n} \leq 2^{n\beta}. \quad (5.6)$$

Label sequences in $\mathcal{T}_{\hat{P}}^n$ for $\hat{P} \in \mathcal{P}_n(> \alpha)$ by $0, 1, \dots, |\mathcal{T}_{\hat{P}}^n| - 1$ and let Ψ_1 be a mapping from \mathcal{U}^n to \mathcal{X}^{ℓ_n} , where $\mathcal{X} = \{0, 1, \dots, q-1\}$ as follows.

If u^n has type \hat{P} in $\mathcal{P}_n(> \alpha)$ and has an index $\text{ind}(u^n)$ with q -ary representation $(x_k, x_{k-1}, \dots, x_2, x_1)$, i.e., $\text{ind}(u^n) = \sum_{i=0}^k x_i q^{i-1}$ for $0 \leq x_i \leq q-1, k = \lceil \log |\mathcal{T}_{\hat{P}}^n| \rceil$, then let

$$\Psi_1(u^n) = (x_1, x_2, \dots, x_{\ell_n}). \quad (5.7)$$

If the type of u^n is not in $\mathcal{P}_n(> \alpha)$, we arbitrarily choose a sequence in \mathcal{X}^{ℓ_n} as $\psi_1(u^n)$.

For any fixed $t \leq \ell_n, \hat{P} \in \mathcal{P}_n(> \alpha)$, and $x^t \in \mathcal{X}^t$, let $\mathcal{U}(\hat{P}, x^t)$ be the set of sequences in $\mathcal{T}_{\hat{P}}^n$ such that x^t is a prefix of $\psi_1(u^n)$. Then it is not hard to see that for all x^t, x'^t with $t \leq \ell_n$

$$|\mathcal{U}(\hat{P}, x^t)| - |\mathcal{U}(\hat{P}, x'^t)| \leq 1.$$

More specifically, for all $t \leq \ell_n$ and $x^t \in \mathcal{X}^t$

$$|\mathcal{U}(\hat{P}, x^t)| = \sum_{j=t+1}^k a_j q^{j-1-t} \quad \text{or} \quad \sum_{j=t+1}^k a_j q^{j-1-t} + 1$$

if $|\mathcal{T}_{\hat{P}}^n| = \sum_{j=1}^k a_j q^{j-1}$ with $a_k \neq 0, 0 \leq a_j \leq q-1$ for $j = 1, 2, \dots, k-1$.

Let $\mathcal{U}(x^t) = \bigcup_{\hat{P} \in \mathcal{P}_n(> \alpha)} \mathcal{U}(\hat{P}, x^t)$ (here it does not matter whether $\hat{P} \in \mathcal{P}_n(> \alpha)$ or not).

Thus, we partition \mathcal{U}^n into q^t parts as $\{\mathcal{U}(x^t) : x^t \in \mathcal{X}^t\}$ for $t \leq \ell_n$.

By the asymptotic equipartition property (AEP), the difference of the conditional probability of the event that the output of U^n is in

$\mathcal{U}(x^t)$ given that the type of U^n is in $\mathcal{P}_n(> \alpha)$ and q^{-1} is not larger than

$$\min_{P \in \mathcal{P}_n(> \alpha)} |\mathcal{I}_P|^{-1} < 2^{-n\alpha}.$$

Recalling that with probability $1 - 2^{-n\theta}$ U^n has type in $\mathcal{P}_n(> \alpha)$ and the assumption that V^n has the same distribution as U^n , we obtain that

$$\Pr(U^n \in \mathcal{U}(x^t)) = \Pr(V^n \in \mathcal{U}(x^t)) = P^n(\mathcal{U}(x^t))$$

and for all $x^t \in \mathcal{X}^t$

$$(1 - 2^{-n\theta})(q^{-t} - 2^{-n\alpha}) \leq P^n(\mathcal{U}(x^t)) \leq (1 - 2^{-n\theta})(q^{-t} + 2^{-n\alpha}) + 2^{-n\theta}$$

which implies that for all $x^t \in \mathcal{X}^t$

$$|P^n(\mathcal{U}(x^t)) - q^{-t}| \leq 2^{-n\theta} + 2^{-n\alpha} < 2^{-2n\beta} \quad (5.8)$$

when $\beta \triangleq \frac{1}{4} \min(\theta, \alpha)$.

