An Achievable Rate Region for Joint Compression and Dispersive Information Routing for Networks

Kumar B. Viswanatha, *Member, IEEE*, Emrah Akyol, *Member, IEEE*, and Kenneth Rose, *Fellow, IEEE*

***Abstract*— This paper considers the problem of minimum cost communication of correlated sources over a network with multiple sinks, which consists of distributed source coding followed by routing. We introduce a new routing paradigm called dispersive information routing (DIR), wherein the intermediate nodes are allowed to split a packet and forward subsets of the received bits on each of the forward paths. This paradigm opens up a rich class of research problems, which focus on the interplay between encoding and routing in a network. Unlike conventional routing methods such as in [1], DIR ensures that each sink receives just the information needed to reconstruct the sources it is required to reproduce. We demonstrate using simple examples that the proposed approach offers better asymptotic performance than conventional routing techniques. We show that, under certain assumptions on the cost function, the problem of finding the minimum cost under DIR essentially reduces to characterizing an achievable rate region for a new multiterminal information theoretic setup. While it is possible to derive an achievable region for this setup using prior results from general multiterminal source coding [3], these techniques do not exploit the underlying problem structure and thereby lead to suboptimal regions. In this paper, we propose a new coding scheme, using principles from multiple descriptions encoding [2], and show that it strictly improves upon a corresponding variant of coding scheme in [3]. We further show that the new coding scheme achieves the complete rate region for certain special cases of the general setup and thereby achieves the minimum communication cost under this routing paradigm.**

***Index Terms*— Distributed source coding, achievable region, minimum cost routing, compression for networks.**

I. Introduction

COMPRESSION of sources within a network has been an important research area, notably with the recent advancements in distributed compression of correlated sources and network (routing) design. Encoding correlated sources in a network, such as a sensor network with multiple nodes and sinks as shown in Fig. 1, has conventionally been approached from two different directions. The first approach is routing the information from different sources in such a way as to efficiently re-compress the data at intermediate nodes, without recourse to distributed source coding (DSC) methods [4], [5] (we refer to this approach as joint coding via 'explicit communication'). Such techniques tend to be wasteful at all but the last hops of the communication path. The second approach performs DSC followed by simple routing [1], [6]. Well designed DSC followed by optimal routing can provide good performance gains. We will focus on the latter category. Relevant background on DSC and route selection in a network is given in the next section.

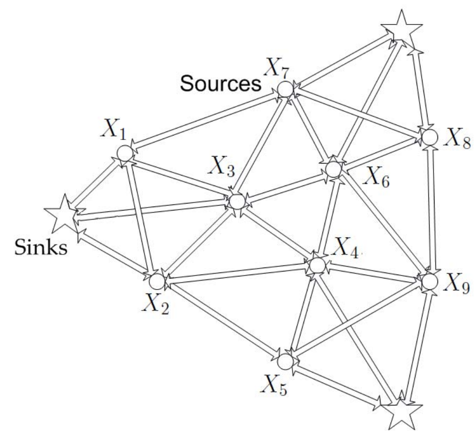

Fig. 1. A general multi-source multi-sink sensor network. The circles denote sources and stars denote sinks. The arrows denote allowed communication links.

This paper focuses on minimum cost communication of correlated sources over a network with multiple-sinks. We introduce a new routing paradigm called Dispersive Information Routing (DIR), wherein intermediate nodes are allowed to "split a packet" and forward a subset of the received bits on each of the forward paths. This paradigm opens up a rich class of research problems which focus on the interplay between encoding and routing in a network. What makes it particularly interesting is the challenge in encoding sources such that exactly the required information is routed to each sink, to reconstruct the prescribed subset of sources. We will show, using simple examples that asymptotically, DIR achieves a lower cost over conventional routing methods,

Manuscript received September 19, 2012; revised November 25, 2013; accepted April 28, 2014. Date of publication July 1, 2014; date of current version August 14, 2014. This work was supported by the National Science Foundation under Grant CCF-1016861, Grant CCF-1118075, and Grant CCF-1320599. The material in this paper was presented in part at the 2010 IEEE International Symposium on Information Theory, 2010 IEEE Information Theory Workshop, and 2011 IEEE International Symposium on Information Theory.

K. B. Viswanatha was with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA. He is now with the Qualcomm Research Center, San Diego, CA 92121 USA (e-mail: kumar@ece.ucsb.edu).

E. Akyol is with the Electrical and Computer Engineering Department, University of California at Santa Barbara, Santa Barbara, CA 93106 USA and with the Electrical Engineering Department, University of Southern California, Los Angeles, CA 90089 USA (e-mail: eakyol@usc.edu).

K. Rose is with the Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: rose@ece.ucsb.edu).

Communicated by C. Fragouli, Associate Editor for Communication Networks.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2014.2334307

wherein the sinks usually receive more information than they need. We then show that, when the cost function is linear in the transmit rates, the problem of finding the minimum communication cost under DIR is equivalent to the problem of establishing the set of achievable rates for a general class of problems in multi-terminal source coding, which have not been studied earlier. In this paper, we formulate this problem and the associated rate region. An achievable region for this new setup can be derived based on results due to Han and Kobayashi for general source networks in [3]. However, this natural extension enforces conditional independence of transmitted messages, which fails to exploit the underlying problem structure and hence leads to sub-optimal achievable regions. In this paper, we introduce a new (random) coding technique using principles from multiple descriptions encoding and Han and Kobayashi decoding, which leads to a new and strictly improved achievable rate region for this problem. We show that this achievable rate region is complete under certain special scenarios.

The rest of the paper is organized as follows. In Section II, we review prior work related to distributed source coding and network routing. Before stating the problem formally, in Section III, we provide a simple example to demonstrate the basic principles underlying DIR. We also demonstrate the sub-optimality of conventional routing methods using this simple example. In Section IV, we formally state the DIR problem and provide an achievable rate region. Next, in Section VII, we show that this achievable rate region is complete for some special cases of the setup and finally in Section VIII, we prove strict improvement in achievable rates.

II. Prior Work

Multi-terminal source coding has one of its early roots in the seminal work of Slepian and Wolf [7]. They showed, in the context of lossless coding, that side-information available only at the decoder can nevertheless be fully exploited as if it were available to the encoder, in the sense that there is no asymptotic performance loss. Extensive work followed considering different network scenarios and obtaining achievable rate regions for them. Cover extended Slepian and Wolf's results to ergodic sources in [8]. Wyner extended Slepian and Wolf's results in [9], to a setup involving two encoders and one decoder, called the decoder side information setup, wherein the goal of the decoder is to reconstruct only one of the sources losslessly. Han and Kobayashi [3] extended the Slepian-Wolf and Wyner's results to general multi-terminal source coding scenarios. For a multi-sink network, with each sink reconstructing a prespecified subset of the sources, they characterized an achievable rate region for lossless reconstruction of the required sources at each sink. Csiszár and Körner [10] provided an alternative characterization of the achievable rate region for the same setup by relating the region to the solution of a class of problems called the "entropy characterization problems".

One of the first lossy coding extensions was derived by Wyner and Ziv in [11], that characterizes the complete rate-distortion region in the presence of decoder side information. Gray and Wyner considered a related network scenario in [12] involving three encoders and two decoders and derived the complete rate-distortion region for this setup. Berger and Tung extended Wyner-Ziv's coding scheme to the distributed lossy source coding setup and derived an achievable rate-distortion region in [13]. Recently, Wagner et.al derived a new coding scheme in [14] based on common components and showed that it strictly improves upon the Berger-Tung region. There has been extensive work on a related problem in multi-terminal source coding, called the 'multiple descriptions' (MD) problem, (see [2], [15], [16]) wherein the encoder sends multiple packets into the network and it is assumed that a subset of packets are lost during the course. The objective of the decoder is to reconstruct the source, upto a distortion constraint, based on the received packets. We note that, although we consider only lossless networks in this paper, the new coding scheme we derive is closely related to the MD problem and derives certain basic principles from MD encoding schemes.

There has also been a considerable amount of work on joint compression-routing for networks. A survey of routing techniques for sensor networks is given in [17]. It was shown in [4] that the problem of finding the optimum route for compression using explicit communication is an NP-complete problem. Pattem et.al compared different joint compression-routing schemes for a correlated sensor grid in [18] and also proposed an approximate, practical, static source clustering scheme to achieve compression efficiency. Much of the above work is related to compression using explicit communication, without recourse to distributed source coding techniques. Cristescu et al. [1] considered joint optimization of Slepian-Wolf coding and a routing mechanism, we call 'broadcasting',[1] wherein each source broadcasts its information to all sinks that intend to reconstruct it. Such a routing mechanism is motivated from the extensive literature on optimal routing for independent sources [19]. The general optimality of that approach for networks with a single sink was proven in [6]. We demonstrated its sub-optimality for the multi-sink scenario, recently in [45]. This paper takes a step further towards finding the best joint compression-routing mechanism for a multi-sink network.

We note the existence of a volume of work on minimum cost network coding for correlated sources, see [20], [21]. But the routing mechanism we introduce in this paper does not require possibly complex network coders at intermediate nodes, and can be realized using simple conventional routers. Throughout the paper, we assume that a 'conventional router' is a router that has the standard capabilities of replicating and/or forwarding a packet. However, the proposed approach, in principle, can be thought of as an extreme special case of network coding, and we will briefly discuss their relations in Section III-D. The approach does have potential implications on network coding in general, but these are beyond the scope of this paper.

[1] Note that we loosely use the term 'broadcasting' instead of 'multicasting' to stress the fact that ***all*** the information transmitted by any source is routed to every sink that reconstructs the source. Also, our approach to routing is in some aspects, a variant of multicasting.

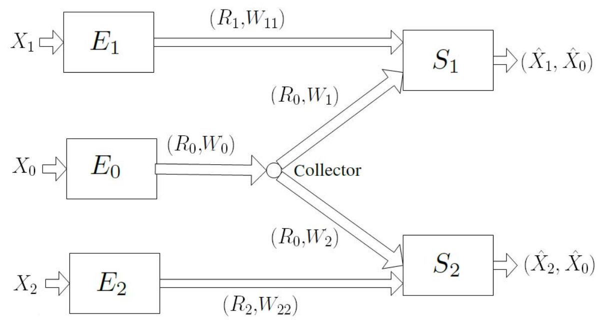

Fig. 2. Example with no helpers - Conventional Routing: Observe that the rates on the branches connecting the collector to the two sinks is same as that from E_0 to the collector.

III. Dispersive Information Routing-Motivating Example

A. Basic Notation

We begin by introducing the basic notation. In what follows, $2^{\mathcal{S}}$ denotes the set of all subsets (power set) of any set $\mathcal{S}$ and $|\mathcal{S}|$ denotes the set cardinality. Note that $|2^{\mathcal{S}}| = 2^{|\mathcal{S}|}$. $\mathcal{S}^c$ denotes the set complement (the universal set will be specified when there is ambiguity) and ϕ denotes the null set. For two sets $\mathcal{S}_1$ and $\mathcal{S}_2$, we denote the set difference by $\mathcal{S}_1 - \mathcal{S}_2 = \{s : s \in \mathcal{S}_1, s \notin \mathcal{S}_2\}$. Random variables are denoted by upper case letters (for example X) and their realizations are denoted by lower case letters (for example x). We also use upper case letters to denote source nodes and sinks and the ambiguity will be clarified wherever necessary. A sequence of n independent and identically distributed (iid) random variables and its realization are denoted by X^n and x^n, respectively. For any set $\mathcal{X}$, $\mathcal{X}^n$ denotes the Cartesian power of set $\mathcal{X}$, i.e., $\mathcal{X}^n = \mathcal{X} \times \mathcal{X} \times \cdots \times \mathcal{X}$, where $\mathcal{A} \times \mathcal{B}$ denotes the Cartesian product of two sets $\mathcal{A}$ and $\mathcal{B}$. The length n, ϵ-typical set is denoted by $\mathcal{T}_{\epsilon}^n$. We refer the readers to Appendix A for a formal definition of a typical set, used throughout this paper. $X \leftrightarrow Y \leftrightarrow Z$ denotes that the three random variables (X, Y, Z) form a Markov chain in that order. Notation in [22] is used to denote standard information theoretic quantities.

B. Illustrative Example - No Helpers Case

We first begin with a very simple example to motivate the new routing paradigm. Consider the network shown in Fig. 2. There are three source nodes, E_0, E_1 and E_2 and two sinks S_1 and S_2. The three source nodes observe correlated memoryless sequences X_0^n, X_1^n and X_2^n, respectively. Sink S_1 reconstructs the pair (X_0^n, X_1^n), while S_2 reconstructs (X_0^n, X_2^n). E_0 communicates with the two sinks through an intermediate node (called the 'collector') which is functionally a simple router. The edge weights on each path in the network are as shown in the figure. The cost of communication through an edge, e, is a function of the bit rate flowing through it, denoted by R_e and the corresponding edge weight, denoted by W_e, which in this paper, we will assume to be a simple product $C(R_e, W_e) = R_e W_e$. We further assume that the total cost is the sum of individual communication cost over each edge. The objective is to find the minimum total communication cost for lossless transmission of sources to the respective sinks.

We first consider the communication cost when broadcast routing is employed [1] wherein the router forwards all the bits received from E_0 to both the decoders. In other words, routers in a network are not allowed to "split" a packet and forward a portion of the received information on the forward paths. Hence the branches connecting the collector to the two sinks carry the same rates as the branch connecting E_0 to the collector. We denote the rate at which X_0, X_1 and X_2 are encoded by R_0, R_1 and R_2, respectively.

