

# Information Masking and Amplification: The Source Coding Setting

Thomas A. Courtade  
Department of Electrical Engineering  
University of California, Los Angeles  
Email: tacourta@ee.ucla.edu

**Abstract**—The complementary problems of masking and amplifying channel state information in the Gel’fand-Pinsker channel have recently been solved by Merhav and Shamai, and Kim et al., respectively. In this paper, we study a related source coding problem. Specifically, we consider the two-encoder source coding setting where one source is to be amplified, while the other source is to be masked. In general, there is a tension between these two objectives which is characterized by the amplification-masking tradeoff. In this paper, we give a single-letter description of this tradeoff.

We apply this result, together with a recent theorem by Courtade and Weissman on multiterminal source coding, to solve a fundamental entropy characterization problem.

## I. INTRODUCTION

The well known source coding with side information problem has an achievable rate region given by

$$R_x \geq H(X|U), \quad R_y \geq I(Y;U)$$

as originally shown by Ahlswede and Körner [1], and independently by Wyner [2]. In this setting, the side information encoder merely serves as a helper with the sole purpose of aiding in the recovery of  $X^n$  at the decoder. However, for given rates  $(R_x, R_y)$ , there may be many different coding schemes which permit recovery of  $X^n$  at the decoder. In some cases, it may be desirable to select a coding scheme that reveals very little information about the side information  $Y^n$  to the decoder. We refer to this objective as *masking* the side information.

To motivate this setting, consider the following example. Suppose  $X$  is an attribute of an online customer that an advertiser would like to specifically target (e.g., gender), and  $Y$  is other detailed information about the same customer (e.g., credit history). Companies A and B separately have databases  $X^n$  and  $Y^n$  corresponding to  $n$  different customers (the databases could be indexed by IP address, for example). The advertiser pays Companies A and B to learn as much about the database  $X^n$  as possible. Now, suppose governing laws prohibit the database  $Y^n$  from being revealed too extensively. In this case, the material given to the advertiser must be chosen so that at most a prescribed amount of information is revealed about  $Y^n$ .

In general, a masking constraint on  $Y^n$  may render near-lossless reconstruction of  $X^n$  impossible. This motivates the study the *amplification-masking tradeoff*. That is, the tradeoff

between amplifying (or revealing) information about  $X^n$  while simultaneously masking the side information  $Y^n$ .

Similar problems have been previously considered in the information theory literature on secrecy and privacy. For example, Sankar et al. determine the utility-privacy tradeoff for the case of a single encoder in [3]. In their setting, the random variable  $X$  is a vector with a given set of coordinates that should be masked and another set that should be revealed (up to a prescribed distortion). In this context, our study of the amplification-masking tradeoff is a distributed version of [3], in which utility is measured by the information revealed about the database  $X^n$ . The problem we consider is distinct from those typically studied in the information-theoretic secrecy literature, in that the masking (i.e., equivocation) constraint corresponds to the intended decoder, rather than an eavesdropper.

We remark that the present paper is inspired in part by the recent, complementary works [4] and [5] which respectively study amplification and masking of channel state information. We borrow our terminology from those works.

This paper is organized as follows. Section II formally defines the problems considered and delivers our main results. The corresponding proofs are given in Section III. Final remarks and directions for future work are discussed in Section IV.

## II. PROBLEM STATEMENT AND RESULTS

Throughout this paper we adopt notational conventions that are standard in the literature. Specifically, random variables are denoted by capital letters (e.g.,  $X$ ) and their corresponding alphabets are denoted by corresponding calligraphic letters (e.g.,  $\mathcal{X}$ ). We abbreviate a sequence  $(X_1, \dots, X_n)$  of  $n$  random variables by  $X^n$ , and we let  $\delta(\epsilon)$  represent a quantity satisfying  $\lim_{\epsilon \rightarrow 0} \delta(\epsilon) = 0$ . Other notation will be introduced where necessary.

For a joint distribution  $p(x, y)$  on finite alphabets  $\mathcal{X} \times \mathcal{Y}$ , consider the source coding setting where separate Encoders 1 and 2 have access to the sequences  $X^n$  and  $Y^n$ , respectively. We make the standard assumption that the sequences  $(X^n, Y^n)$  are drawn i.i.d. according to  $p(x, y)$  (i.e.,  $X^n, Y^n \sim \prod_{i=1}^n p(x_i, y_i)$ ), and  $n$  can be taken arbitrarily large.

The first of the following three subsections characterizes the amplification-masking tradeoff. This result is applied to solve a fundamental entropy characterization in the second

subsection. The final subsection comments on the connection between information amplification and list decoding. Proofs of the main results are postponed until Section III.

#### A. The Amplification-Masking Tradeoff

Formally, a  $(2^{nR_x}, 2^{nR_y}, n)$  code is defined by its encoding functions

$$f_x : \mathcal{X}^n \rightarrow \{1, \dots, 2^{nR_x}\} \text{ and } f_y : \mathcal{Y}^n \rightarrow \{1, \dots, 2^{nR_y}\}.$$

A rate-amplification-masking tuple  $(R_x, R_y, \Delta_A, \Delta_M)$  is achievable if, for any  $\epsilon > 0$ , there exists a  $(2^{nR_x}, 2^{nR_y}, n)$  code satisfying the amplification criterion:

$$\Delta_A \leq \frac{1}{n} I(X^n; f_x(X^n), f_y(Y^n)) + \epsilon, \quad (1)$$

and the masking criterion:

$$\Delta_M \geq \frac{1}{n} I(Y^n; f_x(X^n), f_y(Y^n)) - \epsilon. \quad (2)$$

Thus, we see that the amplification-masking problem is an entropy characterization problem similar to that considered in [6, Chapter 15].