Recall that Ψ_1 is a function from \mathcal{U}^n to \mathcal{X}^{ℓ_n} and that the definition of $\mathcal{U}(x^t)$, $\mathcal{U}(x^{\ell_n})$ is actually the inverse image of \mathcal{X}^{ℓ_n} under Ψ_1 , i.e. $\mathcal{U}(\mathcal{X}^{\ell_n}) = \Psi_1^{-1}(\mathcal{X}^{\ell_n})$.

Let furthermore $\ell^*(x^{\ell_n}) \triangleq \lceil \log |\mathcal{U}(x^n)| \log q \rceil$ and let Ψ_2 be a function on \mathcal{U}^n such that its restriction on $\mathcal{U}(x^{\ell_n})$ is an injection into $\mathcal{X}^{\ell^*(x^{\ell_n})}$ for all x^{ℓ_n} . Then our decoding function is defined as

$$c = (\Psi_1, \Psi_2). \quad (5.9)$$

To estimate $L_C(P^n, P^n)$, we introduce an auxiliary source with alphabet \mathcal{X}^{ℓ_n} and probability distribution P^* such that for all $x^{\ell_n} \in \mathcal{X}^{\ell_n}$

$$P^*(x^{\ell_n}) = P^n(\mathcal{U}(x^{\ell_n})).$$

We divide the waiting time for identification with code \mathcal{C} into two parts according to the two components Ψ_1 and Ψ_2 in (5.9), and we let W_1 and W_2 be the random waiting times of the two parts, respectively. Now let Z be a binary random variable such that

$$Z = \begin{cases} 0, & \text{if } \Psi_1(U^n) \neq \Psi_1(V^n) \\ 1, & \text{otherwise.} \end{cases}$$

Then

$$\begin{aligned} L_C(P^n, P^n) &= \mathbb{E}(W_1 + W_2) = \mathbb{E} W_1 + \mathbb{E}(\mathbb{E}(W_2 | Z)) \\ &= \mathbb{E} W_1 + \Pr(Z=1) \mathbb{E}(W_2 | Z=1) \\ &= \mathbb{E} W_1 + \left[\sum_{x^{\ell_n}} P^n(\Psi_1(U^n) = x^{\ell_1}) P^n(\Psi_1(V^n) = x^{\ell_n}) \right] \\ &\quad \times \mathbb{E}(W_2 | Z=1) \\ &= \mathbb{E} W_1 + \left\{ \sum_{x^{\ell_n}} [P^n(\mathcal{U}(x^{\ell_n}))]^2 \right\} \mathbb{E}(W_2 | Z=1). \end{aligned} \quad (5.10)$$

Let \mathcal{C}^* be the code for the auxiliary source with encoding function $c^* = \Psi_1$. Then we have that

$$\mathbb{E} W_1 = L_{\mathcal{C}^*}(P^*, P^*) \quad (5.11)$$

and with the notation in Corollary 1 $\mathcal{U}(\mathcal{C}^*, x^t) = \mathcal{U}(x^t)$ and $P^*(\mathcal{U}(\mathcal{C}^*, x^t)) = P^n(\mathcal{U}(x^t))$ for $x^t \in \mathcal{X}^t$ with $t \leq \ell_n$. For all $x^t \in \mathcal{X}^t$, $t \leq \ell_n$, we denote

$$\delta(x^t) = q^{-t} - p^n(\mathcal{U}(x^t)).$$

Then we have for all $t \leq \ell_n$ $\sum_{x^t \in \mathcal{X}^t} \delta(x^t) = 0$ and by (5.8) $\delta(x^t) < 2^{-2n\beta}$.

