Commentationes Mathematicae Universitatis Carolinae

Miroslav Katětov

On entropy-like functionals and codes for metrized probability spaces. I.

Commentationes Mathematicae Universitatis Carolinae, Vol. 31 (1990), No. 1, 49--66

Persistent URL: http://dml.cz/dmlcz/106819

On entropy–like functionals and codes for metrized probability spaces I

MIROSLAV KATĚTOV

Dedicated to the memory of Zdeněk Frolík

Abstract. By means of a suitable modification of the concept of code, we introduce certain entropy–like functionals on the class $\mathfrak{W}$ of semimetric spaces equipped with a bounded measure. For finite spaces $P \in \mathfrak{W}$, (1) we prove that these functionals can be characterized in terms not involving codes, (2) we establish some analogues of the well-known connection between the Shannon entropy of a finite probability space P and the average length of the "best" code for P^n, $n \to \infty$.

Keywords: Hamming space, code, regular code, entropic content, pre–entropy, entropy, final entropy

Classification: 94A17

In the author's articles [2] and [3], it has been shown, among other, that the concept of entropy can be extended from finite probability spaces to the class of all probability spaces equipped with a measurable metric. It has also been shown (see [3]) that there are very many different extensions of this kind. At least one of these "extended entropies" (namely that denoted by E in, e.g., [4] and [5]) has certain applications.

The case of E offers a new approach (see [6]) to the differential entropy; with this approach, the conception of entropies as certain measures of information (hence non–negative) is fully compatible with the fact, seemingly contraintuitive, that the differential entropy can assume negative values. The entropy E (and some other entropies) also make possible a fairly broad approach to the concept of dimension of a metrized probability space; the Rényi dimension, i.e. the dimension introduced in [1] and investigated by A. Rényi in [9] and [10], is included as a special case.

However, in the author's papers, no attention has been given to questions concerning the relationship between coding and "extended entropies" or, at least, the entropy E. In particular, it has not been examined whether it is possible to extend to E the well–known basic theorem asserting that the Shannon entropy of a finite probability space P can be obtained from the average length of words of the "best" code for P^n by a certain passage to the limit for $n \to \infty$.

In the present article, we aim at establishing a theorem (see 4.21 and 4.22) of this kind on the basis of an appropriate modification (see 1.14 and 2.4) of the concept of a code. Namely, code words are allowed to consist of "letters" of various length and certain conditions involving the distance, suitably defined, of code words are imposed. It seems plausible that some analogues of other coding theorems can be obtained in a similar way; however, we do not go into these matters here.

Another aim, connected with that just mentioned, consists in investigating certain entropy–like functionals φ defined on the class $\mathfrak{S}$ of semimetric spaces and/or on class $\mathfrak{W}$ of sets equipped with a finite measure and a measurable semimetric (in fact, the investigation is meaningful for totally bounded spaces $P \in \mathfrak{S} \cup \mathfrak{W}$ only, since if P is not totally bounded, then $\varphi(P) = \infty$ for all φ under consideration). These functionals are introduced on the basis of codewords and their length, but it turns out that each of them (including E) can be fully characterized without reference to codes.

The article is divided into two parts. In the present Part I, we are mainly concerned with finite spaces. whereas the general case will be examined in the Part II, in preparation. Results concerning the general case are often obtained from the finite case by a certain kind of passage to the limit; this is the main reason for first examining the finite case.

Part I is organized as follows. Section 1 contains preliminaries, the concept of Hamming space and that of a code f (approximative or exact) of a semimetric space. For every code f, $\delta(f)$ and $\lambda(f)$, the maximal and the weighted (average) length of codewords of f, are introduced. In Section 2, regular codes are considered. By means of these codes, we define $\delta(P)$ and $\lambda(P)$, the entropic content and the pre–entropy of P. In the finite case, $\delta(P)$ and $\lambda(P)$ are, respectively, the minimal value of $\delta(f)$ and $\lambda(f)$ for a regular exact code f of P in a certain fixed Hamming space, denoted by K_∞. In addition, we introduce a functional $\hat{E}(P)$, which is shown to coincide, for finite space, with $E(P)$ and $E^*(P)$ introduced in previous articles ([**2**], [**3**], [**4**]) by the author.

The main results are presented in Sections 3 and 4. Section 3 contains characterization theorems for δ, λ and E on finite spaces, as well as a lower estimate for E, which turns out to give the exact value in the ultrametric case. In Section 4, the functionals Δ and Λ, the final entropic content and the final entropy, are introduced: $\Delta(P)$ is defined as $\inf(\delta(P^n)/n)$, and $\Lambda(P)$ is defined on the basis of $\lambda(P)$ in an analogous way. Characterization theorems (finite case) for Δ and Λ are proved and the inequality $\delta(P) \cdot wP \geq \lambda(P) \geq E(P) \geq \Lambda(P) = \lim(E(P^n)/n)$ is established. It is shown that, in the ultrametric case, we have $\lambda(P) \geq E(P) = \Lambda(P) = \lim(E(P^n)/n)$ in full agreement with the case of a finite probability space P.

1.

1.1. Notation. A) The symbols N, R, R_+, $\overline{R}_+$ have their usual meaning. The letters i, j, k, m, n denote non–negative integers; ε denotes a non–negative real. If S is a set, $|S|$ denotes its cardinality. The first infinite cardinal is denoted by ω. – B) Let $\prec$ be an order on a set S. If $M \subset S$, we put $[M] = [M]_S = \{x \in S : x \prec y$ for some $y \in M\}$. If $x, y \in S$, then $x \wedge y$ denotes the meet of x and y, i.e., the element (provided it exists) $z \in S$ such that $z \prec x$, $z \prec y$, and if $z' \prec x$, $z' \prec y$, then $z' \prec z$; $x \vee y$ denotes the join of x and y. – C) If B is a set, we put $B^* = \bigcap(B^n : n \in N\}$. For any $u = (u_i : i < n) \in B^*$ we put $|u| = n$ and, for any $k \in N$, $u \restriction k = (u_i : i < n \wedge k)$. If $u, v \in B^*$, then $u \prec v$ means that $u = v \restriction k$ for some k. The concatenation $u \cdot v$ of u and v is defined in the usual way. We often write uv instead of $u \cdot v$, $u \cdot a$ or ua instead of $u \cdot (a)$, etc. If $k \geq 1$ and $u_i \in B^*$ for $i < k$, then the concatenation of u_i, $i < k$, is denoted by $\prod_{i<k} u_i$. We put $\prod_{i<0} u_i = \emptyset$ (the void sequence). – D) The completion of a measure μ is denoted by $\overline{\mu}$ (or, for typographical reasons, by $[\mu]$). The product of measures μ_1 and μ_2 is denoted by $\mu_1 \times \mu_2$.

1.2. Conventions. A) We often omit parentheses provided there is no danger of confusion; e.g., if f is a mapping, we write fx instead of $f(x)$, $f^{-1}M$ instead of $f^{-1}(M)$, etc. On the other hand, the symbol for multiplication is often retained to avoid confusion; e.g., if f is a function and $c \in R$, we write $c \cdot fx$ instead of $cf(x)$. – B) A singleton $\{a\}$ is often denoted merely by a. Thus, e.g., if μ is a measure and $\{x\} \in \operatorname{dom} \mu$, we write μx or $\mu(x)$ or else $\mu\{x\}$ instead of $\mu(\{x\})$.

1.3. Notation and conventions. We put $0 \cdot \infty = 0 \cdot (-\infty) = 0$, $0/0 = 0$. We write log instead of $\log_2$ and put $L(x) = -x \log x$ for $x \in R_+$. If $x_i \in R_+$, we put $H(x_1, \ldots, x_n) = \sum_{i=1}^{n} Lx_i - L(\sum_{i=1}^{n} x_i)$.

1.4. A semimetric on a set Q is, by definition, a function $\rho\colon Q \times Q \longrightarrow R_+$ such that $\rho(x,x) = 0$, $\rho(x,y) = \rho(y,x)$ for all x, $y \in Q$. The set of all semimetrics on Q will be denoted by $\mathcal{S}(Q)$. If $\rho \in \mathcal{S}(Q)$ and $T \subset Q$, then $\rho \restriction (T \times T) \in \mathcal{S}(T)$ will be denoted by $\rho \restriction T$. – If $\rho \in \mathcal{S}(Q)$, then $\langle Q, \rho \rangle$ is called a semimetric space or an SM-space (an FSM-space if $|Q| < \omega$). The class of all SM-spaces (all FSM-spaces) will be denoted by $\mathfrak{S}$ (by $\mathfrak{S}_F$). If $P = \langle Q, \rho \rangle \in \mathfrak{S}$, $T \subset Q$, then the subspace $\langle T, \rho \restriction T \rangle$ will often be denoted by $\langle T, \rho \rangle$ or by $T.P$. – If $t \in R_+$ and Q is a set, we put $\langle Q, t \rangle = \langle Q, \rho \rangle$ where $\rho(x,y) = t$ for $x \neq y$, $\rho(x,x) = 0$. – The product of SM-spaces is defined in the usual way. Namely, we put $\langle Q_1, \rho_1 \rangle \times \langle Q_2, \rho_2 \rangle = \langle Q_1 \times Q_2, \rho_1 \times \rho_2 \rangle$ where $(\rho_1 \times \rho_2)((x_1,x_2),(y_1,y_2)) = \rho_1(x_1,y_1) \vee \rho_2(x_2,y_2)$.

1.5. A semimetrized measure space or a W–space is, by definition, a triple $\langle Q, \rho, \mu \rangle$ where $Q \neq \emptyset$, $\langle Q, \rho \rangle \in \mathfrak{S}$, μ is a finite measure on Q and $\rho\colon Q \times Q \longrightarrow R_+$ is $[\mu \times \mu]$-measurable. – Cf. [2], 1.17.

1.6. Let $P = \langle Q, \rho, \mu \rangle$ be a W–space. We put $wP = \mu Q$. If $wP = 1$, P is called a semimetrized probability space or a PW-space. If, for all x, $y \in Q$, $x \neq y$, there is an $M \in \operatorname{dom} \mu$ such that $x \in M$, $y \in Q \setminus M$, we say that P is separated. A finite separated W-space is called an FW-space. The class of all W–spaces (all FW–spaces) will be denoted by $\mathfrak{W}$ (by $\mathfrak{W}_F$). – Cf. [2], 1.17.

1.7. A) If $P = \langle Q, \rho, \mu \rangle \in \mathfrak{W}$, $\emptyset \neq T \in \operatorname{dom} \overline{\mu}$, put $\nu(X) = \overline{\mu}(X \cap T)$ for $X \in \operatorname{dom} \mu$. Then $\langle Q, \rho, \nu \rangle$ will be denoted by $T.P$ and called a subspace of P. – Cf. [2], 1.22 (where the terminology is different). – B) The product of W-spaces is defined in the usual way: $\langle Q_1, \rho_1, \mu_1 \rangle \times \langle Q_2, \rho_2, \mu_2 \rangle = \langle Q_1 \times Q_2, \rho_1 \times \rho_2, \mu_1 \times \mu_2 \rangle$.

1.8. If Q is a set, $(Q_1, \ldots, Q_n)$ is called a partition of Q if $\bigcup Q_i = Q$ and $Q_i \cap Q_j = \emptyset$ for $i \neq j$. If $P \in \mathfrak{S} \cup \mathfrak{W}$, we call $(P_1, \cdots, P_n)$ a partition of P if P_i are subspaces of P and there is a partition $(Q_1, \cdots, Q_n)$ of Q such that $P_i = Q_i \cdot P$ (observe that if $P \in \mathfrak{W}$, then Q_i are necessarily non–void $\overline{\mu}$-measurable). – Cf. [2], 1.30; observe that we use a terminology different from that used in [2].

