Slepian–Wolf Coding Over Broadcast Channels

Ertem Tuncel, *Member, IEEE*

***Abstract*—We discuss reliable transmission of a discrete memoryless source over a discrete memoryless broadcast channel, where each receiver has side information (of arbitrary quality) about the source unknown to the sender. When there are $K = 2$ receivers, the optimum coding strategy using separate and stand-alone source and channel codes is to build two independent binning structures and send bin indices using degraded message sets through the channel, yielding a full characterization of achievable rates. However, as we show with an example, generalization of this technique to multiple binning schemes does not fully resolve the $K > 2$ case. Joint source–channel coding, on the other hand, allows for a much simpler strategy (i.e., with no explicit binning) yielding a successful single-letter characterization of achievable rates for any $K \geq 2$. This characterization, which utilizes a trivial outer bound to the capacity region of general broadcast channels, is in terms of marginal source and channel distributions rather than a joint source–channel distribution. This contrasts with existing results for other multiterminal scenarios and implies that optimal schemes achieve "operational separation." On the other hand, it is shown with an example that an optimal joint source–channel coding strategy is strictly advantageous over the combination of stand-alone source and channel codes, and thus "informational separation" does not hold.**

***Index Terms*—Broadcast channels, common information, degraded message sets, joint source–channel coding, multiterminal, separation theorem, Slepian–Wolf.**

I. Introduction

Consider a group of K sensors, each one measuring the same type of environmental data, e.g., temperature, pressure, or humidity, as a function of time. In addition to their own measurements, the sensors may need the help of a *global* observation in order to make reliable decisions and/or take necessary actions. This global information may be obtained by a more powerful "lead" sensor or by a satellite above. In either case, assume that a broadcast channel is available for conveying this information to the sensors. Since the global observation is in general correlated to the local observations, the latter should be treated as side information unavailable to the lead sensor. In this communication scenario, which we refer to as Slepian–Wolf coding over broadcast channels, we pursue the characterization of the achievable rates in terms of channel uses per source symbol.

A formal definition of the problem is as follows. Let $\{X_t, Y_{1t}, Y_{2t}, \ldots, Y_{Kt}\}_{t=0}^{\infty}$ be a discrete memoryless sequence generated according to the joint probability density function (pdf) $P_{X,Y_1,\ldots,Y_K}$ over the alphabet $\mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_K$. Here, X_t is the source (global observation) sequence available at the lead sensor and Y_{1t} through Y_{Kt} are the local side-information sequences correlated with X_t. Denote by $\{X^n, Y_1^n, Y_2^n, \ldots, Y_K^n\}$ a length-n block from this sequence. The encoder is required to convey X^n losslessly (in the Shannon sense) to *all* the decoders. For the transmission, a discrete memoryless broadcast channel $P_{V_1,\ldots,V_K|U}(v_1, \ldots, v_K|u)$ with input alphabet $\mathcal{U}$ and output alphabets $\mathcal{V}_1, \ldots, \mathcal{V}_K$ is available. We say that *rate* κ (channel uses per symbol) is *achievable* if there exist a sequence of encoders

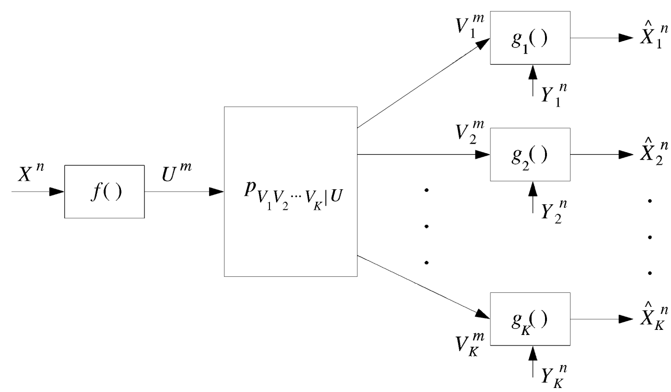

Fig. 1. Slepian–Wolf coding over broadcast channels. The achieved rate is $\kappa = \frac{m}{n}$ channel uses per symbol.

$$f^{(m,n)} : \mathcal{X}^n \longrightarrow \mathcal{U}^m$$

and K sequences of decoders

$$g_k^{(m,n)} : \mathcal{V}_k^m \times \mathcal{Y}_k^n \longrightarrow \mathcal{X}^n$$

such that the probability of error

$$P_k^{(m,n)} = \Pr[X^n \neq g_k(V_k^m, Y_k^n)]$$

vanishes uniformly for $1 \leq k \leq K$ as $n, m \to \infty$ while $\frac{m}{n}$ is fixed at κ. Here, V_k^m denotes the channel-corrupted version of $U^m = f^{(m,n)}(X^n)$ available at the kth sensor node. See Fig. 1 for a pictorial description of the system.

We first analyze the achievable rates for separate source–channel coding in the following classical sense [21].

i) The source coder achieves perfect reconstruction of X^n at every receiver with probability approaching one as $n \to \infty$, assuming that each message at the source encoder output is transmitted without error to all the source decoders it is intended for.
ii) The channel coder ensures, with probability approaching one as $m \to \infty$, that each message at the channel encoder input is reconstructed without error at the output of all the channel decoders it is intended for.

Manuscript received February 23, 2005; revised November 27, 2005. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Adelaide, Australia, September 2005.

The author is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: ertem@ee.ucr.edu).

Communicated by Y. Steinberg, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2006.871609

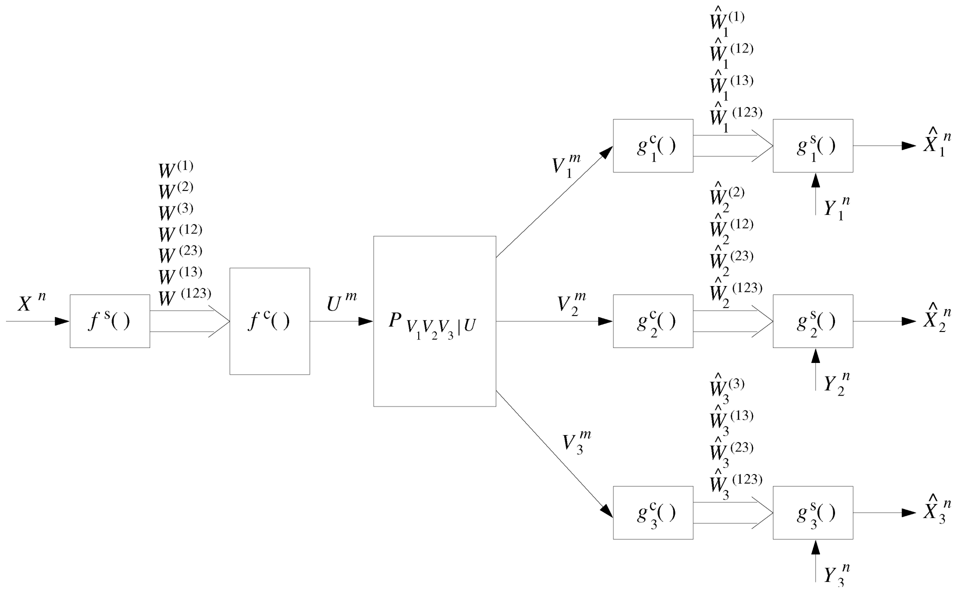

Fig. 2. Separation with stand-alone source and channel coders for $K = 3$. At the output of the source encoder $f^s(\cdot)$, there are seven distinct messages intended for different sets of decoders. For example, message $W^{(13)}$ is for receivers 1 and 3 only, and $\hat{W}_1^{(13)}$ and $\hat{W}_3^{(13)}$ are its reconstructed versions at channel decoders $g_1^c(\cdot)$ and $g_3^c(\cdot)$, respectively.

That is, the source and channel coders are *stand-alone* and they operate separately. This separation is exemplified in Fig. 2 for $K = 3$. As usual, the source and channel coders must be designed using only $P_{X,Y_1,\ldots,Y_K}$ and $P_{V_1,\ldots,V_K|U}$, respectively.

In this paradigm, it turns out that the optimum coding strategy for $K = 2$ is to use two independent binning schemes for X^n and convey the bin indices through the channel as *degraded* messages, i.e., one bin index as a common message $W^{(12)}$ to both receivers and the other as a private message $W^{(2)}$ to the receiver whose side information is of lower quality. Since the necessary and sufficient rates for these messages are easily determined, and the capacity region for any broadcast channel with degraded message sets is known (cf. [14], at least for $K = 2$), this observation immediately reveals a single-letter characterization of achievable rates. On the other hand, complications arise when $K \geq 3$. More specifically, as we show by an example for $K = 3$, sending multiple bin indices as degraded messages is not necessarily optimal, and thus, the problem remains unresolved.

We then turn to joint source–channel coding and show that a coding strategy with no explicit binning at the encoder, along with a converse result, yields for all channels a complete single-letter characterization of achievable rates. Specifically, as the main result of this paper, we prove that κ is an achievable rate if and only if there exists $P_U(u)$ such that

$$H(X|Y_k) < \kappa I(U;V_k) \tag{1}$$

for all k. This result is exciting because, to the best of our knowledge, it constitutes the first example in the literature where a complete single-letter characterization of achievable rates in joint source–channel coding is successfully derived even though that for separate coding is unknown (i.e., for $K > 2$). Also, the characterization is not in terms of a joint distribution of $\{X, Y_1, \ldots, Y_K, U, V_1, \ldots, V_K\}$. Rather, its conditions are based on $\{X, Y_1, \ldots, Y_K\}$ and $\{U, V_1, \ldots, V_K\}$ separately. This contrasts with some existing multiterminal joint source–channel coding results. For example, [4] derived single-letter sufficient conditions for achievability of rate $\kappa = 1$ when correlated sources are transmitted over multiple-access channels. The characterization therein uses joint distributions of source and channel variables. A similar interplay between source and channel variables arises in [12] which treats coding of a pair of correlated sources, both available at the encoder, over broadcast channels. The characterization (1) is also interesting in the sense that it shows a trivial outer bound to the capacity region derived in [3] to be the effective capacity region when the channel is used in our Slepian–Wolf setting.

The separation of source and channel variables in (1) is due to the "operational separation" exhibited by the optimal joint source–channel coding strategy. That is, the optimal strategy can be separated into source and channel components that operate independently. However, neither of these components are stand-alone in the sense mentioned above, i.e., as in Fig. 2. In other words, this observation does not imply "informational separation," i.e., separation in the classical sense.

So when do we have informational separation? By comparing the performances of joint and separate (stand-alone) coding strategies for several well-known broadcast channels, we attempt to answer this question. For degraded broadcast channels, this analysis yields an example where joint coding is strictly advantageous, thereby leading to the conclusion that informational separation does not hold in general.[1] We also prove for a subclass of deterministic channels that separate coding *is* optimal. Finally, for bottleneck, orthogonal, switch-to-talk, and incompatible channels, all of which were introduced in

[1]Although the breakdown of separation for lossy source coding over broadcast channels is well known (see, for example, [10]), to the best of our knowledge, there is no result in the literature for the special case of lossless coding with side information at the receivers.

[3], it is straightforward to show, using our main result (1), that informational separation always holds.