*Definition 1:* The achievable amplification-masking region  $\mathcal{R}_{AM}$  is the closure of the set of all achievable rate-amplification-masking tuples  $(R_x, R_y, \Delta_A, \Delta_M)$ .

*Theorem 1:*  $\mathcal{R}_{AM}$  consists of the rate-amplification-masking tuples  $(R_x, R_y, \Delta_A, \Delta_M)$  satisfying

$$\left. \begin{aligned} R_x &\geq \Delta_A - I(X; U) \\ R_y &\geq I(Y; U) \\ \Delta_M &\geq \max\{I(Y; U, X) + \Delta_A - H(X), I(Y; U)\} \\ \Delta_A &\leq H(X). \end{aligned} \right\} \quad (3)$$

for some joint distribution  $p(x, y, u) = p(x, y)p(u|y)$ , where  $|\mathcal{U}| \leq |\mathcal{Y}| + 1$ .

Observe that  $\mathcal{R}_{AM}$  characterizes the entire tradeoff between amplifying  $X^n$  and masking  $Y^n$ . We remark that maximum amplification  $\Delta_A = H(X)$  does not necessarily imply that  $X^n$  can be recovered near-losslessly at the encoder. However, if an application demands near lossless reproduction of the sequence  $X^n$ , Theorem 1 can be strengthened to include this case. To this end, define a rate-masking triple  $(R_x, R_y, \Delta_M)$  to be achievable if, for any  $\epsilon > 0$ , there exists a  $(2^{nR_x}, 2^{nR_y}, n)$  code satisfying the masking criterion (2), and a decoding function

$$\hat{X}^n : \{1, 2, \dots, 2^{nR_x}\} \times \{1, 2, \dots, 2^{nR_y}\} \rightarrow \mathcal{X}^n$$

which satisfies the decoding-error criterion

$$\Pr[X^n \neq \hat{X}^n(f_x(X^n), f_y(Y^n))] \leq \epsilon.$$

*Definition 2:* The achievable rate-masking region  $\mathcal{R}_M$  is the closure of the set of all achievable rate-masking triples  $(R_x, R_y, \Delta_M)$ .

*Corollary 1:*  $\mathcal{R}_M$  consists of the rate-masking triples  $(R_x, R_y, \Delta_M)$  satisfying

$$\begin{aligned} R_x &\geq H(X|U) \\ R_y &\geq I(Y; U) \\ \Delta_M &\geq I(Y; X, U) \end{aligned}$$

for some joint distribution  $p(x, y, u) = p(x, y)p(u|y)$ , where  $|\mathcal{U}| \leq |\mathcal{Y}| + 1$ .

#### B. An Entropy Characterization Result

As we previously noted, the amplification-masking trade-off solves a multi-letter entropy characterization problem by reducing it to single-letter form. The reader is directed to [6] for an introduction to entropy characterization problems. Here, we apply our results to yield a fundamental characterization of the information revealed about  $X^n$  and  $Y^n$ , respectively, by arbitrary encoding functions  $f_x$  and  $f_y$  (of rates  $R_x, R_y$ ).

*Definition 3:* Define the region  $\mathcal{R}^*(R_x, R_y)$  as follows. The pair  $(\Delta_X, \Delta_Y) \in \mathcal{R}^*(R_x, R_y)$  if and only if, for any  $\epsilon > 0$ , there exists a  $(2^{nR_x}, 2^{nR_y}, n)$  code satisfying

$$\begin{aligned} \left| \Delta_X - \frac{1}{n} I(X^n; f_x(X^n), f_y(Y^n)) \right| &\leq \epsilon, \text{ and} \\ \left| \Delta_Y - \frac{1}{n} I(Y^n; f_x(X^n), f_y(Y^n)) \right| &\leq \epsilon. \end{aligned}$$

Let  $\overline{\mathcal{R}^*}(R_x, R_y)$  be the closure of  $\mathcal{R}^*(R_x, R_y)$ .

Ultimately we obtain a single-letter description of  $\overline{\mathcal{R}^*}(R_x, R_y)$ . However, in order to do so, we require some notation. To this end, let:

$$\mathcal{R}_{AM}(R_x, R_y) = \{(\Delta_X, \Delta_Y) : (R_x, R_y, \Delta_X, \Delta_Y) \in \mathcal{R}_{AM}\}.$$

Symmetrically, let  $\mathcal{R}_{MA}$  be the region where  $X^n$  is subject to masking  $\Delta_X$  and  $Y^n$  is subject to amplification  $\Delta_Y$ . Let

$$\mathcal{R}_{MA}(R_x, R_y) = \{(\Delta_X, \Delta_Y) : (R_x, R_y, \Delta_X, \Delta_Y) \in \mathcal{R}_{MA}\}.$$