1.9. If $P = \langle Q, \rho \rangle \in \mathfrak{S}$ or, respectively, $P = \langle Q, \rho, \mu \rangle \in \mathfrak{W}$, then the infimum of all $b \in \overline{R}_+$ such that $\{(x,y) \in Q \times Q \colon \rho(x,y) > b\} = \emptyset$ (respectively, $[\mu \times \mu]\{(x,y) \in Q \times Q \colon \rho(x,y) > b\} = 0)$ is called the diameter of P and is denoted by $d(P)$. If $M \subset Q$ and $M \cdot P$ is a subspace of P, then $d(M..P)$, the diameter of M in P, is denoted by $d_P(M)$ or simply by $d(M)$.

1.10. Definition. If A is a set, π is a mapping of A onto $\pi(A)$, $|\pi(A)| = m$, $2 \leq m < \omega$, λ is a mapping of A into R_+, $\lambda(A) \neq \{0\}$ and $a \longmapsto (\pi a, \lambda a)$ is a bijection of A onto $\pi(A) \times \lambda(A)$, then $K = \langle A^*, \pi, \lambda \rangle$ will be called an m-ary (binary if $m = 2$ Hamming space. For every $u = (u_i \colon i < n) \in A^*$ we put $\lambda(u) = \sum(\lambda u_i \colon i < n)$; if $u = (u_i \colon i < m)$, $v = (v_i \colon i < n) \in A^*$, we put $\tau(u,v) = \tau_\kappa(u,v) = \sum(\lambda u_i \wedge \lambda v_i \colon i < m \wedge n,\ u_i \neq v_i)$. – Evidently, $\tau_\kappa \in \mathcal{S}(A^*)$; however, τ_κ is not a metric (observe that $\tau_\kappa(u,v) = 0$ if $u \prec v$).

1.11. Notation. If $P = \langle Q, \rho \rangle \in \mathfrak{S}$ or $P = \langle Q, \rho, \mu \rangle \in \mathfrak{W}$, we put $|P| = Q$ and, in accordance with 1.1A, for the cardinality $|Q|$ of Q we have $|Q| = \|P\|$. If $K = \langle A^*, \pi, \lambda \rangle$ is a Hamming space, we put $|K| = A^*$.

1.12. Notation. We put $K_1 = \langle A^*, \pi, \lambda \rangle$ where $A = \{0,1\}$, $\pi(i) = i$, $\lambda(i) = 1$; $K_\infty = \langle A^*, \pi, \lambda \rangle$ where $A = \{0,1\} \times R_+$, $\pi(i,t) = i$, $\lambda(i,t) = t$.

Remark. If $K = K_\infty$, then, for every $n \in N$, $\tau_\kappa \restriction \{0,1\}^n$ is a metric, namely the well–known Hamming distance on $\{0,1\}^n$.

1.13. Convention. In that follows, the letter P, possibly with subscripts, etc., will always denote an SM-space or a W-space, and the letter K, possibly with subscripts, will denote a Hamming space.

1.14. Definition. Let $\varepsilon \geq 0$. A mapping $f \colon |P| \longrightarrow |K|$ will be called an ε-code of P in $K = \langle A^*, \pi, \lambda \rangle$ if the following conditions are satisfied: (1) $|fP| < \omega$, (2) if $P = \langle Q, \rho, \mu \rangle \in \mathfrak{W}$, then all $f^{-1}u$, $u \in fP$, are μ-measurable, (3) if u, $v \in fP$, then $d(f^{-1}\{u,v\}) \leq \tau_\kappa(u,v) \vee \varepsilon$, (4) if $u \cdot (a)$, $u \cdot (b) \in [fP]$, $\pi a = \pi b$, then $a = b$. Every ε-code, $\varepsilon \in R_+$, will be called an approximative code; a 0-code will also be called an exact code (or simply a code).

1.15. Notation. If $P \in \mathfrak{S} \cup \mathfrak{W}$ and $\varepsilon \geq 0$, then $\operatorname{cod}(\varepsilon, P)$ will denote the class of all ε-codes $f \colon P \longrightarrow K$ where K is an arbitrary Hamming space.

1.16. Clearly, the condition (3) in 1.14 is equivalent to the following one: if $P \in \mathfrak{S}$, then $\rho(x,y) \leq \tau_\kappa(fx, fy) \vee \varepsilon$ for all x, $y \in P$, and if $P \in \mathfrak{W}$, then there is a set $Z \subset |P| \times |P|$ such that $[\mu \times \mu](Z) = 0$ and $\rho(x,y) \leq \tau_\kappa(fx,fy) \vee \varepsilon$ for all $(x,y) \in |P| \times |P| \setminus Z$.

1.17. Notation. If ε, $t \in R_+$, we put (1) $\varepsilon * t = 0$ if $t \leq \varepsilon$, $\varepsilon * t = 1$ if $t > \varepsilon$, (2) $\varepsilon \odot t = (\varepsilon * \rho) \cdot t$. If $\rho \in \mathcal{S}(Q)$, we put $(\varepsilon * \rho)(x,y) = \varepsilon * \rho(x,y)$, $(\varepsilon \odot \rho)(x,y) = \varepsilon \odot \rho(x,y)$ for all x, $y \in Q$. If $P = \langle Q, \rho \rangle \in \mathfrak{S}$, we put $\varepsilon * P = \langle Q, \varepsilon * \rho \rangle$, $\varepsilon \odot P = \langle Q, \varepsilon \odot \rho \rangle$. If $P = \langle Q, \rho, \mu \rangle \in \mathfrak{W}$, we put $\varepsilon * P = \langle Q, \varepsilon * \rho, \mu \rangle$, $\varepsilon \odot P = \langle Q, \varepsilon \odot \rho, \mu \rangle$.

1.18. Fact. A mapping $f\colon |P| \longrightarrow |K|$ is an ε-code iff it is an exact code of $\varepsilon \odot P$ in K.

1.19. Facts. A) The following properties of a space $P \in \mathfrak{S} \cup \mathfrak{W}$ are equivalent: (1) for every $\varepsilon > 0$ and every K there is a regular (see 2.4 below) ε-code of P in K, (2) for every $\varepsilon > 0$, P has an ε-code in some K; (3) P is totally bounded. – B) The following properties of $P \in \mathfrak{S} \cup \mathfrak{W}$ are equivalent: (1) for every K there is a regular (see 2.4) exact code of P in K, (2) P has an exact code in some K, (3) $d(P) < \infty$ and there is a partition $(P_1, \dots, P_n)$ of P such that $d(P_i) = 0$, $i = 1, \dots, n$.

The proofs of these facts are easy and can be omitted.

1.20. Notation. A) If f is an ε-code of P in K, we put (1) $\delta(f) = \max\{\lambda(u)\colon u \in fP\}$ if $P \in \mathfrak{S}$, $\delta(f) = \max\{\lambda(u)\colon u \in fP, \overline{\mu}(f^{-1}u) > 0\}$ if $P \in \mathfrak{W}$ (the letter δ stands for Greek $\delta o\lambda\iota\chi\acute{o}\varsigma$=long); (2) $\lambda(f) = \int(\lambda \circ f)d\mu$ if $P = \langle Q, \rho, \mu\rangle \in \mathfrak{W}$. – B) Let $\mathcal{K}$ be a class of approximative codes. If $\varepsilon \geq 0$, $P \in \mathfrak{S} \cup \mathfrak{W}$, we put $\delta(\varepsilon, P, \mathcal{K}) = \inf\{\delta(f)\colon f \in \mathcal{K} \cap \operatorname{cod}(\varepsilon, P)\}$. If $\varepsilon \geq 0$, $P \in \mathfrak{W}$, we put $\lambda(\varepsilon, P, \mathcal{K}) = \inf\{\lambda(f)\colon f \in \mathcal{K} \cap \operatorname{cod}(\varepsilon, P)\}$. If $P \in \mathfrak{S} \cup \mathfrak{W}$, we put $\delta(P, \mathcal{K}) = \sup\{\delta(\varepsilon, P, \mathcal{K})\colon \varepsilon > 0\}$; if $P \in \mathfrak{W}$, we put $\lambda(P, \mathcal{K}) = \sup\{\lambda(\varepsilon, P, \mathcal{K})\colon \varepsilon > 0\}$.

1.21. Fact. Let $P \in \mathfrak{S} \cup \mathfrak{W}$; let $\varepsilon \geq 0$, and let $\varepsilon < d(S)$ whenever S is a subspace of P, $d(S) > 0$. Then every ε-code of P in K is an exact code.

PROOF: Let f be an ε-code of P in P. For every u, $v \in fP$ we have $d(f^{-1}\{u, v\}) \leq d\{u, v\} \vee \varepsilon$. Put $a = d(f^{-1}\{u, v\})$. If $a = 0$, then $a \leq d\{u, v\}$; if $a > 0$, then $a > \varepsilon$, hence $a \leq d\{u, v\}$. Therefore, f is a 0-code. ∎

1.22. Remark. Assume that P is not totally bounded. Then, by 1.19 A, for all sufficiently small $\varepsilon > 0$, we have $\operatorname{cod}(\varepsilon, P) = \emptyset$, hence $\delta(\varepsilon, P, \mathcal{K}) = \lambda(\varepsilon, P, \mathcal{K}) = \infty$ for every class $\mathcal{K}$ of approximative codes. Thus, the theory developed below has a real sense for totally bounded spaces only (though it is formally meaningful for all $P \in \mathfrak{S} \cup \mathfrak{W}$).

1.23. Notation. The class of all approximative codes in K_∞ (in K_1) will be denoted by $\mathcal{K}_\infty$ (by $\mathcal{K}_1$).

1.24. The functionals $\delta(\varepsilon, P, \mathcal{K}_\infty)$, $\delta(P, \mathcal{K}_\infty)$, etc., are of little interest. E.g., it can be shown that if $P = \langle Q, 1, \mu\rangle \in \mathfrak{W}_F$, then $\lambda(P^n, \mathcal{K}_\infty)/n \longrightarrow 0$, whereas we would expect something like $\lambda(P^n, \mathcal{K}_\infty)/n \longrightarrow H(\mu q\colon q \in Q)$, in accordance with the classical result.

1.25. The functionals $\delta(\varepsilon, P, \mathcal{K}_1)$, $\delta(P, \mathcal{K}_1)$, etc., behave better, in some aspects. For instance, if $P \in \mathfrak{S}$, $d(P) \leq 1$ and P is totally bounded, then $\mathcal{H}_\varepsilon(P) \leq \delta(\varepsilon, P, \mathcal{K}_1) < \mathcal{H}_\varepsilon(P) + 1$ where $\mathcal{H}_\varepsilon(P)$ is (a version of) the Kolmogorov ε-entropy (see, e.g. [7] and [8]), namely $\mathcal{H}_\varepsilon(P) = \log \mathcal{N}_\varepsilon(P)$, $\mathcal{N}_\varepsilon(P)$ being the minimal cardinality of a partition of P into sets of diameter $\leq \varepsilon$. – On the other hand, for any $P = \langle Q, \rho\rangle \in \mathfrak{S}_F$, we have $\delta(P^n, \mathcal{K}_1)/n \longrightarrow \log|Q|$; thus, for large n, $\delta(\langle Q, \rho\rangle^n, \mathcal{K}_1)/n$ "depends only slightly" on the semimetric ρ.

2.

