Source Coding with Conditionally Less Noisy Side Information

Roy Timo
Institute for Telecommunications Research
University of South Australia
Adelaide, Australia
roy.timo@unisa.edu.au

Tobias J. Oechtering
ACCESS Linnaeus Center
KTH Royal Institute of Technology
Stockholm, Sweden
oech@kth.se

Michèle Wigger
Comm. and Electr. Department
Telecom ParisTech
Paris, France
michele.wigger@telecom-paristech.fr

***Abstract*—We consider a lossless multi-terminal source coding problem with one transmitter, two receivers and side information. The achievable rate region of the problem is not well understood. In this paper, we characterise the rate region when the side information at one receiver is *conditionally less noisy* than the side information at the other, given this other receiver's desired source. The conditionally less noisy definition includes degraded side information and a common message as special cases, and it is motivated by the concept of less noisy broadcast channels. The key contribution of the paper is a new converse theorem employing a telescoping identity and the Csiszár sum identity.**

I. Introduction and Problem Statement

Consider the multi-terminal source coding problem shown in Fig. 1. A discrete memoryless source emits an independent and identically distributed (iid) sequence of correlated random variables (X, Y, U, V). The Transmitter observes the (X, Y)-component, Receiver 1 observes the U-component, and Receiver 2 observes the V-component. The Transmitter jointly compresses X and Y to a binary stream of rate R, and it sends this stream over a noiseless channel to both receivers. We wish to determine the smallest rate, R^*, at which Receivers 1 and 2 can reliably recover the X and Y-components respectively.

The described problem is a special case of the rate-distortion functions in [1], [2]. Single-letter expressions for R^* are known in the following three special cases: (i) equal source components $X = Y$ [8]; (ii) complementary side information $U = Y$ and $V = X$ [3]; and (iii) degraded side information $(X, Y) -\!\!\circ\!\!- U -\!\!\circ\!\!- V$ [1].

In this paper, we determine R^* for the case where $H(Y|U) \leq H(Y|V)$ and the side information U at Receiver 1 is *conditionally less noisy* than the side information V at Receiver 2 given Y. Our definition of conditionally less noisy side information includes (i) and (iii) as special cases. The definition is motivated by the less noisy condition for discrete memoryless broadcast channels [4], [5]. The key contribution of the paper is a new converse theorem for this class of sources. The converse makes use of a telescoping identity [6] and the Csiszár sum identity [5, Sec. 2.3].

We now describe the problem statement more formally. Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{U}$ and $\mathcal{V}$ denote the finite alphabets of X, Y, U and V respectively. We identify the n-fold Cartesian product of these alphabets using boldfaced notation; for example, $\boldsymbol{\mathcal{X}}$ is the n-fold product of $\mathcal{X}$.

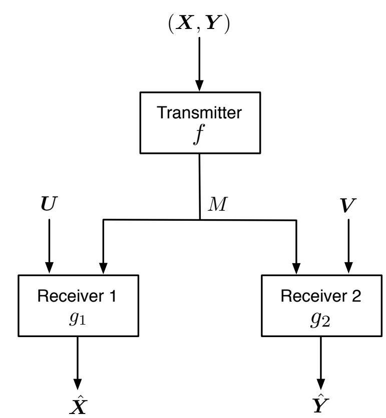

Fig. 1. Almost lossless source coding with side information at two receivers.

Let

$$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{U}, \boldsymbol{V}) \triangleq (X_1, Y_1, U_1, V_1), (X_2, Y_2, U_2, V_2), \ldots, (X_n, Y_n, U_n, V_n)$$

be a string of n-iid drawings of (X, Y, U, V). An n-blockcode consists of three (possibly stochastic) maps

$$f : \boldsymbol{\mathcal{X}} \times \boldsymbol{\mathcal{Y}} \to \mathcal{M} \tag{1a}$$

$$g_1 : \mathcal{M} \times \boldsymbol{\mathcal{U}} \to \boldsymbol{\mathcal{X}} \tag{1b}$$

$$g_2 : \mathcal{M} \times \boldsymbol{\mathcal{V}} \to \boldsymbol{\mathcal{Y}}, \tag{1c}$$

where $\mathcal{M}$ is a finite set whose cardinality depends on n. The Transmitter sends $M \triangleq f(\boldsymbol{X}, \boldsymbol{Y})$, Receiver 1 decodes $\hat{\boldsymbol{X}} \triangleq g_1(M, \boldsymbol{U})$, and Receiver 2 decodes $\hat{\boldsymbol{Y}} \triangleq g_2(M, \boldsymbol{V})$.