A scenario related to ours was recently tackled by Wolf in [24], wherein i) the side information is assumed to be of the same quality at all receivers, i.e., with equal $H(X|Y_k)$, ii) the encoder has access to the whole data $\{X_t, Y_{1t}, Y_{2t}, \ldots, Y_{Kt}\}_{t=0}^{\infty}$, and iii) the broadcast channel is noiseless, i.e., $V_k^m = U^m$ for all k. An interesting observation presented in [24] was that even though the source encoder has full access to the side information, it is not clear how it should use it. Specifically, completely ignoring the side information and implementing a traditional binning scheme seems to be the only option for the encoder. In contrast, in our scenario, no binning at the encoder is required even though the side information is not available. Instead, the optimal joint source–channel coding scheme employs "virtual binning," i.e., binning is performed *in effect* at the decoding stage. More specifically, each local receiver first creates a *list* of possible X^n, and then treats this list as an effective bin and resolves the actual X^n using its local side information. On the other hand, virtual binning cannot be an alternative solution to Wolf's problem because it reduces to traditional binning when the channel is noiseless.

The rest of this paper is organized as follows. We begin with the preliminaries in the next section. We then discuss in Section III the minimum achievable rate when separate coding is adopted. In Section IV, we turn to joint source–channel coding and prove our main result (1). Section V compares the performances of separate and joint coding for several types of broadcast channels. Finally, in Section VI, we discuss the results and point to future directions.

II. Preliminaries

A. Types and Typical Sequences

We heavily use the method of types and strong typicality, and follow the definitions and the notation provided in [6]. The type P_{x^n} of a vector $x^n \in \mathcal{X}^n$ is the empirical distribution given by

$$P_{x^n}(a) = \frac{1}{n} N(a|x^n)$$

where $N(a|x^n)$ denotes the number of occurrences of a in x^n. We denote by T_Q^n the type class Q, i.e., the set of all sequences x^n having type Q, and by $\mathcal{P}_n$ the set of all types for sequences of length n. A well-known crucial fact is that $|\mathcal{P}_n| \le (n+1)^{|\mathcal{X}|}$ (cf. [6, Lemma 1.2.2]). If

$$\left| P_X(a) - \frac{1}{n} N(a|x^n) \right| \le \delta$$

for all $a \in \mathcal{X}$, and also if $N(a|x^n) = 0$ whenever $P_X(a) = 0$, x^n is said to be strongly δ-typical with X. The set of all strongly-typical sequences is denoted by $T_{[X]_\delta}^n$. A notable property of typical sets is that[2]

$$\left| \frac{1}{n} \log |T_{[X]_\delta}^n| - H(X) \right| \le \delta$$

for large enough n, i.e., when $n \ge n_0(\delta)$ (cf. [6, Lemma 1.2.13]). The type and typicality concepts can be generalized to pairs of random variables in a straightforward manner. We refer the reader to [6] for a thorough discussion, and make specific references to lemmas and theorems in [6] whenever necessary in our proofs.

B. The Slepian–Wolf Result and Multiple Binning

The exposition in Section III relies on the extension of the Slepian–Wolf result to *multiple* binning. The original Slepian–Wolf theorem [22] characterizes the minimum achievable rate in a point-to-point lossless communication system where the decoder has side information unknown to the encoder. Observe that this scenario corresponds to $K = 1, \mathcal{U} = \{0, 1\}$ (so that the rate $\kappa = \frac{m}{n}$ is in bits per symbol), and a noiseless channel $P_{V_1|U}$ in our setting. We first repeat the Slepian–Wolf result and sketch its proof for convenience.

Theorem 1: Rate κ is achievable if and only if $H(X|Y_1) < \kappa$.

Proof: The "only if" part is obvious from the fact that the rate cannot fall below $H(X|Y_1)$ even when the encoder also has access to $\{Y_{1t}\}_{t=0}^{\infty}$.

For the "if" part, apply randomized binning: For an arbitrarily small $\delta > 0$, create $2^{n\kappa}$ bins by randomly picking for each bin $2^{n[H(X)+2\delta-\kappa]}$ source vectors from $T_{[X]_\delta}^n$ according to uniform distribution. Inform the encoder and the decoder about the contents of the bins. The encoder then transmits the index of the first bin in which X^n, the actual source vector, appears. If X^n does not appear in any bin, it sends an arbitrary bin index, say 1. This communication requires κ bits per source symbol. Having received only the bin index, the decoder estimates X^n as the first vector $\hat{x}^n$ in that bin which is jointly typical with the side-information vector Y_1^n. If there is no such source vector in the bin, then the decoder outputs an arbitrary $\hat{x}^n$.

There are three sources of error in this strategy. First, X^n may not appear in any bin. The probability of this event vanishes, because i) $\Pr[X^n \notin T_{[X]_\delta}^n]$ vanishes (cf. [6, Lemma 1.2.12]) and ii) during the random binning phase we have generated a total of $2^{n[H(X)+2\delta]}$ source vectors, and the probability q that none of these match $X^n \in T_{[X]_\delta}^n$ is bounded as

$$q \le \left(1 - \frac{1}{|T_{[X]_\delta}^n|}\right)^{2^{n[H(X)+2\delta]}} \le e^{-\frac{2^{n[H(X)+2\delta]}}{|T_{[X]_\delta}^n|}} \le e^{-2^{n\delta}}$$

for large enough n. The second error event occurs when X^n is not typical with Y_1^n. The probability of this also vanishes exponentially fast (cf. [6, Lemma 1.2.12]). Finally, some $x_0^n \ne X^n$ in the selected bin might also be typical with Y_1^n. Using the union bound $\Pr[\mathcal{A} \cup \mathcal{B}] \le \Pr[\mathcal{A}] + \Pr[\mathcal{B}]$, the probability of this event can be upper-bounded by

$$2^{n[H(X)+2\delta-\kappa]} 2^{-n[I(X;Y_1)-2\delta]} = 2^{n[H(X|Y_1)+4\delta-\kappa]}$$

which can be brought to zero as $n \to \infty$ if $\kappa > H(X|Y_1) + 4\delta$. □

We next describe Slepian–Wolf coding with multiple binning. Consider L independent binning schemes, each having $2^{n\kappa_l}$

[2]All logarithms in this paper are base 2.

bins formed by $2^{n[H(X)+2\delta-\kappa_l]}$ independent and uniform drawings from $T^n_{[X]_\delta}$. Let $\sum_{l=1}^{L}\kappa_l = \kappa$. For each binning scheme, the encoder searches for a bin that contains X^n, and transmits its index, thereby expending total rate κ. The decoder then outputs a source vector $\hat{x}^n$ that appears in all L selected bins and is jointly typical with Y_1^n. The next lemma shows that this multiple binning scheme enjoys the same rates as regular Slepian–Wolf coding.

Lemma 1: Rate κ is achievable with multiple binning if and only if $H(X|Y_1) < \kappa$.

Proof: As in the proof of Theorem 1, it suffices to show the "if" part. It is clear that with probability approaching one, X^n appears in at least one bin in each scheme, and is typical with Y_1^n. The only remaining source of error is then the existence of $x_1^n = x_2^n = \cdots = x_L^n \neq X^n$, where x_l^n is a vector in the bin selected by the encoder from the lth scheme and is typical with Y_1^n. Defining $X_1 = X_2 = \cdots = X_L = X$, this is possible if and only if $(x_1^n, x_2^n, \ldots, x_L^n, Y_1^n)$ is jointly typical according to $P_{X_1,X_2,\ldots,X_L,Y_1}$. In a single trial, where x_l^n are independently drawn from $T^n_{[X]_\delta}$, the probability of this joint typicality event cannot be larger than

$$2^{-n[LH(X)-H(X_1,X_2,\ldots,X_L|Y_1)-(L+1)\delta]}$$

or equivalently

$$2^{-n[LH(X)-H(X|Y_1)-(L+1)\delta]}.$$

Thus, for the whole set of trials forming the bins, the union bound yields

$$\left(\prod_{l=1}^{L} 2^{n[H(X)+2\delta-\kappa_l]}\right) 2^{-n[LH(X)-H(X|Y_1)-(L+1)\delta]}$$

as an upper bound to the probability of error, which vanishes as $n \to \infty$ if $\kappa > H(X|Y_1) + (3L+1)\delta$. □

C. Broadcast Channels and Capacity

In the most general broadcast channel scenario, $2^K - 1$ independent messages are to be conveyed through the channel, each one intended for a distinct nonempty subset $\mathcal{S}$ of all receivers $\mathcal{T}$. Denote by $W^{(\overline{\mathcal{S}})}$ the message targeting receivers in $\mathcal{S}$, where notation $\overline{\mathcal{S}}$ indicates a concatenated list of elements of $\mathcal{S}$ and serves as a mnemonic. For example, when $\mathcal{S} = \{1,2,3\}$, $\overline{\mathcal{S}} = 123$, and the message for receivers in $\mathcal{S}$ is denoted by $W^{(123)}$. Similarly, denote by $M^{(\overline{\mathcal{S}})}$ the cardinality of the message set for $W^{(\overline{\mathcal{S}})}$, i.e., $W^{(\overline{\mathcal{S}})} \in \{1,2,\ldots,M^{(\overline{\mathcal{S}})}\}$. Also, let $\hat{W}_k^{(\overline{\mathcal{S}})}$ be the decoded version of $W^{(\overline{\mathcal{S}})}$ at receiver $k \in \mathcal{S}$. The reader is referred to the channel in Fig. 2 as an example for $K = 3$.

Definition 1: The rate vector $\{r^{(\overline{\mathcal{S}})} | \mathcal{S} \subset \mathcal{T}\}$ is achievable if there exists a sequence of channel encoders

$$f^{(m)} : \bigotimes_{\mathcal{S}\subset\mathcal{T}} \{1,2,\ldots,M^{(\overline{\mathcal{S}})}\} \longrightarrow \mathcal{U}^m$$

and K sequences of channel decoders

$$g_k^{(m)} : \mathcal{V}_k^m \longrightarrow \bigotimes_{\mathcal{S}\ni k} \{1,2,\ldots,M^{(\overline{\mathcal{S}})}\}$$

such that for any $\epsilon > 0$ and arbitrarily large m

$$\Pr[\hat{W}_k^{(\overline{\mathcal{S}})} \neq W^{(\overline{\mathcal{S}})}] \leq \epsilon$$

and

$$\frac{1}{n}\log M^{(\overline{\mathcal{S}})} \geq r^{(\overline{\mathcal{S}})} - \epsilon$$

for all $\mathcal{S}, k \in \mathcal{S}$. Here, $\bigotimes$ denotes the Cartesian product, and $\mathcal{S} \ni k$ is a shorthand notation for "all $\mathcal{S}$ such that $k \in \mathcal{S}$."

Definition 2: The capacity region $\mathcal{C}$ is the closure of all achievable $\{r^{(\overline{\mathcal{S}})} | \mathcal{S} \subset \mathcal{T}\}$.

Denoting by R_k the total rate delivered to receiver k, i.e.,

$$R_k \triangleq \sum_{\mathcal{S}\ni k} r^{(\overline{\mathcal{S}})}$$

define the region $\mathcal{R}$ as

$$\mathcal{R} \triangleq \{(R_1,\ldots,R_K) : \exists P_U \text{ with } R_k \leq I(U;V_k),\ \forall k\}.$$

In [3], the following trivial outer bound on $\mathcal{C}$ was derived.