2.1. The facts mentioned in 1.24 and 1.25 lead to the conclusion that the class of codes considered must be restricted if we want to get entropy-like functionals connected with the properties of codes and depending (for W–spaces $P = < Q, \varrho, \mu >$) both on the semimetric ϱ and the measure μ of the space. – To this end, we need some auxiliary concepts introduced below.

2.2. Notation. Let M be a set, $S \subset M^*$, $x \in [S]$. Then (I) $br(x, S)$ will denote the set of all $b \in M$ such that $x.(b) \in [S]$; (II) $Br(x, S)$ will denote the set of all $z \in M^*$ such that (1) $|z| \geq 1$, $x.z \in [S]$, (2) $|br(x.z', S)| = 1$ whenever $z' \prec z$, $\emptyset \neq z' \neq z$, (3) $|br(x.z, S)| = 1$; (III) for every $u \in S$ such that $u \prec x$, $u \neq x$, $\beta(x, u, S)$ will denote the (unique) $z \in M^*$ such that (1) $|z| \geq 1$, $x.z \prec u$, (2) $|br(x.z', S)| = 1$ whenever $z' \prec z$, $\emptyset \neq z' \prec z$, (3) either $|br(x.z, S)| \neq 1$ or $x.z = u$. If $u = x$, we put $\beta(x, u, S) = \emptyset$ (the void sequence)). – Thus, $\beta(x, u, S)$ is, roughly speaking, the "non–branching part" of the sequence $\hat{u}$ defined by $x.\hat{u} = u$.

2.3. Definition. Let M be a set, $S \subset M^*$, $\varrho \in S(M^*)$. We denote by $[\varrho]_S$ or ϱ'_S or simply ϱ' the semimetric on S defined as follows: if $u, v \in S$, we put $\varrho'(u, v) = \varrho(u', v')$ where $u' = \beta(u \wedge v, u, S)$, $v' = \beta(u \wedge v, v, S)$. The semimetric ϱ' will be called the <u>reduction</u> of ϱ with respect to S. If $X \subset S$, we put $d'(X) = d'_S(X) = d\langle X, \varrho' \rangle$. – In the sequel, we shall have $\varrho = \tau = \tau_K$ for some Hamming space K and $S = fP$ for some ε–code of P in K.

2.4. Definition. An ε–code of P in K will be called <u>regular</u> if the following condition is satisfied: (R) if $u, v \in fP$, $s \prec u \wedge v$, $|br(s, fP)| \neq 1$, then $d(f^{-1}\{u, v\}) \leq d'(Br(s, fP)) \vee \varepsilon$. – Observe that the condition (R) implies $d(f^{-1}u) \leq \varepsilon$ for all $u \in fP$.

2.5. Remark. Let f be an ε–code of P in K. Then each of the following conditions is equivalent to (R) introduced above: (1) if $u \in [fP]$, $|br(u, fP)| \neq 1$, then $d\{x \in |P| : u \prec fx\} \leq d'(Br(u, fP)) \vee \varepsilon$, (2) if $u, v, t \in fP$, $u \wedge v \prec u \wedge t$, then $d(f^{-1}\{u, v\}) \leq d'\{u, t\} \vee \varepsilon$.

2.6. Fact. A mapping f of P into K is a regular ε–code of P in K iff it is a regular 0–code of $\varepsilon \odot P$ in K.

2.7. Notation. The class of all $f\colon P \longrightarrow K_\infty$ (all $f\colon P \longrightarrow K_1$) such that $P \in \mathfrak{S} \cup \mathfrak{W}$ and f is a regular approximative code of P in K_∞ (in K_1) will be denoted by $\mathcal{K}^r_\infty$ (by $\mathcal{K}^r_1$).

2.8. Notation. For every $P \in \mathfrak{S} \cup \mathfrak{W}$, we put (1) for every $\varepsilon \geq 0$, $\delta(\varepsilon, P) = \delta(\varepsilon, P, \mathcal{K}^r_\infty)$, (2) $\delta(P) = \delta(P, \mathcal{K}^r_\infty)$. For every $P \in \mathfrak{W}$, we put (1) for every $\varepsilon \geq 0$, $\lambda(\varepsilon, P) = \lambda(\varepsilon, P, \mathcal{K}^r_\infty)$, (2) $\lambda(P) = \lambda(P, \mathcal{K}^r_\infty)$. Thus, $\delta(P)$ is the supremum of all $\inf\{\delta(f) : f \in \mathcal{K}^r_\infty \cap \operatorname{cod}(\varepsilon, P)\}$, $\varepsilon \geq 0$, and $\lambda(P)$ is the supremum of all $\inf\{\lambda(f) : f \in \mathcal{K}^r_\infty \cap \operatorname{cod}(\varepsilon, P)\}$, $\varepsilon > 0$.– Observe that, e.g., if P is the interval $[0, 1]$ with the usual metric, then $\delta(P) \leq 2$ whereas $\delta(0, P) = \infty$ since there exist no 0–codes of P.

2.9. Definition. For every $P \in \mathfrak{S} \cup \mathfrak{W}$(respectively, $P \in \mathfrak{W}$), $\delta(P)$ and $\lambda(P)$ will be called the <u>entropic content</u> and the <u>pre–entropy</u> of P, respectively.

2.10. Fact. If $P \in \mathfrak{S}_F \cup \mathfrak{W}_F$ then $\delta(P) = \inf\{\delta(f) : f \in \mathcal{K}^r_\infty \cap \operatorname{cod}(0, P)\}$. If $P \in \mathfrak{W}_F$, then $\lambda(P) = \inf\{\lambda(f) : f \in \mathcal{K}^r_\infty \cap \operatorname{cod}(0, P)\}$. This follows from 1.21.

2.11. Notation. If $f \in \mathrm{cod}(\varepsilon, P)$, then $B(f)$ will denote the set of all $u \in [fP]$ such that $|br(u, fP)| = 2$.

2.12. Notation. If f is an ε–code of $P = \langle Q, \rho, \mu rangle \in \mathfrak{W}$ in a binary K, then $E(f)$ is defined as follows: for every $u \in B(f)$, we put $Br(u, fP) = \{s, t\}$, $S = \{x \in P : u.s \prec fx\}$, $T = \{x \in P : u.t < fx\}$, $E(u, f) = H(\overline{\mu}S, \overline{\mu}T).\tau'(s, t)$; we put $E(f) = \sum(E(u, f) : u \in B(f))$.

2.13. Definition. For every $P \in \mathfrak{W}$, we put (1) for every $\varepsilon > 0$, $E(\varepsilon, P) = \inf\{E(f) : f \in \mathcal{K}^r_\infty \cap \mathrm{cod}(\varepsilon, P)\}$, (2) $\widehat{E}(P) = \sup\{E(\varepsilon, P) : \varepsilon > 0\}$, and we call $\widehat{E}(P)$ the coding entropy of P (or simply the entropy of P).–
Remark. In 2.22–2.31 the relationship between $\widehat{E}(P)$ and some entropies introduced in [2] will be considered.

2.14. Fact. If $P \in \mathfrak{W}_F$, then $\widehat{E}(P) = \inf\{E(f) : f \in \mathcal{K}^r_\infty \cap \mathrm{cod}(0, P)\}$. This follows from 1.21.

2.15. Fact. Let $P \in \mathfrak{W}$. Then (1) for every approximative code f of P in a binary K, $\delta(f).wP \geq \lambda(f) \geq E(f)$, (2) $\delta P.wP \geq \lambda P \geq \widehat{E}(P)$.

2.16. We are going to show (see 2.20) that every regular ε–code in $\mathcal{K}_\infty$ can be, roughly speaking, replaced by a regular ε–code with certain useful properties (introduced below).

2.17. Definition. An ε–code f of $P \in \mathfrak{S} \cup \mathfrak{W}$ in a binary K will be called (1) strongly branching if $B(f) = [fP] \setminus fP$, (2) well–fitting if, for every $u \in B(f)$, $d\{x \in P : u \prec fx\} = d'(Br(u, fP)) = \lambda(s)$ for each $s \in Br(u, fP)$.

2.18. Fact. Every strongly branching well–fitting ε–code is regular exact.

2.19. Fact. If f is an approximative code of P in K and $u \in [fP]$, then there is exactly one sequence $(z_i : i < m)$ such that $z_i \in |K|$, the concatenation $\prod_{i<m} z_i$ is equal to u, $z_0 \in Br(0, fP)$ and $z_j \in Br(\prod_{i<j} z_i, fP)$ for $1 \leq j < m$.

2.20. Lemma. *If f is a regular ε-code of $P \in \mathfrak{S} \cup \mathfrak{W}$ in K_∞, then there exists a strongly branching regular ε-code g of P in K_∞ such that* (i) $\delta(g) \leq \delta(f)$, (ii) *if $P \in \mathfrak{W}$, then* $\lambda(g) \leq \lambda(f)$, $E(g) \leq E(f)$, (iii) *if $\varepsilon = 0$, then g is well-fitting.*

PROOF : I. Clearly, it is sufficient to consider the case $P \in \mathfrak{W}$. The proof is technically somewhat involved, though the underlying idea is quite simple. It will be performed in two steps: we prove the statements (A) and (B), from which the assertion of the lemma will follow immediately.

Statement (A). For every regular ε–code f of P in $K_\infty h$, there is a regular ε–code h of P in K_∞ such that

(A1) $\emptyset \in B(h)$; if $u \in B(h)$, $Br(u, hP) = \{s, t\}$, $s = (s_i : i < m)$, $t = (t_i : i < m)$, then (a) $m = n$ and, for all $i < m$, $s_i \neq t_i$, $\lambda s_i = \lambda t_i$, (b) $\lambda s = \lambda t = \tau(s, t) = d'(Br(u, hP))$;

(A2) the collections $\{h^{-1}u : u \in hP\}$ and $\{f^{-1}v : v \in fP\}$ coincide, $\lambda(hx) = \lambda(fx)$ for all $x \in P$, $\delta(h) \leq \delta(f)$, $\lambda(h) \leq \lambda(f)$, $E(h) \leq E(f)$;

(A3) there is a bijection $\psi : B(h) \to B(f)$ such that (a) for all $u, v \in B(h)$, $u \prec v$ iff $\psi u \prec \psi v$, (b) $\varphi(hx) = fx$ whenever $hx \in B(h)$, (c) if $u \in B(h)$, $Br(u, hP) = \{s,t\}$, $Br(\psi u, fP) = \{s_1, t_1\}$, then $\lambda(s) = \lambda(t) \leq \lambda s_1 \wedge \lambda t_1$, $\tau(s,t) \leq \tau(s_1, t_1)$.

Statement (B). For every regular ε–code h of P in $\mathcal{K}_\infty$ satisfying (A1)–(A3) with respect to a given f, there is a strongly branching regular ε–code g of P in $\mathcal{K}_\infty$ such that (1) the conditions (A1)–(A3) are satisfied for g with respect to h, (2) for every $u \in gP$, $d\{x \in P : u \prec gx\} = d'(Br(u, gP)) \vee \varepsilon$, hence, in particular, if $\varepsilon = 0$, then g is well–fitting.