A rate $R \geq 0$ is said to be *achievable* if for each $\epsilon > 0$ there exists a code (f, g_1, g_2) for some sufficiently large n such that

$$\frac{1}{n} \log |\mathcal{M}| \leq R + \epsilon$$

and

$$\mathbb{P}\big[\hat{\boldsymbol{X}} \neq \boldsymbol{X} \text{ or } \hat{\boldsymbol{Y}} \neq \boldsymbol{Y}\big] \leq \epsilon.$$

Let

$$R^* \triangleq \inf\big\{R \geq 0 : R \text{ is achievable}\big\}.$$

II. Previous Results and Less Noisy Setups

The best achievability result (upper bound to R^*) can be distilled from [1], [2], [7] and is summarised next.

Lemma 1 (Achievability): We have

$$R^* \leq \min \Big[\max \left\{ I(X,Y;W|U), I(X,Y;W|V) \right\} + H(X|W,U) + H(Y|W,V) \Big],$$

where the minimisation is taken over every discrete finite auxiliary random variable W jointly distributed with (X,Y,U,V) such that

$$W \multimap\!\!- (X,Y) \multimap\!\!- (U,V).$$

The upper bound in Lemma 1 is known to be tight in the following three special cases.

Proposition 1 (Previous Optimality Results):

(i) If $X = Y$, then [7]–[9]

$$R^* = \max \Big\{ H(X|U),\ H(X|V) \Big\}.$$

(ii) If $U = Y$ and $V = X$, then [3], [9], [10]

$$R^* = \max \Big\{ H(X|Y),\ H(Y|X) \Big\}.$$

(iii) If $(X,Y) \multimap\!\!- U \multimap\!\!- V$ is a Markov chain, then the side information is said to be *degraded* and [1], [2]

$$R^* = H(Y|V) + H(X|Y,U). \tag{2}$$

Remarks:

(i) The rate R^* depends on the joint distribution of (X,Y,U,V) only via the marginal distributions of (X,Y,U) and (X,Y,V).

(ii) The side information is said to be *stochastically degraded* if the joint pmf of (X,Y,U,V) is such that there exists some (X',Y',U',V') with degraded side information and marginals (X',Y',U') and (X',Y',V') matching those of (X,Y,U) and (X,Y,V). Proposition 1, (iii), generalises to stochastically degraded side information by the previous remark.

Definition 1: We say that U is *conditionally less noisy* than V given Y if

$$I(C;U|Y) \geq I(C;V|Y)$$

holds for every discrete auxiliary random variable C jointly distributed with (X,Y,U,V) such that

$$C \multimap\!\!- (X,Y) \multimap\!\!- (U,V).$$

The next lemma shows that cases (i) and (iii) of Proposition 1 satisfy our conditionally less noisy Definition 1. The lemma is proved in Section IV.

Lemma 2: If

(i) $X \multimap\!\!- Y \multimap\!\!- V$ or

(ii) $(X,Y) \multimap\!\!- U \multimap\!\!- V$,

then U is conditionally less noisy than V given Y.

The Markov condition in (i) is more general than the equal source components $X = Y$ assumption of Proposition 1, (i). It is also quite natural in practice as it implies, in some sense, that V is closer to Y than it is to X; for example, V might be an old version of Y. The Markov condition in (ii) is precisely that used to define degraded side information.

Definition 1 is motivated by the less noisy condition for discrete memoryless broadcast channels [4], [5]. Recently, Villard and Piantanida [11] introduced a less noisy condition for information-theoretic security for source coding. In our notation, their less noisy condition is expressed as follows: U is said to be *less noisy* than V if [11]

$$I(C;U) \geq I(C;V)$$

holds for all C satisfying $C \multimap\!\!- (X,Y) \multimap\!\!- (U,V)$. Notice that this requirement implies, for example, that $H(Y|U) \leq H(Y|V)$ and $H(X|U) \leq H(X|V)$. In contrast, our conditional less noisy definition implies, for example, that $H(X|Y,U) \leq H(X|Y,V)$.