Using results in [1], it can be shown that the minimum communication cost under broadcast routing is given by the solution to the following linear programming formulation:

$$C_{br} = \min\{(W_0 + W_1 + W_2)R_0 + W_{11}R_1 + W_{22}R_2\} \quad (1)$$

under the constraints:

$$\begin{aligned} R_0 &\geq \max(H(X_0|X_1), H(X_0|X_2)) \\ R_1 &\geq H(X_1|X_0) \\ R_2 &\geq H(X_2|X_0) \\ R_1 + R_0 &\geq H(X_0, X_1) \\ R_2 + R_0 &\geq H(X_0, X_2) \end{aligned} \quad (2)$$

To gain intuition into dispersive information routing, we will later consider a special case of the above network when the branch weights are such that $W_{11}, W_{22} \ll W_0, W_1, W_2$. Let us specialize the above equations for this case. The constraint $W_{11}, W_{22} \ll W_0, W_1, W_2$, implies that X_1 and X_2 should be encoded at rates $R_1 = H(X_1)$ and $R_2 = H(X_2)$, respectively. Therefore the scenario effectively captures the case when X_1 and X_2 are available as side information at the respective decoders. It follows from (1) and (2) that for achieving minimum communication cost, R_0 is:

$$R_0^* = \max\{H(X_0|X_1), H(X_0|X_2)\} \quad (3)$$

and therefore the minimum communication cost is given by:

$$C_{br}^* = (W_0 + W_1 + W_2)R_0^* + W_{11}H(X_1) + W_{22}H(X_2) \quad (4)$$

Is this the best we can do? The collector has to transmit enough information to sink S_1 for it to decode X_0 and therefore the rate is at least $H(X_0|X_1)$. Similarly the rate on the branch connecting the collector to S_2 is at least $H(X_0|X_2)$. But if $H(X_0|X_1) \neq H(X_0|X_2)$, there is excess rate on one of the branches.

Let us now relax this restriction and allow the collector node to "split" the packet and route different subsets of the received bits on the forward paths. We could equivalently think of the source E_0 transmitting 3 smaller packets to the collector; the first packet has a rate $R_{0,12}$ bits and is destined to both sinks. Two other packets have rates $R_{0,1}$ and $R_{0,2}$ and are destined to sinks S_1 and S_2, respectively. Technically, in this case, the collector is again a simple conventional router.

We refer to such a routing mechanism, where each intermediate node transmits a subset of the received bits on each of the forward paths, as "***Dispersive Information Routing***" (DIR). Note that DIR does not require possibly complex coders at

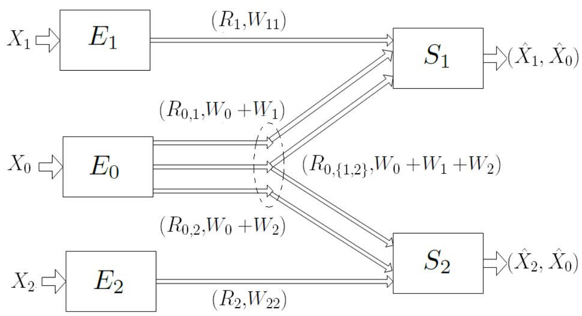

Fig. 3. Example - DIR. Note that the notion of 'packet splitting' is equivalent to the sources transmitting multiple smaller packets.

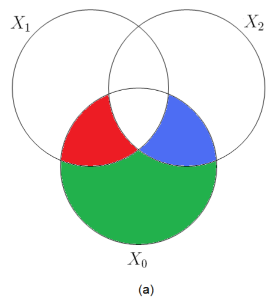

Fig. 4. Venn Diagram based intuition: Amount of information routed using DIR when operating at point P_1. Observe that each of the sinks receive information at the respective minimum rates. Green represents $R_{0,12}$, Blue represents $R_{0,1}$ and Red represents $R_{0,2}$.

intermediate nodes, and can always be realized using conventional routers (which have standard capabilities of replication and/or forwarding), with each source transmitting multiple packets into the network intended to different subsets of sinks. Hereafter, we interchangeably use the ideas of "packet splitting" at intermediate nodes and conventional routing of smaller packets, noting the equivalence in achievable rates and costs. This scenario is depicted in Fig. 3 with the modified cost each packet encounters.

Two obvious questions arise - Does DIR achieve a lower communication cost compared to conventional routing? If so, what is the minimum communication cost under DIR?

We first aim to find the minimum cost using DIR under the special case of $W_{11}, W_{22} \ll W_0, W_1, W_2$ (i.e., $R_1 = H(X_1)$ and $R_2 = H(X_2)$). To establish the minimum communication cost, we need to first establish the complete rate region for the rate tuple $\{R_{0,1}, R_{0,12}, R_{0,2}\}$, for lossless reconstruction of X_0^n at both the decoders and then find the point in the achievable rate region that minimizes the total communication cost, determined using the modified weights shown in Fig. 3.

Before deriving the ultimate solution, it is instructive to consider one operating point, $P_1 \triangleq \{R_{0,1}, R_{0,12}, R_{0,2}\} = \{I(X_1; X_0|X_2), H(X_0|X_1, X_2), I(X_2; X_0|X_1)\}$ and provide the coding scheme that achieves it. Extension to other "interesting points" and to the whole achievable region follows in similar lines. This particular rate point is considered first due to its intuitive appeal as shown in a Venn diagram (Fig. 4).

The complete achievable rate region for this setup can be derived based on a modified version of Slepian and Wolf's random binning technique (see [23]). Every typical sequence of X_0^n is assigned 3 independent bin indices, using uniform pmfs over $[1 : 2^{nR_{0,1}}]$, $[1 : 2^{nR_{0,12}}]$ and $[1 : 2^{nR_{0,2}}]$, respectively. All the sequences with the same first index, $m_{0,1}$, form a bin $\mathcal{B}_{0,1}(m_{0,1})$. Similarly bins $\mathcal{B}_{0,2}(m_{0,2})$ and $\mathcal{B}_{0,12}(m_{0,12})$ are formed for all indices $m_{0,2}$ and $m_{0,12}$, respectively. Upon observing a sequence $x_0^n \in \mathcal{T}_\epsilon^n$ with indices $m_{0,1}, m_{0,2}$ and $m_{0,12}$, the encoder transmits index $m_{0,1}$ to decoder 1 alone, index $m_{0,2}$ to decoder 2 alone and index $m_{0,12}$ to both the decoders. The first decoder receives indices $m_{0,1}$ and $m_{0,12}$. It tries to find a typical sequence $\hat{x}_0^n \in \mathcal{B}_{0,1}(m_{0,1}) \cap \mathcal{B}_{0,12}(m_{0,12})$ which is jointly typical with the decoded information sequence x_1^n. As the indices are assigned independent of each other, every typical sequence has uniform pmf of being assigned to the index pair $\{m_{0,1}, m_{0,12}\}$ over $[1 : 2^{n(R_{0,1}+R_{0,\{1,2\}})}]$. Therefore, having received indices $m_{0,1}$ and $m_{0,12}$, using arguments similar to Slepian-Wolf [7] and Cover [8], the probability of decoding error asymptotically approaches zero if:

$$R_{0,1} + R_{0,12} \geq H(X_0|X_1) \tag{5}$$

Similarly, probability of decoding error approaches zero at the second decoder if:

$$R_{0,2} + R_{0,12} \geq H(X_0|X_2) \tag{6}$$

The above achievable region can easily be shown to satisfy the converse and hence is the complete achievable rate region for this problem. This technique of assigning multiple independent bin indices to each sequence has appeared in several related prior work in the past [23]–[27]. In the context of source-channel coding over broadcast channels, Tuncel derived the above rate region in [23], and called the coding scheme as 'nested binning'. The same region was also obtained by Timo et.al in [28] in the context of deriving the rate-distortion region for the Gray-Wyner network [12] with side information. We term such a binning approach as 'Power Binning' in this paper, as for a more general network involving multiple sinks, an independent index is assigned to each (non-trivial) subset of the decoders - the power set. It is worthwhile to note that the same rate region can also be obtained by applying results of Han and Kobayashi [3], assuming 3 independent encoders at E_0, albeit with a more complicated coding scheme involving multiple auxiliary random variables.

The minimum cost operating point is the point that satisfies equations (5) and (6) and minimizes the cost function:

$$\begin{aligned} C^*_{DIR-SI} = \min\{&(W_0 + W_1)R_{0,1} + (W_0 + W_2)R_{0,2} \\ &+ (W_0 + W_1 + W_2)R_{0,12}\} \end{aligned} \tag{7}$$

The solution to the above formulation is:

$$P_2 = \begin{cases} \{0, h_1, h_2 - h_1\} & h_2 \geq h_1 \\ \{h_1 - h_2, h_2, 0\} & \text{Otherwise} \end{cases}$$

where $h_1 = H(X_0|X_1)$ and $h_2 = H(X_0|X_2)$. Both these points achieve lower total communication cost compared to broadcast routing, C^*_{conv} in (4), for any $W_0, W_1, W_2 \gg W_{11}, W_{22}$, if $H(X_0|X_1) \neq H(X_0|X_2)$.

The above coding scheme can be easily extended to the case of arbitrary edge weights. Then, the rate region for the tuple $\{R_1, R_2, R_{0,1}, R_{0,12}, R_{0,2}\}$ and the cost function to be minimized are given by:

$$C^*_{DIR} = \min\{W_{11}R_1 + W_{22}R_2 + (W_0 + W_1)R_{0,1} + (W_0+W_2)R_{0,2}+(W_0+W_1+W_2)R_{0,12}\} \quad (8)$$

under the constraints:

$$\begin{aligned} R_1 &\geq H(X_1|X_0) \\ R_{0,1} + R_{0,12} &\geq H(X_0|X_1) \\ R_1 + R_{0,1} + R_{0,12} &\geq H(X_0, X_1) \\ R_2 &\geq H(X_2|X_0) \\ R_{0,2} + R_{0,12} &\geq H(X_0|X_2) \\ R_2 + R_{0,2} + R_{0,12} &\geq H(X_0, X_2) \end{aligned} \quad (9)$$

If $R_1 = H(X_1)$ and $R_2 = H(X_2)$, (9) specializes to (5) and (6). Also, it can easily be shown that the total communication cost obtained as a solution to the above formulation is lower than that for conventional routing if $W_0, W_1, W_2 > 0$. This example clearly demonstrates the gains of DIR over broadcast routing to communicate correlated sources over a network.

C. Discussion

For a general network with N sources and M sinks employing DIR, all the packet splitting operations at the intermediate nodes can be equivalently mapped to each source transmitting $2^M - 1$ packets into the network, where each packet is routed to a subset of the sinks. It is important to note that the equivalence in cost between the two notions of 'packet splitting at intermediate nodes' and 'sources transmitting multiple smaller packets' holds when the effective cost is a linear function of the rates on each branch. Quite interestingly, under this linear assumption, the problem of finding the minimum cost under DIR reduces to characterizing the set of achievable rates for a multi-terminal information theoretic setup. The minimum cost then follows from a standard linear programming formulation. In this paper, our objective is to characterize a new achievable rate region for this multi-terminal setup, while noting that the motivation for considering this new setup is its applicability in the context of dispersive information routing. Note that, for more complex cost functions, the problem of establishing the optimum routes under DIR cannot be dealt independently from deriving the achievable rate region. The optimum route then depends on the actual operating rates and hence cannot be solved independently. Nevertheless, the achievable rate region for lossless reconstruction remains the same and results derived in this paper can be extended to handle more complex cost functions. The extensions are not obvious and hence are beyond the scope of this paper.

It was shown in [1] that the two problems of DSC (Slepian-Wolf compression) and optimum *broadcast routing* are separable problems, i.e., the optimum routes can be found without the knowledge of the achievable rates, and vice versa, the rate region can be found without the knowledge of the routes. However, we demonstrated in [45] that such separability holds only under an important assumption - when sinks only receive information from the source nodes they intend to reconstruct.[2] Such a scenario is called the 'No helpers' case in the literature [10]. We also showed that the extent of suboptimality due to separating DSC and broadcast routing is substantial and potentially unbounded when helpers are allowed to communicate. In general the optimum rate region cannot be found without the knowledge of the network costs for broadcast routing. However, for DIR with effective cost being linear in the rates, **the two problems of finding the optimum rate region and finding the optimum routes from the source nodes to the sinks can be separated and dealt independently, without entailing any loss of optimality**.[3] In fact, as we will see in Section IV-A, the problem of finding the optimum costs degenerates to the standard Steiner tree minimization problem, that has been studied extensively [19].

In this paper, our focus is primarily on the associated achievable rate region for the new multi-terminal information theoretic setup that arises in the context of DIR. Although, this new setup has not been explicitly considered in the past, results pertaining to general multi-terminal source networks due to Han and Kobayashi [3], can be extended to derive an achievable region. This extension achieves the complete rate region for networks with no helpers. However, for networks with helpers, applicability of their result requires artificial imposition of independence of certain transmitted messages at each source, which is an unnecessary restriction. We present a more general achievable rate region, which maintains the dependencies between the messages at each encoder. For the no-helpers setting, both the proposed coding scheme and the extended Han and Kobayashi coding scheme degenerate to the Power Binning approach described in section III-B. However, we show that, even for simplest networks with helpers, the proposed coding scheme can achieve a strictly larger region compared to the extended Han and Kobayashi scheme.