II. We prove (A) by induction on the cardinality of fP. Let $|fP| = 2$, $fP = \{s,t\}$, $s = (s(i) : i < m)$, $t = (t(i) : i < n)$. Let $(i_j : j < k)$ be the increasing sequence of all $i < m \wedge n$ such that $s_i \neq t_i$. For $j < k$, let $u_j, v_j \in \{0,1\} \times R_+$, $\lambda u_j = \lambda v_j = \lambda(s(i_j)) \wedge \lambda(t(i_j))$, $\pi(u_j) = 0$, $\pi(v_j) = 1$. Put $u = (u_j : j < k)$, $v = (v_j : j < k)$. For $x \in P$ put $hx = u$ if $fx = s$, $hx = v$ if $fx = t$. It is easy to show that h is a regular ε–code of P in K_∞ satisfying (A1)–(A3) with respect to f.– Assume that the statement (A) holds if $|fP| < n$. Consider an ε–code f of P in K_∞ such that $|fP| = n$. Let z be the least element of $B(f)$, and let $Br(z, fP) = \{\widehat{s}, \widehat{t}\}$; put $s = z.\widehat{x}$, $t = z.\widehat{t}$. Put $Q_0 = \{x \in P : s \prec fx\}$, $Q_1 = \{x \in P : t \prec fx\}$. Put $f'(x) = s$ if $x \in Q_0$, $f'(x) = t$ if $x \in Q_1$; put $c = d(Q_0) \wedge d(Q_1)$. Then f' is a regular c–code of P in K_∞. Since $|f'P| = 2$, there is a c–code h' of P in K_∞ which satisfies (A1)–(A3) with respect to f'.– If $x \in Q_0$ (respectively, $x \in Q_1$), define $f_0(x)$ (respectively, $f_1(x)$) by $f(x) = s.f_0(x)$ (respectively, $f(x) = t.f_1(x)$). It is easy to see that f_i is a regular ε–code of $P_i = Q_i.P$. Since $|f_iP| < n$, there exists, for $i = 0, 1$, an ε–code h_i of P_i which satisfies, with respect to f_i, the conditions (A1)–(A3).– For every $x \in P$, put $h(x) = h'(x).h_i(x)$ if $x \in Q_i$. It is easy to prove that, with respect to f, h has the required properties.

III. We are going to prove (B). Let M consist of all pairs (u, s) such that $u \in B(h)$, $s \in Br(u, hP)$. Let φ be a mapping of M into A such that if (u,s), $(u,t) \in M$, $s \neq t$, then $\varphi(u,s) \neq \varphi(u,t)$, $\lambda(\varphi(u,s)) = \lambda(\varphi(u,t))$. For every $u \in [hP]$, let $(z_i : i < k)$ be the sequence described in 2.19, i.e. the sequence such that $z_i \in A$, the concatenation $\prod_{i<k} z_i$ is equal to u, $z_0 \in Br(\emptyset, hP)$, $z_j \in Br(\prod_{i<j} z_i, hP)$ for $1 \leq j < k$.

Put $\psi(u) = (\varphi(\prod_{i<j} z_i, z_j) : j < k)$. For every $x \in P$ put $h^*(x) = \psi(hx)$. It can be easily proved that h^* is a regular ε–code of P in K_∞ satisfying (A1)–(A3) with respect to h, hence also to f. Evidently, h^* is strongly branching.– Define g as follows: if $h^*(x) = u = (u_i : i < n)$, put $g(x) = v = (v_i : i < n)$ where $\pi v_i = \pi u_i$, $\lambda v_i = \varepsilon.d\{x \in P : u_i < fx\}$. It is easy to show that g has all the required properties. ■

2.21. Fact. For every $P \in \mathfrak{S}_F \cup \mathfrak{W}_F$, there are only finitely many strongly branching well-fitting (hence regular exact) codes of P in K_∞.

PROOF : It is sufficient to consider the case $P \in \mathfrak{W}_F$. For $P \in \mathfrak{W}_F$ let $C(P)$ be the set of all strongly branching well-fitting exact codes of P in K_∞. If $P = \langle Q, \varrho, \mu \rangle \in \mathfrak{W}_F$ and $f \in C(P)$, let a_f denote the (unique) $a \in \{0,1\} \times R_+$ such that $(a) \in [fP]$, $\pi a = 0$, and let $Q_0(f)$ consist of $x \in P$ such that $(a_f) \prec fx$. Clearly, $\lambda(a_f) = d(P)$. It is easy to see that, for every $T \subset Q$, $\emptyset \neq T \neq Q$, we have

$|\{f \in C(P) : Q_0(f) = T\}| = |C(T.P)|.|C((Q \setminus T).P)|$. Since, evidently, $|C(P)| \leq 2$ if $\|P\| \leq 2$, the proof is completed by induction on the cardinality of $|P|$. ∎

2.22. We now turn to the functionals E^* and E, which have been introduced in [2]. More precisely, in [2], 3.4, the concept of gauge functional has been introduced and, for every gauge functional τ, two functionals on $\mathfrak{W}$, C^*_τ and C_τ, called the τ-semientropy and τ-entropy, respectively, have been defined (see [2], 3.17). For a special choice of τ, the functionals C^*_E and C_E are obtained; in subsequent articles, they have been denoted by E^* and E, respectively.

We are not going to state the pertinent definitions. Indeed, we present different but equivalent (see 2.27 and 2.29 below) definitions of E^* and E for FW–spaces. It will turn out that, for every FW–space P, $\hat{E}(P)$, $E^*(P)$ and $E(P)$ coincide.

To state the definitions, we need the concept of a dyadic expansion ([2], 4.3, 4.16), the definition of which is re–stated below. Observe that the terminology is different from that in [2]:we call a dyadic expansion (of a space $P \in \mathfrak{W}$) what was called a pure dyadic expansion in [2], and the term "subspace" is used here instead of "pure subspace" used in [2].

2.23. Notation and definition. A)$\mathcal{D}$ will denote the collection of all $D \subset \{0,1\}^*$ such that $0 < |D| < \omega$, $[D] = D$ and $|br(u, D)| \neq 1$ for all $u \in D$. If $D \in \mathcal{D}$, then we put $D' = \{u \in D : br(u, D) \neq \emptyset\}$, $D'' = D \setminus D'$. – B) If Q is a set, then a collection $(Q_u : u \in D)$ will be called a <u>dyadic expansion</u> (abbreviated d.e.) of Q if $D \in \mathcal{D}$, $Q_0 = Q$ and, for each $u \in D'$, $Q_u = Q_{u_0} \cup Q_{u_1}$, $Q_{u_0} \cap Q_{u_1} = \emptyset$. – C) If $P \in \mathfrak{S} \cup \mathfrak{W}$, then a collection $(P_u : u \in D)$ will be called a <u>dyadic expansion</u> of P in all P_u are subspaces of P and there is a d.e. $(Q_u : u \in D)$ of $|P|$ such that $P_u = Q_u.P$ for all $u \in D$.

2.24. Definition. If $P \in \mathfrak{W}_F$ and $\mathcal{Z} = (P_u : u \in D)$ is a d.e. of P, we put $E(P, \mathcal{Z}) = \sum(H(wP_{u_0}, wP_{u_1})d(P_u) : u \in D')$. If $P \in \mathfrak{W}_F$, then $E^*(P)$ denotes the infimum of all $E(P, \mathcal{Z})$ where $\mathcal{Z} = (P_u : u \in D)$ is a d.e. of P such that (*) $\|P_u\| \leq 1$ for all $u \in D''$. – Evidently, the condition (*) can be replaced by (**) $\|P\| = 1$ for all $u \in D''$.

2.25. Fact. For each $P \in \mathfrak{W}_F$, there exists (i) a d.e. $\mathcal{Z} = (Q_u.P : u \in D)$ of P such that $E^*(P) = E(P, \mathcal{Z})$ and $|Q_u| = 1$ for all $u \in D''$, (ii) a d.e. $\mathcal{T} = (T_v.P : v \in D)$ of P such that $E^*(P) = E(P, \mathcal{T})$ and $d(T_v) = 0$ iff $v \in D''$.

2.26. Proposition. *For every $P \in \mathfrak{W}_F$, $\hat{E}(P) = E^*(P)$.*

PROOF : The equality $\hat{E}(P) = E^*(P)$ is an easy consequence of the following two assertions, the proofs of which can be omitted.

A) Let $P \in \mathfrak{S} \cup \mathfrak{W}$ and let $d(P) < \infty$. Let $\mathcal{Z} = (Q(u) : u \in D)$ be a d.e. of $|P|$ such that $(Q(u).P : u \in D)$ is a d.e. of P (if $P = \langle Q, \varrho, \mu\rangle \in \mathfrak{W}$, this means that $Q(u)$ are $\bar{\mu}$–measurable). Define a mapping f of $|P|$ into $|K_\infty|$ as follows: if $x \in Q(u)$, $u = (u_i : i < n) \in D''$, then $f(x) = (v_i : i < n)$ where $v_i = (u_i, t_i)$, $t_i = d(Q(u \upharpoonright i))$. Then f is a strongly branching regular approximative code of P in K_∞; in addition, if $d(Q(u)) = 0$ for all $u \in D''$, then f is a well–fitting exact code. – B) Let f be a well–fitting strongly branching regular exact code of $P \in \mathfrak{S} \cup \mathfrak{W}$

in K_∞. For $u = (u_i : i < n) \in [fP]$, put $\varphi(u) = (\pi u_i : i < n)$. Let D consist of all $\varphi(u)$, $u \in [fP]$. If $v = \pi(u) \in D$, put $Q(v) = \{x \in P : u \prec fx\}$. Then $(Q(v) : v \in D)$ is a d.e. of $|P|$, $\mathcal{Z} = (Q(v).P : v \in D)$ is a d.e. of P and if $P \in \mathfrak{W}_F$, then $E(P, \mathcal{Z}) = E(f)$. ∎

2.27. Fact. For every $P \in \mathfrak{W}_F$, $E^*(P)$, as defined above (2.24), coincides with $C_E^*(P)$ introduced in [**2**].

This is an easy consequence of 4.15 in [**2**] (see also [**2**], 4.11 and 4.9).

Remark. The fact just mentioned will not be used in what follows. We only want to stress that E^* defined in 2.24 is one of the "extended entropies" examined in [2].

2.28. Definition. A) Let $P = \langle Q, \varrho, \mu \rangle \in \mathfrak{W}_F$. Let $(T_q : q \in Q)$ be a family of disjoint sets, $0 < |T_q| < \omega$. Put $T = \bigcup(T_q : q \in Q)$, $\sigma(s,t) = \varrho(x,y)$ for $s \in T_x$, $t \in T_y$. If $S = < T, \sigma, \nu > \in \mathfrak{W}_F$ and $\nu T_q = \mu q$ for all $q \in Q$, we will say that S is obtained from P by splitting. – B) For every $P \in \mathfrak{W}_F$, $E(P)$ will denote the infimum of all $E^*(S)$ where S is a space obtained from P by splitting.

2.29. Fact. For every $P \in \mathfrak{W}_F$, $E(P)$, as defined above, coincides with $C_E(P)$ introduced in [**2**].

This is an easy consequence of [**2**], 3.23. – Remarks. 1) The functional C_E is one of the functionals C_τ introduced in [**2**], 3.17. – 2) Similarly as with 2.27, the fact stated above will not be used in the sequel. However, it seems useful to point out that E, as defined in 2.28, coincides with one of the "extended entropies".