Now we apply Corollary 1 to estimate

$$\begin{aligned} L_{\mathcal{C}^*}(P^*, P^*) &= \sum_{t=0}^{\ell_n} \sum_{x^t \in \mathcal{X}^t} [P^*(\mathcal{U}(\mathcal{C}^*, x^t))]^2 \\ &= \sum_{t=0}^{\ell_n} \sum_{x^t \in \mathcal{X}^t} [P^n(\mathcal{U}(x^t))]^2 \\ &= \sum_{t=0}^{\ell_n} \sum_{x^t \in \mathcal{X}^t} (q^{-t} - \delta(x^t))^2 \\ &= \sum_{t=0}^{\ell_n} \left[q^t \cdot q^{-2t} - 2q^{-t} \sum_{x^t \in \mathcal{X}^t} \delta(x^t) + \sum_{x^t \in \mathcal{X}^t} \delta(x^t)^2 \right] \\ &\leq \sum_{t=0}^{\ell_n} q^{-t} + \sum_{t=0}^{\ell_n} q^t \cdot 2^{-4n\beta} < \sum_{t=0}^{\infty} q^{-t} + \frac{q^{\ell_n+1} - 1}{q - 1} 2^{-4n\beta} \\ &< \frac{q}{q-1} + \frac{1}{q-1} q^{\ell_n+1} 2^{-4n\beta}. \end{aligned} \quad (5.12)$$

Moreover, by definition of Ψ_2 and W_2

$$\mathbb{E}(W_2 | Z=1) \leq \left\lceil \frac{n \log |\mathcal{U}|}{\log q} \right\rceil$$

and in (5.12) we have shown that

$$\sum_{x^{\ell_n}} [P^n(\mathcal{U}(x^{\ell_n}))]^2 \leq q^{-\ell_n} + q^{\ell_n} \cdot 2^{-4n\beta}.$$

Consequently

$$\begin{aligned} \left\{ \sum_{x^{\ell_n} \in \mathcal{X}^{\ell_n}} [P^n(\mathcal{U}(x^t))]^2 \right\} \mathbb{E}(W_2 | Z=1) &\leq [q^{-\ell_n} + q^{\ell_n} 2^{-4n\beta}] \left\lceil \frac{n \log |\mathcal{U}|}{\log q} \right\rceil. \end{aligned} \quad (5.13)$$

Finally, by combining (5.10), (5.11), (5.12), and (5.13) with the choice of β in (5.6) we have that

$$\overline{\lim}_{n \rightarrow \infty} L_C(P^n, P^n) \leq \frac{q}{q-1};$$

the desired inequality.

It is interesting that the limits of the waiting time of ID-codes on the left-hand side of (5.2) are independent of the generic distributions P and Q and only depend on whether they are equal.

In the case that they are not equal, it is even independent of the alphabet size. In particular, in the case $P \neq Q$, we have seen in the proof that the key step is how to distribute the first symbol and the LUP is applied in the second step. Moreover, for a good code the random user with exponentially vanishing probability needs to wait for the second symbol. So the remaining parts of codewords are not so important.

Similarly, in the case $P = Q$, where we use instead of the LUP the GBP, the key parts of codewords is a relatively small prefix (in the proof it is the ℓ_n th prefix) and after that the user with exponentially small probability has to wait. Thus, again the remaining part of codewords is less important.

APPENDIX I COMMENTS ON GENERALIZED ENTROPIES

After the discovery of ID-entropies in [4] work of Tsallis [13] and also [14] was brought to our attention. The equalities (1) and (2) in [14] are here (A1) and (A2). The letter q used there corresponds to our letter α , because for us q gives the alphabet size. The generalization of Boltzmann's entropy

$$H(P) = -k \sum P_u \ln P_u$$

is

$$S_\alpha(P) = k \frac{1}{\alpha - 1} \left(1 - \sum_{u=1}^N P_u^\alpha \right) \quad (\text{A1})$$

for any real $\alpha \neq 1$. Notice that $\lim_{\alpha \rightarrow 1} S_\alpha(P) = H(P)$, which can be named $S_1(P)$.

One readily verifies that for product distributions $P \times Q$ for independent random variables

$$S_\alpha(P \times Q) = S_\alpha(P) + S_\alpha(Q) - \frac{(\alpha - 1)}{k} S_\alpha(P) S_\alpha(Q). \quad (\text{A2})$$

Since in all cases $S_\alpha \geq 0$, $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$, respectively, correspond to superadditivity, additivity, and subadditivity (also called for the purposes in statistical physics superextensivity, extensivity, and subextensivity).

We recall the grouping identity of [4].

