The Shannon Cipher System With a Guessing Wiretapper: General Sources

Manjesh Kumar Hanawal and Rajesh Sundaresan, *Senior Member, IEEE*

Abstract—**The Shannon cipher system is studied in the context of general sources using a notion of computational secrecy introduced by Merhav and Arikan. Bounds are derived on limiting exponents of guessing moments for general sources. The bounds are shown to be tight for i.i.d., Markov, and unifilar sources, thus recovering some known results. A close relationship between error exponents and correct decoding exponents for fixed rate source compression on the one hand and exponents for guessing moments on the other hand is established.**

Index Terms—**Cipher systems, correct decoding exponent, error exponent, fixed-rate source coding, information spectrum, key rate, large deviations, length function, secrecy, sources with memory.**

I. INTRODUCTION

WE consider the classical cipher system of Shannon [1]. Let $X^n = (X_1, \ldots, X_n)$ be a message where each letter takes values on a finite set $\mathbb{X}$. This message should be communicated securely from a transmitter to a receiver, both of which have access to a common secure key U^k of k purely random bits independent of X^n. The transmitter computes the cryptogram $Y = f_n(X^n, U^k)$ and sends it to the receiver over a public channel. The cryptogram may be of variable length. The encryption function f_n is invertible for any fixed U^k. The receiver, knowing Y and U^k, computes $X^n = f_n^{-1}(Y, U^k)$. The functions f_n and f_n^{-1} are published. A wiretapping attacker has access to the cryptogram Y, knows f_n and f_n^{-1}, and attempts to identify X^n without knowledge of U^k. The attacker can use knowledge of the statistics of X^n. We assume that the attacker has a test mechanism that tells him whether a guess $\hat{X}^n$ is correct or not. For example, the attacker may wish to attack an encrypted password or personal information to gain access to, say, a computer account, or a bank account via internet, or a classified database [2]. In these situations, successful entry into the system provides the natural test mechanism. We assume that the attacker is allowed an unlimited number of guesses. The *key rate*

for the cipher system is $R = k(\ln 2)/n$ nats[1] of secrecy per message (or source) letter.

Merhav and Arikan [2] studied discrete memoryless sources (DMS) in the above setting and characterized the best attainable moments of the number of guesses required by an attacker. In particular, they showed that for a DMS with the governing single letter PMF P on $\mathbb{X}$, the value of the optimal exponent for the ρth moment ($\rho > 0$) is given by

$$E(R, \rho) = \max_Q \{\rho \min\{H(Q), R\} - D(Q \parallel P)\}. \quad (1)$$

The maximization is over all PMFs Q on $\mathbb{X}$, $H(Q)$ is the Shannon entropy of Q, and $D(Q \parallel P)$ is the Kullback-Leibler divergence between Q and P. They also showed that $E(R, \rho)$ increases linearly in R for $R \leq H(P)$, continues to increase in a concave fashion for $R \in [H(P), H']$, where H' is a threshold, and is constant for $R > H'$. Unlike the classical equivocation rate analysis, atypical sequences do affect the behavior of $E(R, \rho)$ for $R \in [H(P), H']$ and perfect secrecy is obtained, i.e., cryptogram is uncorrelated with the message, only for $R > H' > H(P)$. Merhav and Arikan also determined the best achievable performance based on the probability of a large deviation in the number of guesses, and showed that it equals the Legendre-Fenchel transform of $E(R, \rho)$ as a function of ρ. Sundaresan [3], [17] extended the above results to unifilar sources. Hayashi and Yamamoto [4] proved coding theorems for the Shannon cipher system with correlated outputs (X^n, Z^n) where the wiretapper is interested in X^n while the receiver in Z^n.

In this paper, we extend Merhav and Arikan's notion of computational secrecy [2] to general sources. One motivation is that secret messages typically come from the natural languages which are modeled well as sources with memory, for e.g., a Markov source of appropriate order. Another motivation is that the study of general sources clearly brings out the connection between guessing and compression, as discussed next.

As with other studies of general sources, *information spectrum* plays crucial role in this paper. We show that $E(R, \rho)$ is closely related to (a) the error exponent of a rate-R source code, and (b) the correct decoding exponent of a rate-R source code, when exponentiated probabilities are considered (see Section III.B.2). In particular, the exponents in (a) and (b) appear in the first and second terms below when we rewrite $E(R, \rho)$ for a DMS as

$$E(R, \rho) = \max \left\{ \rho R - \min_{Q:H(Q)>R} D(Q \parallel P), \right.$$
$$\left. \min_{Q:H(Q)\leq R} \{\rho H(Q) - D(Q \parallel P)\} \right\}.$$

[1]We shall mostly use *nat* as the unit of information in this paper by taking natural logarithms. $k(\ln 2)/n$ nats per input symbol is the same as k/n bits per input symbol.

This brings out the fundamental connection between source coding exponents and key-rate constrained guessing exponents. Further, unlike the case for the probability of a large deviation in the number of guesses [2, Sec. V], both the error exponent and the correct decoding exponent determine $E(R, \rho)$. We extend the above result to general sources by getting upper and lower bounds on $E(R, \rho)$. We then show that these are tight for DMS, Markov and unifilar sources. The bounds may be of interest even if they are not tight because the upper bound specifies the amount of effort need by an attacker and the lower bound specifies the secrecy strength of the cryptosystem to a designer.

The limiting case as $\rho \downarrow 0$ in (b) yields classical framework for probability of correct decoding. This special case is related to the work of Han [5] and Iriyama [6] who studied the dual problem of rates required to meet a specified error exponent or a specified correct decoding exponent.

The paper is organized as follows. Section II relates our problem to a modification of Campbell's compression problem [7]. Section III gives bounds on the limits of exponential rate of guessing moments, in terms of information spectrum quantities. Section IV evaluates the bounds for some specific examples. Section V concludes the paper with additional remarks. Proofs are given in the Appendices.

II. GUESSING WITH KEY-RATE CONSTRAINTS AND SOURCE COMPRESSION

In this section, we make a precise statement of our problem, and establish a connection between guessing and source compression subject to a new cost criterion.

Let $\mathbb{X}^n$ denote the set of messages and $\mathcal{M}(\mathbb{X}^n)$ the set of PMFs on $\mathbb{X}^n$. By a source, we mean a sequence of PMFs $(P_n : n \in \mathbb{N})$, where[2] $P_n \in \mathcal{M}(\mathbb{X}^n)$. Let X^n denote a message put out by the source and U^k the secure key of k purely random bits independent of X^n. Recall that the transmitter computes the cryptogram $Y = f_n(X^n, U^k)$ and sends it to the receiver over a public channel.

For a given cryptogram $Y = y$, define a *guessing strategy*

$$G_n(\cdot|y) : \mathbb{X}^n \to \{1, 2, \ldots, |\mathbb{X}|^n\}$$

as a bijection that denotes the order in which elements of $\mathbb{X}^n$ are guessed. $G_n(x^n|y) = l$ indicates that x^n is the lth guess, when the cryptogram is y. With knowledge of P_n, the encryption function f_n, and the cryptogram Y, the attacker can exhaustively calculate the posterior probabilities of all plaintexts $P_{X^n|Y}(\cdot|y)$ given the cryptogram. The attacker's optimal guessing strategy is then to guess in the decreasing order of these posterior probabilities $P_{X^n|Y}(\cdot|y)$. Let us denote this optimal attack strategy as G_{f_n}. The key rate for the system is $R = k(\ln 2)/n$ nats of secrecy per source letter. Let $(f_n : n \in \mathbb{N})$ denote the sequence of encryption functions, where $\mathbb{N}$ denotes the set of natural numbers. This sequence is known to the attacker. We assume that the attacker employs the aforementioned optimal guessing strategy.

[2]Sometimes we use P_{X^n} in place of P_n when we refer to the distribution of the random vector X^n.

For a given $\rho > 0$, key rate $R > 0$, define the normalized guessing exponent

$$E_n^g(R, \rho) := \sup_{f_n} \frac{1}{n} \ln \mathbb{E}[G_{f_n}(X^n|Y)^\rho].$$

The supremum is taken over all encryption functions. Further define performance limits of guessing moments as in [2]

$$E_u^g(R, \rho) := \limsup_{n \to \infty} E_n^g(R, \rho) \tag{2}$$

$$E_l^g(R, \rho) := \liminf_{n \to \infty} E_n^g(R, \rho). \tag{3}$$

We next define the related compression quantities. A length function $L_n : \mathbb{X}^n \to \mathbb{N}$ is a mapping that satisfies Kraft's inequality:

$$\sum_{x^n \in \mathbb{X}^n} \exp_2\{-L_n(x^n)\} \le 1,$$

where the code alphabet is taken to be binary and $\exp_2\{a\} = 2^a$. (We shall use exp to denote the inverse of the natural logarithm $\ln$). Every length function yields an attack strategy with a performance characterized as follows.

Proposition 1: Let L_n be any length function on $\mathbb{X}^n$. There is a guessing list G_n such that for any encryption function f_n, we have[3]

$$G_n(x^n|y) \le 2 \exp_2\{\min\{L_n(x^n), nR/(\ln 2)\}\}$$
$$= 2 \exp\{\min\{L_n(x^n) \ln 2, nR\}\}.$$

Proof: We use a technique of Merhav and Arikan [2]. Let G_{L_n} denote the guessing function that ignores the cryptogram and proceeds in the increasing order of L_n lengths. Suppose G_{L_n} proceeds in the order $x_1^n, x_2^n, \ldots$. By [8, Prop. 2], [18], we need at most $\exp_2\{L_n(x^n)\}$ guesses to identify x^n (This is a simple consequence of the fact that there are at most $\exp_2\{L_n(x^n)\}$ strings of length less than or equal to $L_n(x^n)$). As an alternative attack, consider the exhaustive key-search attack defined by the following guessing list:

$$f_n^{-1}(y, u_1^k), f_n^{-1}(y, u_2^k), \ldots$$

where $u_1^k, u_2^k, \ldots$ is an arbitrary ordering of the keys. This strategy identifies x^n in at most $\exp\{nR\} = \exp_2\{nR/(\ln 2)\}$ guesses. Finally, let $G_n(\cdot|y)$ be the list that alternates between the two lists, skipping those already guessed, i.e., the one that proceeds in the order

$$x_1^n, f_n^{-1}(y, u_1^k), x_2^n, f_n^{-1}(y, u_2^k), \ldots. \tag{4}$$

Clearly, for every x^n, we need at most twice the minimum over the two individual lists. ∎

We now look at a weak converse in the expected sense to the above. We first state without proof the following lemma which associates a length function to any guessing function.

Lemma 2: Given a guessing function G_n, there exists a length function L_{G_n} satisfying

$$L_{G_n}(x^n) - 1 - \log_2 c_n \le \log_2 G_n(x^n) \le L_{G_n}(x^n) \tag{5}$$

[3]We reiterate that R is measured in nats.

where

$$c_n = \sum_{i=1}^{|\mathbb{X}|^n} \frac{1}{i}.$$

For a proof, we refer the reader to [8, Prop. 1], [18]. We then have the following proposition.