Lemma 2: If $\{r^{(\overline{\mathcal{S}})} | \mathcal{S} \subset \mathcal{T}\} \in \mathcal{C}$, then $(R_1,\ldots,R_K) \in \mathcal{R}$.

As a matter of fact, it was the convex closure of $\mathcal{R}$ that was proven to be an outer bound to the total rates in [3]. However, as we show next, $\mathcal{R}$ is already a convex region.

Lemma 3: $\mathcal{R}$ is convex.

Proof: For $(R_1^{(i)},\ldots,R_K^{(i)}) \in \mathcal{R}$ for $i = 1,2$, let P_{U_1} and P_{U_2}, respectively, achieve $I(U_1;V_k) \geq R_k^{(1)}$ and $I(U_2;V_k) \geq R_k^{(2)}$ for each k. From concavity of $I(U;V)$ as a function of P_U, the distribution

$$P_{U_\lambda}(u) = \lambda P_{U_1}(u) + (1-\lambda)P_{U_2}(u)$$

satisfies

$$\begin{aligned} I(U_\lambda;V_k) &\geq \lambda I(U_1;V_k) + (1-\lambda)I(U_2;V_k) \\ &\geq \lambda R_k^{(1)} + (1-\lambda)R_k^{(2)} \end{aligned}$$

for each k, proving the lemma. □

Thanks to its convexity, the region $\mathcal{R}$ is in principle easy to compute via the Lagrangian optimization

$$\max_{P_U} \sum_{k=1}^{K} \beta_k I(U;V_k)$$

which is a convex optimization problem. It is well known that $(\beta_1,\beta_2,\ldots,\beta_k)$ is in the normal direction to the boundary of $\mathcal{R}$ at the point

$$\Big(I(U^*;V_1), I(U^*;V_2), \ldots, I(U^*;V_K)\Big)$$

where U^* is the random variable that achieves the above maximum.

As pointed out in [3], the bound in Lemma 2 is generally loose, for it is not always possible to convey a total of $2^{mI(U;V_k)}$ messages to receiver k simultaneously for all k. In this paper, we attach a different meaning to this outer bound. Specifically, as we discuss in Section V, tightness of the outer bound $\mathcal{R}$ implies the separability of the source and the channel coding problems in our scenario for *all* sources.

A single-letter characterization to the capacity region $\mathcal{C}$ is not known, even for the simplest case $K = 2$. The best achievable rate region known to date is derived by Marton [15] (whose proof is later on simplified in [8]). On the other hand, some special cases are well understood. One such case of interest to us is when the message sets are *degraded*, i.e., $M^{(\overline{\mathcal{S}})} = 1$ for all $\mathcal{S}$ other than those of the form $\{k,\ldots,K\}$ with $1 \le k \le K$. This implies a hierarchy of messages $W^{(12\cdots K)}, W^{(23\cdots K)}, \ldots, W^{(K)}$, and thus, of decoders. In particular, the kth decoder resolves the first k messages in the hierarchy. Note that one can always transfer rate from lower levels to higher levels. For example, for $K = 3$, if $(r^{(123)}, r^{(23)}, r^{(3)})$ is achievable, so is $(r^{(123)} - \gamma, r^{(23)} + \gamma, r^{(3)})$, $(r^{(123)} - \gamma, r^{(23)}, r^{(3)} + \gamma)$, or $(r^{(123)} - \gamma, r^{(23)} + \gamma/2, r^{(3)} + \gamma/2)$, provided $\gamma < r^{(123)}$. That is because a lower level decoder can simply ignore some received message bits and let a higher level decoder use them. Therefore, the cross section of the capacity region $\mathcal{C}$ corresponding to degraded message sets can be equivalently characterized by the *cumulative*[3] rates $R_k = \sum_{i=1}^{k} r^{(i\cdots K)}$. We denote by $\mathcal{C}^{dm}$ the capacity region with degraded message sets in terms of cumulative rates. Formally, we define

$$\mathcal{C}^{dm} \triangleq \Big\{(R_1,\ldots,R_K) : R_K \ge R_{K-1} \ge \cdots \ge R_1 \\ \text{and } r^{(k\cdots K)} = R_k - R_{k-1} \text{ are achievable}\Big\}.$$

A single-letter characterization for $\mathcal{C}^{dm}$ when $K = 2$ is given by the following theorem from [14].

Theorem 2: $\mathcal{C}^{dm}$ for $K = 2$ consists of all (R_1, R_2) for which $R_1 \le R_2$ and there exists a joint distribution $P_Z P_{U|Z} P_{V_1,V_2|U}$ satisfying

$$R_1 \le I(Z;V_1)$$
$$R_2 \le \min\Big\{I(U;V_2),\, I(Z;V_1) + I(U;V_2|Z)\Big\}.$$

Remark 1: The original theorem stated the capacity region in terms of marginal rates. See [7] for the cumulative rate version.

A very related case regarding the characterization of capacity is degraded broadcast channels, for which

$$P_{V_1,V_2|U}(v_1,v_2|u) = P_{V_2|U}(v_2|u) P_{V_1|V_2}(v_1|v_2)$$

for some $P_{V_1|V_2}$, or in other words, $U \leftrightarrow V_2 \leftrightarrow V_1$ forms a Markov chain.

Theorem 3: ([1], [9]) For a degraded broadcast channel, $\mathcal{C}$ consists of all $(r^{(1)}, r^{(2)}, r^{(12)})$ for which there exists a joint distribution $P_Z P_{U|Z} P_{V_2|U} P_{V_1|V_2}$ so that

$$r^{(1)} + r^{(12)} \le I(Z;V_1)$$
$$r^{(2)} \le I(U;V_2|Z).$$

Without loss of generality, the message sets can also be assumed degraded, since any private message to receiver 1 can be reliable decoded by receiver 2 as well. Moreover, bits from $W^{(12)}$ and $W^{(1)}$ can always be transferred to $W^{(2)}$. Thus, the equivalent capacity region $\mathcal{C}^{dm}$ in terms of cumulative rates $R_1 = r^{(1)} + r^{(12)}$ and $R_2 = r^{(1)} + r^{(2)} + r^{(12)}$ is obtained by replacing the conditions of Theorem 3 with

$$R_1 \le I(Z;V_1) \tag{2}$$
$$R_2 \le I(Z;V_1) + I(U;V_2|Z). \tag{3}$$

Note that (2) and (3) can also be derived from Theorem 2 using the Markov relation $Z \leftrightarrow U \leftrightarrow V_2 \leftrightarrow V_1$.

We are also interested in the explicit characterization of $\mathcal{C}^{dm}$ for the deterministic, bottleneck, orthogonal, switch-to-talk, and incompatible channels. We defer the discussion on these channels to Section V.

We close this section by discussing another special case where there is a single message to be conveyed to all receivers. In other words, $r^{(\overline{\mathcal{S}})} = 0$ for all $\mathcal{S} \neq \mathcal{T}$. This is equivalent to the *compound* channel scenario where the sender does not know which of $P_{V_k|U}$ is the true channel.[4] The capacity C^{cmp} of this special case was derived by Wolfowitz [25] and Blackwell *et al.* [2] as

$$C^{cmp} = \max_{P_U} \min_{1\le k\le K} I(U;V_k). \tag{4}$$

For $K = 2$, this result could also be obtained by setting $R_1 = R_2$ in Theorem 2.

III. Achievable Rates in Separate Source–Channel Coding

In this section, we analyze the performance of separate and stand-alone source and channel coders. For the source coding part, the Slepian–Wolf result in Theorem 1 implies that it is necessary and sufficient to send $n[H(X|Y_k) + o(1)]$ bits to source decoder k. However, sending these bits (i.e., bin indices) through the broadcast channel as independent information would be naive and an inefficient use of the channel. We discuss in this section more elegant ways of communicating the bin indices.

Before proceeding further, let us assume without loss of generality that

$$H(X|Y_1) \le H(X|Y_2) \le \cdots \le H(X|Y_K).$$

This particular order of conditional entropies is adopted throughout the sequel for the sake of a fair comparison between achievable rates in separate and joint source–channel coding.

A. Universal Binning and Compound Channels

One way of bringing together separate source and channel codes is to use $P_{V_1,V_2,\ldots,V_K|U}$ as a compound channel, that is, by sending only a single common message $W^{(12\cdots K)}$ through the channel. Though suboptimal (as we prove in the next subsection), this technique is worth discussing since it sheds some light on the problem. We first state the achievable set of rates under this regime.

[3]Note that for the case of degraded message sets

$$R_k = \sum_{\mathcal{S}\ni k} r^{(\overline{\mathcal{S}})} = \sum_{i=1}^{k} r^{(i\cdots K)}.$$

[4]Without loss of generality, the receiver may be presumed to know the true channel. Even if it does not initially, some pre-arranged symbols transmitted over the channel with negligible rate (e.g., $\sqrt{m}$ symbols in m uses) will reveal the true channel to the receiver.

Lemma 4: When $P_{V_1,V_2,\ldots,V_K|U}$ is used as a compound channel, rate κ is achievable if and only if

$$H(X|Y_K) \le \kappa C^{cmp}.$$

Proof: That $nH(X|Y_K)$, which is the largest among $nH(X|Y_k)$, is a necessary and sufficient source rate for X^n when the decoders have various levels of side information easily follows by specializing more general lossy source coding scenarios with various levels of side information to lossless coding (see, for example, [13] and [19].) Since the capacity of the compound channel with m uses is mC^{cmp}, the result follows. □

Remark 2: One can directly prove sufficiency of $H(X|Y_K)$ using universal binning arguments. If one creates $H(X|Y_K) + o(1)$ bins, the number of source vectors in each bin is bounded as

$$H(X) - H(X|Y_K) - o(1) \le H(X) - H(X|Y_k) - o(1) < I(X;Y_k)$$

and thus with probability approaching one as $n \to \infty$, only the correct source vector is typical with Y_k^n for any $1 \le k \le K$.

Lemma 4 reveals that the minimum κ in this paradigm is determined by the quality of the worst side information and of the worst channel in the max-min sense of (4). Intuitively, in the fortunate situation where the channel quality improves as the quality of the side information decreases, i.e., if $I(U;V_1) \le I(U;V_2) \le \cdots \le I(U;V_K)$ for some "good" P_U, one must be able to decrease κ further using a more efficient coding scheme. We show in the next subsection that this is indeed possible to some degree with separate coding. On the other hand, as we discuss in Section IV, we can exploit this phenomenon fully with joint coding.

B. Multiple Binning and Degraded Message Sets

The following theorem characterizes the minimum achievable rate in separate source–channel coding in its full generality assuming that the capacity region $\mathcal{C}$ is known.

Theorem 4: Rate κ is achievable using separate source and channel coders if and only if there exists $\{r^{(\mathcal{S})}|\mathcal{S} \subset \mathcal{T}\} \in \mathcal{C}$ such that

$$H(X|Y_k) \le \kappa R_k = \kappa \sum_{\mathcal{S} \ni k} r^{(\mathcal{S})} \tag{5}$$

for all $1 \le k \le K$.

Proof: It is clear that if the channel cannot deliver in m uses a total rate of $nH(X|Y_k)$ to receiver k, it is not possible to have $\Pr[X^n \ne \hat{X}_k^n] \to 0$ as $n, m \to \infty$ with $\kappa = \frac{m}{n}$.