Finally, let  $\mathcal{R}_{AA}(R_x, R_y)$  consist of all pairs  $(\Delta_X, \Delta_Y)$  satisfying

$$\begin{aligned} R_x &\geq I(U_x; X|U_y, Q) \\ R_y &\geq I(U_y; Y|U_x, Q) \\ R_x + R_y &\geq I(U_x, U_y; X, Y|Q) \\ \Delta_X &\leq I(X; U_x, U_y|Q) \\ \Delta_Y &\leq I(Y; U_x, U_y|Q) \end{aligned}$$

for some joint distribution of the form

$$p(x, y, u_x, u_y, q) = p(x, y)p(u_x|x, q)p(u_y|y, q)p(q),$$

where  $|\mathcal{U}_x| \leq |\mathcal{X}|$ ,  $|\mathcal{U}_y| \leq |\mathcal{Y}|$ , and  $|\mathcal{Q}| \leq 5$ .

*Theorem 2:* The region  $\overline{\mathcal{R}^*}(R_x, R_y)$  has a single-letter characterization given by

$$\begin{aligned} \overline{\mathcal{R}^*}(R_x, R_y) &= \\ &\mathcal{R}_{AM}(R_x, R_y) \cap \mathcal{R}_{MA}(R_x, R_y) \cap \mathcal{R}_{AA}(R_x, R_y). \end{aligned}$$

Moreover, restriction of the encoding functions to vector-quantization and/or random binning is sufficient to achieve any point in  $\overline{\mathcal{R}^*}(R_x, R_y)$ .

The second statement of Theorem 2 is notable since it states that relatively simple encoding functions (i.e., vector quantization and/or binning) can asymptotically reveal the same amount of information about  $X^n$  and  $Y^n$ , respectively, as encoding functions that are only restricted in rate. In

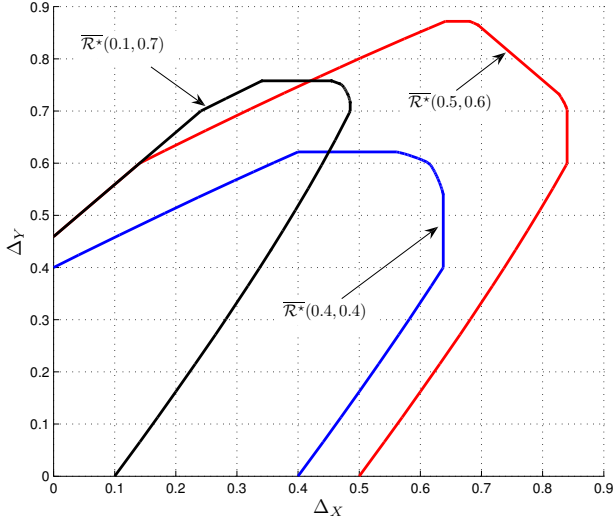


Fig. 1. The region  $\overline{\mathcal{R}}^*(R_x, R_y)$  for joint distribution  $P_{X,Y}$  given by (4) and three different pairs of rates. Rate pairs  $(R_x, R_y)$  equal to (0.1, 0.7), (0.4, 0.4), and (0.5, 0.6) define the convex regions bounded by the black, blue, and red curves, respectively.

contrast, this is not true for the setting of three or more sources, as the modulo-sum problem studied by Körner and Marton [7] provides a counterexample where the Berger-Tung achievability scheme [8] is not optimal. Thus, obtaining a characterization like Theorem 2 for three or more sources represents a formidable challenge.

We remark that the points in  $\overline{\mathcal{R}}^*(R_x, R_y)$  with  $\Delta_X = H(X)$  and/or  $\Delta_Y = H(Y)$  also capture the more stringent constraint(s) of near-lossless reproduction of  $X^n$  and/or  $Y^n$ , respectively. This is a consequence of Corollary 1.

To give a concrete example of  $\overline{\mathcal{R}}^*(R_x, R_y)$ , consider the following joint distribution:

$$\begin{array}{c|cc} P_{X,Y}(x, y) & x = 0 & x = 1 \\ \hline y = 0 & 1/3 & 0 \\ y = 1 & 1/6 & 1/2. \end{array} \quad (4)$$

By performing a brute-force search over the auxiliary random variables defining  $\overline{\mathcal{R}}^*(R_x, R_y)$  for the distribution  $P_{X,Y}$ , we have obtained numerical approximations of  $\overline{\mathcal{R}}^*(\cdot, \cdot)$  for several different pairs of  $(R_x, R_y)$ . The results are given in Figure 1.

### C. Connection to List Decoding

We briefly comment on the connection between an amplification constraint and list decoding. As discussed in detail in [4], the amplification criterion (1) is essentially equivalent to the requirement for a list decoder

$$L_n : \{1, \dots, 2^{nR_x}\} \times \{1, \dots, 2^{nR_y}\} \rightarrow 2^{\mathcal{X}^n}$$

with list size and probability of error respectively satisfying

$$\begin{aligned} \log |L_n| &\leq n(H(X) - \Delta_A + \epsilon), \text{ and} \\ \Pr[X^n \notin L_n(f_x(X^n), f_y(Y^n))] &\leq \epsilon. \end{aligned}$$

Thus maximizing the amplification of  $X^n$  subject to given rate and masking constraints can be thought of as characterizing the best list decoder in that setting.