The next example shows that conditionally less noisy does not imply degraded or less noisy.

Example 1: Let X and U be independent Bernoulli-p and Bernoulli-q random variables, for $p,q \in (0, 0.5)$. Let $Y = V = X \oplus U$. Then, $X \multimap\!\!- Y \multimap\!\!- V$ and by Lemma 2, (i), U is conditionally less noisy than V given Y. In contrast, the setup is not degraded, stochastically degraded or less noisy. To see this last fact, choose $C = Y$ to obtain $I(C;V) = H(Y)$ and $I(C;U) = H(Y) - H(Y|U) = H(Y) - H(X)$.

III. Main Result

The main results of this paper are summarised next in Lemma 3 and Theorem 1. Lemma 3 is proved in Section IV.

Lemma 3 (Converse): If U is conditionally less noisy than V given Y, then

$$R^* \geq H(Y|V) + H(X|Y,U).$$

Theorem 1: If U is conditionally less noisy than V given Y and $H(Y|U) \leq H(Y|V)$, then

$$R^* = H(Y|V) + H(X|Y,U).$$

Proof: Lemma 1 and Lemma 3 together characterise R^* for those conditionally less noisy sources with $H(Y|U) \leq H(Y|V)$. To see this, choose $W = Y$ in Lemma 1 to get

$$R^* \leq \max\{H(Y|U), H(Y|V)\} + H(X|Y,U). \qquad \blacksquare$$

The theorem recovers the result for degraded side information in (2), because by Lemma 2, (ii), this setup satisfies the conditionally less noisy definition and by the data-processing inequality we have $H(Y|U) \leq H(Y|V)$.

Example 2: Let Y, Z be independent Bernoulli $1/2$ and $1/3$ random variables. Let $X = Y \oplus Z$. Let U and V be the outcomes of passing Y through a BEC$(2/3)$ and a BSC$(1/4)$ respectively, see Fig. 2. By Lemma 2, (i), the example satisfies the conditionally less noisy definition 1. Moreover, $H(Y|U) =$

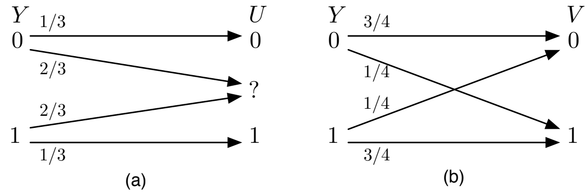

Fig. 2. Binary channels defining the side information in Example 2: (a) Binary Erasure Channel (BEC) with erasure probability $2/3$; and (b) Binary Symmetric Channel (BSC) with crossover probability $1/4$.

$2/3$ is smaller than $H(Y|V) = H_b(1/4) \approx 0.8113$, where $H_b(\cdot)$ denotes the binary entropy function. Therefore, the result in Theorem 1 applies, and

$$R^* = H_b(1/4) + H_b(1/3).$$

This result does not follow from Proposition 1, (iii), because $2/3 > 1/2$ and thus the side information U and V is not (stochastically) degraded with respect to Y [5, p. 121], [12], and hence with respect to (X, Y).

IV. Proof of Lemmas 2 and 3

A. Lemma 2

(i) Suppose that $V \multimap Y \multimap X$ is a Markov chain. Consider any C for which $C \multimap (X,Y) \multimap (U,V)$. We have

$$\begin{aligned} 0 &\le I(C;V|Y) \\ &= H(V|Y) - H(V|C,Y) \\ &\overset{(a)}{\le} H(V|X,Y) - H(V|C,X,Y) \\ &= I(C;V|X,Y) \\ &\overset{(b)}{=} 0, \end{aligned}$$

where (a) follows from $V \multimap Y \multimap X$ and (b) follows from $C \multimap (X,Y) \multimap (U,V)$. Thus, $I(C;V|Y) = 0$ and as a consequence is no larger than $I(C;U|Y)$. ■