As for the achievability of κ satisfying (5), let the encoder create $L = 2^K - 1$ independent binning schemes, each associated with a subset of receivers $\mathcal{S}$ and having $2^{m[r^{(\mathcal{S})}+o(1)]}$ bins. The encoder then sends the bin index associated with $\mathcal{S}$ to all receivers in $\mathcal{S}$ with m uses of the channel. Since $\{r^{(\mathcal{S})}|\mathcal{S} \subset \mathcal{T}\} \in \mathcal{C}$, these bin indices are decoded reliably at each receiver. Thus, each receiver k has access to multiple bin indices with a total rate of κR_k per source symbol. The proof then follows from Lemma 1. □

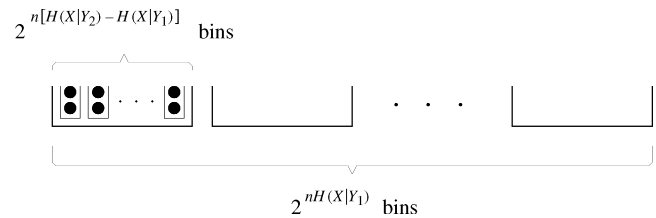

Fig. 3. Nested binning for $K = 2$. One can first create $\approx 2^{nH(X|Y_2)}$ bins and group $\approx 2^{n[H(X|Y_2)-H(X|Y_1)]}$ together to form a coarser binning with $\approx 2^{nH(X|Y_1)}$ bins.

Thus, according to Theorem 4, the achievability of κ is solely determined by the total rates deliverable to each receiver. Exploiting this fact, the next theorem states a major simplification for $K = 2$. More specifically, it provides a single-letter condition on the achievability of κ.

Theorem 5: For $K = 2$, rate κ is achievable using separate source and channel coders if and only if

$$\Big(H(X|Y_1),\ H(X|Y_2)\Big) \in \kappa\mathcal{C}^{dm} \tag{6}$$

where $\kappa\mathcal{C}^{dm} \triangleq \{(\kappa R_1, \kappa R_2) : (R_1, R_2) \in \mathcal{C}^{dm}\}$.

Proof: The "if" part follows from Theorem 4 and the definition of $\mathcal{C}^{dm}$. On the other hand, it also follows from Theorem 4 that if κ is achievable, there exists a rate vector $\{r^{(1)}, r^{(2)}, r^{(12)}\} \in \mathcal{C}$ satisfying

$$H(X|Y_1) \le \kappa(r^{(1)} + r^{(12)})$$
$$H(X|Y_2) \le \kappa(r^{(2)} + r^{(12)}).$$

We handle the two cases $r^{(2)} \ge r^{(1)}$ and $r^{(2)} < r^{(1)}$ separately. If $r^{(2)} \ge r^{(1)}$, then by transferring rate from both private messages to the common message, we observe $\{0, r^{(2)} - r^{(1)}, r^{(12)} + r^{(1)}\} \in \mathcal{C}$. By definition, this implies

$$(r^{(1)} + r^{(12)}, r^{(2)} + r^{(12)}) \in \mathcal{C}^{dm}$$

and therefore, $(H(X|Y_1), H(X|Y_2)) \in \kappa\mathcal{C}^{dm}$. If $r^{(2)} < r^{(1)}$, we similarly observe $\{r^{(1)} - r^{(2)}, 0, r^{(12)} + r^{(2)}\} \in \mathcal{C}$. It then follows that

$$(r^{(2)} + r^{(12)}, r^{(2)} + r^{(12)}) \in \mathcal{C}^{dm}$$

and since

$$H(X|Y_1) \le H(X|Y_2) \le \kappa(r^{(2)} + r^{(12)})$$

this implies $\Big(H(X|Y_1), H(X|Y_2)\Big) \in \kappa\mathcal{C}^{dm}$. □

Remark 3: It follows from this theorem that the encoder needs to implement only two binning schemes instead of three (which was suggested by Theorem 4). Even though the two transmitted bin indices are independent, the one conveyed to both receivers serves as coarse information, and the other, reliably decoded only by receiver 2, as refinement. Thus, one can alternatively implement a *nested* binning structure with two levels, as depicted in Fig. 3.

Remark 4: When C_2, the capacity of the channel $P_{V_2|U}$, is larger than the compound channel capacity C^{cmp}, sending multiple bin indices using degraded message sets is superior over sending a single bin index using the compound channel mode.

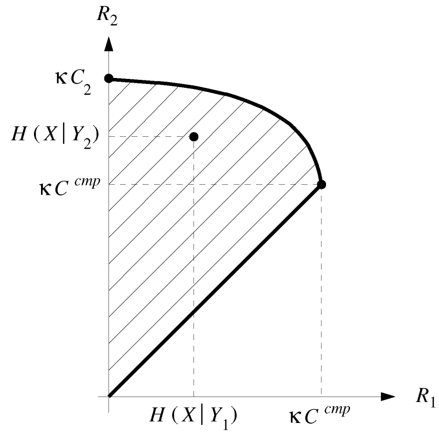

Fig. 4. An example where using the compound channel mode is inferior. The shaded area represents $\mathcal{C}^{dm}$ and according to Theorem 5, κ is an achievable rate even though $H(X|Y_2) > \kappa C^{cmp}$.

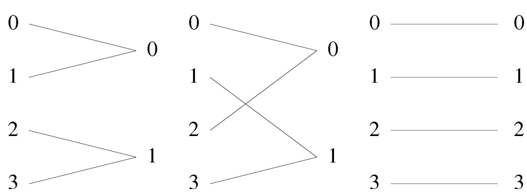

Fig. 5. A deterministic broadcast channel with $\mathcal{U} = \{0,1,2,3\}$, $\mathcal{V}_1 = \mathcal{V}_2 = \{0,1\}$, $\mathcal{V}_3 = \{0,1,2,3\}$.

For a graphical illustration of this fact, see Fig. 4. In fact, the gains can be infinite: Consider a degraded broadcast channel with $U = V_2$ and $I(V_1;V_2) = 0$. Clearly, channel 2 is noiseless and channel 1 cannot convey any messages reliably, and thus $C^{cmp} = 0$ and

$$\mathcal{C}^{dm} = \{(R_1,R_2) : R_1 = 0, 0 \le R_2 \le C_2\}.$$

Now, if $H(X|Y_1) = 0$ and $H(X|Y_2) > 0$, X^n can never be conveyed to both receivers under the compound channel mode no matter how large κ is, whereas any $\kappa \ge \frac{H(X|Y_2)}{C_2}$ is achievable under the degraded message set mode.

Inspired by Theorem 5, one may be inclined to pursue a more general result stating that κ is achievable with separate coding if and only if

$$\Big(H(X|Y_1), H(X|Y_2), \dots, H(X|Y_K)\Big) \in \kappa\mathcal{C}^{dm}$$

for any $K \ge 2$. However, as we now show with an example, that is not true. Let $H(X|Y_1) = H(X|Y_2) = 1$, $H(X|Y_3) = 2$, and $\kappa = 1$. It follows from Theorem 4 that X^n can be reliably decoded at all receivers if and only if one can deliver messages with total rates $R_1 = R_2 = 1$, $R_3 = 2$ to the respective receivers. For the deterministic channel shown in Fig. 5, this is indeed possible by setting $r^{(13)} = r^{(23)} = 1$ and $r^{(\mathcal{S})} = 0$ for all other $\mathcal{S}$. More specifically, the following channel encoder achieves these rates with zero error using the channel only once:

$$\begin{aligned} W^{(13)} &= 1,\ W^{(23)} = 1 \longrightarrow 0\\ W^{(13)} &= 1,\ W^{(23)} = 2 \longrightarrow 1\\ W^{(13)} &= 2,\ W^{(23)} = 1 \longrightarrow 2\\ W^{(13)} &= 2,\ W^{(23)} = 2 \longrightarrow 3. \end{aligned}$$

However, it is not possible to achieve the same total rates with degraded message sets. That is, $(1,1,2) \notin \mathcal{C}^{dm}$. To see this, first observe that one has to use almost all 4^m channel input sequences as codewords to achieve the total rate $R_3 \approx 2$ regardless of whether the message sets are degraded. Since a common message $W^{(123)}$ with rate 1 is to be delivered to all receivers, these 4^m codewords must be partitioned into about 2^m groups, with approximately 2^m elements in each, so that the group index determines $W^{(123)}$, and the specific codeword in the group determines the private message $W^{(3)}$. Now, note that for both channel 1 and channel 2, there are 2^m possible output sequences whose inverse images are of size 2^m uniformly. Thus, the codeword groups must be designed so as to significantly overlap with the inverse images of channel outputs v_1^m and v_2^m simultaneously in order to have $W^{(123)} = \hat{W}_1^{(123)} = \hat{W}_2^{(123)}$ with high probability. On the other hand, such a design is impossible since the intersection of the inverse images for any (v_1^m, v_2^m) pair has *exactly* one element.

In conclusion, for $K \ge 3$, a single-letter expression for the minimum achievable κ in separate source–channel coding remains elusive. However, the following, implied by Lemma 2 and Theorem 4, is a single-letter *necessary* condition for the achievability of κ:

$$\Big(H(X|Y_1), H(X|Y_2), \dots, H(X|Y_K)\Big) \in \kappa\mathcal{R}. \tag{7}$$

Moreover, (7) is sufficient as well if and only if $\mathcal{R}$ is a tight bound on the total rates. In contrast, as we show in the next section, (7) is both *necessary* and *sufficient* in joint source–channel coding.

IV. Achievable Rates in Joint Source–Channel Coding

In this section, we first prove our main result and then discuss its implications.

A. Main Result

Theorem 6: In joint source–channel coding, reliable communication is possible with rate κ if and only if (7) is satisfied.

Proof: We begin with the converse part. Let $P_k^{(m,n)} \to 0$ for a sequence of encoders $f^{(m,n)}$ and decoders $g_k^{(m,n)}$ with a fixed rate $\kappa = \frac{m}{n}$. Let $\hat{X}_k^n = g_k^{(m,n)}(V_k^m, Y_k^n)$.

$$\begin{aligned} \frac{1}{m}\sum_{t=1}^{m} I(U_t;V_{kt}) &\overset{(a)}{\ge} \frac{1}{m} I(U^m;V_k^m)\\ &\overset{(b)}{\ge} \frac{1}{m} I(X^n;V_k^m)\\ &\overset{(c)}{\ge} \frac{1}{m} I(X^n;V_k^m|Y_k^n)\\ &= \frac{1}{m}\Big[H(X^n|Y_k^n) - H(X^n|Y_k^n,V_k^m)\Big]\\ &\overset{(d)}{\ge} \frac{1}{\kappa}H(X|Y_k) - \frac{1}{m}H(X^n|\hat{X}_k^n) \end{aligned}$$

where (a) follows since each channel $P_{V_k|U}$ is memoryless, (b) from the fact that $X^n \leftrightarrow U^m \leftrightarrow V_k^m$ forms a Markov chain, (c) from $Y_k^n \leftrightarrow X^n \leftrightarrow V_k^m$, and finally, (d) from the fact that $\hat{X}_k^n = g_k^{(m,n)}(V_k^m, Y_k^n)$. Using Fano's inequality [5]

$$H(X^n|\hat{X}_k^n) \le 1 + P_k^{(m,n)} n \log|\mathcal{X}|$$

we further obtain

$$\frac{1}{m}\sum_{t=1}^{m} I(U_t;V_{kt}) \geq \frac{1}{\kappa}\left[H(X|Y_k) - \frac{1}{n} - P_k^{(m,n)}\log|\mathcal{X}|\right] \geq \frac{1}{\kappa}[H(X|Y_k)-\epsilon] \quad (8)$$

for any $\epsilon > 0$ and large enough n. Since $\mathcal{R}$ is convex, and

$$\Big(I(U_t,V_{1t}), I(U_t,V_{2t}),\dots,I(U_t,V_{Kt})\Big) \in \mathcal{R}$$

for each t, we have

$$\left(\frac{1}{m}\sum_{t=1}^{m} I(U_t;V_{1t}),\ \dots,\ \frac{1}{m}\sum_{t=1}^{m} I(U_t;V_{Kt})\right) \in \mathcal{R}\,. \quad (9)$$

Combining (8) and (9) proves the converse.