## III. PROOFS OF MAIN RESULTS

*Proof of Theorem 1: Converse Part:* Suppose  $(R_x, R_y, \Delta_A, \Delta_M)$  is achievable. For convenience, define  $F_x = f_x(X^n)$ ,  $F_y = f_y(Y^n)$ , and  $U_i = (F_y, Y^{i-1})$ .

First, note that  $\Delta_A \leq H(X)$  is trivially satisfied. Next, the constraint on  $R_x$  is given by:

$$\begin{aligned} nR_x &\geq H(F_x) \geq H(F_x|F_y) \\ &= \sum_{i=1}^n H(X_i|F_y, X^{i-1}) - H(X^n|F_x, F_y) \\ &\geq \sum_{i=1}^n H(X_i|F_y, Y^{i-1}, X^{i-1}) - H(X^n|F_x, F_y) \\ &= I(X^n; F_x, F_y) - \sum_{i=1}^n I(X_i; U_i) \end{aligned} \quad (5)$$

$$\geq n(\Delta_A - \epsilon) - \sum_{i=1}^n I(X_i; U_i). \quad (6)$$

Equality (5) follows since  $X_i \leftrightarrow F_y, Y^{i-1} \leftrightarrow X^{i-1}$  form a Markov chain, and inequality (6) follows since amplification  $\Delta_A$  is achievable.

The constraint on  $R_y$  is trivial:

$$\begin{aligned} nR_y &\geq H(F_y) \geq I(F_y; Y^n) = \sum_{i=1}^n I(Y_i; F_y|Y^{i-1}) \\ &= \sum_{i=1}^n I(Y_i; F_y, Y^{i-1}) = \sum_{i=1}^n I(Y_i; U_i). \end{aligned}$$

Similarly, we obtain the first lower bound on  $\Delta_M$ :

$$n(\Delta_M + \epsilon) \geq I(Y^n; F_x, F_y) \geq I(Y^n; F_y) = \sum_{i=1}^n I(Y_i; U_i).$$

The second lower bound on  $\Delta_M$  requires slightly more work, and can be derived as follows:

$$\begin{aligned} n(\Delta_M + \epsilon) &\geq I(Y^n; F_x, F_y) \\ &= I(Y^n; X^n, F_y) + I(X^n; F_x, F_y) - I(X^n; F_x, Y^n) \\ &\geq I(Y^n; X^n, F_y) + n\Delta_A - I(X^n; F_x, Y^n) - n\epsilon \\ &\geq \sum_{i=1}^n I(Y_i; X^n, F_y|Y^{i-1}) + n\Delta_A - H(X^n) - n\epsilon \\ &\geq \sum_{i=1}^n I(Y_i; X_i, U_i) + \Delta_A - H(X_i) - \epsilon, \end{aligned} \quad (7)$$

where (7) follows since amplification  $\Delta_A$  is achievable.

Observing that the Markov condition  $U_i \leftrightarrow Y_i \leftrightarrow X_i$  is satisfied for each  $i$ , a standard timesharing argument proves the existence of a random variable  $U$  such that  $U \leftrightarrow Y \leftrightarrow X$  forms a Markov chain and (3) is satisfied.

*Direct Part:* Fix  $p(u|y)$  and suppose  $(R_x, R_y, \Delta_A, \Delta_M)$  satisfy (3) with strict inequality. Next, fix  $\epsilon > 0$  sufficiently small so that it is less than the minimum slack in said inequalities, and set  $\tilde{R} = I(Y; U) + \epsilon$ . Our achievability scheme uses a standard random coding argument which we sketch below.

**Codebook generation.** Randomly and independently, bin the typical  $x^n$ 's uniformly into  $2^{n(\Delta_A - I(X;U) + \epsilon)}$  bins. Let  $b(x^n)$  be the index of the bin which contains  $x^n$ . For  $l \in \{1, \dots, 2^{n\tilde{R}}\}$ , randomly and independently generate  $u^n(l)$ , each according to  $\prod_{i=1}^n p_U(u_i)$ .

**Encoding.** Encoder 1, upon observing the sequence  $X^n$ , sends the corresponding bin index  $b(X^n)$  to the decoder. If  $X^n$  is not typical, an error is declared. Encoder 2, upon observing the sequence  $Y^n$ , finds an  $L \in \{1, \dots, 2^{n\tilde{R}}\}$  such that  $(Y^n, U^n(L))$  are jointly  $\epsilon$ -typical, and sends the unique index  $L$  to the decoder. If more than one such  $L$  exists, ties are broken arbitrarily. If no such  $L$  exists, then an error is declared.

This coding scheme clearly satisfies the given rates. Further, each encoder errs with arbitrarily small probability as  $n \rightarrow \infty$ . Hence, we only need to check that the amplification and masking constraints are satisfied. To this end, let  $\mathcal{C}$  be the random codebook. We first check that the amplification and masking constraints are separately satisfied when averaged over random codebooks  $\mathcal{C}$ .