(ii) Suppose that $(X,Y) \multimap U \multimap V$ is a Markov chain. Consider any C for which $C \multimap (X,Y) \multimap U \multimap V$. We have

$$\begin{aligned} I(C;V|Y) &\le I(C;U,V|Y) \\ &\le I(C;U|Y) + I(C;V|Y,U) \\ &= I(C;U|Y). \end{aligned}$$ ■

B. Lemma 3

We will make use of the following *telescoping* identity: for arbitrarily distributed $(A_1,B_1),(A_2,B_2),\ldots,\ (A_n,B_n)$ we have [6, Sec. G]

$$\sum_{i=1}^{n} I(A_1^i; B_{i+1}^n) = \sum_{i=1}^{n} I(A_1^{i-1}; B_i^n). \tag{3}$$

A consequence of (3), which will also be useful, is the classic *Csiszár sum* identity [5, Sec. 2.4]

$$\sum_{i=1}^{n} I(A_i; B_{i+1}^n | A_1^{i-1}) = \sum_{i=1}^{n} I(B_i; A_1^{i-1} | B_{i+1}^n). \tag{4}$$

Suppose that U is conditionally less noisy than V given Y and (f, g_1, g_2) has a joint error probability $\mathbb{P}[\hat{\boldsymbol{X}} \neq \boldsymbol{X} \text{ or } \hat{\boldsymbol{Y}} \neq \boldsymbol{Y}] \le \epsilon$. We have

$$\begin{aligned} R + \epsilon &\ge \frac{1}{n}\log|\mathcal{M}| \\ &\ge \frac{1}{n}H(M) \\ &\ge \frac{1}{n}H(M|\boldsymbol{V}) \\ &\ge \frac{1}{n}I(\boldsymbol{X},\boldsymbol{Y};M|\mathbf{V}) \\ &= \frac{1}{n}\Big[I(\boldsymbol{Y};M|\boldsymbol{V}) + I(\boldsymbol{X};M|\boldsymbol{Y},\boldsymbol{V})\Big] \\ &= \frac{1}{n}\Big[H(\boldsymbol{Y}|\boldsymbol{V}) - H(\boldsymbol{Y}|M,\boldsymbol{V}) + I(\boldsymbol{X};M|\boldsymbol{Y},\boldsymbol{V})\Big] \\ &\overset{(a)}{\ge} H(Y|V) - \varepsilon(n,\epsilon) + \frac{1}{n}I(\boldsymbol{X};M|\boldsymbol{Y},\boldsymbol{V}), \end{aligned} \tag{5}$$

where (a) follows from the fact that the tuples $\boldsymbol{Y}, \boldsymbol{V}$ are iid, from Fano's inequality and

$$\varepsilon(\epsilon, n) \triangleq \frac{h(\epsilon)}{n} + \epsilon \log|\mathcal{X} \times \mathcal{Y}|.$$

Consider the conditional mutual information term in (5). We have

$$\begin{aligned} & I(\boldsymbol{X};M|\boldsymbol{Y},\boldsymbol{V}) \\ &= H(M|\boldsymbol{Y},\boldsymbol{V}) - H(M|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{V}) \\ &\overset{(a)}{=} H(M|\boldsymbol{Y},\boldsymbol{V}) - H(M|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U}) \\ &= H(M|\boldsymbol{Y}) - I(M;\boldsymbol{V}|\boldsymbol{Y}) - H(M|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U}) \\ &= H(M|\boldsymbol{Y},\boldsymbol{U}) + I(M;\boldsymbol{U}|\boldsymbol{Y}) - I(M;\boldsymbol{V}|\boldsymbol{Y}) \\ &\quad - H(M|\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U}) \\ &= I(\boldsymbol{X};M|\boldsymbol{Y},\boldsymbol{U}) + I(M;\boldsymbol{U}|\boldsymbol{Y}) - I(M;\boldsymbol{V}|\boldsymbol{Y}) \\ &= I(\boldsymbol{X};M|\boldsymbol{Y},\boldsymbol{U}) + I(M;\boldsymbol{Y},\boldsymbol{U}) - I(M;\boldsymbol{Y},\boldsymbol{V}) \\ &= H(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{U}) - H(\boldsymbol{X}|M,\boldsymbol{Y},\boldsymbol{U}) + I(M;\boldsymbol{Y},\boldsymbol{U}) \\ &\quad - I(M;\boldsymbol{Y},\boldsymbol{V}) \\ &\overset{(b)}{\ge} nH(X|Y,U) - \varepsilon(n,\epsilon) + I(M;\boldsymbol{Y},\boldsymbol{U}) \\ &\quad - I(M;\boldsymbol{Y},\boldsymbol{V}), \end{aligned} \tag{6}$$