2.30. Lemma. *Let $\varrho \in \mathcal{S}(Q)$, $a, b \in Q$, $a \neq b$. Let $c \in R_+$, $c > 0$; for $0 \le t \le 1$, let μ_t be a measure on Q, $P_t = \langle Q, \varrho, \mu_t \rangle \in \mathfrak{W}_F$, $\mu_t q = \mu_0 q$ for $q \in Q \setminus \{a, b\}$, $\mu_t a = tc$, $\mu_t b = (1-t)c$. Let $\mathcal{Z} = (Q(u) : u \in D)$ be a dyadic expansion of Q such that $|Q(u)| = 1$ for $u \in D''$. Let $x, y \in D''$, $Q(x) = \{a\}$, $Q(y) = \{b\}$. Then either (I) the diameter of $Q(x \wedge y)$ in some (hence in all) P_t, $0 < t < 1$, is zero and $E(P_0, \mathcal{Z}) \vee E(P_1, \mathcal{Z}) \le E(P_t, \mathcal{Z})$ for $0 < t < 1$, or (II) the diameter mentioned above is positive and $E(P_0, \mathcal{Z}) \wedge E(P_1, \mathcal{Z}) < E(P_t, \mathcal{Z})$ for $0 < t < 1$.*

PROOF : Evidently, for any $X \subset Q$ and $0 < s < t < 1$, the diameters of X in P_s and P_t coincide; their common value will be denoted by $d(X)$. – If $d(Q(x \wedge y)) = 0$, then it is easy to see that all $E(P_t, \mathcal{Z})$, $0 < t < 1$, coincide and $E(P_i) \le E(P_t)$ for $i = 0, 1$, $0 < t < 1$. – Consider the case $d(Q(x \wedge y)) > 0$. Put $h = |x \wedge y|$, $m = |x| - h$, $n = |y| - h$. For $k \le m$ put $u_k = x \restriction (h+k)$; for $k \le n$ put $v_k = y \restriction (h+k)$. For every $t = (t_i : i < p) \in D$, $p > 0$, put $\bar{t} = (\bar{t}_i : i < p)$ where $\bar{t}_i = t_i$ for $i < p-1$, $\bar{t}_{p-1} = 1 - t_{p-1}$. For $X \subset Q$, put $\mu X = \mu_t(X \setminus \{a, b\})$; clearly, μX does not depend on t. For $1 \le k \le m$, put $r_k = \mu Q(u_k)$, $s_k = \mu Q(\bar{u}_k)$, $z_k = d(Q(u_k))$; for $1 \le k \le n$, put $r'_k = \mu(Q(v_k))$, $s'_k = \mu Q(\bar{v}_k)$, $z'_k = d(Q(v_k))$. Put $r'_0 = r_0 = \mu Q(x \wedge y)$, $z'_0 = z_0 = d(Q(x \wedge y))$. ∎

For $0 < t < 1$, put $\varphi(t) = E(P_t, \mathcal{Z})$. It is easy to see that

(1) $E(P_0, \mathcal{Z}) \le \lim\limits_{t \to 0} \varphi(t)$, $E(P_1, \mathcal{Z}) \le \lim\limits_{t \to 1} \varphi(t)$. Clearly, for $0 < t < 1$ we have

(2) $\varphi(t) = \sum\limits_{k=1}^{m} H(r_{k+1} + tc, s_{k+1}) z_k + \sum\limits_{k=1}^{n} H(r'_{k+1} + tc, s'_{k+1}) z'_k + H(r_1 + tc, r_1 + c - tc) z_0 + \kappa$, where κ is a constant, independent of t. Hence,

(3) $\varphi(t) = \sum_{k=1}^{m-1} z_k(L(r_{k+1} + tc - L(r_k + tc)) + \sum_{k=1}^{n-1} z'_k(L(r'_{k+1} + c - tc) - L(r'_k + c - tc)) + z_0(L(r_1 + tc) + L(r'_1 + c - tc)) + \kappa_1$, where κ_1 is a constant. From (3), we easily get

(4) $\varphi(t) = \sum_{k=0}^{m-1} (z_k - z_{k+1})L(r_{k+1} + tc) + \sum_{k=0}^{n-1} (z'_k - z'_{k+1})L(r'_{k+1} + c - tc) + \kappa_1$.

Let $\psi(t)$ denote the derivative of φ at t, $0 < t < 1$. Then

(5) $\psi(t)/\log e = -e - c\sum_{k=0}^{m-1} (z_k - z_{k+1})\log(r_{k+1} + tc) + c\sum_{k=0}^{n-1} (z'_k - z'_{k+1})\cdot \log(r'_{k+1} + c - tc)$.

Since $z'_0 = z_0 = d(Q(x \wedge y)) > 0$ and $z_m = z_n = 0$, some $z_k - z_{k+1}$ (and also some $z'_k - z'_{k+1}$) is positive. Hence ψ is a decreasing function. This implies that $(\lim_{t\to 0} \varphi(t)) \wedge (\lim_{t\to 1} \varphi(t)) < \varphi(t)$ for $0 < t < 1$. By (1), we get $E(P_0, \mathcal{Z}) \wedge E(P_1, \mathcal{Z}) < E(P_t, \mathcal{Z})$ for $0 < t < 1$.

2.31. Proposition. *For any $P = \langle Q, \varrho, \mu\rangle \in \mathfrak{W}_F$, $\hat{E}(P) = E^*(P) = E(P)$.*

PROOF : By 2.26, $\hat{E}(P) = E^*(P)$. Clearly, $E(P) \le E^*(P)$. Thus, we have to show that $E(P) \ge E^*(P)$, i.e., that $E^*(P) \le E^*(S)$ for any FW–space S obtained from P by splitting (see 2.28). To prove this assertion, it is, clearly, sufficient to show that $E^*(S) \ge E^*(P)$ whenever S is of the form $\langle T, \sigma, \nu\rangle$ described in 2.28 and such that $T_p = \{a, b\}$, $a \ne b$, for some $p \in Q$, $T_q = \{q\}$, for $q \in Q \setminus \{p\}$, and $\nu a + \nu b = \mu p$. By 2.30, we get $E^*\langle T, \sigma, \nu\rangle \ge E^*\langle T, \sigma, \nu'\rangle$ where $\nu' q = \mu q$ for $a \ne q \ne b$ and either $\nu' a = \mu p$, $\nu' b = 0$ or $\nu' a = 0$, $\nu' b = \mu p$. Evidently, in both cases, $E^*\langle T, \sigma, \nu'\rangle = E^*(P)$. ∎

2.32.. In view of 2.31, we will write $E(P)$ instead of $\hat{E}(P)$ or $E^*(P)$ provided P is an FW–space, and the fact that $\hat{E}(P) = E^*(P) = E(P)$ will be used without explicit reference to 2.31, as a rule.

3.

3.1. Lemma. *Let $P \in \mathfrak{S}_F \cup \mathfrak{W}_F$. There exist well-fitting strongly branching regular exact codes f_1, f_2, f_3 of P in K_∞ such that* (1) $\delta(f_1) = \delta P$, (2) *if $P \in \mathfrak{W}_F$, then $\lambda(f_2) = \lambda P$, $E(f_3) = E(P)$.*

PROOF : The assertions concerning δ and λ follow easily from 2.20 and 2.21. The assertion concerning E follows from 2.20, 2.21 and the equality $E(P) = E^*(P) = \widehat{E}(P)$. ∎

3.2. Remark. There are very simple FW–spaces P possessing no regular exact code f in K_∞ with both $\lambda(f) = \lambda P$ and $E(f) = E(P)$, as the following example shows. – Let $Q = \{a, b, c\}$, $\rho(a, b) = \rho(a, c) = t > 1$, $\rho(b, c) = 1$. Let $\mu a = \varepsilon$, $0 < \varepsilon < 1/3$, $\mu b = \mu c = (1 - \varepsilon)/2$. Put $P = \langle Q, \rho, \mu\rangle$. An elementary calculation shows that (1) $E(P) = tH(\varepsilon, 1 - \varepsilon) + 1 - \varepsilon$, (2) $E(f) = E(P)$ iff (3) $\{q \in Q\colon f(q) \restriction 1 = (i, t)\} = \{a\}$ for $i = 0$ or for $i = 1$. On the other hand, if f is a well–fitting strongly branching exact code for P in K_∞, then $\lambda(f) = t + 1 - \varepsilon$ if f satisfies (3) whereas $\lambda(f) = t + t(1 + \varepsilon)/2$ if $\{q \in Q\colon f(q) \restriction 1 = (0, t)\}$ is equal to $\{b\}$ or to $\{c\}$. Assume that $t < 2(1 - \varepsilon)/(1 + \varepsilon)$. Then, clearly, $\lambda(f) > \lambda(P)$ if f satisfies (3).

3.3. Proposition. *Let $P \in \mathfrak{S}_F \cup \mathfrak{W}_F$. Then* (1) *for every partition* (P_0, P_1) *of* P, $\delta P \le d(P) + \delta P_0 \vee \delta P_1$, (2) *for some partition* (P_0, P_1) *of* P, $\delta P = d(P) + \delta P_0 \vee \delta P_1$.

PROOF : I. Let (Q_0, Q_1) be a partition of $|P|$, $P_i = Q_i \cdot P$. By 3.1, there are regular codes f_i of P_i, $i = 0, 1$, in K_∞ such that $\delta(f_i) = \delta P_i$. For $i = 0, 1$, put $a_i = (i, t)$ where $t = d(P)$. For $x \in |P|$ put $f(x) = (a_i) \cdot f_i(x)$ if $x \in Q_i$. Clearly, $f \in \mathcal{K}^r_\infty \cap \operatorname{cod}(0, P)$, $\delta(f) = t + \delta(f_0) \vee \delta(f_1)$, hence $\delta P \le d(P) + \delta P_0 \vee \delta P_1$. – Let f be a well-fitting strongly branching regular code of P in K_∞ such that $\delta(f) = \delta(P)$. Clearly, $|\operatorname{br}(\emptyset, fP)| = 2$. Let $\operatorname{br}(\emptyset, fP) = \{a_0, a_1\}$. Since f is well–fitting, $\lambda a_0 = \lambda a_1 = d(P)$. Put $Q_i = \{x \in |P| \colon (a_i) \prec fx\}$, $P_i = Q_i \cdot P$. If $x \in Q_i$, define $f_i(x)$ by $f(x) = (a_i) \cdot f_i(x)$. Clearly, $f_i \in \mathcal{K}^r_\infty \cap \operatorname{cod}(0, P_i)$, $\delta P = \delta(f) = d(P) + \delta(f_0) \vee \delta(f_1)$. In view of $\delta P \le d(P) + \delta P_0 \vee \delta P_1$, this proves $\delta P = d(P) + \delta P_0 \vee \delta P_1$. ∎

3.4. Proposition. *Let $P \in \mathfrak{W}_F$. Then* (1) *for every partition* (P_0, P_1) *of* P, $\lambda P \le d(P) \cdot wP + \lambda P_0 + \lambda P_1$, $E(P) \le d(P) H(wP_0, wP_1) + E(P_0) + E(P_1)$, (2) *there are partitions* (P_0, P_1) *and* (S_0, S_1) *of* P *such that* $\lambda P = d(P) \cdot wP + \lambda P_0 + \lambda P_1$, $E(P) = d(P) H(wS_0, wS_1) + E(S_0) + E(S_1)$.

We omit the proof since it is analogous to that of 3.3.

3.5. Characterization theorem for δ on finite spaces. – *Let $\mathfrak{P} = \mathfrak{S}_F$ or $\mathfrak{P} = \mathfrak{W}_F$. The functional δ defined on $\mathfrak{P}$ is the largest functional φ on $\mathfrak{P}$ such that $\varphi P = 0$ if $\|P\| \le 1$ and, for every partition (P_0, P_1) of a space $P \in \mathfrak{P}$, the inequality $\varphi P \le d(P) + \varphi P_0 \vee \varphi P_1$ is satisfied.*

PROOF : I. By 3.3, δ satisfies the conditions stated in the theorem. – II. Let φ satisfy the conditions in question. We are going to prove that $\varphi P \le \delta P$ for all $P \in \mathfrak{P}$. Suppose this is not true and choose a $P \in \mathfrak{P}$ with $\varphi P > \delta P$ and with the least possible $\|P\|$. By 3.4, there is a partition (P_0, P_1) off P such that $\delta P = d(P) + \delta P_0 \vee \delta P_1$. Then $\varphi P_i = \delta P_i$, hence $\varphi P \le \delta P$, which contradicts the assumption. ∎

3.6. Characterization theorem for λ and E on finite spaces. – *The functional λ (respectively, E), defined on $\mathfrak{W}_F$, is the largest functional φ on $\mathfrak{W}_F$ such that $\varphi P = 0$ if $\|P\| = 1$ and, for every partition (P_0, P_1) of a space $P \in \mathfrak{W}_F$, the inequality $\varphi P \le d(P) \cdot wP + \varphi P_0 + \varphi P_1$ (respectively, $\varphi P \le d(P) H(wP_0, wP_1) + \varphi P_0 + \varphi P_1$) is satisfied.*

The proof is similar to that of 3.5 and can be omitted.

