The Role of the Asymptotic Equipartition Property in Noiseless Source Coding

Sergio Verdú, *Fellow, IEEE*, and Te Sun Han, *Fellow, IEEE*

Abstract— The (noiseless) fixed-length source coding theorem states that, except for outcomes in a set of vanishing probability, a source can be encoded at its entropy but not more efficiently. It is well known that the Asymptotic Equipartition Property (AEP) is a sufficient condition for a source to be encodable at its entropy. This paper shows that the AEP is *necessary* for the source coding theorem to hold for nonzero-entropy finite-alphabet sources. Furthermore, we show that a nonzero-entropy finite-alphabet source satisfies the direct coding theorem if and only if it satisfies the strong converse. In addition, we introduce the more general setting of nonserial information sources which need not put out strings of symbols. In this context, which encompasses the conventional serial setting, the AEP is equivalent to the validity of the strong coding theorem. Fundamental limits for data compression of nonserial information sources are shown based on the flat-top property—a new sufficient condition for the AEP.

Index Terms— Asymptotic equipartition property, entropy, fixed-length source coding, noiseless data compression, Shannon theory, source coding theorem.

I. INTRODUCTION

THE minimum *average* length of a uniquely decodable *variable-length* binary code for an arbitrary random object is equal to its entropy plus at most one bit. In contrast to the generality of this result on the minimum expected length of variable-length coding, the stronger result on *fixed-length* source coding requires to 1) place some conditions on the source and 2) focus on asymptotics, in order to conclude that all source strings outside an event of vanishing probability are optimally encoded at the source entropy. This is satisfied by stationary/ergodic sources but not necessarily by other sources; some sources require more, others require less than the entropy. Typically, the *sufficient condition* imposed on the source so that it can be optimally encoded at its entropy is the *Asymptotic Equipartition Property (AEP):* the probability of the set of n-strings whose log-probabilities are roughly $-H(Z_1, \cdots Z_n)$ goes to 1 as $n \to \infty$.

The main purpose of this paper is to shed further light onto the role of the AEP in noiseless source coding. We show that the AEP is not only a sufficient condition for the validity of the source coding theorem, but it is in fact a *necessary* condition, in the setting of nonzero-entropy finite-alphabet

sources. Within that setting, the AEP turns out to be equivalent to the equality of the minimum achievable source-coding rates in fixed-length coding and in variable-length coding. For more general information sources, the AEP is a necessary condition for the validity of the *strong* source coding theorem (in which the probability of error of any code with rate below the entropy goes to 1). We show that any source that satisfies the strong converse must also satisfy the direct part. For nonzero-entropy finite-alphabet sources, we show that the the (weak) converse is always satisfied and that the strong converse is satisfied if and only if the direct part is satisfied.

Showing that the AEP is equivalent to the validity of the source coding theorem reinforces the prominent role played by the AEP in information theory, which is due to the insight it offers into the behavior of information sources as well as the fact that it is generally much easier to verify whether the AEP holds for a particular source than to check whether the source can be encoded at its entropy but not more efficiently. It is somewhat surprising that the full role of the AEP in noiseless data compression had not been discovered before. A key step in our results is to show that the classical statement of the AEP is redundant, in the sense that the property is equivalent to the probability of the set of atypically big masses vanishing asymptotically.

For the most part, the development will proceed without placing any assumptions on the allowed class of information sources. We even allow a generalization of the conventional setting, where the source does not necessarily output a string of symbols. We refer to those sources as *nonserial information sources.* Consider the following examples of such sources:

1) an image with $n \times k(n)$ pixels;
2) the number of Poisson points with growing mean λ_n;
3) the final value of a random walk $X_1 + \cdots + X_n$;
4) the empirical distribution of an n-string drawn from a finite alphabet A.

In each of these examples, the entropy of the information source grows without bound with n although not linearly as in the case of a conventional serial source. Note that in cases 2) and 3) a single integer-valued random variable is to be encoded; whereas in case 4) the information to be encoded is a vector with fixed dimension $|A|$. Such sources fall outside the scope of the Shannon–MacMillan Theorem and source coding theorems had not been been found previously despite their conceptual, as well as practical, interest. We establish those new source coding theorems thanks to a new simple-to-check condition (the *flat-top* property) which is shown to be sufficient for the AEP to hold.

A nonserial information source is characterized by a sequence of distributions $Z = \{Z^n\}$, where n need not have the connotation of blocklength. Since our emphasis is on noiseless coding, we will require that Ω_n, the set on which Z^n takes values, be finitely or countably infinitely valued. The conventional (serial) setting of a source with alphabet A corresponds to the special case where Ω_n is equal to the nth Cartesian product A^n.

In Section II, we define the main classes of sources we will be concerned with. Those definitions classify information sources according to whether they can be encoded above, below, or at the entropy. Several examples illustrate the various possible behaviors.

Section III is devoted to the AEP and, in particular, to showing that, in a completely general setting, the AEP is *equivalent* to the property that the source is encodable at its entropy but any code operating below the entropy results in probability of error converging to 1. Exact asymptotic results are obtained under the only assumption that the entropy of the object to be encoded is finite and grows without bound. This encompasses those single-sample information sources cited above. Stationary ergodic sources satisfy the AEP according to the Shannon–MacMillan Theorem [2]. Section III shows that the flat-top property (which proves to be particularly useful for nonserial information sources) is another sufficient condition for the AEP.

Section IV focuses on the special, but most important, case of conventional finite-alphabet sources. The additional structure of those sources enables the proof of further results which simplify the general picture. In particular, as long as the entropy of n-strings grows linearly with n, we establish that the AEP is equivalent to the validity of the source coding theorem.

II. ENTROPIC SOURCES

For the purposes of examining whether a source can be encoded below, above, or at its entropy asymptotically, only those sources such that $H(Z^n)$ is finite for every n and grows without bound with n are of interest. This is an underlying assumption throughout the paper. However, the growth of $H(Z^n)$ need not be linear (particularly for nonserial sources) and we shall make no assumptions in that respect. Recall that n is a generic "size" index, which in the special case of conventional serial sources is equal to the string length.

Fixed-length source codes which assign a unique codeword to each of the M most likely source outcomes (and an "error" codeword to all other outcomes) achieve the minimum size. Thus a natural class of information sources are those that can be encoded with a codebook whose log-size grows as the entropy.

Definition 1: A source Z is *subentropic* if the total mass of its most likely $M = \exp((1 + \delta)H(Z^n))$ outcomes[1] (or, *a fortiori*, any other event with no more than M outcomes) goes to 1 as $n \to \infty$ for any $\delta > 0$.

---

[1] The set of the M most likely outcomes is always well-defined (not necessarily uniquely) even if Ω_n is countably infinite.

Conversely, those sources which cannot be encoded at any rate smaller than the entropy can be defined as

Definition 2: A source Z is *superentropic* if the total mass of its most likely $M = \exp((1 - \delta)H(Z^n))$ outcomes does not go to 1 as $n \to \infty$ for any $\delta > 0$.

Notice that a source is either subentropic or superentropic (or both). In the conventional terminology, subentropic sources can be viewed as those that satisfy the direct part of the coding theorem, whereas superentropic sources are those that satisfy the converse part. Thus those sources that satisfy both Definitions 1 and 2 are of major importance:

Definition 3: A source is *entropic* if it is both subentropic and superentropic.