For a partition $(\mathcal{U}_1, \mathcal{U}_2)$ of $\mathcal{U} = \{1, 2, \dots, N\}$, $Q_i = \sum_{u \in \mathcal{U}_i} P_u$, and $P_u^{(i)} = \frac{P_u}{Q_i}$ for $u \in \mathcal{U}_i$ ($i = 1, 2$)

$$H_{I,q}(P) = H_{I,q}(Q) + \sum_i Q_i^2 H_{I,q}\left(\frac{P^{(i)}}{Q_i}\right) \quad (\text{A3})$$

where $Q = (Q_1, Q_2)$. This implies

$$H_{I,q}(P \times Q) = H_{I,q}(Q) + \sum_j Q_j^2 H_{I,q}(P)$$

and since

$$\begin{aligned} \left(1 - \sum_j Q_j^2 \right) &= \frac{q-1}{q} \frac{q}{q-1} \left(1 - \sum_j Q_j^2 \right) \\ &= \frac{q-1}{q} H_{I,q}(Q) \\ &\quad \text{or} \\ \sum_j Q_j^2 &= 1 - \frac{q-1}{q} H_{I,q}(Q) \end{aligned}$$

we get

$$H_{I,q}(P \times Q) = H_{I,q}(Q) + H_{I,q}(P) - \frac{q-1}{q} H_{I,q}(Q) H_{I,q}(P), \quad (\text{A4})$$

which is (A2) for $\alpha = 2$ and $k = \frac{q}{q-1}$.

We have been told by several experts in physics that the operational significance of the quantities S_α (for $\alpha \neq 1$) in statistical physics is not undisputed.

In contrast, the significance of identification entropy was demonstrated in [4] (see Section II), which is formally close to, but essentially different from S_α for two reasons: always $\alpha = 2$ and $k = \frac{q}{q-1}$ is uniquely determined and depends on the alphabet size q !

We also have discussed the coding theoretical meanings of the factors $\frac{q}{q-1}$ and $(1 - \sum_{u=1}^N P_u^2)$.

More recently, we learned from referees that already in 1967 Havrda and Charvát [7] introduced the entropies $\{H_N^\alpha\}$ of **type** α

$$H_N^\alpha(P_1, P_2, \dots, P_N) = (2^{1-\alpha} - 1)^{-1} \left(\sum_{i=1}^N P_i^\alpha - 1 \right) \quad (\text{A5})$$

$$[(P_1, P_2, \dots, P_N) \in \mathcal{P}([N]), N = 2, 3, \dots, 0^\alpha = 0]$$

$$\lim_{\alpha \rightarrow 1} H_N^\alpha(P_1, P_2, \dots, P_N) = H_N(P_1, P_2, \dots, P_N),$$

the Boltzmann/Shannon entropy. So, it is reasonable to define

$$H_N^1(P_1, P_2, \dots, P_N) = H_N(P_1, P_2, \dots, P_N).$$

This is a generalization of the BGS-entropy different from the Rényi entropies of **order** $\alpha \neq 1$ (which according to [2] were introduced by Schützenberger [9]) given by

$${}_\alpha H_N(P_1, P_2, \dots, P_N) = \frac{1}{1-\alpha} \log_2 \sum_{i=1}^N P_i^\alpha$$

$$[(P_1, P_2, \dots, P_N) \in \mathcal{P}([N]), N = 2, 3, \dots].$$

Comparison shows that

$$\begin{aligned} {}_\alpha H_N(P_1, P_2, \dots, P_N) \\ = \frac{1}{1-\alpha} \log_2 [(2^{1-\alpha} - 1) H_N^\alpha(P_1, P_2, \dots, P_N) + 1] \end{aligned}$$

and

$$\begin{aligned} H_N^\alpha(P_1, P_2, \dots, P_N) \\ = (2^{1-\alpha} - 1)^{-1} [2^{(1-\alpha){}_\alpha H_N(P_1, P_2, \dots, P_N)} - 1] \end{aligned}$$

$$[(P_1, P_2, \dots, P_N) \in \mathcal{P}([N]), N = 2, 3, \dots].$$

So, while the entropies of order α and the entropies of type α are different for $\alpha \neq 1$, we see that the bijection

$$t \rightarrow \frac{1}{1-\alpha} \log_2 [(2^{1-\alpha} - 1)t + 1]$$

connects them. Therefore, we may ask what the advantage is in dealing with entropies of type α . We meanwhile also learned that the book [1] gives a comprehensive discussion. Also Daróczy's contribution [6], where "type α " is named "degree α ," gives an enlightening analysis.