where (a) follows because $M \multimap (\boldsymbol{X},\boldsymbol{Y}) \multimap (\boldsymbol{U},\boldsymbol{V})$; and (b) follows from the fact that $\boldsymbol{X},\boldsymbol{Y},\boldsymbol{U}$ are iid and from Fano's inequality.

Consider (5) and (6). If it were the case that

$$I(M;\boldsymbol{Y},\boldsymbol{U}) - I(M;\boldsymbol{Y},\boldsymbol{V}) \ge 0, \tag{7a}$$

or, equivalently,

$$I(M;\boldsymbol{U}|\boldsymbol{Y}) - I(M;\boldsymbol{V}|\boldsymbol{Y}) \ge 0, \tag{7b}$$

then (6) would imply that $R + \epsilon$ can be further lower bound by

$$H(Y|V) + H(X|Y,U) - 2\varepsilon(n,\epsilon),$$

which would complete the converse since $2\varepsilon(n,\epsilon) \to 0$ as $\epsilon \to 0$. Since

$$M \multimap (\boldsymbol{X},\boldsymbol{Y}) \multimap (\boldsymbol{U},\boldsymbol{V})$$

is a Markov chain, the inequality (7) is a multi-letter conditionally less noisy condition. To complete the converse, we convert (7) into a single-letter form by constructing a discrete auxiliary random variable C such that $C -\!\!\circ\!\!- (X,Y) -\!\!\circ\!\!- (U,V)$ and

$$I(M;\boldsymbol{Y},\boldsymbol{U}) - I(M;\boldsymbol{Y},\boldsymbol{V}) = n\left(I(C;Y,U) - I(C;Y,V)\right).$$

The inequality (7) will then follow directly from Definition 1.

Using the telescoping identity (3), we first expand the mutual information $I(M;\boldsymbol{Y},\boldsymbol{U})$:

$$\begin{aligned}
&I(M;\boldsymbol{Y},\boldsymbol{U})\\
&= \sum_{i=1}^{n}\Big[I(M,V_{i+1}^n,Y_{i+1}^n;U_1^i,Y_1^i)\\
&\qquad\qquad - I(M,V_i^n,Y_i^n;U_1^{i-1},Y_1^{i-1})\Big]\\
&= \sum_{i=1}^{n}\Big[I(U_i,Y_i;M,V_{i+1}^n,Y_{i+1}^n|U_1^{i-1},Y_1^{i-1})\\
&\qquad\qquad - I(V_i,Y_i;U_1^{i-1},Y_1^{i-1}|M,V_{i+1}^n,Y_{i+1}^n)\Big]\\
&= \sum_{i=1}^{n}\Big[I(U_i,Y_i;M,U_1^{i-1},V_{i+1}^n,Y_1^{i-1},Y_{i+1}^n)\\
&\qquad\qquad - I(V_i,Y_i;U_1^{i-1},Y_1^{i-1}|M,V_{i+1}^n,Y_{i+1}^n)\Big]\\
&= \sum_{i=1}^{n}\Big[I(U_i,Y_i;C_i)\\
&\qquad\qquad - I(V_i,Y_i;U_1^{i-1},Y_1^{i-1}|M,V_{i+1}^n,Y_{i+1}^n)\Big],
\end{aligned} \tag{8}$$

where we have set

$$C_i = (M,U_1^{i-1},V_{i+1}^n,Y_1^{i-1},Y_{i+1}^n).$$

Using the same telescoping identity, we now expand the mutual information $I(M;\boldsymbol{Y},\boldsymbol{V})$ in the other direction:

$$\begin{aligned}
&I(M;\boldsymbol{Y},\boldsymbol{V})\\
&= \sum_{i=1}^{n}\Big[I(M,U_1^{i-1},Y_1^{i-1};V_i^n,Y_i^n)\\
&\qquad\qquad - I(M,U_1^i,Y_1^i;V_{i+1}^n,Y_{i+1}^n)\Big]\\
&= \sum_{i=1}^{n}\Big[I(V_i,Y_i;M,U_1^{i-1},Y_1^{i-1}|V_{i+1}^n,Y_{i+1}^n)\\
&\qquad\qquad - I(U_i,Y_i;V_{i+1}^n,Y_{i+1}^n|M,U_1^{i-1},Y_1^{i-1})\Big]\\
&= \sum_{i=1}^{n}\Big[I(V_i,Y_i;M,U_1^{i-1},V_{i+1}^n,Y_1^{i-1},Y_{i+1}^n)\\
&\qquad\qquad - I(U_i,Y_i;V_{i+1}^n,Y_{i+1}^n|M,U_1^{i-1},Y_1^{i-1})\Big]\\
&= \sum_{i=1}^{n}\Big[I(V_i,Y_i;C_i)\\
&\qquad\qquad - I(U_i,Y_i;V_{i+1}^n,Y_{i+1}^n|M,U_1^{i-1},Y_1^{i-1})\Big].
\end{aligned} \tag{9}$$

Subtract (9) from (8) and divide by n to get

$$\begin{aligned}
&\frac{1}{n}\Big[I(M;\boldsymbol{Y},\boldsymbol{U}) - I(M;\boldsymbol{Y},\boldsymbol{V})\Big]\\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[I(C_i;U_i,Y_i) - I(C_i;V_i,Y_i)\\
&\qquad + I(U_i,Y_i;V_{i+1}^n,Y_{i+1}^n|M,U_1^{i-1},Y_1^{i-1})\\
&\qquad - I(V_i,Y_i;U_1^{i-1},Y_1^{i-1}|M,V_{i+1}^n,Y_{i+1}^n)\Big]\\
&\overset{(a)}{=} \frac{1}{n}\sum_{i=1}^{n}\Big[I(C_i;U_i,Y_i) - I(C_i;V_i,Y_i)\Big]\\
&\overset{(b)}{=} I(C;U,Y) - I(C;V,Y)\\
&\overset{(c)}{\geq} 0,
\end{aligned} \tag{10}$$

where (a) follows from the Csiszár sum identity (4); (b) follows from standard time-sharing and cardinality-bounding arguments in which C is a discrete finite auxiliary random variable with $C -\!\!\circ\!\!- (X,Y) -\!\!\circ\!\!- (U,V)$; and (c) follows from Definition 1. This establishes the desired Inequality (7). ■

V. Extension to Three Receivers

We now extend[1] the setup in Fig. 1 to include a third source component Z and a third receiver, see Fig. 3. Let

$$\begin{aligned}
(\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z},\boldsymbol{U},\boldsymbol{V}) &\triangleq (X_1,Y_1,Z_1,U_1,V_1),\\
&\qquad \ldots,(X_n,Y_n,Z_n,U_n,V_n)
\end{aligned}$$

denote n-iid drawings of arbitrarily distributed discrete finite alphabet random variables (X,Y,Z,U,V). Suppose that Receiver 1 requires lossless copies of $\boldsymbol{X}$ and $\boldsymbol{Z}$; Receiver 2 requires lossless copies of $\boldsymbol{Y}$ and $\boldsymbol{Z}$; and Receiver 3 requires a lossless copy of $\boldsymbol{Z}$. A code (f,g_1,g_2,g_3) for this setup is defined analogously to (1). Let $(\hat{\boldsymbol{X}},\hat{\boldsymbol{Z}}_1)$, $(\hat{\boldsymbol{Y}},\hat{\boldsymbol{Z}}_2)$ and $\hat{\boldsymbol{Z}}_3$ denote the reconstructions at receivers 1, 2 and 3 respectively. A rate $R \geq 0$ is achievable if there exists a sequence of codes with rate approaching R and vanishing joint error probability. Let $R^\dagger$ denote the smallest achievable rate. The setup of Fig. 1 can be recovered by choosing Z to be constant. The next lemma is a generalisation of Lemma 3.