We note that several publications address the problem of source-channel coding over multi-user networks with correlated sources [23], [29], [30], and the underlying principles therein are somewhat similar to that in dispersive information routing. For example, in [23], Tuncel considers source-channel coding over broadcast channels and the DIR setup shown in Fig. 3 turns out to be a special case involving only deterministic channels. Similarly, deterministic versions of certain multi-user channels considered in [29] and [30] can be viewed as special cases of the general DIR setup. However, in most of these papers, it is assumed that all the decoders reconstruct all the sources in the network, a setting involving no helpers, for which power binning/nested binning achieves the complete rate region.

The main contribution of this work is to derive a new coding scheme for the general DIR setup with helpers. The problem of establishing the complete rate region becomes considerably harder when there are helpers, a scenario, that

[2]Observe that the example considered in Section III-B is a no-helpers scenario.

[3]Note that even though DIR has the inherent advantage of separability, finding the optimum operating point requires optimizing over an $N \times 2^M$ dimensional space and the effective complexity remains the same as that for broadcast routing.

is highly relevant to practical networks. We show that the proposed coding scheme achieves the complete rate region for certain special classes of networks with helpers. Some of the results we derive in this paper can have potential applications in deriving new source-channel coding schemes for networks with helpers. Particularly, the setup we will consider in Section V can be viewed as a generalized Slepian-Wolf problem over a deterministic interference channel. However, these extensions are beyond the scope of this paper.

D. Relations to Network Coding

Following the seminal work by Ahlswede et.al [31], network coding has evolved to be an important research area and has proven to achieve improved network capacity/lower communication cost compared to conventional routing. In network coding, the intermediate nodes are allowed to send arbitrary functions of the information contained in received packets, on each of the forward paths. Although, the original focus of network coding was for networks with independent sources, several researchers have focused on minimum cost network coding for correlated sources, see [20], [21]. It is interesting to note that DIR is indeed an extreme special case of network coding, in the sense that the packet splitting operation can be viewed as a unique function applied to the received packets. It is hence arguable whether it is worth finding a solution to the DIR setup, as it is broadly subsumed in the network coding architecture. However, DIR has important theoretical and practical implications which make this study particularly important.

First, from a practical standpoint, the composite packet splitting operation at all intermediate nodes can equivalently be achieved using sources transmitting multiple smaller packets into the network and hence, DIR can always be achieved using conventional routers, which only need the capability to replicate and/or forward a packet. A generic network code, on the other hand, would require more 'intelligent' intermediate nodes that have the capability to compute functions of the information contained in received packets.

Second, from a theoretical standpoint, the solution to the DIR problem reduces to establishing a set of achievable rate tuples for lossless reconstruction of a prescribed subset of sources at each sink. This is a purely source coding setup and can be addressed using conventional source coding techniques, without recourse to inter-play between source coding and network coding. Moreover, the complete rate region for several special cases of the general DIR setup can be derived (as we will see in Section VII) and hence the minimum cost under DIR can be obtained, while the minimum cost using network coding is quite difficult to establish, even for very simple networks. We do note that DIR is strictly sub-optimal, in general, compared to network coding. A simple counter-example follows directly from the popular butterfly network considered by Ahlswede et.al in their seminal work [31]. It is easy to show that, for general source distributions, for the butterfly network, DIR cannot achieve the minimum communication cost. It is part of our future work to determine the set of networks for which DIR achieves the minimum cost over all network codes. Finally we note that, although DIR is strictly suboptimal in general, it does have potential implications on network coding in general, but these are beyond the scope of this paper.

IV. Dispersive Information Routing-General Problem Setup

Let a network be represented by an undirected connected graph $G = (V, \mathcal{E})$. Each edge $e \in \mathcal{E}$ is associated with an edge weight, W_e. The communication cost is assumed to be a simple product of the edge rate and edge weight, i.e., $C_e = R_e W_e$. The nodes V consist of N source nodes (denoted by $E_1, E_2 \dots E_N$), M sinks (denoted by $S_1, S_2 \dots S_M$), and $|V|$-N-M intermediate nodes. We define the sets $\Sigma = \{1 \dots N\}$ and $\Pi = \{1 \dots M\}$. Source node E_i observes n iid random variables X_i^n, each taking values over a finite alphabet $\mathcal{X}_i$. Sink S_j reconstructs (requests) a subset of the sources specified by $\Sigma_j \subseteq \Sigma$. Conversely, source node E_i is reconstructed at a subset of sinks specified by $\Pi_i \subseteq \Pi$. Each of the intermediate nodes have the capability to split a packet and forward subsets of the received bits on each of the forward paths. For simplicity, we assume that the source nodes only send packets into the network and the sink nodes only receive packets from the network, i.e., the source and the sinks nodes do not behave as routers. The objective is to find the minimum communication cost achievable by dispersive information routing for lossless reconstruction of the requested sources at each sink when every source node can (possibly) communicate with every sink.

A. Obtaining the Effective Costs

Under DIR each source transmits at most $2^M - 1$ packets into the network, each meant for a different subset of sinks. Note that, while Π_i is the subset of sinks reconstructing X_i^n, E_i may be transmitting packets to many other subsets of sinks. Let the packet from source E_i to the subset of sinks $\mathcal{K} \subseteq \Pi$ be denoted by $\mathcal{P}_{i,\mathcal{K}}$ and let it carry information at rate $R_{i,\mathcal{K}}$.

The optimum route for packet $\mathcal{P}_{i,\mathcal{K}}$ from the source to these sinks is determined by a spanning tree optimization (minimum Steiner tree) [19]. More specifically, for each packet $\mathcal{P}_{i,\mathcal{K}}$, the optimum route is obtained by minimizing the cost over all trees rooted at node i which span all sinks $j \in \mathcal{K}$. The minimum cost of transmitting packet $\mathcal{P}_{i,\mathcal{K}}$ with $R_{i,\mathcal{K}}$ bits from source i to the subset of sinks $\mathcal{K}$, denoted by $d_i(\mathcal{K})$ is:

$$d_i(\mathcal{K}) = R_{i,\mathcal{K}} \min_{Q \in \mathcal{E}_{i,\mathcal{K}}} \sum_{e \in Q} w_e \tag{10}$$

where $\mathcal{E}_{i,\mathcal{K}}$ denotes the set of all paths from source i to the subset of sinks $\mathcal{K}$. Note that the minimum Steiner tree problem is NP - hard and requires approximate algorithms to solve in practice. Also note that in theory, each encoder transmits $2^M - 1$ packets into the network. While in practice we might be able to realize improvements over broadcast routing using significantly fewer packets (see [32]).

B. An Achievable Rate Region

Our main objective in this paper is to find an achievable rate region for the tuple ($R_{i,\mathcal{K}}$ $\forall i \in \Sigma, \mathcal{K} \subseteq \Pi$) to achieve lossless reconstruction of sources specified by Σ_j at sink S_j, $\forall j$.

The minimum communication cost then follows directly from a simple linear programming formulation. In this section, we formally introduce the notation and define the achievable rate region for a general setup. We focus on the coding scheme in subsequent sections.

In what follows, we use the shorthand $\{U_i\}_{\mathcal{S}}$ for $\{U_{i,\mathcal{K}} : \mathcal{K} \in \mathcal{S}\}$ and $\{U_\Gamma\}_{\mathcal{S}}$ for $\{U_{i,\mathcal{K}} : i \in \Gamma, \mathcal{K} \in \mathcal{S}\}$. Note the difference between $\{U_i\}_{\mathcal{S}}$ and $U_{i,\mathcal{S}}$. $\{U_i\}_{\mathcal{S}}$ is a set of variables, whereas $U_{i,\mathcal{S}}$ is a single variable. For example, $\{U_1\}_{(1,2,12)}$ denotes the set of variables $(U_{1,1}, U_{1,2}, U_{1,12})$ and $\{U_{(1,2)}\}_{(1,2,12)}$ represents the set $(U_{1,1}, U_{1,2}, U_{1,12}, U_{2,1}, U_{2,2}, U_{2,12})$.

We first give a formal definition of a block code and an associated rate region for DIR. We denote the set $\{1, 2 \ldots L\}$ by I_L for any positive integer L. We assume that the source node E_i observes the random sequence X_i^n. An $(n, P_e, L_{i,\mathcal{K}}; \forall i \in \Sigma, \mathcal{K} \in 2^\Pi - \phi)$ DIR-code is defined by the following mappings:

- *Encoder*s:

$$f_i^E : \mathcal{X}_i^n \to \prod_{\mathcal{K} \in 2^\Pi - \phi} I_{L_{i,\mathcal{K}}} \qquad i = 1, 2, \ldots, N \tag{11}$$

- *Decoder*s:

$$f_j^D : \prod_{i \in \Sigma} \prod_{\mathcal{K} \in 2^\Pi : j \in \mathcal{K}} I_{L_{i,\mathcal{K}}} \to \{\mathcal{X}^n\}_{\Sigma_j} \qquad j = 1, 2, \ldots, M \tag{12}$$

Denoting $f_i^E(X_i^n) = \{T_i\}_{2^\Pi - \phi}$ where $1 \leq T_{i,\mathcal{K}} \leq L_{i,\mathcal{K}}$, the decoder estimates are given by:

$$\{\hat{X}^n\}_{\Sigma_j} = f_j^D(\{T_\Sigma\}_{(\mathcal{K} \in 2^\Pi : j \in \mathcal{K})}) \tag{13}$$

Note the correspondence between the encoder-decoder mappings and dispersive information routing. Observe that packet $\mathcal{P}_{i,\mathcal{K}}$ carries $T_{i,\mathcal{K}}$ at rate $L_{i,\mathcal{K}}$ from source i to the subset of sinks $\mathcal{K}$. The probability of error is defined as the average of the probabilities of reconstruction error at each of the decoders, i.e.,:

$$P_e = \frac{1}{M} \left[\sum_{j \in \Pi} P(\{X^n\}_{\Sigma_j} \neq \{\hat{X}^n\}_{\Sigma_j}) \right] \tag{14}$$

A rate tuple $\{R_{i,\mathcal{K}}; \forall i, \mathcal{K}\}$ is said to be achievable if for any $\eta > 0$ and $0 < \epsilon < 1$, there exists a $(n, P_e, L_{i,\mathcal{K}}; \forall i \in \Sigma, \mathcal{K} \in 2^\Pi - \phi)$ code for n sufficiently large such that,

$$R_{i,\mathcal{K}} \leq \frac{1}{n} \log L_{i,\mathcal{K}} + \eta \tag{15}$$

with the probability of error less than ϵ, i.e.,

$$P_e < \epsilon \tag{16}$$

There is no single-letter characterization of the complete region for this problem, but a non-computable characterization is possible using the results of Han and Kobayashi in [3]. They also provide a single-letter partial achievable rate region. However, applicability of their result requires artificial imposition of independence of transmitted messages, at each source. We will show in section VIII that the proposed coding scheme, which maintains the dependencies across the transmitted messages, performs strictly better than the Han and Kobayashi's coding scheme for networks with helpers. Note that the source coding setup which arises out of the DIR framework is a special case of the general problem of distributed multiple descriptions and therefore the principles underlying the coding schemes for distributed source coding [3] and multiple descriptions encoding [2] play crucial roles in deriving a coding mechanism for dispersive information routing. It is interesting to observe that, unlike the general MD setting, the DIR framework is non-trivial even in the lossless scenario and deriving a complete rate region for lossless reconstruction at all the sinks is a challenging problem.

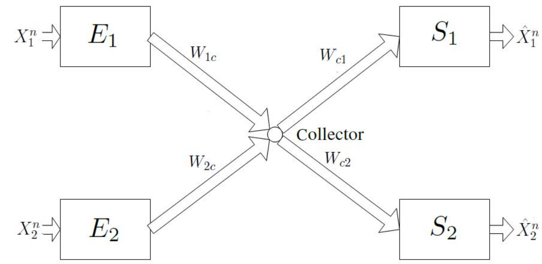

Fig. 5. The 2 Source - 2 Sink example. Each source acts as the principal source for one sink and as a helper for the other.

C. Note on the Network Model

We note that the network model we consider in this paper does not capture the interactive nature of a network, in the sense that the sources are not allowed to receive packets and sinks are not allowed to send feedback or forward packets to neighboring nodes in the network. This drawback can be overcome by allowing the source and sink nodes to also have routing capabilities, or equivalently, by connecting each source/sink node to a routing node through a zero weight link, which is connected to all neighbors through links having the same edge weights. Although, this approach generalizes the model to some extent, the sinks would still be assumed to have the same routing capabilities as all other nodes. The sink nodes can, in principle, forward the decoded version of the source sequence, if it turns out to be more beneficial than forwarding received bits. Our model does not capture this aspect of networks and hence is not completely general. However, we do note that the problem gets significantly harder to solve if we allow the sinks to forward decoded information on the forward links and requires principles both from dispersive information routing and compression using explicit communication, to devise a solution. We therefore consider the simplified model throughout this paper and address the more general setting as part of our future work.

V. A Simple Network With Helpers

We will begin with a simple network with helpers that captures the essence of the new coding scheme and derive the associated achievable region. However, we defer the formal proofs for the general case to section IV-B. Consider the network shown in Fig. 5. Two source nodes E_1 and E_2 observe correlated memoryless sequences X_1^n and X_2^n, respectively.

Two sinks S_1 and S_2 require lossless reconstructions of X_1^n and X_2^n, respectively. The source nodes can communicate with the sinks only through a collector node. The edge weights are as shown in the figure. Observe that each source, while requested by one sink, acts as helper for the other.