3.7. Definition. If $\rho \in S(Q)$ and, for any x, y, $z \in Q$, $\rho(x, y) \le \rho(x, z) \vee \rho(z, y)$, then ρ will be called a U–semimetric. If, in addition, $\rho(x, y) = 0$ implies $x = y$, then ρ is called an ultrametric and $\langle Q, \rho \rangle$ is called an ultrametric space.

3.8. Definition. Let $P = \langle Q, \rho, \mu \rangle \in \mathfrak{W}$. We will say that ρ is (1) a U–semimetric with respect to μ (or simply a U–semimetric) if there is a set $Z \subset Q^3$ such that $[\mu^3](Z) = 0$ and $\rho(x, y) \le \rho(x, z) \vee \rho(z, y)$ whenever $(x, y, z) \in Q^3 \setminus Z$, (2) an ultrametric with respect to μ (or simply an ultrametric) if, in addition, there is a set $Y \subset Q^2$ such that $[\mu^2](Y) = 0$ and $\rho(x, y) > 0$ whenever $(x, y) \in Q^2 \setminus Y$, $x \ne y$.

If ρ is an ultrametric with respect to μ, then $\langle Q, \rho, \mu\rangle$ will be called an ultrametric W-space.

3.9. Lemma. *Let $P = \langle Q, \rho, \mu\rangle \in \mathfrak{W}_F$. Let $a = \min\{\rho(x,y) : x, y \in Q, \rho(x,y) > 0\}$ and let $\rho_1(x,y) = (\rho(x,y) - a) \vee 0$. Then (1) $E(P) \geq aE(0 * P) + E\langle Q, \rho_1, \mu\rangle$, (2) if ρ is a U-semimetric with respect to μ, then $E(P) = aE(0 * P) + E\langle Q, \rho_1, \mu\rangle$.*

PROOF : I. For $x, y \in Q$, put $\rho_2(x,y) = a$ if $\rho(x,y) \geq a$, $\rho_2(x,y) = 0$ if $\rho(x,y) = 0$. Clearly, for every $M \subset P$, $d\langle M, \rho\rangle = d\langle M, \rho_1\rangle + d\langle M, \rho_2\rangle$. Put $P_i = \langle Q, \rho_i, \mu\rangle$. Then, for each d.e. $\mathcal{Z} = (Q_u : u \in D)$ of Q, $E(P, \mathcal{Z}) = E(P_1, \mathcal{Z}) + E(P_2, \mathcal{Z})$. This implies $E(P, \mathcal{Z}) \geq E(P_1) + E(P_2)$, hence $E(P) \geq E\langle Q, \rho_1, \mu\rangle + E(0 * P)$. – II. To prove the assertion, it is sufficient to consider the case when $\mu q > 0$ for all $q \in Q$. By 2.25, there is a d.e. $\mathcal{Z} = (Q_u : u \in D)$ of Q such that $|Q_u| = 1$ for $u \in D''$ and $E(P_1, \mathcal{Z}) = E(P_1)$. It is easy to show that $E(P, \mathcal{Z}) + E(P_1, \mathcal{Z}) + aH(\mu Q_u : u \in T)$ where T consists of $u \in D$ such that $d(Q_u) = 0$ whereas $d(Q_v) > 0$ if $v \prec u$, $v \neq u$. Since $(0 * \rho)(x,y) \in \{0, 1\}$ for all $x, y \in Q$, it is easy to see that $E(0 * P) = H(\mu Q_v : u \in T)$, hence $E(P, \mathcal{Z}) = E(P_1) + aE(0 * P)$. This implies $E(P) \leq E\langle Q, \rho_1, \mu\rangle + aE(0 * P)$ and the assertion follows by (1). ■

3.10. Theorem. *For every FW-space $P = \langle Q, \rho, \mu\rangle$, $E(P) \geq \int_0^\infty E(t * P)dt$, and if ρ is a U-semimetric FW-space, then $E(P) = \int_0^\infty E(t * P)dt$.*

PROOF : Let $(a_i : i < n)$ be the increasing sequence of all $\rho(x,y)$, $x, y \in Q$. From 3.9 we obtain, by induction, the inequality $E(P) \geq \sum_{k=0}^{n-2} E(a_k * P)(a_{k+1} - a_k)$ (respectively, if ρ is a U-semimetric, the corresponding equality). It is easy to see that if $k < n - 1$, $a_k \leq t < a_{k+1}$, then $t * P = a_k * P$, and if $a_{n-1} \leq t$, then $E(t * P) = 0$. Hence $\int_0^\infty E(t * P)dt = \sum_{k=0}^{n-2} E(a_k * P)(a_{k+1} - a_k)$, which proves the theorem. ■

3.11. Lemma. *Let $P = \langle Q, \rho, \mu\rangle \in \mathfrak{W}_F$. Let $a = \min\{\rho(x,y) : x, y \in Q, \rho(x,y) > 0\}$ and let $\rho_1(x,y) = (\rho(x,y) - a) \vee 0$. Then $\lambda(P) \geq a\lambda(0 * P) + \lambda\langle Q, \rho_1, \mu\rangle$.*

The proof is analogous to that of 3.9 and can be omitted.

3.12. Proposition. *For every FW-space P, $\lambda(P) \geq \int_0^\infty \lambda(t * P)dt$.*

This follows from 3.11 in the same way as 3.10 follows from 3.9.

3.13. Examples. A) Let $Q = \{1, 2, 3, 4\}$, $\rho(i,j) = |i - j|$, $\mu\{i\} = 1/4$ for all $i \in Q$. Put $P = \langle Q, \rho, \mu\rangle$. It is easy to see that $E(P) = 4$. On the other hand, $\int_0^\infty E(t * P)dt = E(0 * P) + E(1 * P) + E(2 * P) = 2 + H(1/2, 1/3) + H(3/4, 1/4) < 4$. Thus, if $P = \langle Q, \rho, \mu\rangle \in \mathfrak{W}_F$ and ρ is not a U-semimetric, then the equality $E(P) = \int_0^\infty E(t * P)dt$ need not hold. – B) Let $Q = \{1, 2, 3, 4\}$, $\rho(i,j) = 1$ if $i \neq j$, $i \neq 4$, $j \neq 4$, $\rho(4, i) = 2$ for $i = 1, 2, 3$, $\mu\{i\} = 1/4$ for all $i \in Q$. Put

$P = \langle Q, \rho, \mu\rangle$. Clearly, P is ultrametric. It is easy to see that $\lambda(P) = 13/4$ (this value is obtained for the code $1 \longmapsto ((1,2),(1,1),(1,1))$, $2 \longmapsto ((1,2),(1,1),(0,1))$, $3 \longmapsto ((1,2),(0,1))$, $4 \longmapsto ((0,2))$. Evidently, $\int_0^\infty \lambda(t * P)dt = 3 < 13/4$. Thus, the inequality in 3.12 can be strict even if P is ultrametric.

4.

4.1. Fact. If $P_i \in \mathfrak{S}_F$ or $P_i \in \mathfrak{W}_F$, $i = 1,2$, then $\delta(P_1 \times P_2) \leq \delta P_1 + \delta P_2$. If P_1, $P_2 \in \mathfrak{W}_F$, then $\lambda(P_1 \times P_2) \leq \lambda P_1 \cdot wP_2 + \lambda P_2 \cdot wP_1$, $E(P_1 \times P_2) \leq E(P_1) \cdot wP_2 + E(P_2) \cdot wP_1$.

PROOF : We prove the assertion for δ only, since for λ and E the proof is analogous. Put $P = P_1 \times P_2$. Clearly, $\delta P \leq \delta P_1 + \delta P_2$ holds if $\|P\| \leq 1$. Assume that it holds if $\|P\| \leq n$ and consider the case $\|P\| = n+1$. We can assume $d(P_1) \geq d(P_2)$. By 3.3, there is a partition (P_{10}, P_{11}) of P_1 such that (*) $\delta P_1 = d(P_1) + \delta P_{10} \vee \delta P_{11}$. Since $\|P_{1i} \times P_2\| \leq n$, we have $\delta(P_{1i} \times P_2) \leq \delta P_{1i} + \delta P_2$, $i = 0,1$. By 3.3, $\delta P \leq d(P) + \delta(P_{10} \times P_2) \vee \delta(P_{11} \times P_2) \leq d(P) + \delta(P_1 \times P_2) \leq \delta P_1 + \delta P_2$. ∎

4.2. Remark. None of the inequalities in 4.1 can be replaced by an equality. For δ and λ, this is well known already for FW–spaces of the form $\langle Q, 1, \mu\rangle$. We give an example concerning E. – Let $Q = \{1,2,3\}$, $\rho(i,j) = |i-j|$, $\mu\{i\}1/3$ for $i = 1,2,3$. Put $P = \langle Q, \rho, \mu\rangle$. It is easy to see that $E(P) = 2H(2/3,1/3) + H(1/3,1/3) = 2\log 3 - 2/3$. We are going to show that $E(P^2) < 2E(P)$. Consider a d.e. $\mathcal{Z} = (Q_u : u \in D)$ of Q^2 such that (1) $|Q_u| = 1$ for $u \in D''$, (2) $(Q_{00}, Q_{01}, Q_{10}, Q_{11}) = (A, B \setminus A, \{(1,3)\}, \{(3,1)\})$ where $A = \{1,2\} \times \{1,2\}$, $B = \{2,3\} \times \{2,3\}$. It is easy to see that $E(P^2, \mathcal{Z}) = 2H(4/9,3/9,1/9,1/9) + H(1/9,1/9,1/9,1/9) + H(1/9,1/9,1/9) = (11\log 3)/3 - 8/9$. Hence $E(P^2) \leq E(P^2, \mathcal{Z}) \leq (11\log 3)/3 - 8/9 < 2(2\log 3 - 2/3) = 2E(P)$.