Proposition 3: Fix $n \in \mathbb{N}$, $\rho > 0$. There is an encryption function f_n and a length function L_n such that every guessing strategy G_n (and in particular G_{f_n}) satisfies

$$\mathbb{E}[G_n(X^n|Y)^\rho]$$
$$\geq \frac{1}{(2c_n)^\rho (2+\rho)} \mathbb{E}[\exp\{\rho \min\{L_n(X^n)\ln 2, nR\}\}].$$

Proof: See Appendix A. The proof is an extension of Merhav and Arikan's proof of [2, Th.1] to sources with memory. The idea is to identify an encryption mechanism that maps messages of roughly equal probability to each other. Our proof also suggests an asymptotically optimal encryption strategy for sources with memory. ∎

Remark 1: Note that $c_n \leq 1 + n\ln|\mathbb{X}|$, so that

$$\frac{\log_2 c_n}{n} = O\left(\frac{\log_2 n}{n}\right) = o(1) \tag{6}$$

a fact that will be put to good use in the sequel. ☐

Propositions 1 and 3 naturally suggest the following coding problem: identify

$$E_n^s(R,\rho) := \min_{L_n} \frac{1}{n}$$
$$\times \ln \mathbb{E}[\exp\{\rho \min\{L_n(X^n)\ln 2, nR\}\}]. \tag{7}$$

The minimum is taken over all possible length functions. We may interpret the cost of using length $L_n(x^n)$ as being $\exp\{\min\{L_n(x^n)\ln 2, nR\}\}$, i.e., the cost is exponential in L_n, but saturates at $\exp\{nR\}$ and so all lengths larger than nR nats (i.e., $nR/(\ln 2)$ bits) enjoy the saturated cost. Then $E_n^s(R,\rho)$ is the minimum normalized exponent of the ρth moment of this new compression cost. In analogy with (2) and (3), we define

$$E_u^s(R,\rho) = \limsup_{n\to\infty} E_n^s(R,\rho)$$
$$E_l^s(R,\rho) = \liminf_{n\to\infty} E_n^s(R,\rho).$$

The following is a corollary to Propositions 1 and 3, and relates $E_n^g(R,\rho)$ and $E_n^s(R,\rho)$.

Corollary 4: For a given $R, \rho > 0$, we have

$$|E_n^s(R,\rho) - E_n^g(R,\rho)| \leq \frac{\ln((4c_n)^\rho (2+\rho))}{n}. \tag{8}$$

Proof: Let L_n^* be the length function that achieves $E_n^s(R,\rho)$. Using Proposition 1, and after taking expectation, we have the guessing strategy G_n that satisfies

$$\mathbb{E}[\exp\{\rho \min\{L_n^*(X^n)\ln 2, nR\}\}]$$
$$\geq \sup_{f_n} \frac{1}{2^\rho} \mathbb{E}[G_n(X^n|Y)^\rho]$$
$$\geq \sup_{f_n} \frac{1}{2^\rho} \mathbb{E}[G_{f_n}(X^n|Y)^\rho]$$

$$\geq \frac{1}{(4c_n)^\rho (2+\rho)} \mathbb{E}[\exp\{\rho \min\{L_n(X^n)\ln 2, nR\}\}]$$

for some f_n and L_n, given by Proposition 3,

$$\geq \frac{1}{(4c_n)^\rho (2+\rho)} \mathbb{E}[\exp\{\rho \min\{L_n^*(X^n)\ln 2, nR\}\}].$$

Take logarithms, normalize by n, use $c_n > 1$ and $\rho > 0$ to get (8). ∎

We now state the equivalence between compression and guessing.

Theorem 5 (Guessing-Compression Equivalence): For any $\rho > 0$ and $R > 0$, we have $E_u^s(R,\rho) = E_u^g(R,\rho)$ and $E_l^s(R,\rho) = E_l^g(R,\rho)$.

Proof: From Corollary 4 and (6), magnitude of the difference between $E_n^g(R,\rho)$ and $E_n^s(R,\rho)$ decays as $O((\ln n)/n)$ and vanishes as $n \to \infty$. ∎

Thus, the problem of finding the optimal guessing exponent is the same as that of finding the optimal exponent for the coding problem in (7). When $R \geq \ln|\mathbb{X}|$, the coding problem in (7) reduces to the one considered by Campbell in [7]; this is a case where perfect secrecy is obtained and is studied in [8], [18]. Proposition 1 shows that the optimal length function attaining the minimum in (7) yields an asymptotically optimal attack strategy on the cipher system. Moreover, the encryption strategy in the proof of Proposition 3 (see Appendix A) is asymptotically optimal, from the designer's point of view.

In the rest of the paper we focus on the equivalent compression problem and find bounds on E_u^s and E_l^s.

III. GROWTH EXPONENT FOR THE MODIFIED COMPRESSION PROBLEM

We begin with some words on notation. Recall that $\mathcal{M}(\mathbb{X}^n)$ denotes the set of PMFs on $\mathbb{X}^n$. The Shannon entropy for a $P_n \in \mathcal{M}(\mathbb{X}^n)$ is

$$H(P_n) = -\sum_{x^n \in \mathbb{X}^n} P_n(x^n)\ln P_n(x^n)$$

and the Rényi entropy of order $\alpha \neq 1$ is

$$H_\alpha(P_n) = \frac{1}{1-\alpha} \ln\left(\sum_{x^n \in \mathbb{X}^n} P_n(x^n)^\alpha\right). \tag{9}$$

The Kullback-Leibler divergence or relative entropy between two PMFs Q_n and P_n is

$$D(Q_n \| P_n) = \begin{cases} \sum_{x^n \in \mathbb{X}^n} Q_n(x^n)\ln \frac{Q_n(x^n)}{P_n(x^n)}, & \text{if } Q_n \ll P_n \\ \infty, & \text{otherwise} \end{cases}$$

where $Q_n \ll P_n$ means Q_n is absolutely continuous with respect to P_n. We shall use $(X^n : n \in \mathbb{N})$ to denote a sequence of random variables on $\mathbb{X}^n$, with corresponding sequence of probability measures denoted by $\mathbf{X} := (P_{X^n} : n \in \mathbb{N})$. Thus, $\mathbf{X}$ is a source and X^n its n-letter message output. Abusing notation, we let $\mathcal{M}(\mathbb{X}^{\mathbb{N}})$ denote the set of all sequences $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$ of probability measures, and for each $\mathbf{B} := (B_n \subseteq \mathbb{X}^n : n \in \mathbb{N})$, we define

$$\mathcal{M}(\mathbf{B}) := \left\{\mathbf{Y} \in \mathcal{M}(\mathbb{X}^{\mathbb{N}}) : \lim_{n\to\infty} P_{Y^n}(B_n) = 1\right\}.$$

In the rest of this section $\mathbf{X}$ is a fixed source. For any $\mathbf{Y} \in \mathcal{M}(\mathbf{B})$ and $\rho > 0$, define

$$
\begin{aligned}
E_u(\mathbf{Y}, \mathbf{X}, \rho) := \limsup_{n \to \infty} \frac{1}{n} \{ &\rho H(P_{Y^n}) \\
&- D(P_{Y^n} \| P_{X^n}) \}
\end{aligned}
$$

and

$$
\begin{aligned}
E_l(\mathbf{Y}, \mathbf{X}, \rho) := \liminf_{n \to \infty} \frac{1}{n} \{ &\rho H(P_{Y^n}) \\
&- D(P_{Y^n} \| P_{X^n}) \}.
\end{aligned}
$$

We next state a large deviation result that plays a key role in the derivation of bounds on E_u^s and E_l^s.

Proposition 6: For all $\rho \geq 0$ and $\mathbf{B} = (B_n \subseteq \mathbb{X}^n : n \in \mathbb{N})$, we have

$$
\begin{aligned}
(1+\rho) \limsup_{n \to \infty} \frac{1}{n} \ln \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \\
= \max_{\mathbf{Y} \in \mathcal{M}(\mathbf{B})} E_u(\mathbf{Y}, \mathbf{X}, \rho)
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
(1+\rho) \liminf_{n \to \infty} \frac{1}{n} \ln \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \\
= \max_{\mathbf{Y} \in \mathcal{M}(\mathbf{B})} E_l(\mathbf{Y}, \mathbf{X}, \rho).
\end{aligned}
\tag{11}
$$

The maximum-achieving distribution in (10) and (11) is the source $\mathbf{X}^* = (P_{X^n}^* : n \in \mathbb{N})$ given by

$$
P_{X^n}^*(\cdot) = \frac{P_{X^n}^{\frac{1}{1+\rho}}(\cdot)}{\sum_{y^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(y^n)}.
\tag{12}
$$

Proof: See Appendix B.   ∎

Remark 2: This proposition is a generalization of Iriyama's [6, Prop. 1], which is obtained by setting $\rho = 0$.   □

A. Upper Bound on E_u^s

We first obtain an upper bound on E_u^s. We use $\mathbb{E}_{X^n}[\cdot]$ to denote the expectation with respect to distribution P_{X^n}.

Proposition 7 (Upper Bound): Let $R > 0$ and $\rho > 0$. Then

$$
E_u^s(R, \rho) \leq \min_{0 \leq \theta \leq \rho} \left[(\rho - \theta) R + \max_{\mathbf{Y} \in \mathcal{M}(\mathbb{X}^{\mathbb{N}})} E_u(\mathbf{Y}, \mathbf{X}, \theta) \right].
$$

Proof: We first recall the useful variational formula [9, Prop. 1.4.2]

$$
\begin{aligned}
\ln \mathbb{E}_{X^n}[\exp\{U(X^n)\}] \\
= \sup_{P_{Y^n}} \{ \mathbb{E}_{Y^n}[U(Y^n)] - D(P_{Y^n} \| P_{X^n}) \}
\end{aligned}
\tag{13}
$$

for any $U : \mathbb{X}^n \to \mathbb{R}$, where $\mathbb{R}$ denotes set of real numbers. For notational convenience, let $d(Y^n) := D(P_{Y^n} \| P_{X^n})$. Observe that

$$
\begin{aligned}
\ln \mathbb{E}_{X^n}[\exp\{\rho \min\{L_n(X^n) \ln 2, nR\}\}] \\
= \sup_{P_{Y^n}} [\rho \mathbb{E}_{Y^n}[\min\{L_n(Y^n) \ln 2, nR\}] \\
- d(Y^n)]
\end{aligned}
\tag{14}
$$

$$
\begin{aligned}
\leq \sup_{P_{Y^n}} [\rho \min\{\mathbb{E}_{Y^n}[L_n(Y^n) \ln 2], nR\} \\
- d(Y^n)]
\end{aligned}
\tag{15}
$$

$$
\begin{aligned}
= \sup_{P_{Y^n}} \left\{ \min_{0 \leq \theta \leq \rho} [(\rho - \theta) nR \right. \\
\left. + \theta \mathbb{E}_{Y^n}[L_n(Y^n) \ln 2] - d(Y^n) \right\}
\end{aligned}
\tag{16}
$$

$$
\begin{aligned}
= \min_{0 \leq \theta \leq \rho} \sup_{P_{Y^n}} \{ (\rho - \theta) nR \\
+ \theta \mathbb{E}_{Y^n}[L_n(Y^n) \ln 2] - d(Y^n) \}
\end{aligned}
\tag{17}
$$

$$
\begin{aligned}
= \min_{0 \leq \theta \leq \rho} \left\{ (\rho - \theta) nR \right. \\
\left. + \sup_{P_{Y^n}} \{\theta \mathbb{E}_{Y^n}[L_n(Y^n) \ln 2] - d(Y^n)\} \right\}.
\end{aligned}
$$