To see that the (averaged) amplification constraint is satisfied, consider the following:

$$\begin{aligned} I(X^n; F_x, F_y | \mathcal{C}) &= H(X^n | \mathcal{C}) - H(X^n | b(X^n), L, \mathcal{C}) \\ &\geq nH(X) - n(H(X) - \Delta_A + \delta(\epsilon)) \\ &= n(\Delta_A - \delta(\epsilon)), \end{aligned} \quad (8)$$

where (8) follows since  $X^n$  is independent of  $\mathcal{C}$  and, averaged over codebooks, there are at most  $2^{n(H(X) - \Delta_A + \delta(\epsilon))}$  sequences  $x^n$  in bin  $b(X^n)$  which are typical with  $U^n(L)$ , where  $L \in \{1, \dots, 2^{n\tilde{R}}\}$ . The details are given in the Appendix.

We now turn our attention to the masking criterion. First note the following inequality:

$$\begin{aligned} I(Y^n; F_x, F_y | \mathcal{C}) &= I(Y^n; L | \mathcal{C}) + I(Y^n; b(X^n) | L, \mathcal{C}) \\ &\leq I(Y^n; L | \mathcal{C}) + H(b(X^n) | \mathcal{C}) - H(b(X^n) | Y^n, \mathcal{C}) \\ &= I(Y^n; L | \mathcal{C}) + I(X^n; Y^n) - H(X^n) + H(b(X^n) | \mathcal{C}) \\ &\quad - H(b(X^n) | Y^n, \mathcal{C}) + H(X^n | Y^n) \\ &\leq I(Y^n; L | \mathcal{C}) + I(X^n; Y^n) - H(X^n) + H(b(X^n) | \mathcal{C}) \\ &\quad - I(b(X^n); X^n | Y^n, \mathcal{C}) + H(X^n | Y^n) \\ &= I(Y^n; L | \mathcal{C}) + I(X^n; Y^n) - H(X^n) + H(b(X^n) | \mathcal{C}) \\ &\quad + H(X^n | Y^n, b(X^n), \mathcal{C}) \end{aligned} \quad (9)$$

Two of the terms in (9) can be bounded as follows: First, since  $L \in \{1, \dots, 2^{n\tilde{R}}\}$ , we have

$$I(Y^n; L | \mathcal{C}) \leq n\tilde{R} = n(I(Y; U) + \epsilon).$$

Second, there are  $2^{n(\Delta_A - I(X;U) + \epsilon)}$  bins at Encoder 1 by construction, and hence  $H(b(X^n) | \mathcal{C}) \leq n(\Delta_A - I(X;U) + \epsilon)$ . Therefore, substituting into (9) and simplifying, we have:

$$\begin{aligned} I(Y^n; F_x, F_y | \mathcal{C}) &\leq n(I(Y; U, X) + \Delta_A - H(X)) \\ &\quad + H(X^n | Y^n, b(X^n), \mathcal{C}) + n2\epsilon. \end{aligned} \quad (10)$$

We now consider three separate cases. First, assume  $\Delta_A \leq I(U; X)$ . Then,

$$\begin{aligned} I(Y; X, U) + \Delta_A - H(X) &\leq I(Y; X, U) - H(X | U) \\ &= I(Y; U) - H(X | Y), \end{aligned}$$

and (10) becomes

$$\begin{aligned} I(Y^n; F_x, F_y | \mathcal{C}) &\leq nI(Y; U) - I(X^n; b(X^n) | Y^n, \mathcal{C}) + n2\epsilon \\ &\leq nI(Y; U) + n2\epsilon. \end{aligned}$$

Next, suppose that  $\Delta_A \geq I(X; U) + H(X | Y)$ . In this case, there are greater than  $2^{n(H(X|Y) + \epsilon)}$  bins in which the  $X^n$  sequences are distributed. Hence, knowing  $Y^n$  and  $b(X^n)$  is sufficient to determine  $X^n$  with high probability (i.e., we have a Slepian-Wolf binning at Encoder 1). Therefore,  $H(X^n | Y^n, b(X^n), \mathcal{C}) \leq n\epsilon$ , and (10) becomes

$$I(Y^n; F_x, F_y | \mathcal{C}) \leq n(I(Y; X, U) + \Delta_A - H(X)) + n3\epsilon.$$

Finally, suppose  $\Delta_A = I(X; U) + \theta H(X | Y)$  for some  $\theta \in [0, 1]$ . In this case, we can timeshare between a code  $\mathcal{C}_1$  designed for amplification  $\Delta'_A = I(X; U)$  with probability  $\theta$ , and a code  $\mathcal{C}_2$  designed for amplification  $\Delta''_A = I(X; U) + H(X | Y)$  with probability  $1 - \theta$  to obtain a code  $\mathcal{C}$  with the same average rates and averaged amplification

$$\begin{aligned} I(X^n; F_x, F_y | \mathcal{C}) &= \theta I(X^n; F_x, F_y | \mathcal{C}_1) + (1 - \theta) I(X^n; F_x, F_y | \mathcal{C}_2) \\ &\geq n(I(X; U) + \theta H(X | Y) - \delta(\epsilon)) = n(\Delta_A - \delta(\epsilon)). \end{aligned}$$