For the direct part, we show that for any $\epsilon > 0$, if

$$\Big(H(X|Y_1)+\epsilon, H(X|Y_2)+\epsilon,\dots,H(X|Y_K)+\epsilon\Big) \in \kappa\mathcal{R}$$

then there exists a sequence of encoders $f^{(m,n)}$ and decoders $g_k^{(m,n)}$ with a fixed rate $\kappa = \frac{m}{n}$ with $P_k^{(m,n)} \to 0$ for all k.

Fix $\delta > 0$, $\gamma > 0$. Select P_U such that

$$H(X|Y_k) \leq \kappa I(U;V_k) - \epsilon, \qquad \text{for all } k.$$

Generate $M = 2^{n[H(X)+\epsilon/2]}$ length-n source words and length-m channel words in an independent and identically distributed (i.i.d.) fashion using P_X and P_U, respectively, where $m = \kappa n$. Denote these words by $x^n(i)$ and $u^m(i)$ for $1 \leq i \leq M$, and reveal them to both the encoder and the decoder.

Encoder: Given X^n, find the smallest i such that $X^n = x^n(i)$, and send $u^m(i)$ through the broadcast channel. If no such i is found, declare an error.

Decoder: At receiver k, find the unique i^* that simultaneously satisfies

$$(x^n(i^*), Y_k^n) \in T^n_{[X,Y_k]_\delta}$$

and

$$(u^m(i^*), V_k^m) \in T^m_{[U,V_k]_\gamma}.$$

If i^* is not unique, or if it does not exist, declare an error.

Probability of error: Define the following events:

$$\mathcal{E}_1 = \{X^n \neq x^n(i) \ \ \forall i \text{ when } X_t \text{ i.i.d.} \sim P_X \text{ and } x_t(i) \text{ i.i.d.} \sim P_X\}$$

$$\mathcal{E}_2(k) = \{(X^n, Y_k^n) \notin T^n_{[X,Y_k]_\delta} \text{ when } (X_t, Y_{kt}) \text{ i.i.d.} \sim P_{X,Y_k}\}$$

$$\mathcal{E}_3(k) = \{(U^m, V_k^m) \notin T^m_{[U,V_k]_\gamma} \text{ when } (U_t, V_{kt}) \text{ i.i.d.} \sim P_{U,V_k}\}$$

$$\mathcal{E}_4(k) = \{(X^n, Y_k^n) \in T^n_{[X,Y_k]_\delta} \text{ when } X_t \text{ i.i.d.} \sim P_X \text{ and } Y_{kt} \text{ i.i.d.} \sim P_{Y_k}\}$$

$$\mathcal{E}_5(k) = \{(U^m, V_k^m) \in T^m_{[U,V_k]_\gamma} \text{ when } U_t \text{ i.i.d.} \sim P_U \text{ and } V_{kt} \text{ i.i.d.} \sim P_{V_k}\}\,.$$

Basically, $\mathcal{E}_1$ is the event that the encoder fails to find a single $x^n(i)$ in the randomly generated source codebook that matches with the current source block X^n. Given that at least one such i exists, i.e., $\mathcal{E}_1^c$ has occured, $\mathcal{E}_2(k)$ corresponds to the event that $x^n(i)$ and the side information Y_k^n are not jointly typical. Conversely, $\mathcal{E}_4(k)$ represents the event that a source word $x^n(j)$, with $j \neq i$, *is* jointly typical with Y_k^n. $\mathcal{E}_3(k)$ and $\mathcal{E}_5(k)$ are similarly defined for the transmitted and received channel words $u^m(i)$ and V_k^m.

Taking into account the joint randomness of X^n, Y_k^n, and the codebooks $\{x^n(i)\}_{i=1}^M$, $\{u^m(i)\}_{i=1}^M$, we can bound the probability of error as

$$P_k^{(m,n)} \leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2(k)] + \Pr[\mathcal{E}_3(k)] + M\Pr[\mathcal{E}_4(k)]\Pr[\mathcal{E}_5(k)]$$

where the last term follows from the union bound on probabilities and the observation that for an error to occur, it suffices to have at least one $x^n(j)$ and $u^m(j)$ with $j \neq i$ which are separately typical with the actual side-information block Y_k^n and the actual received channel word V_k^m.

Now, for any $\lambda > 0$ and $n \geq n_0(\lambda)$, we have

$$\begin{aligned}
\Pr[\mathcal{E}_1] &= (1-\Pr[X^n = x^n(1)])^M \\
&= \left(1 - \sum_{x^n\in\mathcal{X}^n}\Pr[X^n = x^n]^2\right)^M \\
&\overset{\text{(a)}}{=} \left(1 - \sum_{x^n\in\mathcal{X}^n} 2^{-2n[H(P_{x^n})+\mathcal{D}(P_{x^n}\|P_X)]}\right)^M \\
&= \left(1 - \sum_{Q\in\mathcal{P}_n} |T_Q^n| 2^{-2n[H(Q)+\mathcal{D}(Q\|P_X)]}\right)^M \\
&\overset{\text{(b)}}{\leq} \left(1 - \sum_{Q\in\mathcal{P}_n} 2^{n[H(Q)-\lambda]-2n[H(Q)+\mathcal{D}(Q\|P_X)]}\right)^M \\
&\overset{\text{(c)}}{\leq} \left(1 - 2^{n[H(X)-2\lambda]-2n[H(X)+2\lambda]}\right)^M \\
&\overset{\text{(d)}}{\leq} e^{-2^{-n[H(X)+6\lambda]}M} \\
&= e^{-2^{n[\frac{\epsilon}{2}-6\lambda]}}
\end{aligned}$$

where (a) follows from [6, Lemma 1.2.6], (b) from [6, Lemma 1.2.3], (c) by choosing a single type in the summation that simultaneously satisfies $|H(X)-H(Q)| \leq \lambda$ and $\mathcal{D}(Q\|P_X) \leq \lambda$, and finally (d) from the well-known inequality

$$(1-t)^s \leq e^{-st}$$

for $0 \leq t \leq 1$ and $s \geq 0$. Choosing $\lambda < \frac{\epsilon}{12}$, the right-hand side vanishes as $n \to \infty$. That $\Pr[\mathcal{E}_2] \to 0$ and $\Pr[\mathcal{E}_3] \to 0$ as $n \to \infty$ and $m \to \infty$ can be also shown in a standard fashion (cf. [6, Lemma 1.2.12]). Now, it is also a matter of folklore to show

$$\Pr[\mathcal{E}_4] \leq 2^{-n[I(X;Y_k)-\lambda]}$$

and

$$\Pr[\mathcal{E}_5] \leq 2^{-m[I(U;V_k)-\lambda]}$$

for any $\lambda > 0$, $n > n_1(\lambda)$, and $m = \kappa n$. Therefore,

$$\begin{aligned}
M\Pr[\mathcal{E}_4]\Pr[\mathcal{E}_5] &\leq 2^{n[H(X)+\frac{\epsilon}{2}]-n[I(X;Y_k)-\lambda]-m[I(U;V_k)-\lambda]} \\
&= 2^{-n[\kappa I(U;V_k)-H(X|Y_k)-(\kappa+1)\lambda-\frac{\epsilon}{2}]} \\
&\leq 2^{-n[\frac{\epsilon}{2}-(\kappa+1)\lambda]}.
\end{aligned}$$

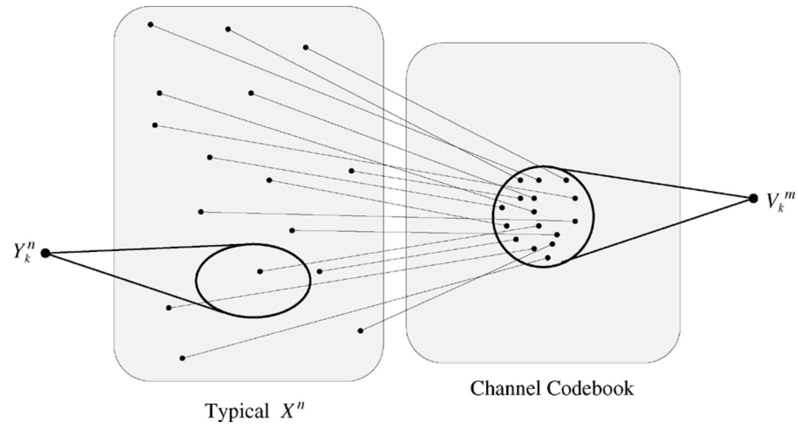

Fig. 6. The virtual binning at the kth receiver. Dots inside the channel codebook and the typical source set represent codewords typical with V_k^m and the corresponding source words, respectively. With high probability, only one source word is jointly typical with Y_k^n.

Choosing

$$\lambda < \frac{\epsilon}{2(\kappa+1)}$$

this last error term also vanishes. □

B. Discussion

A useful interpretation of our main result is that $\mathcal{R}$, a trivial outer bound on the total rates a general broadcast channel can achieve, turns out to be the *effective* capacity region for joint source–channel coding in our Slepian–Wolf setup. The achievability part of the proof of Theorem 6 also reveals other interesting phenomena, which we discuss next.

Our first observation is that even though what is described in the proof is a "joint" coding scheme, the source and the channel variables (and hence words) are statistically independent. That is, because the joint mechanism can be separated into source and channel components which operate independently. More specifically, on the encoder side, the source is directly mapped into a message set of size $\approx 2^{nH(X)}$, and the resultant message is mapped to a random channel word generated according to P_U. On the decoder side, even though the local channel decoders cannot resolve the message, each one can be thought of as providing an exponentially large *list* of possible channel codeword indices, and hence, a list of possible X^n. Then each local source decoder uses this list and the local side information to resolve the actual X^n. This is still a joint source–channel coding scheme, because neither the source nor the channel components are stand-alone, and only when they are used together in the described manner is reliable transmission possible. Thus, we have operational separation, but not necessarily informational separation. Indeed, as we show in the next section, informational separation does not hold in general.