In the above sequence of inequalities, (14) follows from the variational formula (13) with

$$
U(x^n) = \rho \min\{L_n(x^n) \ln 2, nR\}.
$$

Inequality (15) follows from Jensen's inequality because $\min\{\cdot, nR\}$ is concave for a fixed nR. Equality (16) follows from the identity

$$
\rho \min\{a, b\} = \min_{0 \leq \theta \leq \rho} \{\theta a + (\rho - \theta) b\}.
$$

Equality (17) follows because the term within braces is linear in θ for a fixed P_{Y^n}, concave in P_{Y^n} for a fixed θ, and the sets $[0, \rho]$ and $\mathcal{M}(\mathbb{X}^n)$ are compact and convex; these permit an interchange of sup and inf, thanks to a minmax theorem [10, Cor. 2, p. 53]. Taking inf over L_n, and interchanging the inf over L_n and the min over θ, we get

$$
\begin{aligned}
\inf_{L_n} \ln \mathbb{E}_{X^n}[\exp\{\rho \min\{L_n(Y^n) \ln 2, nR\}\}] \\
\leq \min_{0 \leq \theta \leq \rho} \left\{ (\rho - \theta) nR \right. \\
\left. + \inf_{L_n} \sup_{P_{Y^n}} \{\theta \mathbb{E}_{Y^n}[L_n(Y^n) \ln 2] - d(Y^n)\} \right\} \\
= \min_{0 \leq \theta \leq \rho} \left\{ (\rho - \theta) nR \right. \\
+ \sup_{P_{Y^n}} \left\{ \theta \inf_{L_n} \mathbb{E}_{Y^n}[L_n(Y^n) \ln 2] \right. \\
\left. \left. - d(Y^n) \right\} + O(1) \right\}
\end{aligned}
\tag{18}
$$

$$
\begin{aligned}
= \min_{0 \leq \theta \leq \rho} \left\{ (\rho - \theta) nR \right. \\
\left. + \sup_{P_{Y^n}} \{\theta H(P_{Y^n}) - d(Y^n)\} + O(1) \right\}
\end{aligned}
\tag{19}
$$

$$
= \min_{0 \leq \theta \leq \rho} \left\{ (\rho - \theta) nR + \theta H_{\frac{1}{1+\theta}}(P_{X^n}) + O(1) \right\}.
\tag{20}
$$

Equality (18) follows because the function inside the inner braces is concave in P_{Y^n}, asymptotically linear in L_n (see proof of [8, Prop. 6], [18]), and $\mathcal{M}(\mathbb{X}^n)$ is compact; this allows us to interchange inf and sup. Inequality (19) follows because inf of expected compression lengths over all prefix codes is within $\ln 2$ nats (1 bit) of entropy. The last equality follows from the well-known variational characterization of Rényi entropy

$$
\sup_{P_{Y^n}} \{\theta H(P_{Y^n}) - D(P_{Y^n} \| P_{X^n})\} = \theta H_{\frac{1}{1+\theta}}(P_{X^n})
\tag{21}
$$

a fact that can also be gleaned from the variational formula (13). Divide both sides of (20) by n and take limit supremum as $n \to \infty$ to get

$$
\begin{aligned}
&E_u^s(R, \rho) \\
&\leq \limsup_{n \to \infty} \min_{0 \leq \theta \leq \rho} \left\{ (\rho - \theta)R + \frac{\theta}{n} H_{\frac{1}{1+\theta}}(P_{X^n}) \right\} \\
&\leq \min_{0 \leq \theta \leq \rho} \left\{ (\rho - \theta)R + \theta \limsup_{n \to \infty} \frac{1}{n} H_{\frac{1}{1+\theta}}(P_{X^n}) \right\} \\
&= \min_{0 \leq \theta \leq \rho} \left\{ (\rho - \theta)R + \max_{\mathbf{Y} \in \mathcal{M}(\mathbb{X}^{\mathbb{N}})} E_u(\mathbf{Y}, \mathbf{X}, \theta) \right\}
\end{aligned}
$$

where the last inequality follows from Proposition 6 and the formula for Rényi entropy. This completes the proof. ∎

From the above proof, it is clear that the upper bound holds with equality, when Jensen's inequality holds with equality in (15), i.e., the random variable $(1/n)\min\{L_n(X^n)\ln 2, nR\}$ tends asymptotically to a constant. This would happen, for example, when normalized encoded lengths concentrate around the entropy rate of the source.

B. Lower Bound on E_l^s

We now derive a lower bound on E_l^s. For a given distribution P_{Y^n} arrange the elements of set $\mathbb{X}^n$ in the decreasing order of their P_{Y^n}-probabilities as done in Sundaresan [3, Sec. IV], [3]. Enumerate the sequences from 1 to $|\mathbb{X}|^n$. Henceforth refer to a message by its index. Let $T_R(Y^n)$ denote the first $M = \lfloor \exp\{nR\} \rfloor$ elements in the list. We denote the probability of this set by F_{Y^n}, i.e.,

$$
F_{Y^n} = \sum_{x^n \in T_R(Y^n)} P_{Y^n}(x^n)
$$

and the probability of the complement of this set $T_R^c(Y^n)$ by $F_{Y^n}^c$. Let the restriction of P_{Y^n} to this set $T_R(Y^n)$ be P'_{Y^n}. Let L_n^* denote the length function that attains $E_n^s(R, \rho)$ in (7). As the length functions are uniquely decipherable we have $\exp_2\{L_n^*(i)\} \geq i$.

Proposition 8 (Lower Bound): For a given $\rho > 0$ and rate $R > 0$, we have

$$
\begin{aligned}
E_l^s(R, \rho) \geq \max \Bigg\{ &\rho R + \liminf_{n \to \infty} \frac{1}{n} \ln F_{X^n}^c, \\
&(1+\rho) \liminf_{n \to \infty} \frac{1}{n} \ln \sum_{x^n \in T_R(X^n)} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \Bigg\}. \quad (22)
\end{aligned}
$$

Remark 3: The first term contains limit infimum of the error exponent for a rate-R source code. The second exponent is the correct decoding exponent for a rate-R code when $\rho \downarrow 0$. ☐

Proof: The variational formula (13) applied to the function $U(x^n) = \rho \min\{L_n(x^n)\ln 2, nR\}$ gives

$$
\begin{aligned}
&\inf_{L_n} \ln \mathbb{E}_{X^n}[\exp\{\rho \min\{L_n(X^n)\ln 2, nR\}\}] \\
&= \inf_{L_n} \sup_{P_{Y^n}} \{\rho \mathbb{E}_{Y^n}[\min\{L_n(Y^n)\ln 2, nR\}] - d(Y^n)\}
\end{aligned}
$$

$$
\begin{aligned}
\geq \sup_{P_{Y^n}} \Big\{ &\rho \inf_{L_n} \mathbb{E}_{Y^n}[\min\{L_n(X^n)\ln 2, nR\}] \\
&- d(Y^n) \Big\}
\end{aligned}
\quad (23)
$$

where the interchange of inf and sup yields the lower bound in (23). Fix a distribution P_{Y^n} and consider the first term in (23). Using the enumeration indicated above, we may write

$$
\begin{aligned}
&\inf_{L_n} \mathbb{E}_{Y^n}[\min\{L_n(Y^n)\ln 2, nR\}] \\
&= \sum_{i=1}^{|\mathbb{X}|^n} P_{Y^n}(i) \min\{L_n^*(i)\ln 2, nR\} \\
&= \sum_{i=1}^{M} P_{Y^n}(i) \min\{L_n^*(i)\ln 2, nR\} \\
&\quad + \sum_{i=M+1}^{|\mathbb{X}|^n} P_{Y^n}(i) nR \\
&\geq \sum_{i=1}^{M} P_{Y^n}(i) \ln G_n^*(i) + nR F_{Y^n}^c \quad (24) \\
&\geq F_{Y^n} \sum_{i=1}^{M} \frac{P_{Y^n}(i)}{F_{Y^n}} L_{G_n^*}(i) \ln 2 \\
&\quad - \ln 2 - \ln(1 + n \ln |\mathbb{X}|) + nR F_{Y^n}^c \quad (25) \\
&\geq F_{Y^n} H(P'_{Y^n}) - \ln 2 \\
&\quad - \ln(1 + n \ln |\mathbb{X}|) + nR F_{Y^n}^c \quad (26)
\end{aligned}
$$

where in (26), P'_{Y^n} is the conditional distribution of P_{Y^n} given $T_R(Y^n)$. Inequality (24) follows because

$$
L_n^*(i)\ln 2 \geq \ln i = \ln G_n^*(i)
$$

with G_n^* the guessing strategy that guesses in decreasing order of P_{Y^n} probabilities. $L_{G_n^*}$ in (25) denotes the length function given by Lemma 2. Inequality (26) follows from the source coding theorem's lower bound. Substitute (26) in (23), normalize by n, and take limit infimum to get

$$
\begin{aligned}
E_l^s(R, \rho) \geq \liminf_{n \to \infty} \frac{1}{n} \sup_{P_{Y^n}} \{ &\rho F_{Y^n} H(P'_{Y^n}) \\
&+ F_{Y^n}^c \rho nR - d(Y^n) \}.
\end{aligned}
$$

P_{Y^n} may be thought of as a triplet made of P'_{Y^n}, F_{Y^n}, and the restriction of P_{Y^n} to $T_R^c(Y^n)$. We now perform the optimization

$$
\sup_{P_{Y^n}} \{\rho F_{Y^n} H(P'_{Y^n}) + F_{Y^n}^c \rho nR - d(Y^n)\} \quad (27)
$$

in four steps.

Step 1: We first optimize over permutations of probabilities over strings. $F_{Y^n}, F_{Y^n}^c, H(P_{Y^n})$, and $H(P'_{Y^n})$ remain unchanged over these permutations. Observe that

$$
-d(Y^n) = H(P_{Y^n}) + \sum_{y^n} P_{Y^n}(y^n) \ln P_{X^n}(y^n),
$$

and so the maximum for $-d(Y^n)$ is attained when the permutation that orders $P_{X^n}(\cdot)$ in decreasing order also orders $P_{Y^n}(\cdot)$ in decreasing order. In particular, $T_R(Y^n)$ equals $T_R(X^n)$.

Step 2: We now optimize over restriction of P_{Y^n} to $T_R^c(Y^n)$. For a fixed F_{Y^n}, the log-sum inequality yields

$$\sum_{x^n \in T_R^c(X^n)} P_{Y^n}(x^n) \ln \frac{P_{Y^n}(x^n)}{P_{X^n}(x^n)} \geq F_{Y^n}^c \ln \frac{F_{Y^n}^c}{F_{X^n}^c}$$

with equality if and only if $P_{Y^n}(x^n) = P_{X^n}(x^n)\frac{F_{Y^n}^c}{F_{X^n}^c}$ for all $x^n \in T_R^c(P_{X^n})$.