Throughout this paper, the exponential and logarithm functions have a common arbitrary basis. If we denote this basis by a, then an entropic source is one for which the most likely $\alpha^{H(Z^n)}$ outcomes exhaust (respectively, do not exhaust) the probability asymptotically if $\alpha > a$ (respectively, $\alpha < a$). A refinement of the notion of superentropic sources is suggested by the *strong* converse source coding theorem which states that coding below the entropy results in probability of error tending to 1. This is easily captured by the following definition of a subclass of superentropic sources.

Definition 4: A source Z is *strongly superentropic* if for any $\delta > 0$ the total mass of its most likely $M = \exp((1-\delta)H(Z^n))$ outcomes goes to 0 as $n \to \infty$.

Definition 5: A source is *strongly entropic* if it is both subentropic and strongly superentropic.

Even for simple random processes, it is not easy to characterize the probability of the set of the M most likely outcomes where M grows exponentially with the entropy. The development of tools in order to classify information sources according to the foregoing definitions is a main goal of this paper. Prior to developing those tools in Theorems 1–13, we will exhibit several examples in order to gain insight into the various classes of sources.

Example 1: A stationary ergodic source $Z^n = (X_1, \cdots, X_n)$ with finite alphabet is strongly entropic. (This follows from the Shannon–MacMillan Theorem and Theorem 2 in Section III.)

Example 2: The following binary serial source is superentropic but it is neither subentropic nor strongly superentropic:

$$P[Z^n = z^n] = \begin{cases} 1 - q, & z^n = (0, \cdots 0) \\ q/(2^n - 1), & z^n \neq (0, \cdots 0) \end{cases}$$

for any $0 < q < 1$. To check this, note that

$$H(Z^n) = h(q) + qn + q\log_2(1 - 2^{-n}) \qquad (1)$$

where h denotes the binary entropy function. No event whose size grows as α^n exhausts all the probability asymptotically if $\alpha < 2$. Thus the source is not subentropic, and therefore, it is superentropic. The source is not strongly superentropic because the event of M most likely outcomes has probability at least $1 - q$ for any $M \geq 1$. If instead of being constant, $q_n \to 1$, then the source becomes strongly entropic.

Example 3: Let $\Omega_n = \{1, \cdots, K_n\}$ with

$$K_n = 2^n + 4^{n^2}$$

and define the two-level distribution

$$P_{Z^n}(k) = \begin{cases} (1-1/n)2^{-n}, & k = 1, \cdots 2^n \\ \dfrac{1}{n}\, 4^{-n^2}, & k = 2^n + 1, \cdots K_n. \end{cases}$$

The total mass of the outcomes $k = 1, \cdots, 2^n$ is equal to $1-1/n$. Therefore, n bits represent the source with vanishingly small probability of error as $n \to \infty$. On the other hand, the entropy, in bits, of Z^n is

$$H(Z^n) = \left(1 - \frac{1}{n}\right)\log_2 \frac{2^n}{1 - \dfrac{1}{n}} + \frac{1}{n}\log_2(n4^{n^2})$$

$$= 3n + o(n). \tag{2}$$

Therefore, this source can be encoded at a rate which is one-third of its entropy rate. Thus it is subentropic but not superentropic.

Example 4: The Poisson source with mean λ_n, with $\lambda_n \to \infty$ as $n \to \infty$

$$P_{Z^n}(k) = e^{-\lambda_n}\frac{\lambda_n^k}{k!}, \qquad k = 0, 1, 2, \cdots. \tag{3}$$

As we emphasized before, only one integer value is to be encoded. The entropy of Z^n grows as $\log\sqrt{\lambda_n}$. We show in Theorem 6 that this source is strongly entropic.

Example 5: The geometric distribution with parameter $0 < a_n < 1$

$$P_{Z^n}(k) = (1 - a_n)a_n^k, \qquad k = 0, 1, 2\cdots. \tag{4}$$

Its entropy is

$$H(Z^n) = \log\frac{1}{1 - a_n} + \frac{a_n}{1 - a_n}\log\frac{1}{a_n}. \tag{5}$$

If $a_n \to 1$ as $n \to \infty$, then $H(Z_n) \to \infty$. It is shown in Theorem 7 that this source is strongly entropic. In fact, it is easy to show directly that this source is entropic upon noticing that the residual probability (not covered by the M most likely outcomes) is a_n^M.

We show next two common examples of nonserial information sources derived from a conventional independent and identically distributed (i.i.d.) source.

Example 6: Let $\{X_1, X_2, \cdots\}$ be i.i.d. taking values on a finite set A. Under some conditions on the distribution of X_i, it is shown in Section III that $Z^n = X_1 + \cdots + X_n$ is strongly entropic.

Example 7: Let $\{X_1, X_2, \cdots\}$ be i.i.d. taking values on a finite set A. The empirical distribution of $(X_1, \cdots, X_n)$ is defined as the $|A|$-dimensional random vector

$$Q_{X_1 \cdots X_n}(a) = \frac{1}{n}\sum_{i=1}^{n} 1\{X_i = a\}, \qquad a \in A. \tag{6}$$

As we will show, this nonserial source is strongly entropic and can be encoded with $(|A| - 1)\log\sqrt{n}$ bits.

Example 8 (Randomly Chosen Biased Coin): This is a serial stationary nonergodic binary source. Let $\Omega_n = \{H, T\}^n$, and let $Z^n = (Z_1, \cdots, Z_n)$ be such that with probability $\frac{1}{2}$, it is Bernoulli with $P[Z_i = H] = \beta_1 \in (0, \frac{1}{2})$ and with probability $\frac{1}{2}$, it is Bernoulli with $P[Z_i = H] = \beta_2 \in (\beta_1, \frac{1}{2})$. This source is superentropic but it is neither subentropic nor strongly superentropic. The entropy satisfies

$$H(Z^n) = \frac{n}{2}h(\beta_1) + \frac{n}{2}h(\beta_2) + O(1) \tag{7}$$

but the most likely $M = \exp((1 + \delta)H(Z^n))$ outcomes fail to exhaust the probability for sufficiently small $\delta > 0$. This means that the source is not subentropic, so it must be superentropic. However, the probability of the most likely $M = \exp((1 - \delta)H(Z^n))$ outcomes goes to $\frac{1}{2}$ for sufficiently small $\delta > 0$. Thus the source is not strongly superentropic.

In Example 9 below we show a source which is entropic but not strongly entropic. The idea is to have a large set of very unlikely outcomes whose overall probability vanishes but which contributes significantly to the entropy, while at the same time have a single large nonvanishing mass which prevents the source from being strongly entropic. The fact that the source in Example 9 has a number of outcomes which grows superexponentially with n is not accidental, as we will see in Section IV.

Example 9: The source has three types of outcomes:
1) one mass with probability $\frac{1}{2}$;
2) 2^n masses each with probability $(\frac{1}{2} - 1/n)2^{-n}$;
3) $(1/n)2^{n^2/2}$ masses each with probability $2^{-n^2/2}$.