Even though there is no explicit binning in the described joint coding system, an implicit one is performed by the received channel words.[5] In fact, what we refer to as lists above can be thought of as "virtual bins." More specifically, for each typical V_k^m, the list comprises of $\approx 2^{n[H(X)-\kappa I(U;V_k)]}$ channel words jointly typical with V_k^m. Since the mapping between source words and channel words is arbitrary, one can think of the induced source words (i.e., the inverse image of the list) as a randomly created bin. Because $I(X;Y_k) \geq H(X) - \kappa I(U;V_k)$, it follows that with probability approaching one, only the correct source word falls into the intersection of this bin and the set of x^n jointly typical with Y_k^n. Fig. 6 provides an illustration of this phenomenon. Note that we are fully exploiting the fact that the channel quality increases as the quality of the side information decreases. (In Fig. 6, this corresponds to a larger image of Y_k^n on the typical source vectors and a smaller image of V_k^m on the channel codebook.) Note also that there is significant overlap between the virtual bins when $H(V_k|U) > 0$. To see this, observe that there are $\approx 2^{mH(V_k)}$ bins and $\approx 2^{n[H(X)-\kappa I(U;V_k)]}$ source vectors in each bin, but

$$2^{mH(V_k)} 2^{n[H(X)-\kappa I(U;V_k)]} = 2^{nH(X)} 2^{mH(V_k|U)}.$$

Since joint source–channel coding performs no explicit binning at the encoder, can we utilize it as an alternative to the traditional binning solution in the Wolf setup [24] mentioned in Section I? The answer is negative, because when the channel is noiseless, the elements of the codeword list generated at a receiver will be identical (equal to V_k^m), resulting in a fixed partitioning on the source domain. Alternatively, since $H(V_k|U) = 0$, there is almost no overlap between bins and thus virtual binning reduces to traditional binning for noiseless channels.

V. Comparison of Achievable Rates in Separate Versus Joint Source–Channel Coding

For $K = 1$, our setup reduces to a regular point-to-point scenario for which informational source–channel separation is known [20]. In this section, we investigate whether such separation holds in some well-studied broadcast channels for $K = 2$. It follows from Theorems 5 and 6 that it suffices to check whether

$$\mathcal{C}^{dm} = \mathcal{R}'$$

where

$$\mathcal{R}' \triangleq \mathcal{R} \cap \{(R_1, R_2) : R_2 \geq R_1\}$$

[5]We exclusively discuss here the interesting case $H(X) \geq \kappa I(U;V_k)$. Otherwise, with high probability, there is only one channel word jointly typical with V_k^m, and thus X^n can be decoded even without the help of Y_k^n.

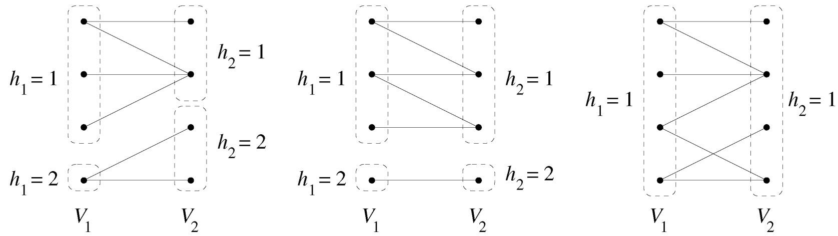

Fig. 7. Examples of common parts of V_1 and V_2. Lines indicate v_1, v_2 pairs with $P_{V_1,V_2}(v_1,v_2) > 0$. Any *connected* component of the graph constitutes $h_1 = h_2 = j$.

or, equivalently, whether all total rate pairs $(R_1, R_2) \in \mathcal{R}'$ are achievable for the broadcast channel.

A. Degraded Broadcast Channels

We analyze here the case where the same choice of U, denoted U^*, simultaneously achieves the maximum of both $I(U;V_1)$ and $I(U;V_2)$, which we assume to be nonzero to eliminate uninteresting cases. This is indeed possible, for example, when we have a binary symmetric broadcast channel, i.e., when $\mathcal{U} = \mathcal{V}_1 = \mathcal{V}_2 = \{0,1\}$ and both $P_{V_1|U}$ and $P_{V_2|U}$ are binary-symmetric channels with respective crossover probabilities p_1 and p_2 satisfying $p_2 < p_1 < 1/2$. In particular, $P_U = \{\frac{1}{2}, \frac{1}{2}\}$ achieves both $I(U;V_1) = C_1 = 1 - \mathcal{H}(p_1)$ and $I(U;V_2) = C_2 = 1 - \mathcal{H}(p_2)$, where $\mathcal{H}(\cdot)$ is the binary entropy function

$$\mathcal{H}(p) = -p\log p - (1-p)\log(1-p)$$

and C_k denotes the ordinary capacity of channel $P_{V_k|U}$. When such U^* exists, we have

$$\mathcal{R}' = \{(R_1,R_2) : 0 \le R_1 \le R_2,\, R_1 \le C_1,\, R_2 \le C_2\}\,.$$

We derive in the next lemma a key condition on $\mathcal{C}^{dm} = \mathcal{R}'$.

Lemma 5: When some U^* achieves both $I(U^*;V_1) = C_1$ and $I(U^*;V_2) = C_2$ simultaneously, $\mathcal{C}^{dm} = \mathcal{R}'$ if and only if there exists Z with $Z \leftrightarrow U^* \leftrightarrow V_2 \leftrightarrow V_1$ satisfying

$$I(U^*;V_1|Z) = 0 \tag{10}$$

$$I(Z;V_2|V_1) = 0. \tag{11}$$

Proof: Let $(R_1,R_2) \in \mathcal{C}^{dm}$. Then there exists Z and U with $Z \leftrightarrow U \leftrightarrow V_2 \leftrightarrow V_1$ such that (2) and (3) are satisfied. From data processing inequality, we have $I(Z;V_1) \le I(U;V_1)$. It also follows from $Z \leftrightarrow U \leftrightarrow V_2 \leftrightarrow V_1$ that

$$\begin{aligned} I(Z;V_1) + I(U;V_2|Z) &\le I(Z;V_2) + I(U;V_2|Z) \\ &= I(U,Z;V_2) \\ &= I(U;V_2)\,. \end{aligned}$$

Since $I(U;V_1) \le C_1$ and $I(U;V_2) \le C_2$, it follows that $\mathcal{C}^{dm} \subset \mathcal{R}'$, as expected. On the other hand, this analysis also implies that $\mathcal{C}^{dm} = \mathcal{R}'$ if and only if there exists Z with $Z \leftrightarrow U^* \leftrightarrow V_2 \leftrightarrow V_1$ satisfying

$$I(Z;V_1) = I(Z;V_2) = I(U^*;V_1) = C_1 \tag{12}$$

which can easily be reformulated as (10) and (11). □

Corollary 1: $\mathcal{C}^{dm} = \mathcal{R}'$ for all degraded channels with deterministic components $P_{V_2|U}$ and $P_{V_1|V_2}$ if some U^* achieves both $I(U^*;V_1) = C_1$ and $I(U^*;V_2) = C_2$ simultaneously.

Proof: It suffices to find Z with $Z \leftrightarrow U^* \leftrightarrow V_2 \leftrightarrow V_1$ satisfying (10) and (11). Create $P_{Z|U^*}$ such that those u with $P_{U^*}(u) > 0$ yielding the same v_1 under $P_{V_1|U^*}(\cdot|u)$ are grouped together and deterministically mapped to the same z. This implies $Z = V_1$ without loss of generality, and both (10) and (11) are immediately satisfied. □

We next introduce a mechanism to create examples of $\mathcal{C}^{dm} \neq \mathcal{R}'$ using the concept of *common part* of two random variables. For P_{V_1,V_2} induced by U^*, let J be the largest integer for which there exists deterministic functions

$$h_1 : \mathcal{V}_1 \longrightarrow \{1,2,\ldots,J\}$$

$$h_2 : \mathcal{V}_2 \longrightarrow \{1,2,\ldots,J\}$$

with $\Pr[h_1(V_1) = j] > 0$ and $\Pr[h_2(V_2) = j] > 0$ for $j = 1,2,\ldots,J$, such that $h_1(V_1) = h_2(V_2)$ with probability 1. The common part W of V_1 and V_2 is then defined as $W = h_1(V_1) = h_2(V_2)$. We say that V_1 and V_2 has no common part if $J = 1$. This concept is illustrated in Fig. 7.

Lemma 6: For P_{V_1,V_2} induced by U^*, if V_1 and V_2 have no common part, $\mathcal{C}^{dm} \neq \mathcal{R}'$.

Remark 5: It immediately follows that $\mathcal{C}^{dm} \neq \mathcal{R}'$ for binary-symmetric broadcast channels, as pointed out in [3].

Proof: Assume $\mathcal{C}^{dm} = \mathcal{R}'$. Then, from Lemma 5, there exists Z with $Z \leftrightarrow U^* \leftrightarrow V_2 \leftrightarrow V_1$ satisfying (10) and (11). But (11) implies that $Z \leftrightarrow V_1 \leftrightarrow V_2$ as well. This, in turn, yields

$$P_{Z|V_1,V_2}(\cdot|v_1,v_2) = P_{Z|V_2}(\cdot|v_2) = P_{Z|V_1}(\cdot|v_1)$$

for all v_1, v_2 such that $P_{V_1,V_2}(v_1,v_2) > 0$. Therefore, for any v_1 with $P_{V_1}(v_1) > 0$, all v_2 satisfying $P_{V_2|V_1}(v_2|v_1) > 0$ must exhibit the same $P_{Z|V_2}(\cdot|v_2)$. In other words, $P_{Z|V_2}(\cdot|v_2)$ must be constant for $v_2 \in h_2^{-1}(j), 1 \le j \le J$, where (h_1,h_2) defines the common part of (V_1,V_2). But since there is no common part, i.e., $J = 1$, this implies the independence of Z and V_2, or $I(Z;V_2) = 0$, which conflicts with (12) since $C_1 > 0$. □

We close this section with an investigation of how large the *gain* in minimum achievable κ can be when a joint source–channel code is used instead of separate source and channel codes for a binary-symmetric broadcast channel. We define the gain as

$$G(p_1,p_2) \triangleq \frac{\kappa_\mathrm{S}}{\kappa_\mathrm{J}}$$

where κ_S and κ_J, respectively, refer to the minimum rates in separate and joint source–channel coding. To evaluate the gain, we use the characterization of $\mathcal{C}^{dm}$ given in [5]

$$R_1 \le [1 - \mathcal{H}(\beta \star p_1)] \tag{13}$$

$$R_2 \le [1 - \mathcal{H}(\beta \star p_1) + \mathcal{H}(\beta \star p_2) - \mathcal{H}(p_2)] \tag{14}$$

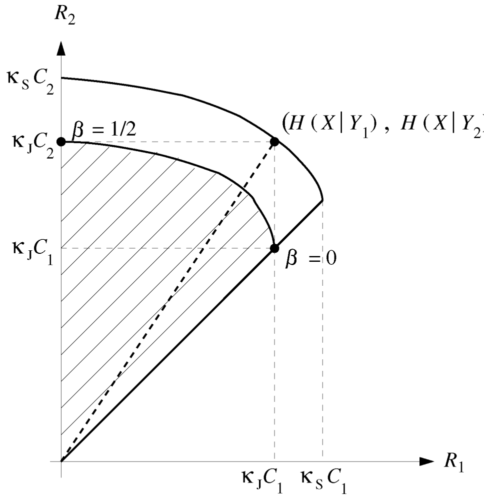

Fig. 8. The scaled version of the capacity region, $\kappa_J \mathcal{C}^{dm}$, for a binary-symmetric broadcast channel is shown as the shaded region. It is assumed that $H(X|Y_1) = \kappa_J C_1$ and $H(X|Y_2) = \kappa_J C_2$, which implies κ_J is indeed the minimum rate in joint source–channel coding. On the other hand, for separate source–channel coding to be possible, the capacity region must be scaled by a larger constant, namely, κ_S.