Lemma 4 (Converse): If U is conditionally less noisy than V given (Y,Z), then

$$R^\dagger \geq H(Z) + H(Y|V,Z) + H(X|U,Y,Z).$$

Proof: The proof mirrors that of Lemma 3. Specifically,

$$\begin{aligned}
R + \epsilon &\geq \frac{1}{n}H(M)\\
&\geq \frac{1}{n}\big[I(M;\boldsymbol{Z}) + H(M|\boldsymbol{Z})\big]\\
&\overset{(*)}{\geq} \frac{1}{n}\big[H(\boldsymbol{Z}) - n\varepsilon^\dagger(n,\epsilon) + H(M|\boldsymbol{Z},\boldsymbol{V})\big]\\
&\geq \frac{1}{n}\big[H(\boldsymbol{Z}) + I(\boldsymbol{X},\boldsymbol{Y};M|\boldsymbol{Z},\boldsymbol{V})\big] - \varepsilon^\dagger(n,\epsilon)
\end{aligned}$$

[1]The extension to three receivers was motivated by the three receiver broadcast channel with degraded message sets [13], [14].

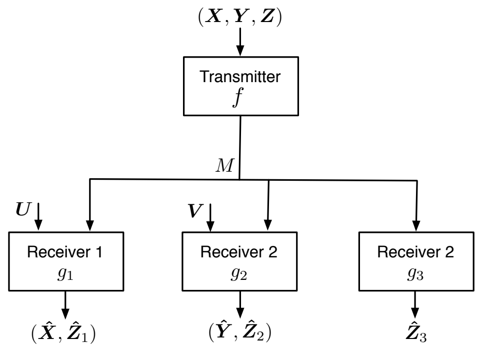

Fig. 3. (Almost) lossless source coding with three receivers.

$$
\begin{aligned}
&\geq \frac{1}{n}\big[H(\boldsymbol{Z}) + I(\boldsymbol{Y}; M|\boldsymbol{Z}, \boldsymbol{V}) \\
&\qquad\qquad + I(\boldsymbol{X}; M|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{V})\big] - \varepsilon^{\dagger}(n, \epsilon) \\
&\overset{(*)}{\geq} \frac{1}{n}\big[H(\boldsymbol{Z}) + H(\boldsymbol{Y}|\boldsymbol{Z}, \boldsymbol{V}) \\
&\qquad\qquad + I(\boldsymbol{X}; M|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{V})\big] - 2\varepsilon^{\dagger}(n, \epsilon) \\
&= H(Z) + H(Y|Z, V) \\
&\qquad\qquad + \frac{1}{n} I(\boldsymbol{X}; M|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{V}) - 2\varepsilon^{\dagger}(n, \epsilon)
\end{aligned} \tag{11}
$$

where both steps marked with a (*) use Fano's inequality and have $\varepsilon^{\dagger}(n, \epsilon)$ vanishing as $n \to \infty$ and $\epsilon \to 0$. The conditional mutual information term in (11) takes the same form as that in (5), with $(\boldsymbol{Y}, \boldsymbol{Z})$ in place of $\boldsymbol{Y}$. In particular, repeating the steps leading to (6), we obtain

$$
\begin{aligned}
I(\boldsymbol{X}; M|\boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{V}) &\geq nH(X|Y, Z, U) - \varepsilon^{\dagger}(n, \epsilon) \\
&\quad + I(M; \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{U}) - I(M; \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{V}).
\end{aligned} \tag{12}
$$

To complete the converse, we need only prove the inequality

$$
\begin{aligned}
I(M; \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{U}) - I(M; \boldsymbol{Y}, \boldsymbol{Z}, \boldsymbol{V}) \\
&= I(M; \boldsymbol{U}|\boldsymbol{Y}, \boldsymbol{Z}) - I(M; \boldsymbol{V}|\boldsymbol{Y}, \boldsymbol{Z}) \\
&\geq 0.
\end{aligned}
$$

As before, this inequality is a multi-letter version of the conditional less noisy definition. We can transform it into a single-letter form by using the telescoping identity (3) and the Csiszár sum identity (4) and by choosing

$$
C_i = (M, U_1^{i-1}, V_{i+1}^n, Y_1^{i-1}, Z_1^{i-1}, Y_{i+1}^n, Z_{i+1}^n). \qquad \blacksquare
$$

The next achievability result can be easily distilled from [2, Thm. 2]. We omit the details.