The entropy of this source satisfies

$$\lim_{n \to \infty} \frac{H(Z^n)}{n} = 1 \text{ bit.}$$

To check that the source is indeed subentropic note that for all $\delta > 0$, the most likely $2^{n(1+\delta)}$ masses have probability greater than $1 - 1/n$. On the other hand, the most likely $2^{n(1-\delta)} + 1$ masses have probability

$$\frac{1}{2} + \left(\frac{1}{2} - \frac{1}{n}\right)2^{-n\delta}$$

which goes to $\frac{1}{2}$ as $n \to \infty$. Thus this source is superentropic but not strongly superentropic.

For many sources (in particular, serial finite-alphabet sources), an easy way to check the validity of the converse of the source coding theorem is the following.

Theorem 1: Let Z^n take values on the set Ω_n. If

$$\limsup_{n \to \infty} \frac{H(Z^n)}{\log|\Omega_n|} > 0 \tag{8}$$

then the source $\mathbf{Z}$ is superentropic.

Proof: Let us define an encoder for the source which is the identity mapping for the M most likely outcomes and assigns a unique element to all the other outcomes. The output of the encoder to Z^n will be denoted by $\hat{Z}^n$. Note that the M most likely outcomes asymptotically exhaust all the probability if and only if

$$P[Z^n \neq \hat{Z}^n] \to 0.$$

In order to lower-bound the error probability we can use Fano's inequality [2]

$$P[Z^n \neq \hat{Z}^n] \geq \frac{1}{\log |\Omega_n|}[H(Z^n) - I(Z^n; \hat{Z}^n) - \log 2]$$
$$\geq \frac{H(Z^n)}{\log |\Omega_n|} \left[1 - \frac{\log(M+1)}{H(Z^n)} - \frac{\log 2}{H(Z^n)}\right]. \quad (9)$$

Now choose $\delta > 0$ and let $M = \exp((1-\delta)H(Z^n))$, then the right side of (9) is lower-bounded for all sufficiently large n by

$$\frac{H(Z^n)}{\log |\Omega_n|} \frac{\delta}{2}.$$

Thus if the condition of the theorem is satisfied, then the probability of error cannot vanish asymptotically, and the source is superentropic. $\square$

The sufficient condition in Theorem 1 is not satisfied by many superentropic nonserial information sources of interest, such as those in Examples 4, 5, 9, or any source with $|\Omega_n| = \infty$. Fano's inequality is not sufficiently powerful to deal with those sources. A simple alternative condition to deal with those sources will be presented at the end of Section III.

III. ASYMPTOTIC EQIPARTITION PROPERTY (AEP)

In the domain of conventional sources, it is common to adopt the notion of the Asymptotic Equipartition Property (AEP), which is satisfied by stationary ergodic sources according to the Shannon–Macmillan Theorem [2]. The definition of the AEP can be immediately extended to any finite-entropy information source (serial or nonserial; with finite or infinite Ω_n).

Definition 6: A source **Z** *satisfies the AEP* if the outcomes whose log-probabilities differ from $-H(Z^n)$ by no more than $\delta H(Z^n)$, exhaust all the probability asymptotically, no matter how small $\delta > 0$. In other words, the AEP states that for all $\delta > 0$, as $n \rightarrow \infty$

$$P[B_n(\delta)] \rightarrow 0 \quad \text{and} \quad P[S_n(\delta)] \rightarrow 0$$

where the subsets of "atypically big" and "atypically small" probability masses are denoted by

$$B_n(\delta) = \{\omega \in \Omega_n: P_{Z^n}(\omega) \geq \exp(-(1-\delta)H(Z^n))\}$$

and

$$S_n(\delta) = \{\omega \in \Omega_n: P_{Z^n}(\omega) \leq \exp(-(1+\delta)H(Z^n))\}$$

respectively.

The dissection of the set of atypical probability masses into atypically big and atypically small will be seen to be crucial for the purposes of this paper.

In the special case of conventional serial sources it is common to state the AEP as

$$P\left[\left|\frac{H(Z^n)}{n} - \frac{1}{n}\log\frac{1}{P_{Z^n}(Z^n)}\right| > \delta\right] \rightarrow 0 \quad (10)$$

for all $\delta > 0$. This is equivalent to Definition 6 as long as the limsup and liminf of $H(Z^n)/n$ are finite and nonzero.

Traditionally, the AEP takes the role of a sufficient condition in fixed-length source coding. The following result (first noticed by Shannon [8, Appendix III] in the context of i.i.d. sources) summarizes the classical role of the AEP.

Theorem 2: A source that satisfies the AEP is strongly entropic.

Proof: (AEP $\Rightarrow$ Subentropic). The cardinality of the event $\Omega_n - S_n(\delta)$ (the set of masses that are not atypically small) is upper-bounded by $\exp((1 + \delta)H(Z^n))$, and its probability goes to 1 because of the AEP. $\square$

Proof: (AEP $\Rightarrow$ Strongly Superentropic).[2] Denote the set of the most likely $\exp((1-\delta)H(Z^n))$ masses in Ω_n by $G_n(\delta)$. Note that a source is strongly superentropic if and only if

$$\lim_{n \rightarrow \infty} P[G_n(\delta)] = 0 \quad (11)$$

for all $\delta > 0$.

Let us show that for all $\delta > 0$, $P[B_n(\delta)] \rightarrow 0$ implies $P[G_n(2\delta)] \rightarrow 0$. To that end we examine the probability of the right side of

$$G_n(2\delta) \subset B_n(\delta) \cup \{G_n(2\delta) - B_n(\delta)\}. \quad (12)$$

The definitions of $B_n(\delta)$ and $G_n(2\delta)$ lead to the following equations, respectively:

$$P[G_n(2\delta) - B_n(\delta)] \leq |G_n(2\delta)| \exp(-(1-\delta)H(Z^n))$$
$$\leq \exp(-\delta H(Z^n)) \quad (13)$$

where the right side vanishes because of the underlying assumption that $H(Z^n)$ grows without bound. Thus (12), (13), and $P[B_n(\delta)] \rightarrow 0$ lead to $P[G_n(2\delta)] \rightarrow 0$ as desired. $\square$

We will strengthen significantly the classical Theorem 2 by showing that not only the AEP is a sufficient condition for any source to be strongly entropic but it is also necessary. Furthermore, we will show the important observation that the strong converse implies the direct part of the source coding theorem. In the proof of Theorem 2 we saw that a sufficient condition for a source to be subentropic (respectively, strongly superentropic) is that the probability of the event of atypically small (respectively, big) masses vanish. To show that the AEP is a necessary condition it is instrumental to prove that its classical statement (Definition 6) is redundant: if the event of atypically big masses has vanishing probability, then, necessarily, the event of atypically small masses must also have vanishing probability (regardless of how we gauge typicality). Thus the absolute value in (10) is unnecessary. Even in the context of i.i.d. sources, it appears that this is a new observation.