Then, applying the inequalities obtained in the previous two cases, we obtain:

$$\begin{aligned} I(Y^n; F_x, F_y | \mathcal{C}) &= \theta I(Y^n; F_x, F_y | \mathcal{C}_1) + (1 - \theta) I(Y^n; F_x, F_y | \mathcal{C}_2) \\ &\leq \theta nI(Y; U) + (1 - \theta) n(I(Y; X, U) + \Delta''_A - H(X)) + 3n\epsilon \\ &= nI(Y; U) + 3n\epsilon. \end{aligned}$$

Combining these three cases proves that

$$\begin{aligned} \frac{1}{n} I(Y^n; F_x, F_y | \mathcal{C}) &\leq \max\{I(Y; U, X) + \Delta_A - H(X), I(Y; U)\} + 3\epsilon \\ &\leq \Delta_M + 3\epsilon. \end{aligned}$$

To show that there exists a code which satisfies the amplification and masking constraints simultaneously, we construct a super-code  $\bar{\mathcal{C}}$  of blocklength  $Nn$  by concatenating  $N$  randomly, independently chosen codes of length  $n$  (each constructed as described above). By the weak law of large numbers and independence of the concatenated coded blocks,

$$\begin{aligned} \Pr \left( \left\{ \bar{c} : \frac{1}{Nn} I(X^{Nn}; \bar{F}_x, \bar{F}_y | \bar{\mathcal{C}} = \bar{c}) > \Delta_A - \delta(\epsilon) \right\} \right) &\geq 3/4 \\ \Pr \left( \left\{ \bar{c} : \frac{1}{Nn} I(Y^{Nn}; \bar{F}_x, \bar{F}_y | \bar{\mathcal{C}} = \bar{c}) < \Delta_M + \delta(\epsilon) \right\} \right) &\geq 3/4 \end{aligned}$$

for  $N$  and  $n$  sufficiently large. Thus, there must exist one super-code which simultaneously satisfies both desired constraints. This completes the proof that  $(R_x, R_y, \Delta_A, \Delta_M)$  is

achievable. Finally, we invoke the Support Lemma [6] to see that  $|\mathcal{Y}| - 1$  letters are sufficient to preserve  $p(y)$ . Plus, we require two more letters to preserve the values of  $H(X|U)$  and  $I(Y; U|X)$ . ■

*Proof of Corollary 1:* By setting  $\Delta_A = H(X)$ , [1, Theorem 2] implies that  $X^n$  can be reproduced near losslessly. A simplified version of the argument in the direct part of the proof of Theorem 1 shows that the masking criterion will be satisfied for the standard coding scheme. The converse of Theorem 1 continues to apply. ■

*Proof of Theorem 2:* First, we remark that the strengthened version of [9, Theorem 6] states that  $\mathcal{R}_{AA}(R_x, R_y)$  is the closure of pairs  $(\Delta_X, \Delta_Y)$  such that there exists a  $(2^{nR_x}, 2^{nR_y}, n)$  code satisfying

$$\begin{aligned}\Delta_X &\leq \frac{1}{n}I(X^n; f_x(X^n), f_y(Y^n)) + \epsilon, \\ \Delta_Y &\leq \frac{1}{n}I(Y^n; f_x(X^n), f_y(Y^n)) + \epsilon\end{aligned}$$

for any  $\epsilon > 0$ .

Suppose  $(\Delta_X, \Delta_Y) \in \mathcal{R}^*(R_x, R_y)$ . By definition of  $\mathcal{R}^*(R_x, R_y)$ , Theorem 1, and the above statement,  $(\Delta_X, \Delta_Y)$  also lies in each of the sets  $\mathcal{R}_{AM}(R_x, R_y)$ ,  $\mathcal{R}_{MA}(R_x, R_y)$ , and  $\mathcal{R}_{AA}(R_x, R_y)$ . Since each of these sets are closed by definition, we must have

$$\overline{\mathcal{R}^*}(R_x, R_y) \subseteq \mathcal{R}_{AM}(R_x, R_y) \cap \mathcal{R}_{MA}(R_x, R_y) \cap \mathcal{R}_{AA}(R_x, R_y).$$

Since each point in the sets  $\mathcal{R}_{AM}(R_x, R_y)$ ,  $\mathcal{R}_{MA}(R_x, R_y)$ , and  $\mathcal{R}_{AA}(R_x, R_y)$  is achievable by vector quantization and/or random binning, the second statement of the Theorem is proved.

To show the reverse inclusion, fix  $\epsilon > 0$  and suppose  $(\Delta_X, \Delta_Y) \in \mathcal{R}_{AM}(R_x, R_y) \cap \mathcal{R}_{MA}(R_x, R_y) \cap \mathcal{R}_{AA}(R_x, R_y)$ . This implies the existence of  $(2^{n_{AM}R_x}, 2^{n_{AM}R_y}, n_{AM})$ ,  $(2^{n_{MA}R_x}, 2^{n_{MA}R_y}, n_{MA})$ , and  $(2^{n_{AA}R_x}, 2^{n_{AA}R_y}, n_{AA})$  codes satisfying:

$$\begin{aligned}\Delta_X &\leq \frac{1}{n_{AM}}I(X^{n_{AM}}; f_x^{AM}(X^{n_{AM}}), f_y^{AM}(Y^{n_{AM}})) + \epsilon, \\ \Delta_Y &\geq \frac{1}{n_{AM}}I(Y^{n_{AM}}; f_x^{AM}(X^{n_{AM}}), f_y^{AM}(Y^{n_{AM}})) - \epsilon, \\ \Delta_X &\geq \frac{1}{n_{MA}}I(X^{n_{MA}}; f_x^{MA}(X^{n_{MA}}), f_y^{MA}(Y^{n_{MA}})) - \epsilon, \\ \Delta_Y &\leq \frac{1}{n_{MA}}I(Y^{n_{MA}}; f_x^{MA}(X^{n_{MA}}), f_y^{MA}(Y^{n_{MA}})) + \epsilon, \\ \Delta_X &\leq \frac{1}{n_{AA}}I(X^{n_{AA}}; f_x^{AA}(X^{n_{AA}}), f_y^{AA}(Y^{n_{AA}})) + \epsilon, \\ \Delta_Y &\leq \frac{1}{n_{AA}}I(Y^{n_{AA}}; f_x^{AA}(X^{n_{AA}}), f_y^{AA}(Y^{n_{AA}})) + \epsilon.\end{aligned}$$

Also, by taking  $f_x^{MM}, f_y^{MM}$  to be constants, we trivially have a  $(2^{n_{MM}R_x}, 2^{n_{MM}R_y}, n_{MM})$  code satisfying

$$\begin{aligned}\Delta_X &\geq \frac{1}{n_{MM}}I(X^{n_{MM}}; f_x^{MM}(X^{n_{MM}}), f_y^{MM}(Y^{n_{MM}})), \\ \Delta_Y &\geq \frac{1}{n_{MM}}I(Y^{n_{MM}}; f_x^{MM}(X^{n_{MM}}), f_y^{MM}(Y^{n_{MM}})).\end{aligned}$$

It is readily verified that, by an appropriate timesharing between these four codes, there exists a  $(2^{nR_x}, 2^{nR_y}, n)$  code satisfying

$$\begin{aligned}\left| \Delta_X - \frac{1}{n}I(X^n; f_x(X^n), f_y(Y^n)) \right| &\leq \delta(\epsilon), \text{ and} \\ \left| \Delta_Y - \frac{1}{n}I(Y^n; f_x(X^n), f_y(Y^n)) \right| &\leq \delta(\epsilon).\end{aligned}$$

This completes the proof of the theorem. ■

#### IV. CONCLUDING REMARKS

In this paper, we considered a setting where two separate encoders have access to correlated sources. We gave a complete characterization of the tradeoff between amplifying information about one source while simultaneously masking another. By combining this result with recent results by Courtade and Weissman [9], we precisely characterized the amount of information that can be revealed about  $X^n$  and  $Y^n$  by any encoding functions satisfying given rates. There are three notable points here: (i) this multi-letter entropy characterization problem admits a single-letter solution, (ii) restriction of encoding functions to vector quantization and/or random binning is sufficient to achieve any point the region, and (iii) this simple characterization does not extend to three or more sources/encoders.

Finally, we remark that in the state amplification and masking problems considered in [4] and [5], the authors obtain explicit characterizations of the achievable regions when the channel state and noise are independent Gaussian random variables. Presumably, this could also be accomplished in our setting using known results on Gaussian multiterminal source coding, however, a complete investigation into this matter is beyond the scope of this paper.

#### ACKNOWLEDGMENT

The author gratefully acknowledges the conversations with Tsachy Weissman and comments by an anonymous reviewer which contributed to this paper.

#### APPENDIX

*Lemma 1:* With all quantities defined as in the proof of Theorem 1,

$$\limsup_{n \rightarrow \infty} \frac{1}{n}H(X^n|L, b(X^n), \mathcal{C}) \leq H(X) - \Delta_A + \delta(\epsilon).$$

*Proof:* We follow the proof strategy of [10, Lemma 22.3] and make adjustments where necessary. For convenience, define  $\tilde{R}_x = \Delta_A - I(X; U) + \epsilon$  and recall that  $\epsilon$  was chosen sufficiently small so that  $\tilde{R}_x < H(X|U)$ . Note that we can express the random codebook  $\mathcal{C}$  as a pair of random codebooks  $\mathcal{C} = (\mathcal{C}_B, \mathcal{C}_{VQ})$ , where  $\mathcal{C}_B$  is the “binning codebook” at Encoder 1, and  $\mathcal{C}_{VQ}$  is the “vector-quantization codebook” at Encoder 2.

Let  $E_1 = 1$  if  $(X^n, U^n(L)) \notin \mathcal{T}_\epsilon^{(n)}$  and  $E_1 = 0$  otherwise. Note that  $\Pr(\{E_1 = 1\})$  tends to 0 as  $n \rightarrow \infty$ . Consider

$$\begin{aligned} & H(X^n|L, b(X^n), \mathcal{C}) \\ & \leq H(X^n, E_1|L, b(X^n), \mathcal{C}) \\ & \leq 1 + n \Pr(\{E_1 = 1\})H(X) \\ & + \sum_{(l, b, c_{VQ})} p(l, b, c_{VQ}|E_1 = 0) \\ & \quad \times H(X^n|L = l, b(X^n) = b, E_1 = 0, \mathcal{C}_{VQ} = c_{VQ}, \mathcal{C}_B). \end{aligned}$$