where $a \star b = a(1-b) + b(1-a)$. Here, β controls the tradeoff between the two constraints. More specifically, when $\beta = 0$, right-hand sides of (13) and (14) both achieve C_1, and when $\beta = 1/2$, they become 0 and C_2, respectively. Fig. 8 shows the capacity region $\mathcal{C}^{dm}$ scaled by κ_J and κ_S, where κ_J and κ_S are determined for a source $\{X, Y_1, Y_2\}$ that satisfies

$$\frac{H(X|Y_2)}{H(X|Y_1)} = \frac{C_2}{C_1}.$$

It can easily be deduced from Theorem 6 that

$$\kappa_J = \frac{H(X|Y_2)}{C_2} = \frac{H(X|Y_1)}{C_1}. \tag{15}$$

In fact, the same minimum rate κ_J would occur if we relax one (and only one) of the equalities in (15) to an inequality of direction $>$. However, it can be rigorously shown that this particular source is the most difficult to encode for a separate source–channel coding system. For an informal proof of this claim, the reader is referred to Fig. 8. When we increase κ further than κ_J, the point $(\kappa_J C_1, \kappa_J C_2)$ is the last to be included in $\kappa \mathcal{C}^{dm}$ among all $(R_1, \kappa_J C_2)$ with $R_1 \le \kappa_J C_1$, and among all $(\kappa_J C_1, R_2)$ with $R_2 \le \kappa_J C_2$.

To find κ_S, it then suffices to find the point (R_1, R_2) on the boundary of $\mathcal{C}^{dm}$ satisfying

$$\frac{R_2}{R_1} = \frac{C_2}{C_1} \tag{16}$$

and use

$$\kappa_S = \kappa_J \frac{C_1}{R_1} = \kappa_J \frac{C_2}{R_2}.$$

Unfortunately, solving (16) analytically proves to be a tedious task. Resorting to numerical solution of (16), we obtain $G(p_1, p_2)$ as shown in Fig. 9, and observe that $G(p_1, p_2)$ reaches its maximum value of ≈ 2 when $p_1 \approx 0.5$ and $p_2 \approx 0.37$.

B. Deterministic Broadcast Channels

In this subsection, we extend Corollary 1 to all deterministic channels. A broadcast channel is called deterministic if $P_{V_1,V_2|U}(v_1, v_2|u)$ only assumes 0 or 1, which implies that V_1 and V_2 are deterministic functions of U, denoted by $f_1(U)$ and $f_2(U)$. Similar to the case in the previous subsection, we consider the cases where the same U, denoted by U^*, maximizes both $I(U;V_1)$ and $I(U;V_2)$ simultaneously. This implies

$$\mathcal{R}' = \{(R_1, R_2) : 0 \le R_1 \le R_2,\ R_1 \le C_1,\ R_2 \le C_2\}.$$

One such example is known as the Blackwell channel, where $\mathcal{U} = \{0, 1, 2\}$, $\mathcal{V}_1 = \mathcal{V}_2 = \{0, 1\}$

$$f_1(u) = \begin{cases} 0 & u = 0 \text{ or } u = 1 \\ 1 & u = 2 \end{cases}$$

and

$$f_2(u) = \begin{cases} 0 & u = 0 \\ 1 & u = 1 \text{ or } u = 2. \end{cases}$$

For this channel, $P_U(u) = \{\frac{1}{2}, 0, \frac{1}{2}\}$ satisfies $I(U;V_1) = C_1 = 1$ and $I(U;V_2) = C_2 = 1$.

As for $\mathcal{C}^{dm}$, we utilize the following theorem which specializes the full characterization of $\mathcal{C}$ derived in [11] to the case of degraded message sets.

Theorem 7: For deterministic broadcast channels with $K = 2$, $\mathcal{C}^{dm}$ is the closure of all (R_1, R_2) for which $R_1 \le R_2$ and there exists a distribution $P_Z P_{U|Z} P_{V_1,V_2|U}$ satisfying

$$R_1 \le \min\{I(Z;V_1), I(Z;V_2)\} \tag{17}$$

$$R_2 \le \left[\min\{I(Z;V_1), I(Z;V_2)\} + H(V_2|Z)\right] \tag{18}$$

$$= \min\{I(Z;V_1) + H(V_2|Z), H(V_2)\}. \tag{19}$$

From data processing inequality, we have $I(Z;V_1) \le I(U;V_1) \le C_1$, and from the fact that $V_2 = f_2(U)$, we observe $H(V_2) = I(U;V_2) \le C_2$. This verifies, as a sanity check, $\mathcal{C}^{dm} \subset \mathcal{R}'$. On the other hand, the next theorem proves that $\mathcal{C}^{dm} = \mathcal{R}'$.

Theorem 8: If U^* achieves

$$I(U^*;V_1) = C_1 \quad \text{and} \quad I(U^*;V_2) = C_2$$

simultaneously, then $\mathcal{C}^{dm} = \mathcal{R}'$.

Proof: We prove $\mathcal{R}' \subset \mathcal{C}^{dm}$ for the cases of $C_2 \le C_1$ and $C_2 > C_1$ separately.

For $C_2 \le C_1$, $\mathcal{R}'$ reduces to the collection of (R_1, R_2) such that $0 \le R_1 \le R_2 \le C_2$. Therefore, it suffices to show $(C_2, C_2) \in \mathcal{C}^{dm}$. Using (17) and (18) with $Z = U^*$ yields the desired result. The Blackwell channel is an example for this case, as it satisfies $C_1 = C_2 = 1$.

For $C_2 > C_1$, we shall show $(C_1, C_2) \in \mathcal{C}^{dm}$. This, according to (17) and (18), is possible if and only if there exists Z with $Z \leftrightarrow U^* \leftrightarrow V_1 V_2$ satisfying

$$\min\{I(Z;V_1), I(Z;V_2)\} = C_1 \tag{20}$$

$$\min\{I(Z;V_1), I(Z;V_2)\} + H(V_2|Z) = C_2 \tag{21}$$

simultaneously. Equivalently, we must have

$$H(V_2|Z) = C_2 - C_1 \tag{22}$$

$$H(V_1|Z) = 0. \tag{23}$$

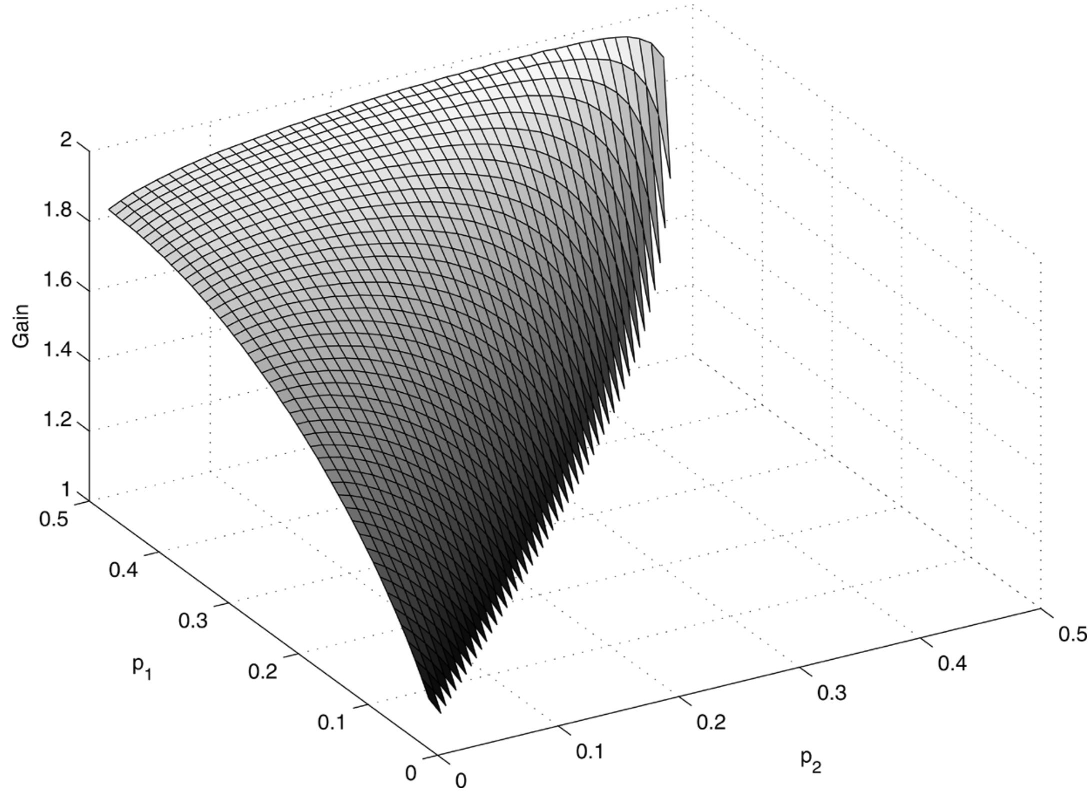

Fig. 9. $G(p_1, p_2)$ obtained by numerical solution of (16).

Here, (22) follows directly from (20) and (21), and (23) follows because $H(V_1) = I(U^*; V_1) = C_1$ and $H(V_2) = I(U^*; V_2) = C_2$, and hence,

$$\begin{aligned}&\min\{I(Z; V_1), I(Z; V_2)\}\\&\quad = \min\{C_1 - H(V_1|Z), C_2 - H(V_2|Z)\}\\&\quad = \min\{C_1 - H(V_1|Z), C_1\}.\end{aligned}$$

Now, $H(V_1|Z) = 0$ implies that $V_1 \leftrightarrow Z \leftrightarrow U^* \leftrightarrow V_2$, and therefore,

$$0 \le H(V_2|Z) \le H(V_2|V_1)$$

where the former and the latter inequalities are tight when $Z = U^*$ and when $Z = V_1$, respectively. Moreover, Z can be properly chosen so that $H(V_2|Z)$ takes any desired value between 0 and $H(V_2|V_1)$. The proof of this claim will essentially finish the proof of the theorem, for

$$H(V_2|V_1) \ge H(V_2) - H(V_1) = C_2 - C_1$$

where $H(V_2|V_1) \ge H(V_2) - H(V_1)$ immediately follows from $H(V_1) \ge I(V_1; V_2)$.