4.3. Proposition. *Let $P_i = \langle Q_i, \rho_i, \mu_i\rangle$, $i = 1,2$, be FW–spaces. If, for $i = 1,2$, ρ_i is a U–semimetric with respect to μ_i (in particular, if P_1 and P_2 are ultrametric), then $E(P_1 \times P_2) = E(P_1) \cdot wP_2 + E(P_2) \cdot wP_1$.*

PROOF : Clearly, we can assume that $wP_i = 1$ and $\mu_i q > 0$ for all $q \in Q_i$. – I. Let Ψ denote the class of all FW–spaces $\langle Q, \rho, \mu\rangle$ such that (1) $\mu Q = 1$, (2) $\mu q > 0$ for all $q \in Q$, (3) ρ is a U–semimetric, (4) $\rho(Q \times Q) \subset \{0,1\}$. For every $T = \langle Q, \rho, \mu\rangle \in \Psi$, let $\mathcal{Z}_T$ consist of all $X \subset Q$ such that $d(X) = 0$ whereas $d(Y) = 1$ whenever $X \subset Y \subset Q$, $X \neq Y$. It is easy to see that $\mathcal{Z}_T$ is a disjoint collection and (*) $E(T) = H(\mu Z : Z \in \mathcal{Z}_T)$. Clearly, if $P, S \in \Psi$, $P = \langle |P|, \rho_P, \mu_P\rangle$, $S = \langle |S|, \rho_S, \mu_S\rangle$, then $T = P \times S \in \Psi$ and $\mathcal{Z}_T = \{U \times V : U \in \mathcal{Z}_P, V \in \mathcal{Z}_S\}$. Write μ_T instead of $\mu_P \times \mu_S$, ρ_T instead of $\rho_P \times \rho_S$. Then, by (*), we have $E(T) = H(\mu_T Z : Z \in \mathcal{Z}_T) = H(\mu_P U : U \in \mathcal{Z}_P) + H(\mu_S V : V \in \mathcal{Z}_S)$, hence $E(T) = E(P) + E(S)$. – II. Clearly it is sufficient to consider the case when $P_i = \langle Q_i, \rho_i, \mu_i\rangle$, $i = 1,2$, are FW–spaces such that $wP_i = 1$, $\mu_i q > 0$ for all $q \in Q_i$ and ρ_i is a U–semimetric μ_i. Then, $\rho_1 \times \rho_2$ is a U–semimetric and therefore, by 3.10, $E((P_1 \times P_2) = \int_0^\infty E(t * (P_1 \times P_2))dt$. By I, $E(t * (P_1 \times P_2)) = E(t * P) + E(t * P_2)$ for all $t \in R_+$. Hence $E(P_1 \times P_2) =$

$\int\limits_0^\infty E(t * P_1)dt + \int\limits_0^\infty E(t * P_2)dt = E(P_1) + E(P_2)$, by 3.10. ■

4.4. Fact. Let m, $n \in N$, $m > 0$, $n > 0$. If $P \in \mathfrak{S}_F \cup \mathfrak{W}_F$, then $\delta(P^{m+n}) \leq \delta(P^m) + \delta(P^n)$. If $P \in \mathfrak{W}_F$, $wP = 1$, then $\lambda(P^{m+n}) \leq \lambda(P^m) + \lambda(P^n)$.

This is a consequence of 4.1.

4.5. Fact. Let $x_k \in R_+$ for $k \in N$, $k \geq 1$. Assume that for all m, $n \in N \setminus \{0\}$, $x_{m+n} \leq x_m + x_n$. Then $\lim(x/n) = \inf(x/n\colon n > 0)$.

This is well known.

4.6. Definition. If $P \in \mathfrak{S} \cup \mathfrak{W}$, then $\inf(\delta(P^n)/n\colon n \in N, n > 0)$ will be denoted by $\Delta(P)$ and will be called the final entropic content of P. If $P \in \mathfrak{W}$, then $\inf(\lambda(P^n)/n(wP)^{n-1}\colon n \in N, n > 0)$ will be denoted by $\Lambda(P)$ and will be called the final entropy of P.

4.7. Fact. If $P \in \mathfrak{S}_F \cup \mathfrak{W}_F$, then $\Delta(P) = \lim(\delta(P^n)/n)$. If $P \in \mathfrak{W}_F$, $wP > 0$, then $\Lambda(P) = \lim(\lambda(P^n)/n(wP)^{n-1})$; in particular, $\Lambda(P) = \lim(\lambda(P^n)/n)$ if $wP = 1$.

This is a consequence of 4.4 and 4.5.

4.8. Remarks. 1) The equalities in 4.7 do hold for all $P \in \mathfrak{S}$, respectively $P \in \mathfrak{W}$. This will be proved in the forthcoming Part II. – 2) It will be proved below (4.21) that if $P \in \mathfrak{W}_F$, $wP = 1$, then $\Lambda(P) = \inf(E(P^n)/n\colon n \in N, n > 0) = \lim(E(P^n)/n)$, which justifies the term "final entropy".

4.9. Proposition. *If $P \in \mathfrak{S}_F \cup \mathfrak{W}_F$, then $\Delta(P^m) = m\,\Delta\,(P)$ for every $m \in N$, $n > 0$. If P, $S \in \mathfrak{S}_F$ or P, $S \in \mathfrak{W}_F$, then $\Delta(P \times S) \leq \Delta(P) + \Delta(S)$.*

PROOF : From 4.7, $\Delta(P^m) = m\,\Delta\,(P)$ follows at once. By 4.1, $\delta(P^m \times S^n) \leq \delta(P^m) + \delta(S^n)$, from which the inequality for Δ follows by 4.7. ■

4.10. Proposition. *If $P \in \mathfrak{S}_F \cup \mathfrak{W}_F$ and (P_0, P_1) is a partition of P, then $\Delta(P) \leq d(P) + \Delta(P_0) \vee \Delta(P_1)$.*

PROOF : It is easy to see that, for every $n \in N$, $n > 0$, $\delta(P^n) \leq n \cdot d(P) + \max\{\delta(P_0^k \times P_1^m)\colon k + m = n\}$ where we put $P_0^n \times P_1^0 = P_0^n$, $P_0^0 \times P_1^n = P_1^n$. Let $\varepsilon > 0$. By 4.7, there is an $n_0 \in N$ such that if $j > n_0$, then $\delta(P_0^j)/j < \Delta(P_0) + \varepsilon$, $\delta(P_1^j)/j < \Delta(P_1) + \varepsilon$. Choose $n_1 \in N$ such that $n_0(\delta P_0 \vee \delta P_1) < \varepsilon n_1$. Let $n > n_1$, $n = k + m$. Then either (I) $k > n_0$, $m > n_0$ or (II) $m \leq n_0$ or $k \leq n_0$. If $k > n_0$, $m > n_0$, then $\delta(P_0^k \times P_1^m) \leq k(\Delta(P_0) + \varepsilon) + m(\Delta(P_1) + \varepsilon) \leq n(\Delta(P_0) \vee \Delta(P_1)) + n\varepsilon$. – If, e.g., $k \leq n_0$, then $\delta(P_0^k \times P_1^m) \leq k \cdot \delta P_0 + m(\Delta(P_1) + \varepsilon) \leq (\varepsilon n + n)\,\Delta(P_1) + \varepsilon \leq n(\Delta(P_0) \vee \Delta(P_1)) + 2n\varepsilon$. Thus, in both cases, $\delta(P^n)/n \leq d(P) + \Delta(P_0) \vee \Delta(P_1) + 2\varepsilon$. This proves the proposition. ■

4.11. Characterization theorem for Δ on finite spaces. *Let $\mathfrak{P}$ be either the class of all finite semimetric spaces or that of all FW–spaces. The functional Δ defined on $\mathfrak{P}$ is the largest of all functionals φ on $\mathfrak{P}$ such that $\varphi P = 0$ if $\|P\| \leq 1$, $\varphi(P^n) = n\varphi(P)$ for every $P \in \mathfrak{P}$ and every $n \in N$, $n > 0$, and $\varphi P \leq d(P) + \varphi(P_0) \vee \varphi(P_1)$ for every partition (P_0, P_1) of a space $P \in \mathfrak{P}$.*

PROOF : I. By 4.9 and 4.10, Δ satisfies the conditions stated in the theorem. – II. Let φ satisfy the conditions. Then, by 3.5, $\varphi(S) \leq \delta(S)$ for every $S \in \mathfrak{P}$ and therefore $n\varphi(P) = \varphi(P^n) \leq \delta(P^n)$, $\varphi(P) \leq \delta(P^n)/n$ for all $P \in \mathfrak{P}$ and $n \in N$, $n > 0$. This implies $\varphi(P) \leq \Delta(P)$. ■

4.12. Facts. I) For every $P \in \mathfrak{W}$ and every $m \in N$, $m > 0$, $\Lambda(P^m) = m(wP)^{m-1}\Lambda(P)$; in particular, $\Lambda(P^m) = m\Lambda(P)$ if $wP = 1$. – II) If P, $S \in \mathfrak{W}_F$, then $\Lambda(P \times S) \leq \Lambda(P) \cdot wS + \Lambda(S) \cdot wP$.
PROOF : I. We can assume that $wP = 1$. Then, by 4.7, $\Lambda(P^m) = \lim_{n\to\infty}(\lambda(P^{nm}/n) = m \cdot \lim_{n\to\infty}(\lambda(P^{nm})/nm)$. Hence, again by 4.7, $\Lambda(P^m) = m\Lambda(P)$. – II. We can assume that $wP = wS = 1$. By 4.7, $\Lambda(P \times S) = \lim(\lambda(P^n \times S^n)/n)$, $\Lambda(P) = \lim(\lambda(P^n)/n)$, $\Lambda(S) = \lim(\lambda(S^n)/n)$,. This implies $\Lambda(P \times S) \leq \Lambda(P) + \Lambda(S)$, since $\lambda(P^n \times S^n) \leq \lambda(P^n) + \lambda(S^n)$, by 4.1. ■

4.13. In 4.14, 4.15 and 4.18 below we prove some propositions concerning those classes $\mathfrak{P} \subset \mathfrak{W}$ which satisfy the following conditions: (1) if $\langle Q, a\rho, b\mu\rangle \in \mathfrak{P}$ and a, $b \in R_+$, then $\langle Q, a\rho, b\mu\rangle \in \mathfrak{W}$, (2) if S is a subspace of $P \in \mathfrak{P}$, then $S \in \mathfrak{P}$, (3) if (P_0, P_1) is a partition of $P \in \mathfrak{P}$, then $\lambda P \leq d(P) \cdot wP + \lambda P_0 + \lambda P_1$, (4) if P_1, $P_2 \in \mathfrak{P}$, $wP_1 = wP_2 = 1$, then $P_1 \times P_2 \in \mathfrak{P}$ and $\lambda P \leq \lambda P_1 + \lambda P_2$. – By 3.4 and 4.1, the class $\mathfrak{W}_F$ satisfies (1)–(4). In the forthcoming Part II, it will be shown that (1)–(4) are satisfied by $\mathfrak{W}$ as well.

4.14. Lemma. *Let $\mathfrak{P} \subset \mathfrak{W}$ satisfy* (1)–(4) *from* 4.13. *Let $P \in \mathfrak{P}$ and let $(P_1, \ldots, P_n)$ be a partition of P. Put $S = \langle\{1, \ldots, n\}, t, \nu\rangle$ where $t = d(P)$, $\nu\{k\} = wP_k$. Then $\lambda P \leq \lambda S + \sum(\lambda P_i \colon i = 1, \ldots, n)$.*

PROOF : By 3.4, the assertion is true for $n = 2$. Assume that it holds for all $n < m$. Let $(Q_1, \ldots, Q_m)$ be a partition of $|P|$ such that $P_i = Q_i \cdot P$ are subspaces of P. By 3.4, there is a partition (X_0, X_1) of $\{1, \ldots, m\} = |S|$ such that $\lambda S = d(S) \cdot wS + \lambda S_0 + \lambda S_1$. Where $S_i = X_i \cdot S$. Put $Y_j = \bigcup(Q_i \colon i \in X_j)$, $j = 0, 1$, $P^{(j)} = Y_j \cdot P$. By the assumption, we have $\lambda P^{(j)} \leq \lambda S_j + \sum(\lambda P_i \colon i \in X_j)$, $j = 0, 1$. By 3.4, $\lambda P \leq d(P) \cdot wP + \lambda P^{(0)} + \lambda P^{(1)}$. Hence $\lambda P \leq d(S) \cdot wS + \lambda S_0 + \lambda S_1 + \sum(\lambda P_i \colon i \in |S|) = \lambda S + \sum(\lambda P_i \colon i \in |S|)$. ■

4.15. Fact. Let $P \subset \mathfrak{W}$ satisfy (1)–(4) from 4.13. Then for every $P \in \mathfrak{P}$, $\lambda(P^n)/n(wP)^{n-1}$ converges to $\Lambda(P)$ for $n \to \infty$.
PROOF : We can assume $wP = 1$. By (4) from 4.13, $\lambda(P^{m+n}) \leq \lambda(P^m) + \lambda(P^n)$ for all positive m, $n \in N$. By 4.5, this proves the assertion. ■

4.16. Fact. Let ν be a probability measure on $\{0, 1\}$, $\nu 0 > 0$, $\nu 1 > 0$. For every $n \in N$, $n > 0$, and ε, let $B_n(\varepsilon)$ consist of all $x = (x(i) \colon i < n) \in \{0, 1\}^n$ such that $\nu 0 - \varepsilon < |\{i < n \colon x(i) = 0\}|/n < \nu 0 + \varepsilon$. Then, for every sufficiently small $\varepsilon > 0$, (1) $\lim \nu^n(B_n(\varepsilon)) = 1$, (2) $\lim(\log |B_n(\varepsilon)|/hn) = 1$ where $h = H(\nu 0, \nu 1)$.