Step 3: To optimize over P'_{Y^n}, rewrite (27) as

$$\sup_{P_{Y^n}} \left\{ \rho F_{Y^n} H(P'_{Y^n}) + F_{Y^n}^c \rho nR \right.$$

$$- \sum_{i=1}^{M} P_{Y^n}(i) \ln \frac{P_{Y^n}(i)}{P_{X^n}(i)}$$

$$\left. - \sum_{i=M+1}^{|\mathbb{X}|^n} P_{Y^n}(i) \ln \frac{P_{Y^n}(i)}{P_{X^n}(i)} \right\}$$

$$= \sup_{P'_{Y^n}, F_{Y^n}} \left\{ \rho F_{Y^n} H(P'_{Y^n}) + F_{Y^n}^c \rho nR \right.$$

$$\left. - \sum_{i=1}^{M} P_{Y^n}(i) \ln \frac{P_{Y^n}(i)}{P_{X^n}(i)} - F_{Y^n}^c \ln \frac{F_{Y^n}^c}{F_{X^n}^c} \right\} \quad (28)$$

$$= \sup_{P'_{Y^n}, F_{Y^n}} \left\{ \rho F_{Y^n} H(P'_{Y^n}) + F_{Y^n}^c \rho nR \right.$$

$$\left. - F_{Y^n} D(P'_{Y^n} \| P'_{X^n}) - D(F_{Y^n} \| F_{X^n}) \right\} \quad (29)$$

$$= \sup_{F_{Y^n}} \left\{ \rho F_{Y^n} H_{\frac{1}{1+\rho}}(P'_{X^n}) + F_{Y^n}^c \rho nR \right.$$

$$\left. - D(F_{Y^n} \| F_{X^n}) \right\}. \quad (30)$$

Equality (28) is obtained by substituting the attained lower bound in Step 2. In (29), P'_{Y^n} and P'_{X^n} denote conditional distributions of P_{Y^n} and P_{X^n} given $T_R(Y^n)$ and $T_R(X^n)$, respectively, where $T_R(Y^n) = T_R(X^n)$ as argued in Step 1. $D(F_{Y^n} \| F_{X^n})$ denotes the divergence between binary random variables whose probabilities are $\{F_{Y^n}, 1 - F_{Y^n}\}$ and $\{F_{X^n}, 1 - F_{X^n}\}$ respectively. Finally we used variational characterization of Rényi entropy given in (21) to arrive at (30).

Step 4: We now optimize over $F_{Y^n} \in [0, 1]$. Let Z be a binary random variable defined as

$$Z = \begin{cases} \rho H_{\frac{1}{1+\rho}}(P'_{X^n}) & \text{with probability } F_{Y^n} \\ \rho nR & \text{with probability } 1 - F_{Y^n} \end{cases}.$$

By $\mathbb{E}_{F_{Y^n}}[Z]$ we mean the expectation of Z with respect to the above distribution. Since Z is a positive random variable, the variational formula yields

$$\sup_{F_{Y^n}} \left\{ \mathbb{E}_{F_{Y^n}}[Z] - D(F_{Y^n} \| F_{X^n}) \right\} = \ln \mathbb{E}_{F_{X^n}}[\exp\{Z\}].$$

Continuing with the chain of equalities from (30), we get

$$\sup_{F_{Y^n}} \left\{ F_{Y^n} \rho H_{\frac{1}{1+\rho}}(P'_{X^n}) + F_{Y^n}^c \rho nR - D(F_{Y^n} \| F_{X^n}) \right\}$$

$$= \ln \left\{ F_{X^n}^c \exp\{nR\rho\} + F_{X^n} \left(\sum_{i=1}^{M} P'_{X^n}{}^{\frac{1}{1+\rho}}(i) \right)^{1+\rho} \right\}$$

$$= \ln \left\{ F_{X^n}^c \exp\{nR\rho\} + \left(\sum_{i=1}^{M} P_{X^n}^{\frac{1}{1+\rho}}(i) \right)^{1+\rho} \right\}. \quad (31)$$

Finally, normalize both sides of (31) by n, take limit infimum, and apply [11, Lemma 1.2.15], which states that the exponential rate of a sum is governed by the maximum of the individual terms' exponential rates, to get the desired result. ∎

In the subsequent subsections we further lower bound each of the two terms under max on the right-hand side of (22). For an arbitrary source, we first recall the source coding error exponent. We also identify the growth rate of sum of exponentiated probabilities of the correct decoding set. We then relate these to the terms in the lower bound obtained in (22). We largely follow the approach and notation of Iriyama [6], which we now describe.

Given $\mathbf{X} = (P_{X^n} : n \in \mathbb{N})$ and $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$, we define the upper divergence $D_u(\cdot \| \cdot)$ and lower divergence $D_l(\cdot \| \cdot)$ by

$$D_u(\mathbf{Y} \| \mathbf{X}) := \limsup_{n \to \infty} \frac{1}{n} D(P_{Y^n} \| P_{X^n})$$

$$D_l(\mathbf{Y} \| \mathbf{X}) := \liminf_{n \to \infty} \frac{1}{n} D(P_{Y^n} \| P_{X^n}).$$

For a $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N})$, denote the *spectral sup-entropy-rate* [5, Sec. II], [12] as

$$\overline{H}(\mathbf{Y}) := \inf \left\{ \theta : \lim_{n \to \infty} \Pr \left\{ \frac{1}{n} \ln \frac{1}{P_{Y^n}(Y^n)} > \theta \right\} = 0 \right\}$$

and the *spectral inf-entropy-rate* as

$$\underline{H}(\mathbf{Y}) := \sup \left\{ \theta : \lim_{n \to \infty} \Pr \left\{ \frac{1}{n} \ln \frac{1}{P_{Y^n}(Y^n)} < \theta \right\} = 0 \right\}.$$

Also define, as in [6, Sec. II], the following quantity which determines the performance under mismatched compression

$$\underline{R}(\mathbf{Y}, \mathbf{X}) := \sup \left\{ \theta : \right.$$

$$\left. \lim_{n \to \infty} \Pr \left\{ \frac{1}{n} \ln \frac{1}{P_{X^n}(Y^n)} < \theta \right\} = 0 \right\}.$$

1) Decoding Error Exponent: In this subsection, we recall the decoding error exponent for fixed-rate encoding of an arbitrary source. We identify the first term in (22) as composed of the exponent of minimum probability of decoding error, and obtain a lower bound for it, or alternatively an upper bound on the error exponent. This is made precise in the following definitions.

By an (n, M_n, ϵ_n)-code we mean an encoding mapping

$$\phi_n : \mathbb{X}^n \to \{1, 2, \ldots, M_n\}$$

and a decoding mapping

$$\psi_n : \{1, 2, \ldots, M_n\} \to \mathbb{X}^n$$

with probability of error $\epsilon_n := \Pr\{\psi_n(\phi_n(X^n)) \neq X^n\}$. R is r-achievable if for all $\eta > 0$ there exists a sequence of (n, M_n, ϵ_n)-codes such that

$$\limsup_{n \to \infty} \frac{1}{n} \ln \frac{1}{\epsilon_n} \geq r \qquad (32)$$

$$\limsup_{n \to \infty} \frac{1}{n} \ln M_n \leq R + \eta. \qquad (33)$$

The *infimum fixed-length coding rate* for exponent r is

$$\hat{R}(r|\mathbf{X}) = \inf\{R : R \text{ is } r\text{–achievable}\}.$$

On the other hand, the *supremum fixed-length coding exponent* for rate R is

$$\hat{E}(R|\mathbf{X}) = \sup\{r : R \text{ is } r\text{–achievable}\}.$$

See Iriyama [6] and Han [12, Sec. 1.9] for a pessimistic definition for fixed rate source coding, i.e., the liminf in place of limsup in (32). See also Iriyama and Ihara [13] for both the pessimistic and optimistic definitions. These works obtained bounds on the infimum coding rate. In particular, Iriyama [6, eq. (13)], Iriyama and Ihara [13, eq. (12)] obtained lower bounds on the infimum coding rate $\hat{R}(r|\mathbf{X})$ under the optimistic definition, the definition of interest to us. We however work with the error exponent, and obtain an upper bound on supremum coding exponent. This suffices to lower bound the first term in (22).

Clearly, $M_n = \lfloor \exp\{nR\} \rfloor$ satisfies (33), and with

$$r_0 = \limsup_{n \to \infty} \frac{1}{n} \log \frac{1}{F_{X^n}^c}$$

R is r_0-achievable. It follows from the definition of $\hat{E}(R|\mathbf{X})$ that

$$\limsup_{n \to \infty} \frac{1}{n} \ln \frac{1}{F_{X^n}^c} \leq \hat{E}(R|\mathbf{X})$$

so that

$$\liminf_{n \to \infty} \frac{1}{n} \ln F_{X^n}^c \geq -\hat{E}(R|\mathbf{X}).$$

The following proposition upper bounds the supremum coding exponent.

Proposition 9: For any rate $R > 0$

$$\hat{E}(R|\mathbf{X}) \leq \inf_{\mathbf{Y}: \underline{H}(\mathbf{Y}) > R} D_u(\mathbf{Y} \| \mathbf{X}). \qquad (34)$$

Proof: See Appendix C. ∎

Remark 4: When $R \geq \ln |\mathbb{X}|$, the probability of decoding error $\epsilon_n = 0$, so that $\hat{E}(R|\mathbf{X}) = +\infty$. The right-hand side is an infimum over an empty set and is $+\infty$ by convention, and the proposition holds for such R as well.

One can also show the alternative bound

$$\hat{E}(R|\mathbf{X}) \leq \inf_{\mathbf{Y}: \underline{R}(\mathbf{Y},\mathbf{X}) - D_u(\mathbf{Y}\|\mathbf{X}) > R} D_u(\mathbf{Y} \| \mathbf{X}). \qquad (35)$$

See the end of Appendix C on how to prove this. This result would be the functional inverse of Iriyama's [6, eq. (13)], while Proposition 9 is the functional inverse of Iriyama & Ihara's [13, eq. (12)]. Proposition 9, as we will soon see, provides a more natural extension of Arikan & Merhav's expression for $E(R, \rho)$ to general sources. ∎

2) Correct Decoding Exponent: We now study a generalization of the exponential rate for probability of correct decoding.

For a given (n, M_n, ϵ_n)-code, let

$$A_n := \{x^n \in \mathbb{X}^n : \psi_n(\phi_n(x^n)) = x^n\}$$

denote the set of correctly decoded sequences. For a given $\rho > 0$, R is (r, ρ)-admissible if for every $\eta > 0$ there exists a sequence of (n, M_n, ϵ_n)-codes such that

$$(1 + \rho) \liminf_{n \to \infty} \frac{1}{n} \ln \sum_{x^n \in A_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \geq r \qquad (36)$$

$$\limsup_{n \to \infty} \frac{1}{n} \ln M_n \leq R + \eta. \qquad (37)$$

Unlike the exponent for the probability of error, here r can be positive or negative. The *infimum fixed-length admissible rate* for a given r and $\rho > 0$ is

$$R^*(r, \rho|\mathbf{X}) = \inf\{R : R \text{ is } (r, \rho) \text{–admissible}\}.$$

It is easy to see that the set $\{R : R \text{ is } (r, \rho)\text{–admissible}\}$ is closed and so $R^*(r, \rho|\mathbf{X})$ is (r, ρ)-admissible.
The *supremum fixed-length coding exponent* for a given R and ρ is

$$E^*(R, \rho|\mathbf{X}) = \sup\{r : R \text{ is } (r, \rho)\text{–admissible}\}.$$

Remark 5: The choice of limit infimum in (36) makes the definition of admissibility pessimistic. For $\rho \downarrow 0$, the above definitions reduce to the special case of exponential rate for probability of correct decoding (see [12, Sec. 1.10]). ∎

Clearly, A_n should be $T_R(X^n)$ to maximize the left-hand side of (36), and hence

$$E^*(R, \rho|\mathbf{X}) = (1 + \rho) \liminf_{n \to \infty} \frac{1}{n} \ln \sum_{x^n \in T_R(X^n)} P_{X^n}^{\frac{1}{1+\rho}}(x^n).$$

The following proposition gives an expression for $E^*(R, \rho|\mathbf{X})$ and generalizes [6, Thm. 4] to any arbitrary $\rho > 0$. En route to its derivation, we find the expression for $R^*(r, \rho|\mathbf{X})$.