Under dispersive information routing, each source transmits a packet to every subset of sinks. In this example, E_1 sends 3 packets to the collector at rates $(R_{1,1}, R_{1,2}, R_{1,12})$, respectively. The collector forwards the first packet to S_1, the second to S_2 and the third to both S_1 and S_2. Similarly, E_2 sends 3 packets to the collector at rates $(R_{2,1}, R_{2,2}, R_{2,12})$ which are forwarded to the corresponding sinks. Our objective is to determine the set of achievable rate tuples $(R_{1,1}, R_{1,2}, R_{1,12}, R_{2,1}, R_{2,2}, R_{2,12})$ that allows for lossless reconstruction at the two sinks.

It is instructive to first consider a trivial outer-bound and illustrate the sub-optimality of power binning. A trivial outerbound follows using standard cut-set bounding techniques (see [33]), and is given by:

$$\begin{aligned} R_{1,1} + R_{1,12} &\geq H(X_1|X_2) \\ R_{2,2} + R_{2,12} &\geq H(X_2|X_1) \\ \Big\{R_{1,1} + R_{1,12} + R_{1,2} + R_{2,1} + R_{2,12} + R_{2,2}\Big\} &\geq H(X_2, X_1) \end{aligned} \tag{17}$$

The achievable region due to power binning can be easily obtained using techniques similar to Section III-B. To state the achievable region using power binning explicitly, we define the following rate regions:

$$\begin{aligned} \mathcal{R}_1^1 &:= \Big\{R_{1,1} + R_{1,12} \geq H(X_1)\Big\} \\ \mathcal{R}_2^2 &:= \Big\{R_{2,2} + R_{2,12} \geq H(X_2)\Big\} \\ \mathcal{R}_1^{12} &:= \begin{cases} R_{1,1} + R_{1,12} \geq H(X_1|X_2) \\ R_{2,1} + R_{2,12} \geq H(X_2|X_1) \\ R_{1,1} + R_{1,12} + R_{2,1} + R_{2,2} \geq H(X_1, X_2) \end{cases} \\ \mathcal{R}_2^{12} &:= \begin{cases} R_{2,2} + R_{2,12} \geq H(X_2|X_1) \\ R_{1,2} + R_{1,12} \geq H(X_1|X_2) \\ R_{2,2} + R_{2,12} + R_{1,2} + R_{1,12} \geq H(X_1, X_2) \end{cases} \end{aligned} \tag{18}$$

Note that conditions $\mathcal{R}_1^1$ allow for lossless decoding of X_1 at S_1. Similarly, conditions $\mathcal{R}_2^2$ allow for lossless decoding of X_2 at S_2. However, conditions $\mathcal{R}_1^{12}$ and $\mathcal{R}_2^{12}$ allow for lossless decoding of both X_1 and X_2 at sinks S_1 and S_2, respectively. The power binning region is now given by:

$$\mathcal{R}_{PB} = \{\mathcal{R}_1^1 \cap \mathcal{R}_1^{12}\} \cup \{\mathcal{R}_2^2 \cap \mathcal{R}_2^{12}\} \tag{19}$$

This is clearly not equal to the outerbound (17). The primary difficulty with power binning is that, it allows for only one of the two extreme situations. For example, at sink S_1, on the one hand, the information received from E_2 can be used to reconstruct x_2^n, and then the sequence x_1^n can be reconstructed based on x_2^n and the bin indices received from E_1. On the other hand, the information received from E_2 can be discarded and the sequence x_1^n can be reconstructed only based on the bin indices received from E_1. The power binning scheme does not allow the information from E_2 to be useful at S_1, unless it is large enough to decode x_2^n. This sub-optimality with random binning approach is quite well known in the literature and is in fact the main motivation for the encoding schemes in [3] and [9].

We now provide a new coding scheme and an associated achievable region for the setup in Fig. 5, wherein the information from the helpers is utilized without actually decoding the helper's source sequence. Suppose we are given random variables $(U_{1,12}, U_{1,1}, U_{1,2}, U_{2,12}, U_{2,1}, U_{2,2})$ jointly distributed with (X_1, X_2) such that the following Markov chain conditions hold:

$$\begin{aligned} (U_{1,12}, U_{1,1}) \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow (U_{2,12}, U_{2,1}) \\ (U_{1,12}, U_{1,2}) \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow (U_{2,12}, U_{2,2}) \end{aligned} \tag{20}$$

Then the following rates are achievable:

$$\begin{aligned} R_{1,12} &\geq I(X_1; U_{1,12}|U_{2,12}) \\ R_{2,12} &\geq I(X_2; U_{2,12}|U_{1,12}) \\ R_{1,12} + R_{2,12} &\geq I(X_1, X_2; U_{1,12}, U_{2,12}) \\ R_{1,1} &\geq H(X_1|U_{1,12}, U_{2,12}, U_{2,1}) \\ R_{2,1} &\geq I(X_2; U_{2,1}|U_{1,12}, U_{2,12}, U_{1,1}) \\ R_{1,1} + R_{2,1} &\geq I(X_1, X_2; U_{1,1}, U_{2,1}|U_{1,12}, U_{2,12}) \\ &\quad + H(X_1|U_{1,12}, U_{2,12}, U_{1,1}, U_{2,1}) \\ R_{2,2} &\geq H(X_2|U_{1,12}, U_{2,12}, U_{1,2}) \\ R_{1,2} &\geq I(X_2; U_{1,2}|U_{1,12}, U_{2,12}, U_{2,2}) \\ R_{2,2} + R_{1,2} &\geq I(X_1, X_2; U_{2,2}, U_{1,2}|U_{1,12}, U_{2,12}) \\ &\quad + H(X_2|U_{1,12}, U_{2,12}, U_{1,2}, U_{2,2}) \end{aligned} \tag{21}$$

The convex closure of these rates over all random variables, jointly distributed with (X_1, X_2), satisfying the conditions in (20), leads to an achievable region for the setup in Fig. 5.

The encoding is divided into 3 stages.

Encoding: We first focus on the encoding at E_1. In the first stage, $2^{nR'_{1,12}}$ codewords of $U_{1,12}$, each of length n are generated independently, with elements drawn according to the marginal density $P(U_{1,12})$. Conditioned on each of these codewords, $2^{nR'_{1,1}}$ and $2^{nR'_{1,2}}$ codewords of $U_{1,1}$ and $U_{1,2}$ are generated according to the conditional densities $P(U_{1,1}|U_{1,12})$ and $P(U_{1,2}|U_{1,12})$, respectively. Codebooks for $U_{2,12}, U_{2,1}$ and $U_{2,2}$ are generated at E_2 in a similar fashion. On observing a sequence x_1^n, E_1 first tries to find a codeword tuple from the codebooks of $(U_{1,12}, U_{1,1}, U_{1,2})$ such that $(x_1^n, u_{1,12}^n, u_{1,1}^n) \in \mathcal{T}_\epsilon^n$ and $(x_1^n, u_{1,12}^n, u_{1,2}^n) \in \mathcal{T}_\epsilon^n$. The probability of finding such a codeword tuple approaches 1 if,

$$\begin{aligned} R'_{1,12} &\geq I(X_1; U_{1,12}) \\ R'_{1,1} &\geq I(X_1; U_{1,1}|U_{1,12}) \\ R'_{1,2} &\geq I(X_1; U_{1,2}|U_{1,12}) \end{aligned} \tag{22}$$

Let the codewords selected be denoted by $(u_{1,12},u_{1,1},u_{1,2})$. Similar constraints on $(R'_{2,1}, R'_{2,2}, R'_{2,12})$ must be satisfied for encoding at E_2. Denote the codewords selected at E_2 by $(u_{2,12}, u_{2,1}, u_{2,2})$. It follows from (20) and the 'Conditional Markov Lemma' in [14] that $(x_1^n, x_2^n, u_{1,12}, u_{1,1}, u_{2,12}, u_{2,1}) \in \mathcal{T}_\epsilon^n$ and $(x_1^n, x_2^n, u_{1,12}, u_{1,2}, u_{2,12}, u_{2,2}) \in \mathcal{T}_\epsilon^n$ with high probability.

In the second stage of encoding, each encoder uniformly divides the $2^{nR'_{i,\mathcal{S}}}$ codewords of $U_{i,\mathcal{S}}$ into $2^{nR''_{i,\mathcal{S}}}$ bins $\forall\, i \in \{1,2\}$, $\mathcal{S} \in \{1,2,12\}$. All the codewords which have the same bin index m are said to fall in the bin $\mathcal{C}_{i,\mathcal{S}}(m)$ $\forall m \in (1\ldots 2^{nR''_{i,\mathcal{S}}})$. Note that the number of codewords in bin $\mathcal{C}_{i,\mathcal{S}}(m)$ is $2^{n(R'_{i,\mathcal{S}}-R''_{i,\mathcal{S}})}$. If E_1 selects the codewords $(u_{1,12}, u_{1,1}, u_{1,2})$ in the first stage and if the bin indices associated with $(u_{1,12}, u_{1,1}, u_{1,2})$ are $(m_{1,12}, m_{1,1}, m_{1,2})$, then index $m_{1,1}$ is routed to sink S_1, $m_{1,2}$ to sink S_2 and $m_{1,12}$ to both the sinks S_1 and S_2. Similarly, bin indices $(m_{2,12}, m_{2,1}, m_{2,2})$ are routed from E_2 to the corresponding sinks.

The third stage of encoding, resembles the 'Power Binning' scheme described in Section III-B. Every typical sequence of X_1^n is assigned a random bin index uniformly chosen over $[1 : 2^{n\tilde{R}_{1,1}}]$. All sequences with the same index, $l_{1,1}$, form a bin $\mathcal{B}_{1,1}(l_{1,1})$ $\forall l_{1,1} \in \{1\ldots 2^{n\tilde{R}_{1,1}}\}$. Upon observing a sequence $X_1^n \in \mathcal{T}_\epsilon^n$ with bin index $l_{1,1}$, in addition to $m_{1,1}$ (from the second stage of encoding), encoder E_1 also routes index $l_{1,1}$ to sink S_1. Similarly bin index $l_{2,2}$ is routed from E_2 to S_2 in addition to $m_{2,2}$. These bin indices are used to reconstruct X_1^n and X_2^n losslessly at the respective decoders. Note that, in a general setup, if source i is to be reconstructed at a subset of sinks Π_i, the source assigns $2^{|\Pi_i|}-1$ independently generated indices, each being routed to a subset of Π_i. We also note that $U_{1,1}$ and $U_{2,2}$ can be conveniently set to constants without changing the overall rate region. However, we continue to use them to avoid complex notation.

Decoding: We again focus on the first sink S_1. It receives the indices $(m_{1,12}, m_{1,1}, m_{2,12}, m_{2,1}, l_{1,1})$. It first looks for a pair of unique codewords from $\mathcal{C}_{1,12}(m_{1,12})$ and $\mathcal{C}_{2,12}(m_{2,12})$ which are jointly typical. Obviously, there is at least one pair, $(u_{1,12}, u_{2,12})$, which is jointly typical. The probability that no other pair of codewords are jointly typical approaches 1 if:

$$(R'_{1,12} - R''_{1,12}) + (R'_{2,12} - R''_{2,12}) \le I(U_{1,12}; U_{2,12}) \quad (23)$$

Noting that $(R'_{1,12} - R''_{1,12}) \ge 0$ and $(R'_{2,12} - R''_{2,12}) \ge 0$, and applying the constraints on $R'_{1,12}$ and $R'_{2,12}$ from (22) we get the following constraints for $R''_{1,12}$ and $R''_{2,12}$:

$$\begin{aligned} R''_{1,12} &\ge I(X_1; U_{1,12}|U_{2,12}) \\ R''_{2,12} &\ge I(X_2; U_{2,12}|U_{1,12}) \\ R''_{1,12} + R''_{2,12} &\ge I(X_1, X_2; U_{1,12}, U_{2,12}) \end{aligned} \quad (24)$$

The decoder at S_1 next looks at the codebooks of $U_{1,1}$ and $U_{2,1}$ which were generated conditioned on $u_{1,12}$ and $u_{2,12}$, respectively, to find a unique pair of codewords from $\mathcal{C}_{1,1}(m_{1,1})$ and $\mathcal{C}_{2,1}(m_{2,1})$ which are jointly typical with $(u_{1,12}, u_{2,12})$. We again have one pair, $(u_{1,1}, u_{2,1})$, which is jointly typical with $(u_{1,12}, u_{2,12})$. It can be shown using arguments similar to [3] that the probability of finding no other jointly typical pair approaches 1 if:

$$\begin{aligned} (R'_{1,1} - R''_{1,1}) &\le I(U_{1,1}; U_{2,1}, U_{2,12}|U_{1,12}) \\ (R'_{2,1} - R''_{2,1}) &\le I(U_{2,1}; U_{1,1}, U_{1,12}|U_{2,12}) \\ \{(R'_{1,1} - R''_{1,1}) + (R'_{2,1} - R''_{2,1})\} &\le H(U_{1,1}|U_{1,12}) + H(U_{2,1}|U_{2,12}) \\ &\quad - H(U_{1,1}, U_{2,1}|U_{1,12}, U_{2,12}) \end{aligned} \quad (25)$$