Now, let  $N(l, b, c_{VQ}, \mathcal{C}_B)$  be the number of sequences  $x^n \in \mathcal{B}(b) \cap \mathcal{T}_\epsilon^{(n)}(X|u^n(l))$ , where  $\mathcal{B}(b)$  denotes the bin of  $x$ -sequences which is labeled by index  $b$  and  $u^n(l)$  is the codeword in the (fixed) codebook  $c_{VQ}$  with index  $l$ . Note that  $N(l, b, c_{VQ}, \mathcal{C}_B)$  is a binomial random variable, where the source of randomness comes from the random codebook  $\mathcal{C}_B$ . Define

$$E_2(l, b, c_{VQ}, \mathcal{C}_B) = \begin{cases} 1 & \text{if } N(l, b, c_{VQ}, \mathcal{C}_B) \geq 2\mathbb{E}[N(l, b, c_{VQ}, \mathcal{C}_B)], \\ 0 & \text{otherwise.} \end{cases}$$

Due to the binomial distribution of  $N(l, b, c_{VQ}, \mathcal{C}_B)$ , it is readily verified that

$$\begin{aligned} \mathbb{E}[N(l, b, c_{VQ}, \mathcal{C}_B)] &= 2^{-n\tilde{R}_x} \left| \mathcal{T}_\epsilon^{(n)}(X|u^n(l)) \right|, \\ \text{Var}(N(l, b, c_{VQ}, \mathcal{C}_B)) &\leq 2^{-n\tilde{R}_x} \left| \mathcal{T}_\epsilon^{(n)}(X|u^n(l)) \right|. \end{aligned}$$

Then, by the Chebyshev lemma [10, Appendix B],

$$\begin{aligned} \Pr(\{E_2(l, b, c_{VQ}, \mathcal{C}_B) = 1\}) &\leq \frac{\text{Var}(N(l, b, c_{VQ}, \mathcal{C}_B))}{(\mathbb{E}[N(l, b, c_{VQ}, \mathcal{C}_B)])^2} \\ &\leq 2^{-n(H(X|U) - \tilde{R}_x - \delta(\epsilon))}, \end{aligned}$$

which tends to zero as  $n \rightarrow \infty$  if  $\tilde{R}_x < H(X|U) - \delta(\epsilon)$ , which is satisfied for  $\epsilon$  sufficiently small. Now consider

$$\begin{aligned} & H(X^n|L = l, b(X^n) = b, E_1 = 0, \mathcal{C}_{VQ} = c_{VQ}, \mathcal{C}_B) \\ & \leq H(X^n, E_2|L = l, b(X^n) = b, E_1 = 0, \mathcal{C}_{VQ} = c_{VQ}, \mathcal{C}_B) \\ & \leq 1 + n \Pr(\{E_2 = 1\})H(X) \\ & + H(X^n|L = l, b(X^n) = b, E_1 = 0, E_2 = 0, \mathcal{C}_{VQ} = c_{VQ}, \mathcal{C}_B) \\ & \leq 1 + n \Pr(\{E_2 = 1\})H(X) \\ & + n(H(X|U) - \tilde{R}_x + \delta(\epsilon)), \end{aligned}$$

which implies that

$$\begin{aligned} & H(X^n|L, b(X^n), \mathcal{C}) \\ & \leq 2 + n(\Pr(\{E_1 = 1\}) + \Pr(\{E_2 = 1\}))H(X) \\ & \quad + n(H(X|U) - \tilde{R}_x + \delta(\epsilon)) \\ & \leq 2 + n(\Pr(\{E_1 = 1\}) + \Pr(\{E_2 = 1\}))H(X) \\ & \quad + n(H(X) - \Delta_A + \delta(\epsilon)). \end{aligned}$$

Taking  $n \rightarrow \infty$  completes the proof.  $\blacksquare$

## REFERENCES

- [1] R. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *Information Theory, IEEE Transactions on*, vol. 21, no. 6, pp. 629 – 637, Nov 1975.
- [2] A. Wyner, "On source coding with side information at the decoder," *Inf. Theory, IEEE Trans. on*, vol. 21, no. 3, pp. 294 – 300, May 1975.
- [3] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "A theory of privacy and utility in databases," *CoRR*, vol. abs/1102.3751v1, 2011.
- [4] Y.-H. Kim, A. Suvivong, and T. Cover, "State amplification," *Information Theory, IEEE Transactions on*, vol. 54, no. 5, pp. 1850 – 1859, may 2008.
- [5] N. Merhav and S. Shamai, "Information rates subject to state masking," *Inf. Theory, IEEE Trans. on*, vol. 53, no. 6, pp. 2254 – 2261, june 2007.
- [6] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.
- [7] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (corresp.)," *Information Theory, IEEE Transactions on*, vol. 25, no. 2, pp. 219 – 221, Mar 1979.
- [8] T. Berger, *Multiterminal Source Coding*. In G. Longo (Ed.), *The Information Theory Approach to Comms*. New York: Springer-Verlag, 1977.
- [9] T. A. Courtade and T. Weissman, "Multiterminal source coding under logarithmic loss," *CoRR*, vol. abs/1110.3069v2, 2011.
- [10] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.