Proposition 10: For any $\rho > 0$, we have

$$R^*(r, \rho|\mathbf{X}) = \inf_{\mathbf{Y}: E_l(\mathbf{Y},\mathbf{X},\rho) \geq r} \bar{H}(\mathbf{Y}) \qquad (38)$$

$$E^*(R, \rho|\mathbf{X}) = \sup_{\mathbf{Y}: \bar{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho). \qquad (39)$$

Proof: See Appendix D. ∎

C. Summary of Bounds on E_u^s and E_l^s

We now combine Propositions 7–10 of the previous subsections to obtain the main result of the paper.

Theorem 11: For a given $\rho > 0$ and $R > 0$

$$
\max\left\{ \rho R - \inf_{\mathbf{Y}:\underline{H}(\mathbf{Y})>R} D_u(\mathbf{Y} \| \mathbf{X}), \right.
$$
$$
\left. \sup_{\mathbf{Y}:\bar{H}(\mathbf{Y})\leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho)\right\}
$$
$$
\leq E_l^s(R, \rho) \leq E_u^s(R, \rho)
$$
$$
\leq \min_{0\leq\theta\leq\rho}\left\{(\rho-\theta)R + \max_{\mathbf{Y}} E_u(\mathbf{Y}, \mathbf{X}, \theta)\right\}. \quad (40)
$$

Proof: The last inequality was proved in Proposition 8. Proposition 7 indicates that

$$
E_l^s(R, \rho)
$$
$$
\geq \max\left\{ \rho R + \liminf_{n\to\infty}\frac{1}{n}\ln F_{X^n}^c, \right.
$$
$$
\left. (1+\rho)\liminf_{n\to\infty}\frac{1}{n}\ln\sum_{x^n\in T_R(X^n)} P_{X^n}^{\frac{1}{1+\rho}}(x^n)\right\} \quad (41)
$$
$$
\geq \max\left\{\rho R - \hat{E}(R|\mathbf{X}), E^*(R, \rho|\mathbf{X})\right\}
$$
$$
\geq \max\left\{\rho R - \inf_{\mathbf{Y}:\underline{H}(\mathbf{Y})>R} D_u(\mathbf{Y} \| \mathbf{X}), \right.
$$
$$
\left. \sup_{\mathbf{Y}:\bar{H}(\mathbf{Y})\leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho)\right\} \quad (42)
$$

where (41) follows from the lower bound on $\hat{E}(R|\mathbf{X})$ and the definition of $E^*(R, \rho|\mathbf{X})$, and (42) from Propositions 9 and 10. ∎

IV. EXAMPLES

In this section, we evaluate the bounds for some examples where they are tight, and recover some known results.

Example 1 (Perfect Secrecy): First consider the perfect secrecy case, for example, $R \geq \ln|\mathbb{X}|$. Because of Remark 4 and because we may take $\theta = \rho$ in the upper bound in (40), the limiting exponential rate of guessing moments simplifies to

$$
\sup_{\mathbf{Y}} E_l(\mathbf{Y}, \mathbf{X}, \rho) \leq E_l^s(R, \rho)
$$
$$
\leq E_u^s(R, \rho) \leq \max_{\mathbf{Y}} E_u(\mathbf{Y}, \mathbf{X}, \rho).
$$

On account of (11) in Proposition 6, sup in the left-most term is achieved. From Proposition 6, upper and lower bounds are ρ times the liminf and limsup Rényi entropy rates of order $\frac{1}{1+\rho}$. In a related work, we proved in [8, Prop. 7], [18] that whenever the

information spectrum of the source satisfies the large deviation property with rate function I, the Rényi entropy rate converges and limiting guessing exponent equals the Legendre-Fenchel dual of the scaled rate function $I_1(t) := (1+\rho)I(t)$, i.e.,

$$
E_u^s(R, \rho) = E_l^s(R, \rho) = \sup_{t\in\mathbb{R}}\{\rho t - I_1(t)\}.
$$

In the next examples, we consider the case $R < \ln|\mathbb{X}|$.

Example 2 (An iid Source): This example was first studied by Merhav and Arikan [2]. Recall that an iid source is one for which $P_n(x^n) = \prod_{i=1}^n P_1(x_i)$, where P_1 denotes the marginal of X_1. We will now evaluate each term in (40).

We first argue that

$$
\inf_{\mathbf{Y}:\underline{H}(\mathbf{Y})>R} D_u(\mathbf{Y} \| \mathbf{X}) = \inf_{P_Y:H(P_Y)>R} D(P_Y \| P_1). \quad (43)
$$

To prove that the left-hand side in (43) is less than or equal to the right-hand side, let $P_Y \in \mathcal{M}(\mathbb{X})$ be such that $H(P_Y) > R$. Construct an iid source $\hat{\mathbf{Y}} = (P_{\hat{Y}^n} : n \in \mathbb{N})$ such that $P_{\hat{Y}_i} = P_Y$ for all $1 \leq i \leq n$. The i.i.d. property easily implies that

$$
D_u(\hat{\mathbf{Y}} \| \mathbf{X}) = D(P_Y \| P_1)
$$

and the law of large numbers for i.i.d. random variables yields

$$
\underline{H}(\hat{\mathbf{Y}}) = H(P_Y) > R. \quad (44)
$$

From (44), we have that the infimum on the left-hand side of (43) is over a larger set. We can therefore conclude that "$\leq$" holds in (43).

To prove "$\geq$" in (43) we use the result (see [12, Th. 1.7.2])

$$
\underline{H}(\mathbf{Y}) \leq H_l(\mathbf{Y}) := \liminf_{n\to\infty}\frac{1}{n}H(P_{Y^n})
$$

to get that the infimum over a larger set is smaller, i.e.,

$$
\inf_{\mathbf{Y}:\underline{H}(\mathbf{Y})>R} D_u(\mathbf{Y} \| \mathbf{X}) \geq \inf_{\mathbf{Y}:H_l(\mathbf{Y})>R} D_u(\mathbf{Y} \| \mathbf{X}). \quad (45)
$$

Because of (45), it is sufficient to prove

$$
\inf_{\mathbf{Y}:H_l(\mathbf{Y})>R} D_u(\mathbf{Y} \| \mathbf{X}) \geq \inf_{P_Y:H(P_Y)>R} D(P_Y \| P_1). \quad (46)
$$

Let $\mathbf{Y}$ be such that $H_l(\mathbf{Y}) > R$. Construct a source $\hat{\mathbf{Y}}$ such that, $P_{\hat{Y}_i} = P_{Y_i}$ for $1 \leq i \leq n$ and $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ are independent. Let $\mathbf{Z}$ be another source such that $Z_1, Z_2, \ldots, Z_n$ is an iid sequence with distribution

$$
P_{Z_j} = \frac{1}{n}\sum_{i=1}^n P_{Y_i}, \quad j = 1, 2, \ldots, n.
$$

As the marginals of Y^n and $\hat{Y}^n$ with independent components are the same, it easily follows from the formula for Kullback-Leibler divergence that

$$
\begin{aligned}
D(P_{Y^n} \parallel P_{X^n}) &= D(P_{Y^n} \parallel P_{\hat{Y}^n}) \\
&\quad + D(P_{\hat{Y}^n} \parallel P_{X^n}) \\
&\geq D(P_{\hat{Y}^n} \parallel P_{X^n}) \\
&= \sum_{i=1}^{n} D(P_{\hat{Y}_i} \parallel P_1) \\
&\geq n D(P_{Z_1} \parallel P_1), \quad (47)
\end{aligned}
$$

where (47) follows from the convexity of divergence. From the concavity of Shannon entropy, we also have

$$
H(P_{Y^n}) \leq \sum_{i=1}^{n} H(P_{Y_i}) \leq n H(P_{Z_1}). \quad (48)
$$

Normalize by n take limsup in (47) and liminf in (48) to get $D_u(\mathbf{Y} \parallel \mathbf{X}) \geq D(P_{Z_1} \parallel P_1)$ and $H(P_{Z_1}) > R$ for a P_{Z_1} that is a limit point of the sequence $(n^{-1} \sum_{i=1}^{n} P_{Y_i}, n \in \mathbb{N})$. From these, we conclude that (46) holds. This proves (43).

Following a similar procedure as above, we can bound the other terms in (40) for an i.i.d. source as

$$
\begin{aligned}
&\sup_{\mathbf{Y}: \bar{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho) \\
&\geq \sup_{P_Y: H(P_Y) \leq R} \{\rho H(P_Y) - D(P_Y \parallel P_1)\} \quad (49)
\end{aligned}
$$

and

$$
\sup_{\mathbf{Y}} E_u(\mathbf{Y}, \mathbf{X}, \theta) = \sup_{P_Y} \{\theta H(P_Y) - D(P_Y \parallel P_1)\}. \quad (50)
$$

Substitution of (43) and (49) in the lower bound of (40) yields

$$
\begin{aligned}
E_l^s(R, \rho) &\geq \max \Bigg\{ \rho R - \inf_{P_Y: H(P_Y) > R} D(P_Y \parallel P_1), \\
&\quad \sup_{P_Y: H(P_Y) \leq R} \{\rho H(P_Y) - D(P_Y \parallel P_1)\} \Bigg\} \\
&= \sup_{P_Y} \{\rho \min\{H(P_Y), R\} - D(P_Y \parallel P_1)\}. \quad (51)
\end{aligned}
$$

Similarly, substitution of (50) in the upper bound of (40) yields

$$
\begin{aligned}
&E_u^s(R, \rho) \\
&\leq \min_{0 \leq \theta \leq \rho} \Bigg\{ (\rho - \theta) R + \sup_{P_Y} \{\theta H(P_Y) \\
&\quad - D(P_Y \parallel P_1)\} \Bigg\} \\
&= \sup_{P_Y} \Bigg\{ \min_{0 \leq \theta \leq \rho} \{(\rho - \theta) R + \theta H(P_Y)\} \\
&\quad - D(P_Y \parallel P_1) \Bigg\} \quad (52) \\
&= \sup_{P_Y} \{\rho \min\{H(P_Y), R\} - D(P_Y \parallel P_1)\} \quad (53)
\end{aligned}
$$

where the interchange of sup and min in (52) holds because the function within braces is linear in θ and concave in P_Y. From

(51) and (53), we recover Merhav and Arikan's result (1) for an i.i.d. source [2, eq. (3)].

Example 3 (Markov Source): In this example we focus on an irreducible stationary Markov source taking values on $\mathbb{X}$ and having a transition probability matrix π.