On substituting the constraints for $R'_{1,1}$ and $R'_{1,2}$ from (22), and using the Markov chain condition in (20) we get:

$$\begin{aligned} R''_{1,1} &\ge I(X_1; U_{1,1}|U_{1,12}, U_{2,12}, U_{2,1}) \\ R''_{2,1} &\ge I(X_2; U_{2,1}|U_{1,12}, U_{2,12}, U_{1,1}) \\ R''_{1,1} + R''_{2,1} &\ge I(X_1, X_2; U_{1,1}, U_{2,1}|U_{1,12}, U_{2,12}) \end{aligned} \quad (26)$$

After successfully decoding the codewords $(u_{1,12}, u_{1,1}, u_{2,12}, u_{2,1})$, the decoder at S_1 looks for a unique sequence from $\mathcal{B}_{1,1}(l_{1,1})$ which is jointly typical with $(u_{1,12}, u_{1,1}, u_{2,12}, u_{2,1})$. We again have x_1^n satisfying this property. It can be shown that the probability of finding no other sequence which is jointly typical with $(u_{1,12}, u_{1,1}, u_{2,12}, u_{2,1})$ approaches 1 if:

$$\tilde{R}_{1,1} \ge H(X_1|U_{1,12}, U_{2,12}, U_{1,1}, U_{2,1}) \quad (27)$$

Similar conditions at sink S_2 lead to the following constraints:

$$\begin{aligned} R''_{2,2} &\ge I(X_2; U_{2,2}|U_{1,12}, U_{2,12}, U_{1,2}) \\ R''_{1,2} &\ge I(X_2; U_{1,2}|U_{1,12}, U_{2,12}, U_{2,2}) \\ R''_{2,2} + R''_{1,2} &\ge I(X_1, X_2; U_{2,2}, U_{1,2}|U_{1,12}, U_{2,12}) \\ \tilde{R}_{2,2} &\ge H(X_2|U_{1,12}, U_{2,12}, U_{1,2}, U_{2,2}) \end{aligned} \quad (28)$$

The first packet from E_1, destined to only S_1, carries indices $(m_{1,1}, l_{1,1})$ at rate $R_{1,1} = R''_{1,1} + \tilde{R}_{1,1}$. The second and third packets carry $m_{1,2}$ and $m_{1,12}$ at rates $R_{1,2} = R''_{1,2}$ and $R_{1,12} = R''_{1,12}$, respectively and are routed to the corresponding sinks. Similarly, 3 packets are transmitted from E_2 carrying indices $\{m_{2,1}, m_{2,12}, (m_{2,2}, l_{2,2})\}$ at rates $(R_{2,1}, R_{2,12}, R_{2,2}) = (R''_{2,1}, R''_{2,12}, R''_{2,2} + \tilde{R}_{2,2})$ to sinks $\{S_1, S_2, (S_1, S_2)\}$, respectively. Constraints for $(R_{1,1}, R_{1,2}, R_{1,12}, R_{2,1}, R_{2,2}, R_{2,12})$ can now be obtained by applying a standard Fourier-Motzkin elimination procedure on (24), (26), (27) and (28), and the achievable rate region simplifies to (21). The convex closure of achievable rates over all such random variables $(U_{1,12}, U_{1,1}, U_{1,2}, U_{2,12}, U_{2,1}, U_{2,2})$ gives the achievable rate region for the 2 source - 2 sink DIR problem. Observe that in the above illustration, we assumed that the decoding is performed in a sequential manner, i.e., the codewords of $U_{1,12}$ are decoded first followed by the codewords of $(U_{1,1})$ and $(U_{1,2})$, respectively. This was done only for the ease of understanding. In Theorem 1, we derive the conditions on rates for the decoders to find typical sequences from all the codebooks jointly (at once). Note that conditions on the rates for joint decoding is generally weaker (the region is larger) than that for sequential decoding. We also note that it is yet unknown if the above achievable rate region is complete or if there is an alternate coding scheme that achieves the trivial outerbound given by (17). Establishing the converse result for this setup in part of our future work.

VI. ACHIEVABLE REGION FOR THE GENERAL SETUP

We extend the coding scheme described in section V to derive an achievable rate region for the tuple

($R_{i,\mathcal{K}}$ $\forall i \in \Sigma, \mathcal{K} \in 2^{\Pi} - \phi$) using principles from multiple descriptions encoding [2], [15], [16] and Han and Kobayashi decoding [3], albeit with more complex notation. Without loss of generality, we assume that every source can send packets to every sink. For ease of understanding the notation, we use an arbitrary network with 2 encoders and 3 decoders as a running example throughout this section and explain the notation and the coding scheme for this example along with the general framework. Note that, for any network with two encoders and three decoders, $\Sigma = \{1, 2\}$, $\Pi = \{1, 2, 3\}$. For this example, we set $\Sigma_1 = \{1\}$, $\Sigma_2 = \{1, 2\}$ and $\Sigma_3 = \{2\}$.

Before stating the achievable rate region in Theorem 1, we define the following subsets of 2^{Π}:

$$\begin{aligned} \mathcal{I}_W &= \{\mathcal{K} : \mathcal{K} \in 2^{\Pi},\ |\mathcal{K}| = W\} \\ \mathcal{I}_{W+} &= \{\mathcal{K} : \mathcal{K} \in 2^{\Pi},\ |\mathcal{K}| > W\} \end{aligned} \tag{29}$$

Let $\mathcal{B}$ be any subset of Π with $|\mathcal{B}| \leq W$. We define the following subsets of $\mathcal{I}_W$ and $\mathcal{I}_{W+}$:

$$\begin{aligned} \mathcal{I}_W(\mathcal{B}) &= \{\mathcal{K} : \mathcal{K} \in \mathcal{I}_W,\ \mathcal{B} \subseteq \mathcal{K}\} \\ \mathcal{I}_{W+}(\mathcal{B}) &= \{\mathcal{K} : \mathcal{K} \in \mathcal{I}_{W+},\ \mathcal{B} \subseteq \mathcal{K}\} \end{aligned} \tag{30}$$

We also define:

$$\mathcal{J}(\mathcal{S}) = \{\mathcal{K} :\ \mathcal{K} \in 2^{\Pi},\ |\mathcal{K} \bigcap \mathcal{S}| > 0\} \tag{31}$$

Note that $\mathcal{J}(\Pi) = 2^{\Pi} - \phi$. To make the understanding simpler, we explicitly state a few of these subsets for the two encoder-three decoder running example: $2^{\Pi} = \{\phi, 1, 2, 3, 12, 13, 23, 123\}$, $\mathcal{I}_1 = \{1, 2, 3\}$, $\mathcal{I}_2 = \{12, 13, 23\}$, $\mathcal{I}_3 = \{123\}$, $\mathcal{I}_{1+} = \mathcal{I}_2 \cup \mathcal{I}_3$ and $\mathcal{I}_{2+} = \mathcal{I}_3$. Let $\mathcal{B} = \{12\}$, then, $\mathcal{I}_1(\mathcal{B}) = \phi$, $\mathcal{I}_2(\mathcal{B}) = \{12\}$, $\mathcal{I}_3(\mathcal{B}) = \{123\}$, $\mathcal{I}_{1+}(\mathcal{B}) = \{12, 123\}$, $\mathcal{I}_{2+}(\mathcal{B}) = \{123\}$ and $\mathcal{I}_{3+}(\mathcal{B}) = \phi$. Let $\mathcal{S} = \{1\}$, then $\mathcal{J}(\mathcal{S}) = \{1, 12, 13, 123\}$.

Let $\mathcal{Q}$ be any subset of $2^{\Pi} - \phi$. We say that $\mathcal{Q} \in \mathcal{Q}^*$ if it satisfies the following property $\forall \mathcal{K} \in \mathcal{Q}$:

$$\text{if } \mathcal{K} \in \mathcal{Q} \ \Rightarrow\ \mathcal{I}_{|\mathcal{K}|+}(\mathcal{K}) \subset \mathcal{Q} \tag{32}$$

For example, one of the elements of $\mathcal{Q}^*$ for the two encoder-three decoder running example is $\mathcal{J}(\{1\}) = \{1, 12, 13, 123\}$.

Let $\{U_\Sigma\}_{\mathcal{J}(\Pi)}$ be any set of $N(2^M - 1)$ random variables defined on arbitrary finite alphabets, jointly distributed with $\{X\}_\Sigma$ satisfying the following: $\forall j \in \Pi$,

$$P(\{X\}_\Sigma, \{U_\Sigma\}_{\mathcal{J}(j)}) = P(\{X\}_\Sigma) \prod_{i \in \Sigma} P(\{U_i\}_{\mathcal{J}(j)} | X_i) \tag{33}$$

The above Markov condition ensures that all the codewords which reach a sink are jointly typical with $\{X\}_{\Sigma_j}$. For the running example, this implies that the following Markov chain conditions hold:

$$\begin{aligned} &\{U_1\}_{\{123,12,13,1\}} \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow \{U_2\}_{\{123,12,13,1\}} \\ &\{U_1\}_{\{123,12,23,2\}} \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow \{U_2\}_{\{123,12,23,2\}} \\ &\{U_1\}_{\{123,13,23,3\}} \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow \{U_2\}_{\{123,13,23,3\}} \end{aligned} \tag{34}$$

We next define $\alpha(i, \mathcal{Q})$ as:

$$\begin{aligned} \alpha(i, \mathcal{Q}) = &-H\left(\{U_i\}_{\mathcal{Q}} | X_i\right) \\ &+ \sum_{\mathcal{K} \in \mathcal{Q}} H\left(U_{i,\mathcal{K}} | \{U_i\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}\right) \end{aligned} \tag{35}$$

$\forall i \in \Sigma, \mathcal{Q} \subseteq \mathcal{J}(\Pi)$. We further define $\beta(k, \mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N)$ $\forall k \in \Pi,\ \mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N \subseteq \mathcal{J}(k)$ as:

$$\begin{aligned} \beta(k, \mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N) = &\ H(\{U_i\}_{\mathcal{Q}_i^c} \forall i | \{U_i\}_{\mathcal{Q}_i} \forall i) \\ &- \sum_{i \in \Sigma} \sum_{\mathcal{K} \in \mathcal{Q}_i^c} H(U_{i,\mathcal{K}} | \{U_i\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}) \end{aligned} \tag{36}$$

where $\mathcal{Q}_i^c = \mathcal{J}(k) - \mathcal{Q}_i$ and define $\gamma_k(\Gamma)$ as:

$$\gamma_k(\Gamma) = H\left(\{X\}_\Gamma | \{X\}_{\Gamma^c}, \{U_\Sigma\}_{\mathcal{J}(k)}\right) \forall k \in \Pi, \Gamma \subseteq \Sigma_k \tag{37}$$

where $\Gamma^c = \Sigma_k - \Gamma$.

Just to illustrate, for the two encoder-three decoder running example, if $i = 1$, $k = 1$, $\mathcal{Q} = \mathcal{Q}_1 = \{123, 12, 13\}$, $\mathcal{Q}_2 = \{123\}$ and $\Gamma = \{1\}$, then:

$$\begin{aligned} \alpha(i, \mathcal{Q}) &= -H\left(\{U_1\}_{\{123,12,13\}} | X_1\right) + H\left(U_{i,12} | U_{i,123}\right) \\ &\quad + H\left(U_{i,123}\right) + H\left(U_{i,12} | U_{i,123}\right) \\ \beta(k, \mathcal{Q}_1, \mathcal{Q}_2) &= -H(U_{1,1} | \{U_1\}_{\{123,12,13\}}) - H(U_{2,12} | U_{2,123}) \\ &\quad - H(U_{2,13} | U_{2,123}) - H(U_{2,1} | \{U_2\}_{\{12,13,123\}}) \\ &\quad + H\left(U_{1,1}, \{U_2\}_{\{12,13,1\}} | \{U_1\}_{\{123,12,13\}}, U_{2,123}\right) \\ \gamma_k(\Gamma) &= H\left(X_1 | \{U_{\{1,2\}}\}_{\{123,12,13,1\}}\right) \end{aligned} \tag{38}$$

We state our main result in the following Theorem.

Theorem 1: Achievable Rate Region for DIR :Let $\{U_\Sigma\}_{2^{\Pi} - \phi}$ be any set of random variables satisfying (33). Let ($R'_{i,\mathcal{K}}$ $\forall i \in \Sigma, \mathcal{K} \in 2^{\Pi} - \phi$) be any set of auxiliary rate tuples such that:

$$\sum_{\mathcal{K} \in \mathcal{Q}} R'_{i,\mathcal{K}} \geq \alpha(i, \mathcal{Q}) \tag{39}$$

$\forall \mathcal{Q} \in \mathcal{Q}^*$. *Further, let ($R''_{i,\mathcal{K}}$ $\forall i \in \Sigma, \mathcal{K} \in 2^{\Pi} - \phi$) be any set of rate tuples* such that*:*

$$\sum_{i \in \Sigma} \sum_{\mathcal{K} \in \mathcal{Q}_i^c} R''_{i,\mathcal{K}} \geq \sum_{i \in \Sigma} \sum_{\mathcal{K} \in \mathcal{Q}_i^c} R'_{i,\mathcal{K}} + \beta(k, \mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N) \tag{40}$$

for each $k \in \Pi$, $\forall \mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N \subseteq \mathcal{J}(k)$ satisfying *(32) such that $\exists i \in \{1, \ldots, N\} : \mathcal{Q}_i \neq \mathcal{J}(k)$. Let ($\tilde{R}_{i,\mathcal{K}}$ $\forall i \in \Sigma, \mathcal{K} \in 2^{\Pi_k} - \phi$) satisfy:*

$$\sum_{i \in \Gamma} \sum_{\mathcal{K} : k \in \mathcal{K}} \tilde{R}_{i,\mathcal{K}} \geq \gamma_k(\Gamma) \tag{41}$$

$\forall k \in \Pi, \Gamma \in 2^{\Sigma_k} - \phi$. *Then, the achievable rate region for the tuple ($R_{i,\mathcal{S}}$ $\forall i \in \Sigma, \mathcal{S} \in 2^{\Pi} - \phi$)* contains all rates such that,

$$R_{i,\mathcal{K}} \geq \begin{cases} R''_{i,\mathcal{K}} + \tilde{R}_{i,\mathcal{K}} & \text{if } \mathcal{K} \subseteq 2^{\Pi_i} - \phi \\ R''_{i,\mathcal{K}} & \text{if } \mathcal{K} \nsubseteq 2^{\Pi_i} - \phi \end{cases} \tag{42}$$

The convex closure of the achievable tuples over all such $N(2^M - 1)$ random variables satisfying (33) is the achievable rate region for DIR and is denoted by $\mathcal{R}_{DIR}$.