Theorem 3: A source **Z** satisfies the AEP if and only if for all $\delta > 0$

$$\lim_{n \rightarrow \infty} P[B_n(\delta)]$$
$$= \lim_{n \rightarrow \infty} P\left[\log\frac{1}{P_{Z^n}(Z^n)} \leq (1-\delta)H(Z^n)\right] = 0. \quad (14)$$

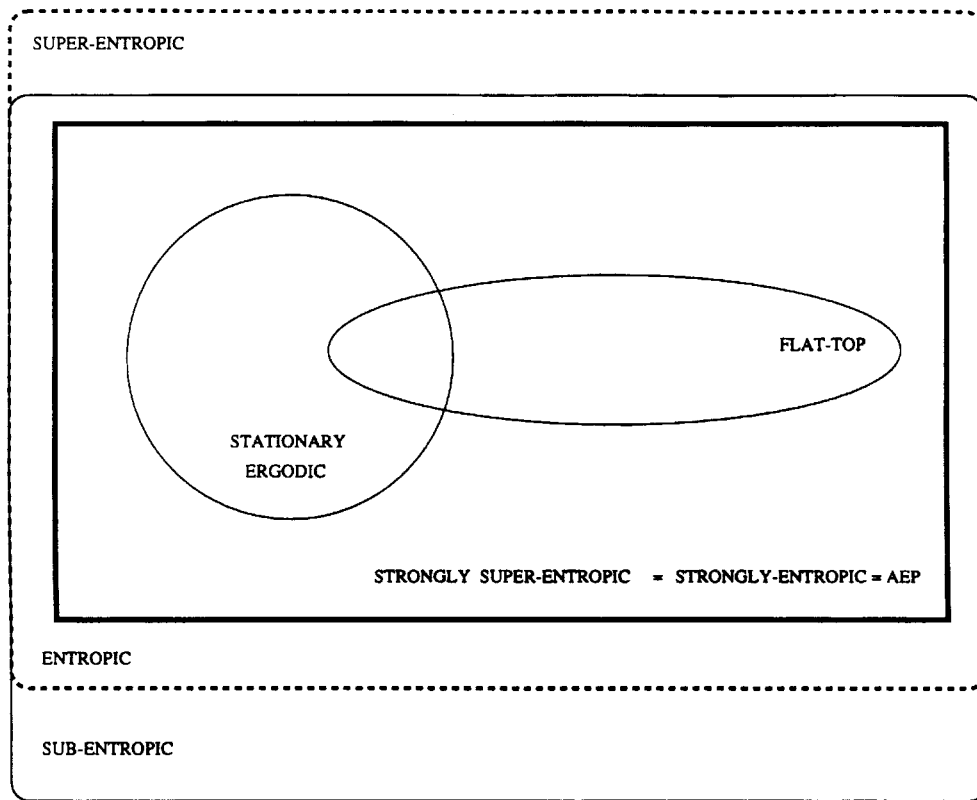[2] See [5, pp. 42–43] and [3, pp. 16–17] for similar proofs in the context of i.i.d. sources.

Fig. 1. Classes of information sources.

Proof: We need to show that if $P[B_n(\delta)] \to 0$ for all $\delta > 0$, then $P[S_n(\delta)] \to 0$ for all $\delta > 0$. The idea is to notice that the contribution of the "big" masses to the entropy is negligible if $P[B_n(\delta)] \to 0$. Then, the total probability of the "small" masses must go to zero as well, for otherwise the average of $\log 1/P_{Z^n}(Z^n)$ would exceed $H(Z^n)$.

Choose arbitrary $\delta_1 > 0$ and $\delta_2 > 0$. Recalling the definitions of $B_n(\delta_1)$ and $S_n(\delta_2)$ and noticing that they are nonoverlapping sets, we can write

$$
\begin{aligned}
H(Z^n) &\geq \sum_{\omega \in B_n^c(\delta_1) \cap S_n^c(\delta_2)} P_{Z^n}(\omega) \log \frac{1}{P_{Z^n}(\omega)} \\
&\quad + \sum_{\omega \in S_n(\delta_2)} P_{Z^n}(\omega) \log \frac{1}{P_{Z^n}(\omega)} \\
&\geq H(Z^n)(1 - \delta_1) P[B_n^c(\delta_1) \cap S_n^c(\delta_2)] \\
&\quad + H(Z^n)(1 + \delta_2) P[S_n(\delta_2)] \\
&= H(Z^n)(1 - \delta_1)[1 - P[B_n(\delta_1)] - P[S_n(\delta_2)]] \\
&\quad + H(Z^n)(1 + \delta_2) P[S_n(\delta_2)].
\end{aligned}
\tag{15}
$$

Eventually, $H(Z^n) > 0$. Thus we may divide (15) by $H(Z^n)$ to yield, after rearranging terms,

$$
P[S_n(\delta_2)] \leq \frac{\delta_1 + (1 - \delta_1) P[B_n(\delta_1)]}{\delta_1 + \delta_2}.
\tag{16}
$$

But no matter how small $\delta_2 > 0$, the upper bound in (16) is as as small as desired for sufficiently large n because the choice of δ_1 was arbitrary and $P[B_n(\delta_1)] \to 0$. Thus the proof is complete. □

Note that the only property of the random variables $\log 1/P_{Z^n}(Z^n)$ used in the proof of Theorem 3 is that they are nonnegative and their mean is eventually positive.

We are now in a position to state and prove the following equivalence:

Theorem 4: The following conditions are equivalent for an information source:

a) AEP,
b) strongly entropic,
c) strongly superentropic.

Proof: Theorem 2 gives a) ⇒ b) and Definition 5 gives b) ⇒ c). Thus we need to show that c) ⇒ a). Thanks to Theorem 3, it will be enough to show that any strongly superentropic source satisfies (14) for all $\delta > 0$. But this follows immediately from (cf. proof of Theorem 2 for notation)

$$
B_n(\delta) \subset G_n(\delta)
$$

which in turn is a consequence of the fact that

$$
|B_n(\delta)| \leq \exp((1 - \delta) H(Z^n)).
\qquad \square
$$

Fig. 1 shows the classification of sources resulting from the foregoing results. It also shows a new sufficient condition for the AEP which is particularly useful for nonserial information sources:

Definition 7: Denote the infinite-order Rényi entropy by

$$
R_\infty(Z^n) = \log \frac{1}{\max_{\omega \in \Omega^n} P_{Z^n}(\omega)}.
$$

Note that $R_\infty(Z^n) \leq H(Z^n)$ for any Z^n. A source $\mathbf{Z}$ is a *flat-top* source if

$$\lim_{n \to \infty} \frac{R_\infty(Z^n)}{H(Z^n)} = 1. \tag{17}$$

The term *flat top* recalls the asymptotic shape of the probability mass function of Z^n in a logarithmic scale when (17) is satisfied.

Theorem 5: A flat-top source satisfies the AEP.

Proof: According to Theorem 3 we just need to check that for any $\delta > 0$, (14) is satisfied. In fact, even more can be shown: under the flat-top property eventually there are no atypically big masses. According to (17), for all sufficiently large n

$$H(Z^n) \leq \left(1 + \frac{\delta}{2}\right) R_\infty(Z^n).$$

Therefore, for those n, we have

$$P\left[\log \frac{1}{P_{Z^n}(Z^n)} \leq (1 - \delta)H(Z^n)\right]$$
$$\leq P\left[\log \frac{1}{P_{Z^n}(Z^n)} \leq \left(1 + \frac{\delta}{2}\right)(1 - \delta)R_\infty(Z^n)\right]$$
$$\leq P\left[\log \frac{1}{P_{Z^n}(Z^n)} \leq \left(1 - \frac{\delta}{2}\right)R_\infty(Z^n)\right]$$
$$= 0$$

where the equality follows from the definition of $R_\infty(Z^n)$. $\square$

A source of fair coin flips is flat-top. However, a source that satisfies the AEP need not be flat-top, when the most likely outcomes are not "typical."