Let $\mathcal{M}_s(\mathbb{X}^2)$ denote the set of *stationary* PMFs defined by

$$
\mathcal{M}_s(\mathbb{X}^2) = \Bigg\{ Q \in \mathcal{M}(\mathbb{X}^2) : \\
\sum_{x_1 \in \mathbb{X}} Q(x_1, x) = \sum_{x_2 \in \mathbb{X}} Q(x, x_2), \forall x \in \mathbb{X} \Bigg\}.
$$

Denote the common marginal by q and let

$$
\eta(\cdot | x_1) := \begin{cases} Q(x_1, \cdot)/q(x_1), & \text{if } q(x_1) \neq 0 \\ 1/|\mathbb{X}|, & \text{otherwise} \end{cases}.
$$

We may then denote $Q = q \times \eta$, where q is the distribution of X_1 and η the conditional distribution of X_2 given X_1. Following steps similar to the i.i.d. case, we have

$$
E_u^s = E_l^s = \sup_{Q \in \mathcal{M}_s(\mathbb{X}^2)} \{\rho \min\{H(\eta | q), R\} - D(\eta \parallel \pi | q)\}
$$

where

$$
H(\eta | q) := \sum_{x \in \mathbb{X}} q(x) H(\eta(\cdot | x))
$$

is the conditional one-step entropy, and

$$
D(\eta \parallel \pi | q) = \sum_{x_1 \in \mathbb{X}} q(x_1) D(\eta(\cdot | x_1) \parallel \pi(\cdot | x_1)).
$$

For a unifilar source the underlying state space forms a Markov chain and the entropy and divergence of the source equals those of the underlying Markov state space source [14, Thm. 6.4.2]. The arguments for the Markov source are now directly applicable to a unifilar source.

V. CONCLUSION

We saw the close connection between the problem of guessing a source realization given a cryptogram and the problem of compression with saturated exponential costs. The latter is a modification of a problem posed by Campbell [7]. Moreover, the exponents for both these problems coincide. This exponent is determined by the error exponent and a generalization of correct decoding exponent for fixed length block source codes.

We end this paper with some open questions.
- The equivalence between guessing and compression exploits the finite alphabet size assumption. Can this be relaxed?
- How do the results of this paper extend to the case with receiver side information? Can the result of Hayashi and Yamamoto be extended to general sources?
- If guessing to within a distortion is allowed, can the result of Merhav and Arikan [15] be extended to general sources? Both cases of perfect secrecy and key-rate constrained secrecy remain open.

APPENDIX A
PROOF OF PROPOSITION 3

Let P_n be any PMF on $\mathbb{X}^n$. Enumerate the elements of $\mathbb{X}^n$ from 1 to $|\mathbb{X}|^n$ in the decreasing order of their P_n-probabilities. Let $M = \exp\{nR\}$ denote the number of distinct key strings. For convenience, we shall assume that M is a power of 2 so that the number of key bits $k = nR/(\ln 2)$ is an integer. The general case will be easily handled towards the end of this section.

If M does not divide $|\mathbb{X}|^n$, append a few dummy messages of zero probability to make the number of messages N a multiple of M. Further, index the messages from 0 to $N-1$. Henceforth, we identify a message x^n by its index.

Divide the messages into groups of M so that message m belongs to group T_j, where $j = \lfloor m/M \rfloor$, and $\lfloor \cdot \rfloor$ is the floor function. Enumerate the key streams from 0 to $M-1$, so that $0 \leq u \leq M-1$. The function f_n is now defined as follows. For $m = jM + i$ set

$$f_n(jM + i, u) \triangleq jM + (i \oplus u)$$

where $i \oplus u$ is the bit-wise XOR operation. Thus messages in group T_j are encrypted to messages in the same group. The index i identifying the specific message in group T_j, i.e., the last $k = nR/(\ln 2)$ bits of m, are encrypted via bit-wise XOR with the key stream. Given u and the cryptogram, decryption is clear—perform bit-wise XOR with u on the last $nR/(\ln 2)$ bits of y.

Given a cryptogram y, the only information that the attacker gleans is that the message belongs to the group determined by y. Indeed, if $y \in T_j$, then

$$P_n\{Y = y\} = \frac{1}{M} P_n\{X^n \in T_j\}$$

and therefore

$$P_n\{X^n = m | Y = y\} = \begin{cases} \frac{P_n\{X^n = m\}}{P_n\{X^n \in T_j\}}, & \lfloor m/M \rfloor = j, \\ 0, & \text{otherwise} \end{cases}$$

which decreases with m for $m \in T_j$, because of our enumeration in the decreasing order of probabilities, and is 0 for $m \notin T_j$. The attacker's best strategy $G_{f_n}(\cdot | y)$ is therefore to restrict his guesses to T_j and guess in the order $jM, jM+1, \ldots, jM + M - 1$. Thus, when $x^n = jM + i$, the optimal attack strategy requires $i + 1$ guesses.

We now analyze the performance of this attack strategy as follows:

$$\mathbb{E}[G_{f_n}(X^n | Y)^\rho]$$

$$= \sum_{j=0}^{N/M-1} \sum_{i=0}^{M-1} P_n\{X^n = jM + i\}(i+1)^\rho$$

$$\geq \sum_{j=0}^{N/M-1} \sum_{i=0}^{M-1} P_n\{X^n = (j+1)M - 1\}(i+1)^\rho$$

$$\tag{54}$$

$$\geq \sum_{j=0}^{N/M-1} P_n\{X^n = (j+1)M - 1\}\frac{M^{1+\rho}}{1+\rho} \tag{55}$$

$$\geq \frac{1}{1+\rho} \sum_{j=0}^{N/M-1} \sum_{i=0}^{M-1} P_n\{X^n = (j+1)M + i\}M^\rho$$

$$\tag{56}$$

$$= \frac{1}{1+\rho} \sum_{m=M}^{N-1} P_n\{X^n = m\}M^\rho \tag{57}$$

where (54) follows because the arrangement in the decreasing order of probabilities implies that

$$P_n\{X^n = jM + i\} \geq P_n\{X^n = (j+1)M - 1\}$$

for $i = 0, \ldots, M - 1$. Inequality (55) follows because

$$\sum_{i=0}^{M-1} (i+1)^\rho = \sum_{i=1}^{M} i^\rho \geq \int_0^M z^\rho dz = \frac{M^{1+\rho}}{1+\rho}.$$

Inequality (56) follows because the decreasing probability arrangement implies

$$P_n\{X^n = (j+1)M - 1\}$$

$$\geq \frac{1}{M} \sum_{i=0}^{M-1} P_n\{X^n = (j+1)M + i\}.$$

Inequality (57) follows because we take $P_n(X^n = m) = 0$ for all the further dummy messages with indices $m > N$. Thus, (57) implies that

$$\sum_{m=0}^{N-1} P_n\{X^n = m\}(\min\{m+1, M\})^\rho$$

$$= \sum_{m=0}^{M-1} P_n\{X^n = m\}(m+1)^\rho$$

$$+ \sum_{m=M}^{N-1} P_n\{X^n = m\}M^\rho$$

$$\leq \mathbb{E}[G_{f_n}(X^n|Y)^\rho]$$

$$+ (1+\rho)\mathbb{E}[G_{f_n}(X^n|Y)^\rho]$$

$$= (2+\rho)\mathbb{E}[G_{f_n}(X^n|Y)^\rho]. \tag{58}$$

Let G be the guessing function that guesses in the decreasing order of P_n-probabilities without regard to Y, i.e., $G(m) = m+1$. Let L_G be the associated length function, given in Lemma 2. Now use (58) and Lemma 2 to get

$$\mathbb{E}[G_{f_n}(X^n|Y)^\rho]$$

$$\geq \frac{1}{2+\rho}\mathbb{E}[(\min\{G(X^n), M\})^\rho]$$

$$\geq \frac{1}{2+\rho}\mathbb{E}\left[\left(\min\left\{\frac{\exp_2\{L_G(X^n)\}}{2c_n}, M\right\}\right)^\rho\right]$$

$$\geq \frac{1}{(2c_n)^\rho(2+\rho)}\mathbb{E}[\exp\{\rho \min\{L_G(X^n)\ln 2, nR\}\}]$$

$$\tag{59}$$

where the last inequality follows by pulling out $2c_n$ and recognizing that $2c_n M \geq M \geq \exp\{nR\}$. Since G_{f_n} is the strategy that minimizes $\mathbb{E}[G(X^n|Y)^\rho]$, the proof is complete for the cases when $k = nR/(\ln 2)$ is an integer.

When $nR/(\ln 2)$ is not an integer, choose $k = \lceil nR/(\ln 2) \rceil$. Then $M = \exp_2\{k\} \geq \exp\{nR\}$, and it immediately follows that inequality (59) continues to hold. This completes the proof. ∎

APPENDIX B
PROOF OF PROPOSITION 6

We begin with the following lemma. Recall that $\mathcal{M}(\mathbb{X})$ is the set of all probability measures on $\mathbb{X}$ and $\mathcal{M}(B)$ the subset of $\mathcal{M}(\mathbb{X})$ with support set $B \subseteq \mathbb{X}$

$$\mathcal{M}(B) = \{\nu \in \mathcal{M}(\mathbb{X}) : \nu(B) = 1\}.$$

Lemma 12: For any $\rho > 0, \mu \in \mathcal{M}(\mathbb{X})$ and $B \subseteq \mathbb{X}$

$$(1 + \rho)\ln \sum_{x \in B} \mu^{\frac{1}{1+\rho}}(x) = \max_{\nu \in \mathcal{M}(B)} \{\rho H(\nu) - D(\nu \parallel \mu)\}.$$

Remark 6: [6, Lemma 1] is the special case when $\rho = 0$. □

Proof: Let $\mu_B(x) = \frac{\mu(x)}{\mu(B)} 1\{x \in B\}$. We then have

$$(1 + \rho)\ln \sum_{x \in B} \mu^{\frac{1}{1+\rho}}(x)$$
$$= (1 + \rho)\ln \sum_{x \in B} \mu_B^{\frac{1}{1+\rho}}(x) + \ln \mu(B)$$
$$= (1 + \rho)\max_{\nu \in \mathcal{M}(B)} \left\{ \sum_{x \in B} \frac{\rho}{1 + \rho}\nu(x)\ln \frac{1}{\mu_B(x)} \right.$$
$$\left. -D(\nu \parallel \mu_B) \right\} + \ln \mu(B) \tag{60}$$
$$= (1 + \rho)\max_{\nu \in \mathcal{M}(B)} \left\{ \frac{\rho}{1 + \rho}\{H(\nu) + D(\nu \parallel \mu)\} \right.$$
$$\left. -D(\nu \parallel \mu) \right\} \tag{61}$$
$$= \max_{\nu \in \mathcal{M}(B)} \{\rho H(\nu) - D(\nu \parallel \mu)\} \tag{62}$$

where (60) follows from the variational formula for Rényi entropy of μ_B. The maximum achieving distribution in (62) is $\mu^* \in \mathcal{M}(B)$ given by

$$\mu^*(x) = \frac{\mu^{\frac{1}{1+\rho}}(x)}{\sum_{y \in B} \mu^{\frac{1}{1+\rho}}(y)} 1\{x \in B\}$$

a fact that is easily verified via direct substitution. ∎

We now prove (11); proof of (10) is similar and therefore omitted. We begin by showing "≤" in (11). Let $\mathbf{X}^* = (P_{X^n}^* :$