Remark 1: The converse to this achievability region does not hold in general. A simple counter example follows from the famous binary modulo two sum problem proposed by Körner and Marton for the 2 helper setup in [34]. However, in section VII we prove the converse for certain special cases.

Remark 2: The coding scheme in Theorem 1 can be easily specialized to 'power binning' by setting $\{U_\Sigma\}_{2^\Pi-\phi}$ to constants. This effectively becomes the 'no-helpers' scenario as setting $\{U_\Sigma\}_{2^\Pi-\phi}$ to constants implies that $R_{i,\mathcal{S}} = 0\ \forall \mathcal{S} \notin 2^{\Pi_i}$.

Remark 3: If paths from certain source nodes to sink nodes do not exist in the network, then several of the auxiliary random variables in the above theorem can be set to constants, without any loss of optimality. However, in general, it is not obvious how to eliminate these auxiliary random variables without compromising the rate region. We will be considering this and related optimization problems on computability of the rate region, as part of our future work.

Remark 4: Lastly, we note that deriving cardinality bounds to the auxiliary random variables in the above theorem is not straight-forward and application of standard Carathéodry theorem based arguments leads to indefinite cardinality bounds, similar to the Marton's achievable region for general broadcast channels. Just to illustrate, for the two-encoder-three-decoder running example, we obtain the following bounds on the cardinalities of $\{U_1\}_{\{123,12,13,23\}}$:

$$\begin{aligned} |\mathcal{U}_{1,123}| &\le |\mathcal{X}_1| + C_{123} \\ |\mathcal{U}_{1,12}| &\le |\mathcal{X}_1||\mathcal{U}_{1,13}||\mathcal{U}_{1,23}| + C_{12} \\ |\mathcal{U}_{1,13}| &\le |\mathcal{X}_1||\mathcal{U}_{1,12}||\mathcal{U}_{1,23}| + C_{13} \\ |\mathcal{U}_{1,23}| &\le |\mathcal{X}_1||\mathcal{U}_{1,12}||\mathcal{U}_{1,13}| + C_{23} \end{aligned} \tag{43}$$

where C_{123}, C_{12}, C_{13} and C_{23} are constants,[4] independent of $|\mathcal{X}_1|$. Observe that these bounds are indefinite, in the sense that $|\mathcal{U}_{1,12}|$, $|\mathcal{U}_{1,23}|$ and $|\mathcal{U}_{1,13}|$ depend on each other. However, it is possible to derive definite bounds for certain special cases of the general DIR setup, as we will see in Section VII. We further note that it may be possible to apply recent results on cardinality bounding for Marton's region, based on perturbation techniques [35], to the DIR setup. However, the details are quite involved and orthogonal to the rest of this paper. Hence, we will consider deriving such bounds as part of our future work.

Proof: We refer to Appendix A for formal definitions and basic Lemmas associated with typicality.

Encoding: Suppose we are given $\{U_\Sigma\}_{2^\Pi-\phi}$ satisfying (33). As in section V, the encoding at each node is divided into 3 stages:

1) *Stage 1:* We focus on the encoding at source node E_i. The codebook generation is done following the order of $U_{i,\mathcal{K}}$, $|\mathcal{K}| = M, M-1, M-2\ldots, 1$ as shown in Fig. 6. First, $2^{nR'_{i,\Pi}}$ independent codewords of $U_{i,\Pi}$, $u^n_{i,\Pi}(j)$ $j \in \{1\ldots 2^{nR'_{i,\Pi}}\}$, are generated according to the density $\prod_{t=1}^n P_{U_{i,\Pi}}(u^{(t)}_{i,\Pi})$. Conditioned on each codeword $u^n_{i,\Pi}(j)$, $2^{nR'_{i,\mathcal{K}}}$ codewords of $U_{i,\mathcal{K}} : |\mathcal{K}| = M-1$ are generated independent of each other according to the conditional density $\prod_{t=1}^n P_{U_{i,\mathcal{K}}|U_{i,\Pi}}(u^{(t)}_{i,\mathcal{K}}|u^{(t)}_{i,\Pi})$. Similarly, $\forall \mathcal{K} : |\mathcal{K}| < M$, $2^{nR'_{i,\mathcal{K}}}$ codewords of $U_{i,\mathcal{K}}$ are independently generated conditioned on each codeword tuple of $\{U_i\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}$ according

[4]Applying Carathéodry theorem, without accounting for several repetitions in terms, leads to values of $C_{123} = 87$, $C_{12} = C_{13} = C_{23} = 58$.

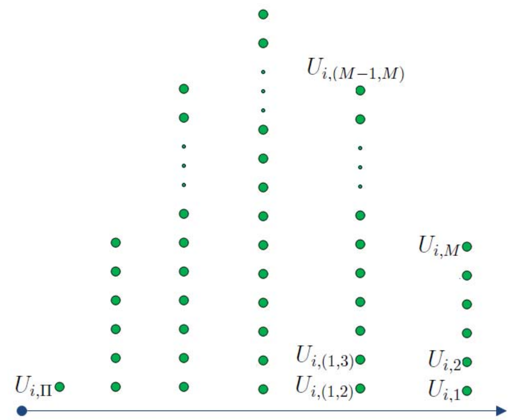

Fig. 6. Illustrates the order of codebook generation at source i.

to $\prod_{t=1}^n P_{U_{i,\mathcal{K}}|\{U_i\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}}(u^{(t)}_{i,\mathcal{K}}|\{u_i\}^{(t)}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})})$. Note that to generate the codewords of $U_{i,\mathcal{K}}$, we first need all the codebooks of $\{U_i\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}$. On observing a sequence, x^n_i, the encoder at E_i attempts to find a set of codewords, one for each variable, such that they are all jointly typical. If it fails to find such a set, it declares an error. Codebooks are generated similarly at all the source nodes. Note that all the random variables $U_{i,i} \forall i \in \Sigma$ can be set to constants without changing the rate region of Theorem 1. However, we continue to use them to avoid more complex notation.

For the two encoder-three decoder running example, we focus on the encoding at source node E_1. First, $2^{nR'_{1,123}}$ independent codewords of $U_{1,123}$ are generated according to the density $\prod_{t=1}^n P_{U_{1,123}}(u^{(t)}_{1,\Pi})$. These codewords are denoted by $u^n_{1,123}(j)$ $j \in \{1,\ldots,2^{nR'_{1,123}}\}$. Next, conditioned on each codeword of $U_{1,123}$, independent codewords of $U_{1,12}$, $U_{1,13}$ and $U_{1,23}$ are generated according to the respective conditional densities at rates $R'_{1,12}$, $R'_{1,13}$ and $R'_{1,23}$, respectively. Finally, codewords of $U_{1,1}$, $U_{1,2}$ and $U_{1,3}$, at rates $R'_{1,1}$, $R'_{1,2}$ and $R'_{1,3}$, a generated conditioned on each codeword tuple of $\{U_1\}_{\{123,12,13\}}$, $\{U_1\}_{\{123,12,23\}}$ and $\{U_1\}_{\{123,13,23\}}$, respectively. On observing a sequence x^n_1, the encoder attempts to find a set of codeword tuples from the codebooks of $\{U_1\}_{\{123,12,13,23,1,2,3\}}$, such that they are all jointly typical.

2) *Stage 2:* In stage 2, the codewords in each codebook are divided into uniform bins. Specifically, the $2^{nR'_{i,\mathcal{K}}}$ codewords in any codebook of $U_{i,\mathcal{K}}$ are subdivided into $2^{nR''_{i,\mathcal{K}}}$ bins, with each bin containing $2^{n(R'_{i,\mathcal{K}}-R''_{i,\mathcal{K}})}$ codewords. All the codewords which have the same bin index m are said to fall in the bin $\mathcal{C}_{i,\mathcal{K}}(m)$ $\forall m \in (1\ldots 2^{nR''_{i,\mathcal{K}}})$. If in stage 1, the encoder succeeds in finding a jointly typical set of codewords, the bin index of the codeword of $U_{i,\mathcal{K}}$ is sent as part of packet $\mathcal{P}_{i,\mathcal{K}}$.

3) *Stage 3 Power Binning:* In this stage, each typical sequence of X_i is assigned $2^{|\Pi_i|}-1$ indices, randomly generated using uniform pmfs over $(1,\ldots,2^{\tilde{R}_{i,\mathcal{K}}})$ $\forall \mathcal{K} \in 2^{\Pi_i} - \phi$, respectively. All the sequences of i which have the same bin index l are said to fall in the bin $\mathcal{B}_{i,\mathcal{K}}(l)$ $\forall l \in (1\ldots 2^{n\tilde{R}_{i,\mathcal{K}}})$. On observing a sequence x^n_i, if it is typical, the encoder sends the corresponding bin indices in the packets $\mathcal{P}_{i,\mathcal{K}}$:

$\mathcal{K} \in 2^{\Pi_i} - \phi$, in addition to the bin indices in stage 2. If it is not typical, the encoder declares an error. Note that all packets from source node E_i to a subset of sinks $\mathcal{K}$ such that $\mathcal{K} \subseteq 2^{\Pi_i} - \phi$, carry two bin indices, one each from stages 2 and 3, respectively.

Just to illustrate, for the example setup considered, the following bin indices are sent from encoder 1 to decoder 1: Bin indices of $\{U_{1,123}, U_{1,12}, U_{1,13}, U_{1,1}\}$ from the second stage of encoding at rates $\{2^{R''_{1,123}}, 2^{R''_{1,12}}, 2^{R''_{1,13}}, 2^{R''_{1,1}}\}$, respectively and two independently assigned bin indices from the third stage of encoding at rates $\{2^{\tilde{R}_{1,1}}, 2^{\tilde{R}_{1,12}}\}$, respectively. Similarly, the following indices are sent from encoder 2 to decoder 2 : Bin indices of $\{U_{2,123}, U_{2,12}, U_{2,23}, U_{2,2}\}$ from the second stage of encoding at rates $\{2^{R''_{2,123}}, 2^{R''_{2,12}}, 2^{R''_{2,13}}, 2^{R''_{2,1}}\}$, respectively and two independently assigned bin indices from the third stage of encoding at rates $\{2^{\tilde{R}_{2,2}}, 2^{\tilde{R}_{2,23}}\}$, respectively.

In Appendix B, we show that, if the rates $R'_{i,\mathcal{K}}$ satisfy (39), then the probability of encoding error asymptotically approaches zero, i.e., we can, with probability approaching 1, find a codeword tuple, one from each codebook such that all the codewords are jointly typical if the rates satisfy (39). Let the codewords, which are jointly typical with x_i^n, be denoted as $u^*_{i,\mathcal{K}}$ $\forall \mathcal{K} \in \mathcal{J}(\Pi) = 2^{\Pi} - \phi$. To ensure joint typicality of $(\{x\}^n_\Sigma, \{u^*_\Sigma\}_{\mathcal{J}(k)})$, we require a stronger version of the "conditional Markov lemma" in [14]. We state and prove this stronger version, called the "conditional Markov lemma for mutual covering" in Appendix C. From this lemma, it follows that $\left(\{x\}^n_\Sigma, \{u^*_\Sigma\}_{\mathcal{J}(k)}\right) \in \mathcal{T}^n_\epsilon\left(\{X\}^n_\Sigma, \{U^*_\Sigma\}_{\mathcal{J}(k)}\right)$ with very high probability given that the encoding at all the source nodes is error free. Let the bin indices of $u^*_{i,\mathcal{K}}$ (assigned in stage 2) be denoted by $m_{i,\mathcal{K}}$ $\forall \mathcal{K} \in 2^{\Pi} - \phi$ and let the bin indices of x_i^n (assigned in stage 3) be denoted by $l_{i,\mathcal{K}} \forall \mathcal{K} \in 2^{\Pi_i} - \phi$.

Decoding: We focus on a particular sink S_k. Sink S_k receives all the indices $\{m_\Sigma\}_{\mathcal{J}(k)}$ of stage 2 of encoding from all source nodes. It also receives $\{l_{\Sigma_k}\}_{\mathcal{J}(k)}$ of stage 3 of encoding from source nodes Σ_k. In the first stage of decoding, it begins decoding $u^*_{i,\mathcal{J}(k)}$ $\forall i \in \Sigma$ by looking for a unique jointly typical codeword tuple from $\{\mathcal{C}_{i,\mathcal{J}(k)}(m_{i,\mathcal{J}(k)}); \forall i \in \Sigma\}$. Clearly $\{u^*_\Sigma\}_{\mathcal{J}(k)}$ satisfies this property. If the decoder finds another such jointly typical codeword tuple in the received bins, it declares an error. In Appendix B, we show that if conditions (40) are satisfied by $R''_{i,\mathcal{K}}$, then the probability that the decoder finds another such jointly typical codeword tuple approaches zero.