Lemma 5:

$$
\begin{aligned}
R^{\dagger} \leq \min \Big[& I(X, Y, Z; W_{123}) + I(X, Y, Z; W_{12}|W_{123}) \\
& - \min\left\{I(W_{12}; U|W_{123}), I(W_{12}; V|W_{123})\right\} \\
& + I(X, Y, Z, W_{12}; W_{13}|W_{123}) \\
& + I(X, Y, Z, W_{12}, W_{13}; W_{23}|W_{123}) \\
& - \min\left\{I(W_{23}; W_{12}, V|W_{123}), I(W_{23}; W_{13}|W_{123})\right\} \\
& + H(X|W_{123}, W_{12}, W_{13}, U) \\
& + H(Y|W_{123}, W_{12}, W_{23}, V) \\
& + H(Z|W_{123}, W_{13}, W_{23})\Big],
\end{aligned}
$$

where the minimisation is taken over all discrete finite auxiliary random variables $(W_{123}, W_{12}, W_{13}, W_{23})$ for which $(W_{123}, W_{12}, W_{13}, W_{23}) -\!\!\circ\!\!- (X, Y, Z) -\!\!\circ\!\!- (U, V)$.

The next, and final, result of the paper is a generalisation of Theorem 1 to three receivers.

Theorem 2: If U is conditionally less noisy than V given (Y, Z) and $H(Y|U, Z) \leq H(Y|V, Z)$, then

$$
R^{\dagger} = H(Z) + H(Y|Z, V) + H(X|Y, Z, U).
$$

Proof: The upper bound of Lemma 5 is equal to the lower bound of Lemma 4 on selecting W_{13} and W_{23} to be constant, $W_{123} = Z$ and $W_{12} = (Y, Z)$. $\blacksquare$

ACKNOWLEDGEMENTS

The work of M. Wigger has partly been supported by the city of Paris under the programme "Emergences". The work of R. Timo was supported by the Australian Research Council Discovery Grant DP120102123.

REFERENCES

[1] C. Heegard and T. Berger, "Rate distortion when side information may be absent," *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 727–734, 1985.

[2] R. Timo, T. Chan, and A. Grant, "Rate distortion with side-information at many decoders," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 5240–5257, 2011.

[3] R. Timo, A. Grant, and G. Kramer, "Lossy broadcasting with complementary side information," *accepted IEEE Trans. Inform. Theory*, 2012.

[4] J. Körner and K. Marton, "Comparison of two noisy channels," in *Topics in Inform. Theory*, Keszthely, Hungry, 1977.

[5] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge University Press, 2011.

[6] G. Kramer, "Teaching IT: an identity for the Gelfand-Pinsker converse," *IEEE Inform. Theory Society Newsletter*, vol. 61, no. 4, pp. 4–6, 2012.

[7] R. Timo, A. Grant, T. Chan, and G. Kramer, "Source coding for a simple network with receiver side information," in *IEEE Intl. Symp. Inform. Theory*, Toronto, Canada, 2008.

[8] A. Sgarro, "Source coding with side information at several decoders," *IEEE Trans. Inform. Theory*, vol. 23, no. 2, pp. 179–182, 1977.

[9] E. Tuncel, "Slepian-Wolf coding over broadcast channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1469–1482, 2006.

[10] A. Wyner, J. Wolf, and F. Willems, "Communicating via a processing broadcast satellite," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1243–1249, 2002.

[11] J. Villard and P. Piantanida, "Secure multiterminal source coding with side information at the eavesdropper," *ArXiV*, vol. 1105.1658v1, 2012.

[12] C. Nair, "Capacity regions of two new classes of two-receiver broadcast channels," *IEEE Trans. Inform. Theory*, vol. 56, no. 9, pp. 4207–4214, 2010.

[13] T. J. Oechtering, M. Wigger and R. Timo, "Broadcast capacity regions with three receivers and message cognition," in *IEEE Intl. Symp. Inform. Theory (ISIT)*, MIT, Cambridge, MA, July, 2012.

[14] C. Nair and Z. V. Wang, "The capacity region of the three receiver less noisy broadcast channel," *IEEE Trans. Inform. Theory*, vol. 57, no. 7, pp. 4058–4062, 2011.