$n \in \mathbb{N}) \in \mathcal{M}(\mathbf{B})$ be as defined in (12). It is straightforward to verify by direct substitution that

$$(1 + \rho)\ln \sum_{x^n \in B_n} P_{X^n}^{*\frac{1}{1+\rho}}(x^n) = \rho H(P_{X^n}^*) - D(P_{X^n}^* \parallel P_{X^n}).$$

Normalize by n and take limit infimum, and use the definition of $E_l(\mathbf{X}^*, \mathbf{X}, \rho)$ to get

$$(1 + \rho)\liminf_{n \to \infty} \frac{1}{n}\ln \sum_{x^n \in B_n} P_{X^n}^{*\frac{1}{1+\rho}}(x^n)$$
$$= E_l(\mathbf{X}^*, \mathbf{X}, \rho)$$
$$\leq \max_{\mathbf{Y} \in \mathcal{M}(\mathbf{B})} E_l(\mathbf{Y}, \mathbf{X}, \rho). \tag{63}$$

To prove "≥" in (11), let $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N}) \in \mathcal{M}(\mathbf{B})$ be an arbitrary sequence. We may assume that for all sufficiently large n, $P_{Y^n} \ll P_{X^n}$ holds; otherwise $E_l(\mathbf{Y}, \mathbf{X}, \rho) = -\infty$ and the inequality "≥" holds automatically. Define $\mathbf{Y}^* = (P_{Y^n}^* : n \in \mathbb{N}) \in \mathcal{M}(\mathbf{B})$ by

$$P_{Y^n}^*(y^n) = \frac{P_{Y^n}(y^n)}{P_{Y^n}(B_n)} 1\{y^n \in B_n\}.$$

It is clear that $P_{Y^n}^* \in \mathcal{M}(B_n)$ for every n. From Lemma 12, we have

$$(1 + \rho)\ln \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n)$$
$$= \max_{P_{Y^n} \in \mathcal{M}(B_n)} \{\rho H(P_{Y^n}) - D(P_{Y^n} \parallel P_{X^n})\}$$
$$\geq \rho H(P_{Y^n}^*) - D(P_{Y^n}^* \parallel P_{X^n}). \tag{64}$$

We now study each term on the right-hand side of (64). The entropy term is lower bounded as follows:

$$\rho H(P_{Y^n}^*)$$
$$= \frac{\rho}{P_{Y^n}(B_n)} \left\{ \sum_{x^n \in B_n} P_{Y^n}(x^n)\ln \frac{1}{P_{Y^n}(x^n)} \right\}$$
$$+ \rho \ln P_{Y^n}(B_n)$$
$$= \frac{\rho}{P_{Y^n}(B_n)} \left\{ H(P_{Y^n}) - \sum_{x^n \in B_n^c} P_{Y^n}(x^n)\ln \frac{1}{P_{Y^n}(x^n)} \right\}$$
$$+ \rho \ln P_{Y^n}(B_n)$$
$$= \frac{\rho}{P_{Y^n}(B_n)} \{H(P_{Y^n}) - P_{Y^n}(B_n^c) H(P_{Y^n}|B_n^c)$$
$$+ P_{Y^n}(B_n^c)\ln P_{Y^n}(B_n^c)\}$$
$$+ \rho \ln P_{Y^n}(B_n)$$
$$\geq \frac{\rho}{P_{Y^n}(B_n)} \{H(P_{Y^n}) - P_{Y^n}(B_n^c) n \ln |\mathbb{X}|$$
$$+ P_{Y^n}(B_n^c)\ln P_{Y^n}(B_n^c)\} + \rho \ln P_{Y^n}(B_n). \tag{65}$$

The divergence term is upper bounded, as in the proof of Iriyama's [6, Prop. 1], as follows:

$$
\begin{aligned}
&D(P_{Y^n}^* \| P_{X^n}) \\
&= -\ln P_{Y^n}(B_n) \\
&\quad + \frac{1}{P_{Y^n}(B_n)} \sum_{x^n \in B_n} P_{Y^n}(x^n) \ln \frac{P_{Y^n}(x^n)}{P_{X^n}(x^n)} \\
&= -\ln P_{Y^n}(B_n) + \frac{1}{P_{Y^n}(B_n)} D(P_{Y^n} \| P_{X^n}) \\
&\quad - \frac{1}{P_{Y^n}(B_n)} \sum_{x^n \in B_n^c} P_{Y^n}(x^n) \ln \frac{P_{Y^n}(x^n)}{P_{X^n}(x^n)} \\
&\leq -\ln P_{Y^n}(B_n) + \frac{1}{P_{Y^n}(B_n)} D(P_{Y^n} \| P_{X^n}) \\
&\quad - \frac{P_{Y^n}(B_n^c) - P_{X^n}(B_n^c)}{P_{Y^n}(B_n)} \quad (66) \\
&\leq -\ln P_{Y^n}(B_n) + \frac{1}{P_{Y^n}(B_n)} D(P_{Y^n} \| P_{X^n}) \\
&\quad + \frac{1}{P_{Y^n}(B_n)}. \quad (67)
\end{aligned}
$$

To get (66), we used the fact that $\ln x \geq 1 - \frac{1}{x}$ for all $x > 0$ and in inequality (67), we used the relation

$$
P_{Y^n}(B_n^c) - P_{X^n}(B_n^c) \geq -1.
$$

Substitution of (65) and (67) in (64) and the fact that $\lim_{n\to\infty} P_{Y^n}(B_n) = 1$ yield

$$
\begin{aligned}
&(1+\rho) \liminf_{n\to\infty} \frac{1}{n} \ln \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \\
&\geq \liminf_{n\to\infty} \frac{1}{n} \{ \rho H(P_{Y^n}) - D(P_{Y^n} \| P_{X^n}) - O(1) \} \\
&= E_l(\mathbf{Y}, \mathbf{X}, \rho).
\end{aligned}
$$

Since the choice of $\mathbf{Y} = (P_{Y^n} : n \in \mathbb{N}) \in \mathcal{M}(\mathbf{B})$ was arbitrary, we have proved "$\geq$" in (11).

From (63) and (11), the maximum is attained by $\mathbf{X}^*$, the distribution defined in (12). This completes the proof. ∎

APPENDIX C
PROOF OF PROPOSITION 9

Iriyama and Ihara showed the following lower bound on the infimum coding rate ([13, Th.3, (12)])

$$
\sup_{\mathbf{Y}:D_u(\mathbf{Y}\|\mathbf{X})<r} \underline{H}(\mathbf{Y}) \leq \hat{R}(r|\mathbf{X}). \quad (68)
$$

We claim that (68) is equivalent to (34). This proves the proposition.

We first show that (68) implies (34). Fix the source $\mathbf{X}$. Let R be a given rate. Consider an arbitrary candidate exponent r and an arbitrary source $\mathbf{Y}$. We argue that

$$
\begin{aligned}
R \text{ is } r\text{-achievable and } \underline{H}(\mathbf{Y}) > R \Longrightarrow r \\
\leq D_u(\mathbf{Y} \| \mathbf{X}). \quad (69)
\end{aligned}
$$

Taking the infimum on the right-hand side of (69) over $\mathbf{Y}$ with $\underline{H}(\mathbf{Y}) > R$, and then the supremum over r will yield (34).

To argue (69) by contraposition, we shall show that

$$
\begin{aligned}
&r > D_u(\mathbf{Y} \| \mathbf{X}) \\
&\quad \Longrightarrow \text{ either } R \text{ is not } r\text{-achievable or } \underline{H}(\mathbf{Y}) \leq R
\end{aligned}
$$

or equivalently, we shall show that

$$
\begin{aligned}
&r > D_u(\mathbf{Y} \| \mathbf{X}) \text{ and } \underline{H}(\mathbf{Y}) > R \\
&\quad \Longrightarrow R \text{ is not } r\text{-achievable.}
\end{aligned}
$$

But the conditions on the left-hand side imply

$$
\sup_{\mathbf{Y}:D_u(\mathbf{Y}\|\mathbf{X})<r} \underline{H}(\mathbf{Y}) > R
$$

which together with (68) yields $\hat{R}(r|\mathbf{X}) > R$, and this is the same as saying R is not r-achievable. This completes the proof of (68) $\Rightarrow$ (34). (This direction suffices to prove Proposition 9). The proof of the other direction is analogous. ∎

To prove the upper bound in (35), we begin with Iriyama's [6, eq. (13)], which is

$$
\sup_{\mathbf{Y}:D_u(\mathbf{Y}\|\mathbf{X})<r} \{ \underline{R}(\mathbf{Y}, \mathbf{X}) - D_u(\mathbf{Y} \| \mathbf{X}) \} \leq \hat{R}(r|\mathbf{X})
$$

instead of (68). The rest of the proof is completely analogous to the proof of Proposition 9.

APPENDIX D
PROOF OF PROPOSITION 10

We use the following notations in this proof. For each $\mathbf{B} = (B_n : n \in \mathbb{N})$ define

$$
|\mathbf{B}| := \limsup_{n\to\infty} \frac{1}{n} \ln |B_n|
$$
and
$$
S(\mathbf{Y}) := \left\{ \mathbf{B} : \lim_{n\to\infty} P_{Y^n}(B_n) = 1 \right\}.
$$

Note that $\mathbf{B} \in S(\mathbf{Y}) \Leftrightarrow \mathbf{Y} \in \mathcal{M}(\mathbf{B})$. We will first prove (38). Define a set

$$
\mathcal{B}(r, \rho|\mathbf{X}) = \left\{ \mathbf{B} := (B_n : n \in \mathbb{N}) : \right.
$$
$$
\left. (1+\rho) \liminf_{n\to\infty} \frac{1}{n} \ln \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n) \geq r \right\}. \quad (70)
$$

Then, by definition

$$
R^*(r, \rho|\mathbf{X}) = \inf\{|\mathbf{B}| : \mathbf{B} \in \mathcal{B}(r, \rho|\mathbf{X})\}. \quad (71)
$$

Fix a $\mathbf{B} \in \mathcal{B}(r, \rho|\mathbf{X})$. Proposition 6 then implies

$$
(1+\rho) \liminf_{n\to\infty} \frac{1}{n} \ln \sum_{x^n \in B_n} P_{X^n}^{\frac{1}{1+\rho}}(x^n)
$$
$$
= \max_{\mathbf{Y}:\mathbf{B}\in S(\mathbf{Y})} E_l(\mathbf{Y}, \mathbf{X}, \rho).
$$

We can therefore conclude using (70) that the following set equivalence holds:

$$\mathcal{B}(r, \rho|\mathbf{X}) = \bigcup_{E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r} S(\mathbf{Y}). \tag{72}$$

From (71) and (72), we get

$$R^*(r, \rho|\mathbf{X}) = \inf \left\{ |\mathbf{B}| : \mathbf{B} \in \bigcup_{E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r} S(\mathbf{Y}) \right\}$$
$$= \inf_{\mathbf{Y}} \{|\mathbf{B}| : E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r, \mathbf{B} \in S(\mathbf{Y})\}$$
$$= \inf_{\mathbf{Y}: E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r} \bar{H}(\mathbf{Y})$$

where last equality follows because

$$\bar{H}(\mathbf{Y}) = \inf\{|\mathbf{B}| : \mathbf{B} \in S(\mathbf{Y})\}$$

as proved by Han and Verdú [16]. This proves (38).