In the last stage of decoding, after having decoded all $\{u^*_\Sigma\}_{\mathcal{J}(k)}$, the decoder looks for unique source sequences from $\bigcap\{\mathcal{B}_{i,\mathcal{K}}(l_{i,\mathcal{K}}) : i \in \Sigma_k, \mathcal{K} \ni k\}$ which are jointly typical with $\{u^*_\Sigma\}_{\mathcal{J}(k)}$. Hence what remains is to find conditions on $\tilde{R}_{i,\mathcal{K}}$ to ensure lossless reconstruction of the respective sources at each sink. Following similar steps as in [3] and [7], it is easy to show that this probability can be made arbitrarily small if (41) is satisfied $\forall \Gamma \in 2^{\Sigma_k} - \phi$. We have shown that if the rates satisfy the conditions in Theorem 1, the probability of decoding error at each sink can be made arbitrarily small. Arbitrarily small decoding error ensures that the decoder decodes the correct sequence with very high probability. Hence, if the rate constraints are satisfied, for any $\epsilon > 0$, we can find a sufficiently large n such that:

$$P(\hat{X}^n_{\sum_j} \neq X^n_{\sum_j}) < \epsilon \tag{44}$$

For the running example, let us focus on sink 1. The decoder at sink 1 receives the following bin indices from the second stage of encoding at both the sources : $\{U_{1,2}\}_{\{123,12,13,1\}}$. During the first stage of decoding, it attempts to recover the codeword tuple $\{u^n_{\{1,2\}}\}_{\{123,12,13,1\}}$, from their respective bin indices. The decoder can recover the codeword tuple if there is no other tuple, that is jointly typical, within the bins. It follows from Appendix B that, if the codebook and binning rates satisfy certain conditions, the decoder can recover the codeword indices with probability 1, as $n \to \infty$. Ofcourse, the task of the decoder is to recover the source sequence x_1^n with vanishing probability of error. Hence, during the second stage of decoding, the decoder attempts to find a unique source sequence from the bin indices received from the third stage of encoding, that is jointly typical with all the recovered codewords. It is easy to show that if $\tilde{R}_{1,1}$ and $\tilde{R}_{1,12}$ satisfy (41), then the decoder succeeds in recovering the source sequence with probability approaching 1, as $n \to \infty$.

Recall that packets from source node E_i to sinks $\mathcal{K} \subseteq \Pi_i$ carry both $m_{i,\mathcal{K}}$ (at rate $R''_{i,\mathcal{K}}$) and $l_{i,\mathcal{K}}$ (at rate $\tilde{R}_{i,\mathcal{K}}$). While the other packets carry only $m_{i,\mathcal{K}}$ (at rate $R''_{i,\mathcal{K}}$). Hence, the rates of each packet must satisfy the following constraints for lossless decoding of the requested sources:

$$R_{i,\mathcal{K}} \geq \begin{cases} R''_{i,\mathcal{K}} + \tilde{R}_{i,\mathcal{K}} & \text{if } \mathcal{K} \subseteq 2^{\Pi_i} - \phi \\ R''_{i,\mathcal{K}} & \text{if } \mathcal{K} \nsubseteq 2^{\Pi_i} - \phi \end{cases} \tag{45}$$

proving the theorem. ∎

Remark 5 (Note on Joint Typicality of Codewords): In the above theorem, we imposed joint typicality of all codewords, $\{u^*_i\}_{\mathcal{J}(\Pi)}$, with x_i^n. However, this is an unnecessary restriction. The problem setup only requires joint typicality of codewords within subsets $\{u^*_i\}_{\mathcal{J}(k)}$, i.e., it is sufficient if $\{\{u^*_i\}_{\mathcal{J}(k)}, x_i^n\} \in \mathcal{T}^n_\epsilon$, $\forall k \in \Pi$, $\forall i \in \Sigma$. Imposing joint typicality only within subsets can potentially lead to strictly larger achievable region compared to that in the above theorem. Deriving such bounds in a challenging and important problem of its own right and is beyond the scope of this paper. It will be considered as part of our future work. Nevertheless, we would like to point out that, in Section V, for the setup shown in Fig. **5,** we actually imposed joint typicality only within the required subsets and did not impose joint typicality across all codebooks. For this setup however, it is easy to verify that both approaches lead to the same achievable region and imposing subset wise joint typicality does not provide any improvement. In fact, it is possible to show that for any DIR setup with just two sinks, imposing typicality only within subsets leads to the same region as imposing joint typicality across all codewords. This fact will be implicitly exploited in the converse theorem in Section VII-B.

Remark 6: We note that the above coding scheme utilizes conditional codebooks to exploit correlation between the transmitted packets from a given source while using random binning techniques to exploit the dependencies across sources.

An alternative encoding scheme is to generate all the codebooks independently and to exploit all the dependencies only using random binning (see [36]). However, it is possible to show that the resulting rate region is generally subsumed in the rate region derived in this paper. It may be possible for certain scenarios that random binning over independently generated codebooks achieves the same rate region as Theorem 1. Extensions in this direction, including lossy compression are currently under investigation. Lastly, we note that this approach of generating a combinatorial number of codewords has been shown in precursor work to be useful in related applicational contexts of multiple descriptions coding [2], [37]–[39] and successive refinement with side information [40], [41].

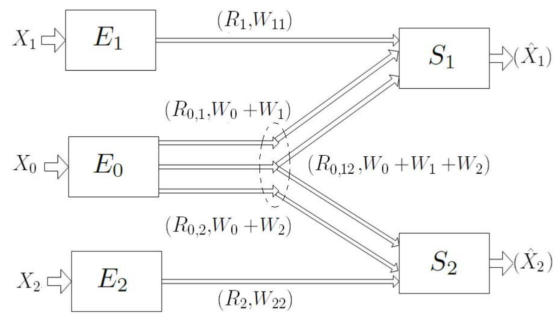

Fig. 7. Example of a 2-sink, 1 Helper DIR.

VII. Outerbounds to Certain Special Scenarios

We note that the converse to the achievability region does not hold in general. However, we can prove the converse for two important special cases.

A. When There Are no Helpers

Theorem 2: When each sink is allowed to receive packets only from sources it intends to reconstruct, the complete rate region for dispersive information routing is given by: $\forall j \in \Pi$ *and* $\forall \mathcal{S} \in 2^{\Sigma_j} - \phi$:

$$\sum_{i \in \mathcal{S}} \sum_{\mathcal{K} \in 2^{\Pi_i} - \phi,\ \mathcal{K} \ni j} R_{i,\mathcal{K}} \geq H\left(\{X\}_{\mathcal{S}} | \{X\}_{\Sigma_j - \mathcal{S}}\right) \tag{46}$$

It is achieved by 'Power Binning'.

Proof: In the achievable rate region of Theorem 1, setting $U_{i,\mathcal{S}} = \Phi\ \forall i \in \Sigma, \mathcal{S} \in 2^{\Pi} - \phi$, where Φ is a constant, leads to the above rate region. The converse to this rate region follows directly from the converse to the lossless source coding theorem [22] (section 15.4.2). In fact, the above rate region is the cut-set upper bound for the DIR setup. Fix some $j \in \Pi$ and $\mathcal{S} \in 2^{\Sigma_j} - \phi$. To derive the upper bound, let us assume that the sink j can observe all sources $i \in \{\Sigma_j - \mathcal{S}\}$, i.e., all sources $\{\Sigma_j - \mathcal{S}\}$ are available to sink j as side information. The sink must reconstruct $\{X\}_{\mathcal{S}}$ from all the packets it receives from the remaining sources. From the lossless source coding theorem, this is possible only if:

$$\sum_{i \in \mathcal{S}} \sum_{\mathcal{K} \in 2^{\Pi_i} - \phi,\ \mathcal{K} \ni j} R_{i,\mathcal{K}} \geq H\left(\{X\}_{\mathcal{S}} | \{X\}_{\Sigma_j - \mathcal{S}}\right) \tag{47}$$

These conditions must be satisfied $\forall j \in \Pi$ and $\forall \mathcal{S} \in 2^{\Sigma_j} - \phi$. Hence, the converse follows. ■

B. A 2-Sink Network With a Single Helper

The converse can be proven in general for any 2 sink network with a single helper. However, to avoid complex notation, we just give a simple example of a 2 sink network with a single helper and prove the converse to the rate region. The proof of converse for a general 2 sink network with a single helper follows in similar lines.

Consider the network shown in Fig. 7, with 3 source nodes and 2 sinks. The three source nodes E_1, E_0, E_2 observe three correlated memoryless random sequences X_1^n, X_0^n, X_2^n, respectively. The two sinks S_1 and S_2 respectively reconstruct X_1^n and X_2^n losslessly. Note that E_0 acts as a helper to both the sinks. Our objective is to find the rate region for the tuple $(R_1, R_2, R_{0,1}, R_{0,2}, R_{0,\{1,2\}})$ for lossless reconstruction of the respective sources. The following theorem establishes the complete rate region.

Theorem 3: Let (U_0, U_1, U_2) *be random variables distributed over arbitrary finite sets* $\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2$*, jointly distributed with* (X_1, X_0, X_2) *such that the following hold:*

$$\begin{aligned} X_1 &\leftrightarrow X_0 \leftrightarrow (U_0, U_1, U_2) \\ X_2 &\leftrightarrow X_0 \leftrightarrow (U_0, U_1, U_2) \end{aligned} \tag{48}$$

Then any rate tuple satisfying the following constraints is achievable for the 2-Sink 1-Helper DIR problem:

$$\begin{aligned} R_{0,12} &\geq I(X_0; U_0) \\ R_{0,1} &\geq I(X_0; U_1|U_0) \\ R_{0,2} &\geq I(X_0; U_2|U_0) \\ R_{1,1} &\geq H(X_1|U_0, U_1) \\ R_{2,2} &\geq H(X_2|U_0, U_2) \end{aligned} \tag{49}$$

The closure of the achievable rates over all such (U_0, U_1, U_2) *is the complete rate region for this setup.*

Remark 7: Bounds on cardinalities of (U_0, U_1, U_2) follow using standard techniques, based on Carathéodry theorem [42] (see also [33] Appendix C):

$$\begin{aligned} |\mathcal{U}_0| &\leq |\mathcal{X}_0| + 4 \\ |\mathcal{U}_1| &\leq |\mathcal{X}_0|(|\mathcal{X}_0| + 4) + 1 \\ |\mathcal{U}_2| &\leq |\mathcal{X}_0|(|\mathcal{X}_0| + 4) + 1 \end{aligned} \tag{50}$$

Proof Achievability: Let (U_0, U_1, U_2) be any random variables satisfying (48). The following achievable rate region is obtained by setting $U_{0,12} = U_0$, $U_{0,1} = U_1$, $U_{0,12} = U_2$ and all the remaining random variables to constants in the general achievable rate region of Theorem 1:

$$\begin{aligned} R_{0,12} &\geq I(X_0; U_0) \\ R_{0,12} + R_{0,1} &\geq I(X_0; U_0) + I(X_0; U_1|U_0) \\ R_{0,12} + R_{0,2} &\geq I(X_0; U_0) + I(X_0; U_2|U_0) \\ R_{0,12} + R_{0,1} + R_{0,2} &\geq I(X_0; U_1, U_2, U_0) + I(U_1; U_2|U_0) \\ R_{1,1} &\geq H(X_1|U_0, U_1) \\ R_{2,2} &\geq H(X_2|U_0, U_2) \end{aligned} \tag{51}$$

We further restrict the joint density to satisfy the following Markov condition in addition to (48):

$$U_1 \leftrightarrow (X_0, U_0) \leftrightarrow U_2 \tag{52}$$

On using this Markov condition in (51), the sum rate constraint on $R_{0,12} + R_{0,1} + R_{0,2}$ becomes:

$$\begin{aligned} R_{0,12} + R_{0,1} + R_{0,2} \geq I(X_0; U_0) + I(X_0; U_1|U_0) \\ +I(X_0; U_2|U_0) \end{aligned} \tag{53}$$

Observe that if a rate tuple satisfies (49), then it also satisfies (51) and hence the region given by (49) is achievable for the 2-Sink 1-Helper problem shown in Fig. 7.

Converse: Recall the notation in the definition of an achievable rate region in Section IV-B. The output of encoder 1 is denoted $f_1^E(X_1^n) = T_1$ and the output of encoder 2 is $f_2^E(X_2^n) = T_2$. Remember that $0 \leq T_1 \leq 2^{nR_1}$ and $0 \leq T_2 \leq 2^{nR_2}$. Similarly the encoder at E_0 transmits 3 indices denoted by $(T_{0,1}, T_{0,2}, T_{0,12})$ which are routed to the respective sinks. Sink S_1 receives $(T_1, T_{0,1}, T_{0,12})$ and reconstructs X_1^n with vanishing probability of error. Similarly sink S_2 receives $(T_2, T_{0,2}, T_{0,12})$ and reconstructs X_2^n losslessly. We need to prove that for any code with vanishing probability of error, the rates must satisfy (49) for some (U_0, U_1, U_2) satisfying (48).