Example 10 (Biased-Coin Flips): A serial Bernoulli source with parameter $P[Z_i = 1] = p > 1/2$. It satisfies the AEP, but it is not flat-top.

$$H(Z^n) = nh(p)$$

and

$$R_\infty(Z^n) = n \log(1/p).$$

Note that the sequence $(1, \cdots, 1)$ is not typical, even though it is much more likely to occur than any one of the typical sequences.

Accustomed to thinking of the AEP as a property satisfied by stationary ergodic sequences, it may be surprising that it is satisfied by many nonserial sources, even single-sample sources, which at first sight look decidedly not almost-equiprobable. In order to illustrate the utility of Theorem 5, we will show that the nonserial sources in Examples 4–7 are flat-top, and thus satisfy the AEP and are strongly entropic.

Theorem 6: The Poisson source (Example 4) with mean $\lambda_n \to \infty$ satisfies

$$\lim_{n \to \infty} \frac{R_\infty(Z^n)}{\log \sqrt{\lambda_n}} = \lim_{n \to \infty} \frac{H(Z^n)}{\log \sqrt{\lambda_n}} = 1.$$

Proof: The maximum probability mass is located at $k = \lfloor \lambda_n \rfloor$, and

$$R_\infty(Z^n) = \lambda_n \log e - \lfloor \lambda_n \rfloor \log \lambda_n + \log(\lfloor \lambda_n \rfloor!).$$

Using Stirling's approximation it can be checked that $R_\infty(Z^n)$ grows as $\log \sqrt{\lambda_n}$. Furthermore, it can be checked that the differential entropy bound on discrete entropy [2] yields

$$\limsup_{n \to \infty} \frac{H(Z^n)}{\frac{1}{2} \log \lambda_n} \leq 1$$

and the proof is complete in view of $R_\infty(Z^n) \leq H(Z^n)$. $\square$

Theorem 7: The geometric distribution (Example 5) with mean $a_n/(1 - a_n) \to \infty$ is flat-top and

$$R_\infty(Z^n) = \log \frac{1}{1 - a_n}. \tag{18}$$

Proof: The expression for $R_\infty(Z^n)$ follows immediately from (4). Furthermore, using (5) it can be checked that

$$\lim_{n \to \infty} \frac{H(Z^n)}{\log \frac{1}{1 - a_n}} = 1. \tag{19}$$

Therefore, the geometric source is flat-top as long as $a_n \to 1$ when $n \to \infty$. $\square$

Theorem 8: The empirical distribution of a finite alphabet i.i.d. source with distribution P_X (Example 7) is flat-top.

$$\lim_{n \to \infty} \frac{R_\infty(Q_{X_1 \cdots X_n})}{(|A| - 1) \log \sqrt{n}} = \lim_{n \to \infty} \frac{H(Q_{X_1 \cdots X_n})}{(|A| - 1) \log \sqrt{n}} = 1 \quad (20)$$

where $|A|$ is the number of nonzero probability masses in P_X.

Proof: For any n-type distribution Q, the probability that the empirical distribution is equal to Q is given by [3, pp. 32, 39]

$$-\log P[Q_{X_1 \cdots X_n} = Q] = nD(Q\|P_X) + \frac{w(Q) - 1}{2} \log n$$
$$- \frac{1}{2} \sum_{a:Q(a)>0} \log \frac{1}{Q(a)} + \eta(n, Q) \tag{21}$$

where $w(Q)$ is the number of nonzero masses in Q, and $\eta(n, Q)$ lies in a closed interval independent of n and Q. The Shannon entropy and the infinite-order Rényi entropy of $Q_{X_1 \cdots X_n}$ are equal to the average and minimum, respectively, of (21) over the set of n-types. In fact, we will only lower-bound $R_\infty(Q_{X_1 \cdots X_n})$ and upper-bound $H(Q_{X_1 \cdots X_n})$ since we know that

$$R_\infty(Q_{X_1 \cdots X_n}) \leq H(Q_{X_1 \cdots X_n}).$$

Fix a sufficiently small $\epsilon > 0$. We will analyze the minimum value of (21) achievable by Q according to the following two cases:

1) $D(Q\|P_X) > \epsilon$.

Since for an n-type either $Q(a) = 0$ or $Q(a) \geq 1/n$, the right side of (21) can be lower-bounded, in this case, by

$$n\epsilon - \frac{1}{2} \log n + O(1).$$

2) $D(Q\|P_X) \leq \epsilon$.

If we indeed chose ϵ to be sufficiently small, then this case implies that for all $a \in A$

$$Q(a) \geq \frac{P_X(a)}{2}$$

because the Csiszár–Kullback–Kemperman inequality [3, p. 58] dictates that

$$\frac{\log e}{2}\left(\sum_{a\in A}|P_X(a) - Q(a)|\right)^2 \leq D(Q\|P_X).$$

So the right side of (21) can be lower-bounded by

$$\frac{|A| - 1}{2}\log n - \frac{1}{2}\sum_{a\in A}\log\frac{2}{P_X(a)} + O(1).$$

Joining the conclusions obtained in both cases we can state that

$$R_\infty(Q_{X_1\cdots X_n}) \geq \frac{|A| - 1}{2}\log n + O(1). \tag{22}$$

Let us now upper-bound $H(Q_{X_1\cdots X_n})$. By definition

$$H(Q_{X_1\cdots X_n}) = -\sum_Q P[Q_{X_1\cdots X_n} = Q]\log P[Q_{X_1\cdots X_n} = Q] \tag{23}$$

where the sum ranges over all n-type distributions Q. It should be noted that this entropy reflects the randomness in the outcome of the empirical distribution; in particular, it is not an average of the entropies of the possible empirical distributions. Expressions (21) and (23) lead to the upper bound

$$H(Q_{X_1\cdots X_n}) \leq \frac{|A| - 1}{2}\log n + nE[D(Q_{X_1\cdots X_n}\|P_X)] + O(1) \tag{24}$$

where the expectation is with respect to $(X_1, \cdots, X_n)$. This average divergence between the empirical and the true distributions is a quantity of interest in itself: it is equal to the input/output mutual information of an empirical-distribution computer driven by $(X_1, \cdots, X_n)$. Using Lemma 1 below we see that (24) and (25) result in

$$H(Q_{X_1\cdots X_n}) \leq \frac{|A| - 1}{2}\log n + O(1). \tag{25}$$

Together with (22), this concludes the proof of Theorem 8. $\square$

Lemma 1:[3] For every n

$$E[D(Q_{X_1\cdots X_n}\|P_X)] \leq \frac{|A| - 1}{n}\log e. \tag{26}$$

---

[3]Lemma 1 is equivalent to [7, Lemma 1]. We give the proof for completeness.