Note that Rényi entropies ($\alpha \neq 1$) are **additive**, but not subadditive (except for $\alpha = 0$) and not **recursive**, and they have not the **branching** property nor the **sum** property, that is, there exists a measurable function g on $(0, 1)$ such that

$$H_N^\alpha(P_1, P_2, \dots, P_N) = \sum_{i=1}^N g(P_i).$$

Entropies of type α , on the other hand, are **not additive** but do have the subadditivity property and the **sum property** and furthermore are **additive of degree** α :

$$\begin{aligned} H_{MN}^\alpha(P_1 Q_1, P_1 Q_2, \dots, P_1 Q_N, P_2 Q_1, P_2 Q_2, \dots, P_2 Q_N, \\ \dots, P_M Q_1, P_M Q_2, \dots, P_M Q_N) \\ = H_M^\alpha(P_1, P_2, \dots, P_M) + H_N^\alpha(Q_1, Q_2, \dots, Q_N) \end{aligned}$$

$$+ (2^{1-\alpha} - 1) H_M^\alpha(P_1, P_2, \dots, P_M) H_N^\alpha(Q_1, Q_2, \dots, Q_N)$$

$[(P_1, P_2, \dots, P_M) \in \mathcal{P}([M]), (Q_1, Q_2, \dots, Q_N) \in \mathcal{P}([N]); M = 2, 3, \dots; N = 2, 3, \dots]$.

strong additive of degree α :

$$\begin{aligned} & H_{MN}^\alpha(P_1 Q_{11}, P_1 Q_{12}, \dots, P_1 Q_{1N}, P_2 Q_{21}, P_2 Q_{22}, \\ & \dots, P_2 Q_{2N}, \dots, P_M Q_{M1}, P_M Q_{M2}, \dots, P_M Q_{MN}) \\ &= H_M^\alpha(P_1, P_2, \dots, P_M) + \sum_{j=1}^M P_j^\alpha H_N^\alpha(Q_{j1}, Q_{j2}, \dots, Q_{jN}) \end{aligned}$$

$[(P_1, P_2, \dots, P_M) \in \mathcal{P}([M]), (Q_{j1}, Q_{j2}, \dots, Q_{jN}) \in \mathcal{P}([N]); j = 1, 2, \dots, M; M = 2, 3, \dots; N = 2, 3, \dots]$.

recursive of degree α :

$$\begin{aligned} H_N^\alpha(P_1, P_2, \dots, P_N) &= H_{N-1}^\alpha(P_1 + P_2, P_3, \dots, P_N) \\ &+ (P_1 + P_2)^\alpha H_2^\alpha\left(\frac{P_1}{P_1 + P_2}, \frac{P_2}{P_1 + P_2}\right) \end{aligned}$$

$[(P_1, P_2, \dots, P_N) \in \mathcal{P}([N]), N = 3, 4, \dots \text{ with } P_1 + P_2 > 0]$.

(In consequence, entropies of type α also have the branching property.)

It is clear now that for binary alphabet the ID-entropy is exactly the entropy of type $\alpha = 2$.

However, prior to [13] there are hardly any applications or operational justifications of the entropy of type α .

Moreover, the q -ary case did not exist at all and therefore the name ID-entropy is well justified.

We feel that it must be said that in many papers (with several coauthors) Tsallis at least developed ideas to promote non-standard-equilibrium theory in Statistical Physics using generalized entropies S_α and generalized concepts of inner energy.

Our attention has been drawn also to the papers [5], [11], [12] with possibilities of connections to our work.

Recently, clear-cut progress was made by C. Heup in his forthcoming thesis with a generalization of ID-entropy motivated by L-identification.