We now prove (39). We first show that if R is (r, ρ)-admissible then $r \leq \sup_{\bar{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho)$.

Since R is (r, ρ)-admissible, the definition of $R^*(r, \rho|\mathbf{X})$ and (38) imply that

$$R \geq R^*(r, \rho|\mathbf{X}) = \inf_{\mathbf{Y}: E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r} \bar{H}(\mathbf{Y})$$

which is the same as saying the following:

(A): For all $\eta > 0$ there exists a $\hat{\mathbf{Y}}(\eta)$ such that $E_l(\hat{\mathbf{Y}}(\eta), \mathbf{X}, \rho) \geq r$ and $\bar{H}(\hat{\mathbf{Y}}(\eta)) < R + \eta$. $\square$

Lemma 13 below implies that there exists a $\hat{\mathbf{Y}}$ that satisfies $E_l(\hat{\mathbf{Y}}, \mathbf{X}, \rho) \geq r$ and $\bar{H}(\hat{\mathbf{Y}}) \leq R$. It follows that

$$r \leq \sup_{\bar{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho)$$

and the converse part is proved.

For the direct part, fix any R, ρ, and set

$$r := \sup_{\bar{H}(\mathbf{Y}) \leq R} E_l(\mathbf{Y}, \mathbf{X}, \rho).$$

We should show that R is (r, ρ)-admissible.

With r set as above, we have

(B): For all $\eta > 0$, there exists a $\hat{\mathbf{Y}}(\eta)$ such that $E_l(\hat{\mathbf{Y}}(\eta), \mathbf{X}, \rho) > r - \eta$ and $\bar{H}(\hat{\mathbf{Y}}(\eta)) \leq R$. $\square$

Lemma 13 below implies that there exists a $\hat{\mathbf{Y}}$ that satisfies $E_l(\hat{\mathbf{Y}}, \mathbf{X}, \rho) \geq r$ and $\bar{H}(\hat{\mathbf{Y}}) \leq R$. It follows that

$$R \geq \inf_{E_l(\mathbf{Y}, \mathbf{X}, \rho) \geq r} \bar{H}(\mathbf{Y}) = R^*(r, \rho|\mathbf{X}),$$

i.e., R is (r, ρ)-admissible. This completes the proof. ∎

Lemma 13: If either **(A)** or **(B)** holds, then there is a $\hat{\mathbf{Y}}$ that satisfies $E_l(\hat{\mathbf{Y}}, \mathbf{X}, \rho) \geq r$ and $\bar{H}(\hat{\mathbf{Y}}) \leq R$.

Proof: We will argue that condition **(A)** implies the existence of an appropriate $\hat{\mathbf{Y}}$. The other argument is analogous and omitted. Our argument is similar to a diagonal argument.

Let **(A)** hold. Then, for each $\eta > 0$, we have a $\hat{\mathbf{Y}}(\eta)$ that satisfies (a) $E_l(\hat{\mathbf{Y}}(\eta), \mathbf{X}, \rho) \geq r$, which is the same as saying

$$\liminf_{n \to \infty} \frac{1}{n} \left(\rho H \left(P_{\hat{Y}^n(\eta)} \right) - D \left(P_{\hat{Y}^n(\eta)} \| P_{X^n} \right) \right) \geq r \tag{73}$$

and (b) $\bar{H}(\hat{\mathbf{Y}}(\eta)) < R + \eta$, which implies

$$\lim_{n \to \infty} \Pr \left\{ \frac{1}{n} \ln \frac{1}{P_{\hat{Y}^n(\eta)}(\hat{Y}^n(\eta))} > R + \eta \right\} = 0. \tag{74}$$

Let $\eta = 1/m$ and consider $m \in \mathbb{N}$. Let us now find an subsequence $\{n_m\}_{m \geq 0}$ in order to define the desired $\hat{\mathbf{Y}}$. Let $n_0 = 0$. Recursively, for each $m \geq 1$, we have the following.

i) From (73) with $\eta = 1/m$, we get that there is an N_m such that for any $n \geq N_m$,

$$\frac{1}{n} \left(\rho H \left(P_{\hat{Y}^n(\frac{1}{m})} \right) - D \left(P_{\hat{Y}^n(\frac{1}{m})} \| P_{X^n} \right) \right) > r - \frac{1}{m}. \tag{75}$$

ii) From (74) with $\eta = 1/m$, we get that there is an N'_m such that for any $n \geq N'_m$,

$$\Pr \left\{ \frac{1}{n} \ln \frac{1}{P_{\hat{Y}^n(\frac{1}{m})} \left(\hat{Y}^n\left(\frac{1}{m}\right) \right)} > R + \frac{1}{m} \right\} < \frac{1}{m}. \tag{76}$$

By taking an $n_m > \max\{N_m, N'_m, n_{m-1}\}$, it follows that we can find an $n_m > n_{m-1}$ such that (75) and (76) hold for all $n \geq n_m$.

Define $\hat{\mathbf{Y}} = (P_{\hat{Y}^n} : n \in \mathbb{N})$ by setting

$$\hat{Y}^n \stackrel{d}{=} \hat{Y}^n(1/m), \quad n_{m-1} < n \leq n_m. \tag{77}$$

We now show that this $\hat{\mathbf{Y}}$ has the required properties.

We first show $E_l(\hat{\mathbf{Y}}, \mathbf{X}, \rho) \geq r$. Take any arbitrary $\delta > 0$, fix $m > 1/\delta$, and consider n_m. For any $n \geq n_m$, there is an $m' \geq m$ such that $n_{m'-1} < n \leq n_{m'}$. Use the definition in (77) and the fact that (75) holds for all $n \geq n_m$ (with $\hat{Y}^n \stackrel{d}{=} \hat{Y}^n(\frac{1}{m'})$), to get

$$\frac{1}{n}(\rho H(P_{\hat{Y}^n}) - D(P_{\hat{Y}^n} \| P_{X^n})) > \quad r - \frac{1}{m'}$$
$$\geq \quad r - \frac{1}{m}$$
$$> \quad r - \delta.$$

We can thus conclude that the inequality

$$\frac{1}{n}(\rho H(P_{\hat{Y}^n}) - D(P_{\hat{Y}^n} \| P_{X^n})) > r - \delta$$

holds for all $n \geq n_m$. Since $\delta > 0$ was arbitrary, this establishes that $E_l(\hat{\mathbf{Y}}, \mathbf{X}, \rho) \geq r$.

We next show $\bar{\bar{H}}(\hat{\mathbf{Y}}) \leq R$. Using the definition of $\bar{H}(\hat{\mathbf{Y}})$, it suffices to show that for every $\delta > 0$, the following limit holds:

$$\lim_{n \to \infty} P_{\hat{Y}^n} \left(\frac{1}{n} \ln \frac{1}{P_{\hat{Y}^n}(\hat{Y}^n)} > R + \delta \right) = 0. \tag{78}$$

Take any arbitrary $\varepsilon > 0$, fix an $m > 1/\min\{\varepsilon, \delta\}$, and consider $n \geq n_m$. There is an $m' \geq m$ for which $n_{m'-1} \leq n \leq n_{m'}$. Use the definition in (77) and the fact that (76) holds to get

$$
\begin{aligned}
P_{\hat{Y}^n} & \left(\frac{1}{n} \ln \frac{1}{P_{\hat{Y}^n}(\hat{Y}^n)} > R + \delta \right) \\
& \leq P_{\hat{Y}^n} \left(\frac{1}{n} \ln \frac{1}{P_{\hat{Y}^n}(\hat{Y}^n)} > R + \frac{1}{m'} \right) \\
& < \frac{1}{m'} \\
& \leq \frac{1}{m} \\
& < \varepsilon,
\end{aligned}
$$

for any $n \geq n_m$. Consequently (78) holds, and thence $\bar{H}(\hat{Y}) \leq R$.

We have thus shown the existence of an appropriate $\hat{Y}$ if **(A)** holds. A similar argument holds under condition **(B)**. ∎

REFERENCES

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell Syst. Tech. J.*, vol. 28, no. 3, pp. 565–715, Oct. 1949.

[2] N. Merhav and E. Arikan, "The Shannon cipher system with a guessing wiretapper," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 1860–1866, Sep. 1999.

[3] R. Sundaresan, "Guessing based on length functions," in *Proce. Conf. Managing Complexity in a Distributed World (MCDES)*, Bangalore, India, May 2008.

[4] Y. Hayashi and H. Yamamoto, "Coding theorems for the Shannon cipher system with a guessing wiretapper and correlated source outputs," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2808–2817, Jun. 2008.

[5] T. S. Han, "The reliability functions of the general source with fixed-length coding," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2117–2132, Sep. 2000.

[6] K. Iriyama, "Probability of error for the fixed-length source coding of general sources," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1537–1543, May 2001.

[7] L. L. Campbell, "A coding theorem and Rényi's entropy," *Inform. and Control*, vol. 8, pp. 423–429, 1965.

[8] M. K. Hanawal and R. Sundaresan, "Guessing revisited: A large deviations approach," *IEEE Trans. Inf. Theory*, to be published.

[9] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. New York: Wiley, 1997.

[10] D. Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*. New York: Wiley, 1954.

[11] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*, 2nd ed. New York: Springer-Verlag, 1998.

[12] T. S. Han, *Information-Spectrum Methods in Information Theory*. : Springer-Verlag, 2003.

[13] K. Iriyama and S. Ihara, "The error exponent and minimum achievable rates for the fixed-length coding of general sources," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E84-A, no. 10, pp. 2466–2473, Oct. 2001.

[14] R. Ash, *Information Theory*. New York: Interscience Publishers, 1965.

[15] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1041–1056, May 1998.

[16] T. S. Han and S. Verdú, "Approximation theory of of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[17] DRDO-IISc Programme in Mathematical Engineering Tech. Rep. TR-PME-2007-02, Feb. 2007 [Online]. Available: http://pal.ece.iisc.ernet.in/PAM/tech_rep07/TR-PME-2007-02.pdf

[18] DRDO-IISc Programme in Mathematical Engineering Tech. Rep. TR-PME-2008-08, Dec. 2008 [Online]. Available: http://pal.ece.iisc.ernet.in/PAM/tech_rep08/TR-PME-2008-08.pdf

Manjesh Kumar Hanawal received the B.E. degree in electronics and communication from the National Institute of Technology, Bhopal, India, and the M.Sc. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 2009. He is currently pursuing the Ph.D. degree at INRIA, Sophia Antipolis, France, and the University of Avignon, Avignon, France.

From 2004 to 2007, he was with the Centre for Artificial Intelligence and Robotics (CAIR), Bangalore, as Scientist-B on various cryptographic projects.

Rajesh Sundaresan (S'96–M'00–SM'06) received the B.Tech. degree in electronics and communication from the Indian Institute of Technology, Madras, India, and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1996 and 1999, respectively.

From 1999 to 2005, he was with Qualcomm Inc., Campbell, CA, working on the design of communication algorithms for WCDMA and HSDPA modems. Since 2005, he has been with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore. His research interests are in the areas of wireless communication networks and information theory.