Proof:

$$E[D(Q_{X_1\cdots X_n}\|P_X)]$$
$$= \sum_{a\in A}\sum_{a_1\in A}\cdots\sum_{a_n\in A}P_X(a_1)\cdots P_X(a_n)Q_{a_1\cdots a_n}(a)$$
$$\cdot \log\frac{Q_{a_1\cdots a_n}(a)}{P_X(a)}$$
$$\leq \sum_{a\in A}\sum_{a_1\in A}\cdots$$
$$\sum_{a_n\in A}P_X(a_n)\cdot\left[\frac{Q^2_{a_1\cdots a_n}(a)}{P_X(a)} - Q_{a_1\cdots a_n}(a)\right]\log e \tag{27}$$

where we upper-bounded $\log z$ by $(z - 1)\log e$. Straightforward computations yield

$$\sum_{a_1\in A}\cdots\sum_{a_n\in A}P_X(a_1)\cdots P_X(a_n)Q_{a_1\cdots a_n}(a) = P_X(a) \tag{28}$$

and

$$\sum_{a_1\in A}\sum_{a_n\in A}P_X(a_1)\cdots P_X(a_n)Q^2_{a_1\cdots a_n}(a)$$
$$= \frac{1}{n}P_X(a) + \left(1 - \frac{1}{n}\right)P^2_X(a). \tag{29}$$

The desired result follows by inserting (28) and (29) in the right side of (27) $\square$

We now show that the random walk with independent increments distributed according to P_X (Example 6) is flat-top, and thus strongly entropic. Encoding the whole evolution of the random walk $X_1 + \cdots + X_i, i = 1, 2, \cdots, n$, is equivalent to encoding its increments, which takes (asymptotically) $nH(X)$ bits according to classical information theory. The result below shows that if only the final value of the random walk is to be encoded, $\beta\log n$ bits suffice where β depends on the structure of the alphabet.

Theorem 9: Consider the random walk $Z^n = X_1 + \cdots + X_n$ where $\{X_1, X_2, \cdots\}$ are independent with identical distribution P_X on a finite set $A = \{a_1, \cdots, a_K\}$.

a) Suppose that A is *incongruent*, in the sense that the equation

$$m_1 a_1 + \cdots + m_K a_K = 0$$

has no integer solutions satisfying $\Sigma_{i=1}^K m_i = 0$ other than $m_1 = \cdots = m_K = 0$. Then

$$\lim_{n\to\infty}\frac{R_\infty(X_1 + \cdots + X_n)}{(|A| - 1)\log\sqrt{n}}$$
$$= \lim_{n\to\infty}\frac{H(X_1 + \cdots + X_n)}{(|A| - 1)\log\sqrt{n}} = 1. \tag{30}$$

b) Suppose that A is a *lattice* ($a_i = b + m_i h$, where m_i is an integer). Then

$$\lim_{n\to\infty}\frac{R_\infty(X_1 + \cdots + X_n)}{\log\sqrt{n}}$$
$$= \lim_{n\to\infty}\frac{H(X_1 + \cdots + X_n)}{\log\sqrt{n}} = 1. \tag{31}$$

Proof:

a) In this case, two sequences with identical sum must have identical empirical distribution. Since the converse is always true, knowing the value of $X_1 + \cdots + X_n$ is equivalent to knowing the empirical distribution $Q_{X_1 \cdots X_n}$ and (30) is given by Theorem 8.

b) Both the Shannon entropy and the infinite-order Rényi entropy are invariant to scaling and shifting. Thus we are free to work with the normalized version of the random walk

$$\overline{Z}^n = \frac{X_1 + \cdots + X_n - nE[X]}{\sigma_X \sqrt{n}}$$

which is known [4, p. 517] to satisfy the central limit theorem

$$\frac{\sqrt{n}}{h} P[\overline{Z}^n = z] \to \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

if z is a possible value of $\overline{Z}^n$. According to this uniform-convergence result, the largest mass of $\overline{Z}^n$ times $\sqrt{n}$ is asymptotically upper-bounded by $h/\sqrt{2\pi}$. Thus

$$1 \le \liminf_{n \to \infty} \frac{R_\infty(\overline{Z}^n)}{\log \frac{\sqrt{2\pi n}}{h}} = \liminf_{n \to \infty} \frac{R_\infty(X_1 + \cdots + X_n)}{\log \sqrt{n}}.$$

To bound the Shannon entropy, we will not rely on the central limit theorem but on the differential entropy bound on discrete entropy [2, p. 235]

$$H(\overline{Z}^n) \le \frac{1}{2} \log \left[2\pi e \left(\frac{n\sigma_X^2}{h^2} + \frac{1}{12} \right) \right] \qquad (32)$$

which is justified on the basis that, asymptotically, the masses of $\overline{Z}^n$ become equispaced with span $h/(\sigma_X \sqrt{n})$, while its variance is always 1. Thus if we were to stretch the masses to unit distance the variance would become $n\sigma_X^2/h^2$. The bound in (32) implies

$$H(\overline{Z}^n) \le \tfrac{1}{2} \log n + O(1)$$

and the proof of (31) is complete.   □

The cases considered in Theorem 9, namely, lattice and incongruent alphabets do not exhaust all possibilities (e.g., the union of two lattices); however, they provide two extreme cases where we clearly see the dependence of the entropy of the random walk on the structure of the alphabet.

Returning to Fig. 1, note that we have given an example of every set and intersection therein, except for the class of sources which satisfy the AEP (and thus are strongly entropic) but are neither stationary/ergodic nor flat-top. Using Chebychev's weak law of large numbers (for uncorrelated sequences with uniformly bounded second moments [1, Theorem 5.1.1]), the reader can verify that an example of such a source is

Example 11: A finite-alphabet nonstationary memoryless source.

IV. FINITE-ALPHABET SERIAL SOURCES

So far, our results have applied in full generality to serial and nonserial sources. We now investigate whether sharper results are possible within the domain of conventional finite-alphabet serial sources. That is, henceforth, Z^n takes values on the nth Cartesian product $\Omega_n = A^n$, with $|A| < \infty$. Indeed, the general picture in Fig. 1 wil be shown to simplify considerably for those sources. Let us first state the following immediate corollary to Theorem 1.

Theorem 10: A finite-alphabet source with

$$\limsup_{n \to \infty} \frac{H(Z^n)}{n} > 0 \qquad (33)$$

is superentropic.

Note that Theorem 10 gives a rather general converse fixed-length source coding theorem for conventional serial sources, by means of a simple proof based on Fano's inequality (cf. proof of Theorem 1).

The reason we will be able to show sharper results for conventional finite-alphabet sources is that events with small probability have a small contribution to the normalized entropy because their cardinality cannot be superexponential. We state this property in the following simple result.

Lemma 2:

a) For any $D \subset \Omega_n$

$$\sum_{z^n \in D} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)}$$

$$\le P_{Z^n}(D) \log |D| + P_{Z^n}(D) \log \frac{1}{P_{Z^n}(D)}. \qquad (34)$$

b) For any finite-alphabet source, if $P_{Z^n}(D_n) \to 0$, then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{z^n \in D_n} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)} = 0. \qquad (35)$$

Proof: To prove part a) first note that if $P_{Z^n}(D) = 0$, then both sides are equal to zero. Otherwise, denote by Q the distribution P_{Z^n} conditioned to D, then replace P_{Z^n} by $P_{Z^n}(D)Q$ in the left side of (34) and use the fact that the entropy of Q is at most $\log |D|$. Part b) follows from a) and $|D_n| \le |A^n|$.   □

Of course, Lemma 2 ceases to be useful if the entropy of the source grows more slowly than n. This is reflected in the assumption of our main result in this section.