REFERENCES

- [1] S. Abe, "Axioms and uniqueness theorem for Tsallis entropy," *Phys. Lett.*, vol. A 271, no. 1–2, pp. 74–79, 2000.
- [2] J. Aczél and Z. Daróczy, *On Measures of Information and their Characterizations, Mathematics in Science and Engineering*. New York/London: Academic, 1975, vol. 115.
- [3] R. Ahlswede, "General theory of information transfer," *Discr. Appl. Math.*, updated Special issue "General Theory of Information Transfer and Combinatorics", to be published.
- [4] R. Ahlswede, Identification Entropy, General Theory of Information Transfer and Combinatorics, Oct. 1, 2002–Aug. 31, 2004, Report on a Research Project at the ZIF (Center of Interdisciplinary Studies) in Bielefeld, edited by R. Ahlswede with the assistance of L. Bäumer and N. Cai, *Lecture Notes in Computer Science*, no. 4123, to be published..
- [5] L. L. Campbell, "A coding theorem and Rényi's entropy," *Inf. Contr.*, vol. 8, pp. 423–429, 1965.
- [6] Z. Daróczy, "Generalized information functions," *Inf. Contr.*, vol. 16, pp. 36–51, 1970.
- [7] J. Havrda and F. Charvát, "Quantification method of classification processes, concept of structural α -entropy," *Kybernetika (Prague)*, vol. 3, pp. 30–35, 1967.
- [8] A. Rényi, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Mathematical Statistics and Probability*, vol. I, pp. 547–561, Berkeley, CA: Univ. California Press, 1961.
- [9] M. P. Schützenberger, *Contribution aux applications statistiques de la théorie de l'information*. Paris, France: Publ. Inst. Statist. Univ. Paris 3, 1954, vol. 3, 1–2, pp. 3–117.
- [10] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.
- [11] B. D. Sharma and H. C. Gupta, "Entropy as an optimal measure, Information theory," in *Proc. Int. CNRS Colloq., Cachan, 1977*, Paris, French, 1978, pp. 151–159, Colloq. Internat. CNRS, 276, CNRS.
- [12] F. Topsøe, "Game-theoretical equilibrium, maximum entropy and minimum information discrimination, Maximum entropy and Bayesian methods," *Fund. Theor. Phys.*, vol. 53, pp. 15–23, 1992, Published by Kluwer Acad., Paris, France, 1993.
- [13] C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *J. Statist. Phys.*, vol. 52, no. 1–2, pp. 479–487, 1988.
- [14] C. Tsallis, R. S. Mendes, and A. R. Plastino, "The role of constraints within generalized nonextensive statistics," *Phys. A*, vol. 261, pp. 534–554, 1998.

Coding for the Feedback Gel'fand-Pinsker Channel and the Feedforward Wyner-Ziv Source

Neri Merhav, *Fellow, IEEE*, and Tsachy Weissman, *Member, IEEE*

Abstract—We consider both channel coding and source coding, with perfect past feedback/feedforward, in the presence of side information. It is first observed that feedback does not increase the capacity of the Gel'fand-Pinsker channel, nor does feedforward improve the achievable rate-distortion performance in the Wyner-Ziv problem. We then focus on the Gaussian case showing that, as in the absence of side information, feedback/feedforward allows to efficiently attain the respective performance limits. In particular, we derive schemes via variations on that of Schalkwijk and Kailath. These variants, which are as simple as their origin and require no binning, are shown to achieve, respectively, the capacity of Costa's channel, and the Wyner-Ziv rate distortion function. Finally, we consider the finite-alphabet setting and derive schemes for both the channel and the source coding problems that attain the fundamental limits, using variations on schemes of Ahlswede and Ooi and Wornell, and of Martinian and Wornell, respectively.

Index Terms—Dirty paper, feedback, feedforward, Schalkwijk-Kailath scheme, side information, source-channel duality.

I. INTRODUCTION

That feedback does not increase the capacity of a memoryless channel, yet can dramatically simplify the schemes for achieving it, is a well known fact (cf. [6] and the literature survey therein). More recently, an analogous phenomenon was shown to hold for the dual problem of lossy source coding with perfect past feedback, also known as "feedforward," at the decoder [12], [7], [5], a problem arising in contexts as diverse as prediction theory, remote sensing, and control.

In this work, we revisit these problems to accommodate the presence of side information. As is the case for problems without feedback/feed-

Manuscript received September 1, 2005. The material in this correspondence was presented in part at the IEEE International Symposium on Information theory, Adelaide, Australia, September 2005.

N. Merhav is with the Department of Electrical Engineering, Technion-Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: merhav@ee.technion.ac.il).

T. Weissman is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9510 USA (e-mail: tsachy@stanford.edu).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2006.880040