This is well known: the first assertion is an elementary fact, the second one is easily proved using the Stirling formula.

4.17. Fact. If ν is a probability measure on $\{0, 1\}$, $S = \langle\{0, 1\}, 1, \nu\rangle$, then $\lambda(S^n)/n \to H(\nu 0, \nu 1)$.

This is an easy consequence of 4.16.

4.18. Proposition. *Let $\mathfrak{P} \subset \mathfrak{W}$ satisfy* (1)–(4) *from* 4.13. *Then for every $P \in \mathfrak{P}$ and every partition (P_0, P_1) of P, $\Lambda(P) \leq d(P)H(wP_0, wP_1) + \Lambda(P_0) + \Lambda(P_1)$.*

PROOF : I. We can assume that $d(P) = 1$, $wP_i > 0$, $\lambda P_i < \infty$. Put $a = wP_0$, $b = wP_1$, $c = (\lambda(P_0)/a) \vee (\lambda(P_1)/b)$. Put $S = \langle\{0,1\}, 1, \nu\rangle$ where $\nu 0 = a$, $\nu 1 = b$. If $x = (x(i) \colon i < n) \in \{0,1\}^n$, $n > 0$, put $u(x) = |\{i < n \colon x(i) = 0\}|$, $v(x) = |\{i < n \colon x(i) = 1\}|$, $P(x) = \prod_{i<n} P_{x(i)}$; clearly $wP(x) = \nu^n(x)$. By 4.14, we have (1) $\lambda(P^n) \leq \lambda(S^n) + \sum(\lambda(P(x)) \colon x \in \{0,1\}^n)$ for each $n > 0$. If $n > 0$, $\varepsilon > 0$, put $B_n(\varepsilon) = \{x \in \{0,1\}^n \colon |u(x)/n - a| < \varepsilon\}$. Clearly, for every $\varepsilon > 0$, (2) $\nu^n(B_n(\varepsilon)) \to 1$ for $n \to \infty$. – II. If $n \in N$, $n > 0$, $x \in \{0,1\}^n$, then, by (4) in 4.13, $\lambda(P(x)) \leq wP(x)(u(x) \cdot \lambda P_0/a + v(x) \cdot \lambda P_1/b)$, hence (3) $\lambda(P(x)) \leq \nu^n(x)nc$. – III. Let $\varepsilon > 0$, $\varepsilon < a \wedge b$. By 4.15, there is an $n_0 \in N$ such that (4) $n > n_0$ implies $\lambda(P_0^n)/na^{n-1} < \Lambda(P_0) + \varepsilon, \lambda(P_1^n)/nb^{n-1} < \Lambda(P_1) + \varepsilon$. By (2) and 4.17, there is an $n_1 \in N$, $n_1 > n_0$, such that (5) $n > n_1$ implies (i) $\nu^n(B_n(\varepsilon)) > 1 - \varepsilon$, (ii) $u(x) > n_0$, $v(x) > n_0$ for all $x \in B_n(\varepsilon)$, (iii) $\lambda(S^n) < n(H(a,b) + \varepsilon)$. Let $n > n_1$, $x \in B_n(\varepsilon)$; put $u = u(x)$, $v = v(x)$. Then $\lambda(P(x)) = \lambda(P_0^u \times P_1^v)$ and therefore, by (4) from 4.13 and the inequalities (4) above, $\lambda(P(x)) \leq (\Lambda(P_0) + \varepsilon)ua^{u-1}b^v + (\Lambda(P_1) + \varepsilon)vb^{v-1}a^u$, hence (6) $\lambda(P(x)) \leq (\Lambda(P_0) + \varepsilon)a^{-1} \cdot u(x)\nu^n(x) + (\Lambda(P_1) + \varepsilon)b^{-1} \cdot v(x)\nu^n(x)$. – IV. Let $n > n_1$. By (3) and (5i), we have $\sum(\lambda(P(x)) \colon x \in \{0,1\}^n \setminus B_n(\varepsilon)) \leq \varepsilon nc$. Since $\sum(u(x)\nu^n(x) \colon x \in \{0,1\}) = na$, we get $\sum(\lambda(P(x)) \colon x \in B_{.}(\varepsilon)) \leq n(\Lambda(P) + \varepsilon + \Lambda(P_1) + \varepsilon)$. Hence $\sum(\lambda(P(x)) \colon x \in \{0,1\}^n) \leq (\Lambda(P_0) + \Lambda(P_1) + 2\varepsilon + c\varepsilon)$ and therefore, by (5iii) and (1), $\lambda(P^n)/n \leq H(a,b) + \Lambda(P_0) + \Lambda(P_1) + c\varepsilon + 2\varepsilon$.

4.19. Proposition. *If (P_0, P_1) is a partition of an FW-space P, then $\Lambda(P) \leq d(P)H(wP_0, wP_1) + \Lambda(P_0) + \Lambda(P_1)$.*

This is an immediate consequence of 4.18 and the fact that $\mathfrak{W}_F$ satisfies the conditions (1)–(4) stated in 4.13.

4.20. Characterization theorem for Λ on finite spaces. *The functional Λ defined on the class $\mathfrak{W}_F$ of al FW-spaces is* (A) *the largest of all functionals φ on $\mathfrak{W}_F$ satisfying* (1) *$\varphi P = 0$ if $\|P\| = 1$,* (2) *$\varphi(P^n) = n(wP)^{n-1} \cdot \varphi P$ for all $P \in \mathfrak{W}_F$ and $n \in N$, $n > 0$,* (3) *$\varphi P \leq d(P) \cdot wP + \varphi P_0 + \varphi P_1$ for all $P \in \mathfrak{W}_F$ and all partitions (P_0, P_1) of P,* (B) *the largest of all functionals φ on $\mathfrak{W}_F$ satisfying* (1), (2) *and* (3') *$\varphi P \leq d(P)H(wP_0, wP_1) + \varphi P_0 + \varphi P_1$ for all $P \in \mathfrak{W}_F$ and all partitions (P_0, P_1) of P.*

PROOF : I. Clearly, Λ satisfies (1). It satisfies (2) by 4.12, and (3'), hence also (3), by 4.19. – II. Let φ satisfy (1), (2) and (3). By 3.6, $\varphi S \leq \lambda S$ for all $S \in \mathfrak{W}_F$. Hence, if $P \in \mathfrak{W}_F$, $wP = 1$, $n \in N$, $n > 0$, then, by (2), $n\varphi(P) = \varphi(P^n) \leq \lambda(P^n)$, $\varphi P \leq \lambda(P^n)/n$ and therefore $\varphi P \leq \Lambda(P)$. ∎

4.21. Theorem. *If P is a finite separated semimetrized measure space, $wP > 0$, then $\delta P \cdot wP \geq \lambda P \geq E(P) \geq \Lambda(P) = \lim(E(P^n)/n(wP)^{n-1})$; in particular, if $wP = 1$, then $\delta P \geq \lambda P \geq E(P) \geq \Lambda(P) = \lim(E(P^n)/n)$.*

PROOF : The first two inequalities follow from 2.5, and 2.31. The inequality $E(P) \geq \Lambda(P)$ follows from 4.20 and 3.6. If $wP = 1$, $n \in N$, $n > 0$, then $E(P^n) \geq$

$\Lambda(P^n) = n\Lambda(P)$, hence $E(P^n)/n \geq \Lambda(P)$. On the other hand, $\underline{E}(P^n) \leq \lambda(P)$, hence $E(P^n)/n \leq \lambda(P^n)/n$ for all $n \in N$, $n > 0$, and therefore $\overline{\lim}(E(P^n)/n) \leq \Lambda(P)$. This proves the theorem. ∎

4.22. Theorem. *If P is a finite separated probability space equipped with an ultrametric, then $\lambda P \geq E(P) = \Lambda(P)$.*

PROOF : By 4.3, we have $E(P^n) = nE(P)$ for all $n \in N$, $n > 0$, hence $\lim(E(P^n)/n) = E(P)$ and therefore, by 4.21, $\Lambda(P) = E(P)$. ∎

4.23. Remarks. 1) Clearly, 4.21 and 4.22 correspond to a rather special version of the well-known theorems (for finite probability spaces) on coding in the absence of noise. In fact, 4.22 extends to finite probability spaces equipped with an ultrametric the basic theorem asserting that if $\langle Q, \mu \rangle$ is a finite probability space, the sequences $(x_i : i < n) \in Q^n$ can be coded, provided n is large, in $\{0, 1\}^*$ in such a way that the average length of codewords is less than $nH(\mu q : q \in Q) + \varepsilon)$, ε being any given positive number. - 2) If P is not ultrametric, then $E(P) = \Lambda(P)$ does not hold, in general.

REFERENCES

[1] J. Balatoni, A. Rényi, *On the notion of entropy* (Hungarian), Publ. Math. Inst. Hungarian Acad. Sci. 1 (1956), 9–40. English translation: Selected papers of A. Rényi, vol. I, pp. 558–584, Akadémiai Kiadó, Budapest, 1976

[2] M. Katětov, *Extended Shannon entropies I*, Czechoslovak Math. J. 33(108) (1983), 564–601.

[3] ———, *Extended Shannon entropies II*, Czechoslovak Math. J. 35(110) (1985), 565–616.

[4] ———, *On the Rényi dimension*, Comment. Math. Univ. Carolinae 27 (1986), 741–753.

[5] ———, *On dimensions of semimetrized measure spaces*, Comment. Math. Univ. Carolinae 28 (1987), 399–411.

[6] ———, *On the differential and residual entropy*, Comment. Math. Univ. Carolinae 29 (1988), 319–349.

[7] A. Kolmogorov, *On some asymptotic characteristics of totally bounded metric spaces* (Russian), Doklady Akad. Nauk SSSR 108 (1956), 385–389.

[8] A. Kolmogorov, V. Tihomirov, *ε-entropy and ε-capacity of sets in function spaces* (Russian), Uspehi Mat. Nauk 14, no.2 (1959), 3–86.

[9] A. Rényi, *On the dimension and entropy of probability distributions*, Acta Math. Acad. Sci. Hung. 10 (1959), 193–215.

[10] ———, *Dimension, entropy and information*, Trans. 2nd Prague Conf. Information Theory, pp. 545–556, Prague, 1960.

Mathematical Institute of Charles University, Sokolovská 83, 186 00 Praha 8, Czechoslovakia

(*Received* December 18, 1989)