Theorem 11: Assume a finite-alphabet source satisfies

$$\liminf_{n \to \infty} \frac{H(Z^n)}{n} > 0. \qquad (36)$$

Then, the following conditions are equivalent:

a) subentropic
b) strongly superentropic.

Proof: Any strongly superentropic source is subentropic (Theorem 4). Thus we just need to show that the reverse holds assuming that (36) holds. To that end, let us start by fixing $\delta_1 > 0$ and $\delta_2 > 0$. Noticing that $G_n(\delta_2) \subset G_n(-\delta_1)$ (using the notation introduced in the proof of Theorem 4), we can

decompose the entropy as

$$\frac{1}{n}H(Z^n) = \frac{1}{n}\sum_{z^n \in G_n(-\delta_1)-G_n(\delta_2)} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)}$$
$$+ \frac{1}{n}\sum_{z^n \in G_n(\delta_2)} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)}$$
$$+ \frac{1}{n}\sum_{z^n \in G_n^c(-\delta_1)} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)}. \quad (37)$$

Lemma 2a) can be used to upper-bound the first and second terms in the right side of (37)

$$\frac{1}{n}\sum_{z^n \in G_n(-\delta_1)-G_n(\delta_2)} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)}$$
$$\leq \frac{1}{n}P[G_n(-\delta_1) - G_n(\delta_2)]\log|G_n(-\delta_1)|$$
$$+ \frac{1}{n}P[G_n(-\delta_1) - G_n(\delta_2)]$$
$$\cdot \log \frac{1}{P[G_n(-\delta_1) - G_n(\delta_2)]}$$
$$= \frac{1}{n}P[G_n(-\delta_1) - G_n(\delta_2)]$$
$$\cdot (1+\delta_1)H(Z^n) + o(1) \quad (38)$$

and

$$\frac{1}{n}\sum_{z^n \in G_n(\delta_2)} P_{Z^n}(z^n) \log \frac{1}{P_{Z^n}(z^n)}$$
$$\leq \frac{1}{n}P[G_n(\delta_2)]\log|G_n(\delta_2)|$$
$$+ \frac{1}{n}P[G_n(\delta_2)]\log \frac{1}{P[G_n(\delta_2)]}$$
$$= \frac{1}{n}P[G_n(\delta_2)](1-\delta_2)H(Z^n) + o(1) \quad (39)$$

where we have used the fact that if $z \in [0,1]$ then $-z\log z \in [0, e^{-1}\log e]$. Now let us insert (38) and (39) into (37). Applying Lemma 2b) to the third term in (37), we can see that it vanishes because $P[G_n(-\delta_1)] \to 1$ since the source is subentropic. So we end up with

$$\frac{1}{n}H(Z^n) \leq \frac{1}{n}H(Z^n)[(1+\delta_1)P[G_n(-\delta_1)]$$
$$- (\delta_1+\delta_2)P[G_n(\delta_2)]] + o(1). \quad (40)$$

We will now reach a contradiction by assuming that the source is not strongly superentropic. This means that there exists $\alpha > 0$ and a subsequence of n along which

$$P[G_n(\delta_2)] > \alpha. \quad (41)$$

If δ_2 and δ_1 are chosen so that

$$\alpha\delta_2 - (1-\alpha)\delta_1 = \eta > 0 \quad (42)$$

then (40)–(42) imply the existence of a subsequence along which

$$\frac{1}{n}H(Z^n) \leq \frac{1-\eta}{n}H(Z^n) + o(1)$$

which is impossible because of (36). □

As an immediate consequence of Theorem 11, we see that for any nonzero entropy-rate finite-alphabet source, we can sharpen Theorem 4 to conclude that the AEP is a necessary and sufficient condition for the source to be encodable at its entropy and not more efficiently:

Corollary: For any finite-alphabet source that satisfies (36), the following conditions are equivalent:

1) AEP
2) entropic
3) strongly entropic.

In the context of finite-alphabet sources satisfying (36), the relationships depicted in Fig. 1 simplify to those shown in Fig. 2.

We conclude this section by giving further results for finite-alphabet sources for which the entropy rate exists.

A result on fixed-length source coding for general sources was given in [6] where it was shown that the minimum achievable fixed-length source coding rate for $\boldsymbol{Z}$ is its *sup-entropy rate*, $\overline{H}(\boldsymbol{Z})$, defined as the infimum of the numbers $h > 0$ for which $P[(1/n)\log 1/P_{Z^n}(Z^n) \geq h]$ vanishes as $n \to \infty$. Analogously, the *inf-entropy rate*, $\underline{H}(\boldsymbol{Z})$, is defined as the supremum of the numbers $h > 0$ for which $P[(1/n)\log 1/P_{Z^n}(Z^n) \leq h]$ vanishes as $n \to \infty$. We will see several relevant results that relate $\overline{H}(\boldsymbol{Z})$ and $\underline{H}(\boldsymbol{Z})$ to the results in Sections II and III, in the important special case in which the entropy rate exists.

Theorem 12: Assume that the entropy rate

$$\lim_{n\to\infty} \frac{H(Z^n)}{n}$$

of a finite alphabet source exists. Then, either

$$\underline{H}(\boldsymbol{Z}) = \lim_{n\to\infty} \frac{H(Z^n)}{n} = \overline{H}(\boldsymbol{Z}) \quad (43)$$

or

$$\underline{H}(\boldsymbol{Z}) < \lim_{n\to\infty} \frac{H(Z^n)}{n} < \overline{H}(\boldsymbol{Z}). \quad (44)$$

Furthermore, if the source is subentropic, then (43) is satisfied.

Proof: If $\underline{H}(\boldsymbol{Z}) = \overline{H}(\boldsymbol{Z})$, then (43) follows from [6, Lemma 1]. If $\underline{H}(\boldsymbol{Z}) < \overline{H}(\boldsymbol{Z})$, then choose $\epsilon > 0$, and define the sets

$$T_1 = \{z^n \in A^n : P_{Z^n}(z^n) \geq \exp(-n\underline{H}(\boldsymbol{Z}) + n\epsilon)\}$$
$$T_2 = \{z^n \in A^n - T_1 : P_{Z^n}(z^n)$$
$$\geq \exp(-n\overline{H}(\boldsymbol{Z})/2 - n\underline{H}(\boldsymbol{Z})/2)\}$$
$$T_3 = \{z^n \in A^n - T_2 : P_{Z^n}(z^n) \geq \exp(-n\overline{H}(\boldsymbol{Z}) - n\epsilon)\}$$
$$T_4 = \{z^n \in A^n : P_{Z^n}(z^n) < \exp(-n\overline{H}(\boldsymbol{Z}) - n\epsilon)\}$$
$$= A^n - T_1 - T_2 - T_3.$$