Toward the goal of proving that $H(V_2|Z)$ can take any value between 0 and $H(V_2|V_1)$ with an appropriate choice of Z, we first observe that $Z \leftrightarrow U^* \leftrightarrow V_1V_2$ is automatically satisfied for any Z due to the deterministic nature of the channel. Thus, we can solely focus on the condition $V_1 \leftrightarrow Z \leftrightarrow U^* \leftrightarrow V_2$ when choosing Z. Second, $P_{Z|V_1}(z|v_1)$ and $P_{U^*|Z}(u|z)$ can take arbitrary values, provided they form legitimate probability distributions and satisfy

$$\sum_{z \in \mathcal{Z}} P_{Z|V_1}(z|v_1) P_{U^*|Z}(u|z) = P_{U^*|V_1}(u|v_1). \tag{24}$$

In other words, for each $v_1 \in \mathcal{V}_1$, the (known) vector $P_{U^*|V_1}(\cdot|v_1)$ must be in the convex closure of the (chosen) vectors $P_{U^*|Z}(\cdot|z)$, where $P_{Z|V_1}(z|v_1)$ plays the role of convex combination coefficients. Two possible extreme choices mentioned before are i) to use a single z for each v_1, i.e., create a deterministic $P_{Z|V_1}$, and let $P_{U^*|Z}(\cdot|z) = P_{U^*|V_1}(\cdot|v_1)$ whenever $P_{Z|V_1}(z|v_1) = 1$, thus yielding $H(V_2|Z) = H(V_2|V_1)$, and ii) to use as many as $|f_1^{-1}(v_1)|$ different z values for each v_1, and let $P_{U^*|Z}$ be deterministic, so that $P_{Z|V_1}(z|v_1) = P_{U^*|V_1}(u|v_1)$ whenever $P_{U^*|Z}(u|z) = 1$, resulting in $H(V_2|Z) = 0$. See Fig. 10(a) and (b) for an illustration of this discussion for the Blackwell channel. Fig. 10(c) shows a general choice of $P_{Z|V_1}$ and $P_{U^*|Z}$. Since this general choice affects $H(V_2|Z)$ *continuously*, we conclude that there must exist some choice so that $H(V_2|Z)$ attains the desired value in between 0 and $H(V_2|V_1)$. □

C. Other Channels Where Separation Holds

In [3], it was shown that for some classes of channels $\mathcal{R}$ is a tight outer bound on the total rates. In this section, we discuss these channels briefly.

The Bottleneck Channel: The bottleneck channel satisfies

$$P_{V_1|U}(v|u) = P_{V_2|U}(v|u)$$

for all $u \in \mathcal{U}$ and $v \in \mathcal{V}_1 = \mathcal{V}_2$. Thus, both receivers "see" the same channel, and are able to reliably decode the same message set with rate $C^{cmp} = C$, where C is the point-to-point capacity of $P_{V_1|U}$. Based on this observation, it follows that

$$\mathcal{R}' = \mathcal{C}^{dm} = \{(R_1, R_2) : 0 \le R_1 \le R_2 \le C\}.$$

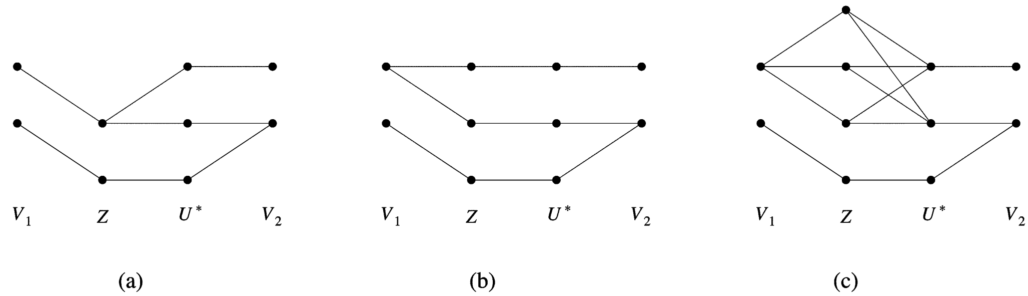

Fig. 10. Possible ways to satisfy (24) for the Blackwell channel. (a) $P_{U^*|Z} = P_{U^*|V_1}$ and deterministic $P_{Z|V_1}$, (b) deterministic $P_{U^*|Z}$ and $P_{Z|V_1} = P_{U^*|V_1}$, and (c) more general choice of $P_{U^*|Z}$.

Orthogonal Channels: Let $\mathcal{U} = \{1,2,3,4\}$, $\mathcal{V}_1 = \mathcal{V}_2 = \{1,2\}$, and

$$P_{V_1|U} = \begin{bmatrix} \alpha & 1-\alpha \\ \alpha & 1-\alpha \\ 1-\alpha & \alpha \\ 1-\alpha & \alpha \end{bmatrix} \quad P_{V_2|U} = \begin{bmatrix} \beta & 1-\beta \\ 1-\beta & \beta \\ 1-\beta & \beta \\ \beta & 1-\beta \end{bmatrix}.$$

Since $P_U(u) = \frac{1}{4}$ maximizes both $I(U;V_1)$ and $I(U;V_2)$ simultaneously, we have that

$$\mathcal{R}' = \{(R_1, R_2) : 0 \le R_1 \le R_2,\, R_1 \le C_1,\, R_2 \le C_2\}$$

where $C_1 = 1 - \mathcal{H}(\alpha)$ and $C_2 = 1 - \mathcal{H}(\beta)$ are the ordinary capacities of $P_{V_1|U}$ and $P_{V_2|U}$, respectively. From the discussion in [3], it also follows that

$$\mathcal{C}^{dm} = \Big\{(R_1,R_2) : 0 \le R_1 \le R_2, R_1 \le \min(C_1,C_2),\, R_2 \le C_2\Big\}.$$

Optimality of separate coding is then obvious when $C_2 \ge C_1$. When $C_1 > C_2$, the conditions $R_1 \le C_1$ and $R_1 \le \min(C_1, C_2)$ for $\mathcal{R}'$ and $\mathcal{C}^{dm}$, respectively, become vacuous, and

$$\mathcal{R}' = \mathcal{C}^{dm} = \{(R_1, R_2) : 0 \le R_1 \le R_2 \le C_2\}.$$

Switch-to-Talk Channel: Let

$$P_{V_1|U} = \begin{bmatrix} * & * & \cdots & * & 0 \\ * & * & \cdots & * & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & \cdots & * & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

and

$$P_{V_2|U} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ * & * & \cdots & * & 0 \\ * & * & \cdots & * & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ * & * & \cdots & * & 0 \end{bmatrix}.$$

Equivalently, there are two separate channels, with capacities C_1^* and C_2^*, between the sender and each of the receivers, and both receivers have indicators that signal when the sender is communicating with the other receiver. It was shown in [3] that (R_1, R_2) is an achievable total rate pair if and only if there exists α satisfying

$$R_1 \le \alpha C_1^* + \mathcal{H}(\alpha)$$
$$R_2 \le (1-\alpha) C_2^* + \mathcal{H}(\alpha).$$

It is also straightforward to show $\mathcal{R}$ consists of the same rate pairs, as pointed out in [3].

Incompatible Channels: Let

$$P_{V_1|U} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad P_{V_2|U} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

which implies that if the sender wishes to communicate with one particular receiver, all the other receiver sees is pure noise. In this case, by simple time sharing arguments one can show the achievability of a pair of total rates (R_1, R_2) when $R_1 + R_2 \le 1$. In [3], it was shown that

$$\mathcal{R} = \{(R_1, R_2) : R_1 \ge 0, R_2 \ge 0, R_1 + R_2 \le 1\}.$$

Therefore, $\mathcal{R}' = \mathcal{C}^{dm}$ follows.

VI. Conclusion and Future Directions

We considered the problem of Slepian–Wolf coding over broadcast channels, where each decoder at the receiving end of the channel has access to side information (possibly of different quality) about the source. For both separate and joint source–channel coding strategies, we characterized the minimum achievable rate in terms of channel uses per source symbol. The characterization for joint source–channel coding is particularly powerful, because it is not only of single-letter nature, but also easy to compute. On the contrary, the minimum separate coding rate is computable only if the region of achievable *total* message rates at each receiver is known. For $K = 2$, this is the same as the capacity region with degraded message sets. However, as we have shown with an example, for $K \ge 3$, the capacity region with degraded message sets does not fully characterize the whole set of achievable total rates.

We discussed whether separate coding with stand-alone source and channel coders is optimal for some well-studied channels with $K = 2$. Since there exists at least one case where it is not, the general conclusion is that this communication scenario is not separable.

The main strength of the simple joint source–channel coding approach of Theorem 6 is that it employs independent (albeit not stand-alone) source and channel components leading to statistically unrelated source and channel variables in the characterization of achievable rates. It would certainly be interesting to find other multiterminal communication scenarios in which the separation of variables could be repeated. We shall note here that our first candidate, lossless coding of correlated sources over multiple-access channels as in [4], turned out not to be such a scenario, because the simple approach of Theorem 6 results in the same minimum rate as in separate source–channel coding.

Another future research goal is to generalize our result to lossy coding with a fidelity criterion. Of special interest would be when side information is absent at the decoders. Even in that simplified case, the minimum achievable rate in channel uses per symbol is not known despite increased activity in the last decade [16]–[18], [23].

ACKNOWLEDGMENT

The author is grateful to Ram Zamir and the Associate Editor for their comments which resulted in significant improvement of the content and the presentation of this paper.

REFERENCES

[1] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 2, pp. 197–207, Mar. 1973.

[2] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, 1959.

[3] T. M. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 2–14, Jan. 1972.

[4] T. M. Cover, A. El Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 6, pp. 648–657, Nov. 1980.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.

[7] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 2, pp. 166–169, Mar. 1979.

[8] A. El Gamal and E. Van der Meulen, "A proof of Marton's coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 120–122, Jan. 1981.

[9] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Probl. Pered. Inform.*, vol. 10, no. 3, pp. 3–14, 1974.

[10] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source–channel communication revisited," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.

[11] T. S. Han, "The capacity region for the deterministic broadcast channel with a common message," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 122–125, Jan. 1981.

[12] T. S. Han and M. H. M. Costa, "Broadcast channels with arbitrarily correlated sources," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 5, pp. 641–650, Sep. 1987.

[13] C. Heegard and T. Berger, "Rate-distortion when the side information may be absent," *IEEE Trans. Inf. Theory*, vol. IT-31, no. 5, pp. 727–734, Nov. 1985.

[14] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. IT-23, no. 1, pp. 60–64, Jan. 1977.

[15] K. Marton, "A coding theorem for the discrete broadcast channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 3, pp. 306–311, May 1979.

[16] U. Mittal and N. Phamdo, "Hybrid digital–analog (HDA) joint source–channel codes for broadcasting and robust communications," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1082–1102, May 2002.

[17] Z. Reznic, R. Zamir, and M. Feder, "Joint source–channel coding of a Gaussian mixture source over the Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 776–781, Mar. 2002.

[18] ——, "Distortion bounds for broadcasting with bandwidth expansion," *IEEE Trans. Inf. Theory*, submitted for publication.

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

[20] S. Shamai (Shitz) and S. Verdú, "Capacity of channels with side information," *Europ. Trans. Commun.*, vol. 6, no. 5, pp. 587–600, Sep./Oct. 1995.

[21] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

[22] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.

[23] E. Tuncel, "On optimal multiresolution source-channel coding across degraded broadcast channels," in *Proc. IEEE Int. Symp. Information Theory*, Chicago, IL, Jun./Jul. 2004, p. 32.

[24] J. K. Wolf, "Source coding for a noiseless broadcast channel," in *Proc. Conf. Information Science and Systems*, Princeton, NJ, Mar. 2004, pp. 666–671.

[25] J. Wolfowitz, *Coding Theorems of Information Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1964.