Note that by the definitions of $\underline{H}(\boldsymbol{Z})$ and $\overline{H}(\boldsymbol{Z})$, $P[T_1] \to 0$, and $P[T_4] \to 0$. Using Lemma 2b), we can bound the normalized entropy by

$$(\underline{H}(\boldsymbol{Z}) - \epsilon)P[T_2] + \frac{\underline{H}(\boldsymbol{Z}) + \overline{H}(\boldsymbol{Z})}{2}P[T_3] + o(1)$$
$$\leq \frac{H(Z^n)}{n}$$
$$\leq \frac{\underline{H}(\boldsymbol{Z}) + \overline{H}(\boldsymbol{Z})}{2}P[T_2] + (\overline{H}(\boldsymbol{Z}) + \epsilon)P[T_3] + o(1).$$
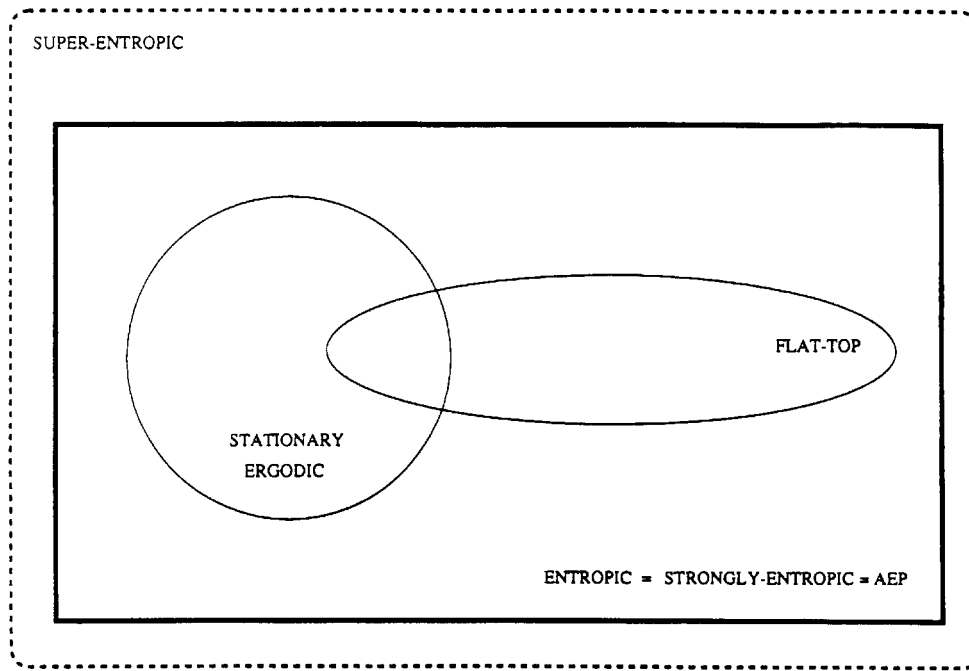
Fig. 2.   Classes of finite-alphabet nonzero-entropy rate sources.

Now (44) follows upon recalling that ϵ is as small as desired, $P[T_2] + P[T_3] \to 1$, and neither $P[T_2]$ nor $P[T_3]$ vanish.

Now, let us show that (43) is the only possible alternative when the source is subentropic, i.e., its most likely $\exp(H(Z^n)(1 + \delta))$ outcomes exhaust all the probability asymptotically. This implies that if $\alpha > \lim_{n\to\infty} H(Z^n)/n$, then the $\exp(n\alpha(1 + \delta))$ most likely outcomes exhaust all the probability asymptotically. But, for a finite-alphabet source, $\overline{H}(Z)$ is necessarily finite and according to [6, Theorem 3] it is the smallest number β such that for all $\delta > 0$, the $\exp(n\beta(1 + \delta))$ most likely outcomes exhaust all the probability asymptotically. Thus the right inequality in (44) is impossible. ☐

We are finally in a position to state and prove the general equivalence result for finite-alphabet nonzero entropy-rate sources.

Theorem 13: The following properties are equivalent for a finite-alphabet source for which the entropy rate $\lim_{n\to\infty} H(Z^n)/n$ exists and is positive.

a) subentropic
b) strongly superentropic
c) entropic
d) strongly entropic
e) AEP
f) $\underline{H}(Z) = \lim_{n\to\infty} H(Z^n)/n$
g) $\lim_{n\to\infty} H(Z^n)/n = \overline{H}(Z)$
h) $\underline{H}(Z) = \overline{H}(Z)$.

Proof: The equivalence of a), b), c), d), e) has been established above under a weaker condition. The equivalence of f), g), h) follows from Theorem 12. Also, a) implies f) according to Theorem 12. To conclude the proof we will show that f) implies e).

Choose $\delta > 0$. Assuming f) and $\underline{H}(Z) > 0$, the following bound holds for all sufficiently large n:

$$
\begin{aligned}
P[B_n(\delta)] &= P\left[\frac{1}{n}\log\frac{1}{P_{Z^n}(Z^n)} \le \frac{1}{n}H(Z^n)(1 - \delta)\right]\\
&\le P\left[\frac{1}{n}\log\frac{1}{P_{Z^n}(Z^n)} \le \underline{H}(Z)\left(1 + \frac{\delta}{2}\right)(1 - \delta)\right]\\
&\le P\left[\frac{1}{n}\log\frac{1}{P_{Z^n}(Z^n)} \le \underline{H}(Z)\left(1 - \frac{\delta}{2}\right)\right]
\end{aligned}
$$

which vanishes asymptotically by definition of $\underline{H}(Z)$.   ☐

We note that the results of [6] imply one more equivalent condition in Theorem 13: that the minimum achievable fixed-length and variable-length source coding rates are equal. The equivalence of a slightly different version of the AEP to conditions reminiscent of f) and g) in Theorem 13 was shown in [10].

Example 2 is a conventional finite-alphabet source with positive entropy rate which does not satisfy a)–h). The following example illustrates that the assumption of nonzero entropy rate in Theorem 13 is not superfluous.

Example 12: Consider a binary source such that all n n-strings with one "0" have probability α/n $(0 < \alpha < 1)$, and all $(n^2 - n)/2$ n-strings with two "0" have probability $(2 - 2\alpha)/(n^2 - n)$. The entropy of an n-string is

$$
H(Z^n) = (2 - \alpha)\log n + O(1).
$$

Thus for sufficiently small $\delta > 0$, the probability of the atypically big masses is

$$
\begin{aligned}
P&\left[\log\frac{1}{P_{Z^n}(Z^n)} \le (1 - \delta)H(Z^n)\right]\\
&\ge P\left[\log\frac{1}{P_{Z^n}(Z^n)} \le (1 - 2\delta)(2 - \alpha)\log n\right] = \alpha
\end{aligned}
$$

and, hence, this source does not satisfy the AEP even though

$$\underline{\boldsymbol{H}}(\boldsymbol{Z}) = \lim_{n \to \infty} \frac{H(Z^n)}{n} = 0.$$

Finally, we mention that for *infinite-alphabet* serial sources, the entropy rate many not lie between the inf-entropy rate $\underline{\boldsymbol{H}}(\boldsymbol{Z})$ and the sup-entropy rate $\overline{\boldsymbol{H}}(\boldsymbol{Z})$; however, the equality of the latter two quantities is equivalent to the validity of the strong-converse fixed-length coding theorem in parallel to the results reported in [9] for channel capacity.

ACKNOWLEDGMENT

REFERENCES

[1] K. L. Chung, *A Course in Probability Theory*, 2nd ed. New York: Academic, 1974.
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[3] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[4] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2, 2nd. ed. New York: Wiley, 1968.
[5] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[6] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, pp. 752–772, May 1993.
[7] R. E. Krichevsky and V. K. Trofimov, "The performance of universal encoding," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 199–206, Mar. 1981.
[8] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 623–656, July–Oct. 1948.
[9] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.