We follow standard converse techniques to prove the above claim. We begin with the following series of inequalities:

$$\begin{aligned} nR_{0,12} &\geq H(T_{0,12}) \geq I(X_0^n; T_{0,12}) \\ &= \sum_{i=1}^n I(X_0^i; T_{0,12}|X_0^{1,i-1}) \\ &=^{(a)} \sum_{i=1}^n I(X_0^i; T_{0,12}, X_0^{1,i-1}) \\ &=^{(b)} \sum_{i=1}^n I(X_0^i; U_{0,12}^i) \end{aligned} \tag{54}$$

where (a) follows from the memoryless property of the sources and (b) follows by setting $U_{0,12}^i = (T_{0,12}, X_0^{1,i-1})$. Here X_0^i denotes the i'th realization of X_0^n and $X_0^{1,i-1}$ denotes the first $i-1$ realizations of X_0^n. Next we have:

$$\begin{aligned} nR_{0,1} &\geq H(T_{0,1}) \geq H(T_{0,1}|T_{0,12}) \\ &\geq I(X_0^n; T_{0,1}|T_{0,12}) \\ &= \sum_{i=1}^n I(X_0^i; T_{0,1}|T_{0,12}, X_0^{1,i-1}) \\ &= \sum_{i=1}^n I(X_0^i; U_{0,1}^i|U_{0,12}^i) \end{aligned} \tag{55}$$

Where $U_{0,1}^i = (T_{0,1})\ \forall i$. Similarly, we can show that:

$$nR_{0,2} \geq \sum_{i=1}^n I(X_0^i; U_{0,2}^i|U_{0,12}^i) \tag{56}$$

where $U_{0,2}^i = (T_{0,2})\ \forall i$. Note that as $(T_{0,1}, T_{0,2}, T_{0,12}, X_0^{1,i-1})$ depends on (X_1^i, X_2^i) only through X_0^i, we have the following two Markov chain conditions:

$$\begin{aligned} X_1^i \leftrightarrow X_0^i \leftrightarrow (U_0^i, U_1^i, U_2^i) \\ X_2^i \leftrightarrow X_0^i \leftrightarrow (U_0^i, U_1^i, U_2^i) \end{aligned} \tag{57}$$

Further, we need lossless reconstruction of X_1^n at S_1. The following series of inequalities hold:

$$\begin{aligned} nR_1 &\geq H(T_1) \\ &\geq H(T_1|T_{0,12}, T_{0,1}) \\ &= H(T_1|T_{0,12}, T_{0,1}) + H(X_1^n|T_{0,12}, T_{0,1}, T_1) \\ &\quad -H(X_1^n|T_{0,12}, T_{0,1}, T_1) \\ &\geq^{(a)} H(X_1^n, T_1|T_{0,12}, T_{0,1}) - n\epsilon_n \\ &= H(X_1^n|T_{0,12}, T_{0,1}) - n\epsilon_n \\ &= \sum_{i=1}^n H(X_1^i|X_1^{i-1}, T_{0,12}, T_{0,1}) - n\epsilon_n \\ &= \sum_{i=1}^n H(X_1^i|U_{0,12}^i, U_{0,1}^i) - n\epsilon_n \end{aligned} \tag{58}$$

where (a) follows from Fano's inequality, i.e., $H(X_1^n|T_1, T_{0,1}, T_{0,12}) < n\epsilon_n$. Similarly, for lossless reconstruction at S_2, we have:

$$nR_2 \geq \sum_{i=1}^n H(X_2^i|U_{0,12}^i, U_{0,2}^i) - n\epsilon_n \tag{59}$$

We next introduce a time sharing random variable $Q \sim \text{Unif}[1:n]$, independent of $(X_0^n, X_1^n, X_2^n, U_{0,1}^n, U_{0,2}^n, U_{0,12}^n)$, so that we can rewrite (54), (55), (56), (58) and (59) as:

$$\begin{aligned} nR_{0,12} &\geq I(X_0^Q; U_{0,12}^Q|Q) = I(X_0^Q; U_{0,12}^Q, Q) \\ nR_{0,1} &\geq I(X_0^Q; U_{0,1}^Q|U_{0,12}^Q, Q) \\ &= I(X_0^Q; U_{0,1}^Q, Q|U_{0,12}^Q, Q) \\ nR_{0,2} &\geq I(X_0^Q; U_{0,2}^Q|U_{0,12}^Q, Q) \\ &= I(X_0^Q; U_{0,2}^Q, Q|U_{0,12}^Q, Q) \\ nR_1 &\geq H(X_1^Q|U_{0,12}^Q, U_{0,1}^Q, Q) \\ nR_2 &\geq H(X_2^Q|U_{0,12}^Q, U_{0,2}^Q, Q) \end{aligned} \tag{60}$$

Setting $(U_{0,12}^Q, Q) = U_{0,12}$, $(U_{0,1}^Q, Q) = U_{0,1}$, $(U_{0,2}^Q, Q) = U_{0,2}$ and observing that (X_0^Q, X_1^Q, X_2^Q) has the same density as (X_0, X_1, X_2) we get the rate region given in (49). ■

VIII. Strict Improvement for Networks With Helpers

In this section, we consider a variant of Han and Kobayashi's rate region for the DIR setup and show that the rate region in Theorem 1 is strictly larger. Note that the original characterization by Han and Kobayashi assumes that each encoder transmits a single packet that is sent to a subset of sinks. Application of Han and Kobayashi's coding scheme to the DIR setup requires artificial imposition of independent encoders at each source, which leads to strict sub-optimality in the rate region, for networks with helpers. Specifically, it would require us to assume $2^{|\Pi|} - 1$ independent encoders at each source that generate the $2^{|\Pi|} - 1$ packets that are routed to the respective sinks. We show that, for the 2-sink-1-helper setup shown in Fig. 7, the rate region in Theorem 1 strictly improves upon the extended Han and Kobayashi rate region derived assuming three independent encoders at E_0. We denote the Han and Kobayashi rate region obtained by assuming independent encoders at each source by $\mathcal{R}_{HK-I}$ and the rate region in Theorem 1 by $\mathcal{R}_{DIR}$. Before stating the result, we specialize the Han and Kobayashi rate region to the DIR setup and state the rate region explicitly.

Let $\{U_\Sigma\}_{\mathcal{J}(\Pi)}$ be any set of $N(2^M-1)$ random variables defined on arbitrary finite alphabets, jointly distributed with $\{X\}_\Sigma$, satisfying the following conditions $\forall i \in \Sigma$, $j \in \Pi$:

$$P(\{X\}_\Sigma, \{U_\Sigma\}_{\mathcal{J}(j)}) = P(\{X\}_\Sigma) \prod_{i \in \Sigma} P(\{U_i\}_{\mathcal{J}(j)}|X_i) \quad (61)$$

$$P\left(\{U_i\}_{\mathcal{J}(\Pi)}|X_i\right) = \prod_{\mathcal{K}\in 2^\Pi-\phi} P\left(U_{i,\mathcal{K}}\Big|X_i\right) \quad (62)$$

Let $(R''_{i,\mathcal{K}}\ \forall i \in \Sigma, \mathcal{K} \in 2^\Pi - \phi)$, be any set of auxiliary rate tuples such that:

$$\sum_{i\in\Sigma}\sum_{\mathcal{K}\in\mathcal{Q}_i^c} R''_{i,\mathcal{K}} \geq I(\{X\}_\Sigma; \{U_i\}_{\mathcal{Q}_i^c}\ \forall i | \{U_i\}_{\mathcal{Q}_i}\ \forall i) \quad (63)$$

for each $k \in \Pi$, $\forall \mathcal{Q}_1, \mathcal{Q}_2, \dots \mathcal{Q}_N \subseteq \mathcal{J}(k)$ satisfying (32), such that $\exists i \in \{1,\dots,N\} : \mathcal{Q}_i \neq \mathcal{J}(k)$. Let $(\tilde{R}_{i,\mathcal{K}}\ \forall i \in \Sigma, \mathcal{K} \in 2^{\Pi_k} - \phi)$ satisfy:

$$\sum_{i\in\Gamma}\sum_{\mathcal{K}:k\in\mathcal{K}} \tilde{R}_{i,\mathcal{K}} \geq \gamma_k(\Gamma) \quad (64)$$

$\forall k \in \Pi$, $\Gamma \in 2^{\Sigma_k} - \phi$. Then, an achievable rate region for the tuple $(R_{i,\mathcal{S}}\ \forall i \in \Sigma, \mathcal{S} \in 2^\Pi - \phi)$ contains all rates such that,

$$R_{i,\mathcal{K}} \geq \begin{cases} R''_{i,\mathcal{K}} + \tilde{R}_{i,\mathcal{K}} & \text{if } \mathcal{K} \subseteq 2^{\Pi_i} - \phi \\ R''_{i,\mathcal{K}} & \text{if } \mathcal{K} \nsubseteq 2^{\Pi_i} - \phi \end{cases} \quad (65)$$

The convex closure of these achievable tuples over all such $N(2^M-1)$ random variables satisfying (62) is denoted by $\mathcal{R}_{HK-I}$. It is easy to verify that that the above region can also be obtained from the region in Theorem 1, by restricting the auxiliary random variables to satisfy (62), in addition to (33). We next formally state the strict improvement result as part of the following theorem.

Theorem 4: (a) For any network, we have:

$$\mathcal{R}_{HK-I} \subseteq \mathcal{R}_{DIR}$$

(b) For any network with no helpers:

$$\mathcal{R}_{HK-I} = \mathcal{R}_{DIR}$$

(c) There exist network scenarios with helpers for which:

$$\mathcal{R}_{HK-I} \subset \mathcal{R}_{DIR}$$

Specifically, for the 2-sink-1-helper setup shown in Fig. 7, $\mathcal{R}_{HK-I} \subset \mathcal{R}_{DIR}$.

Proof: Parts (a) and (b) are rather straightforward to prove. (a) follows by noting that $\mathcal{R}_{HK-I}$ can be obtained from $\mathcal{R}_{DIR}$ by restricting the auxiliary random variables to satisfy (62), in addition to (33). (b) also follows directly, as for any network with no helpers, all the auxiliary random variables in the characterizations of both $\mathcal{R}_{HK-I}$ and $\mathcal{R}_{DIR}$, can be set to constants without any loss of optimality. In fact, both the regions degenerate to the power binning region, which is complete for networks with no helpers.

To prove (c), we consider the 2-sink-1-helper setup with binary symmetric sources and show that $\mathcal{R}_{DIR}$ is strictly larger than $\mathcal{R}_{HK-I}$. We first restate the Han and Kobayashi rate region assuming independent encoders at E_0. Let (U_0, U_1, U_2) be random variables distributed over arbitrary finite sets $\mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2$, jointly distributed with (X_1, X_0, X_2) such that the following conditions hold:

$$\begin{aligned} X_1 \leftrightarrow X_0 &\leftrightarrow (U_0, U_1, U_2) \\ X_2 \leftrightarrow X_0 &\leftrightarrow (U_0, U_1, U_2) \\ P(U_0, U_1, U_2|X_0) &= P(U_0|X_0)P(U_1|X_0)P(U_2|X_0) \end{aligned} \quad (66)$$

Then, the following rate tuples are achievable:

$$\begin{aligned} R_{0,12} &\geq I(X_0; U_0|U_1) \\ R_{0,1} &\geq I(X_0; U_1|U_0) \\ R_{0,12} + R_{0,1} &\geq I(X_0; U_0, U_1) \\ R_{0,12} &\geq I(X_0; U_0|U_2) \\ R_{0,2} &\geq I(X_0; U_2|U_0) \\ R_{0,12} + R_{0,2} &\geq I(X_0; U_0, U_2) \\ R_1 &\geq H(X_1|U_0, U_1) \\ R_2 &\geq H(X_2|U_0, U_2) \end{aligned} \quad (67)$$

The convex closure of the above rate region over all random variables satisfying (66) is $\mathcal{R}_{HK-I}$. Our objective is to show that $\mathcal{R}_{HK-I}$ is strictly suboptimal. Recall from Theorem 3 that $\mathcal{R}_{DIR}$ is complete for this setup.

We consider a particular example where (X_0, X_1, X_2) are binary symmetric sources such that $X_1 \leftrightarrow X_0 \leftrightarrow X_2$ holds. The transition probabilities are such that X_1 and X_2 are obtained as outputs of two independent binary symmetric channels (BSC) with X_0 as input and cross-over probabilities of P_1 and P_2, respectively. Without loss of generality, let us assume that $0.5 > P_1 > P_2 > 0$. We consider one operating point and show that it is not part of $\mathcal{R}_{HK-I}$. Fix some $\Delta > 0$, such that $\max(H_b(P_1) + \Delta, H_b(P_2) + \Delta) < 1$. Consider the following operating point:

$$\begin{aligned} R_1 &= H_b(P_1) + \Delta \\ R_2 &= H_b(P_2) + \Delta \\ R_{0,12} &= 1 - H_b(P_{01}) \\ R_{0,1} &= 0 \\ R_{0,2} &= H_b(P_{01}) - H_b(P_{02}) \end{aligned} \quad (68)$$

where P_{01} and P_{02} solve the respective equations $H_b(P_1 \bullet P_{01}) = H_b(P_1) + \Delta$ and $H_b(P_2 \bullet P_{02}) = H_b(P_2) + \Delta$ where $P_1 \bullet P_2 = P_1(1-P_2) + (1-P_1)P_2$. We will show that this point is part of $\mathcal{R}_{DIR}$, but is not in $\mathcal{R}_{HK-I}$.

To prove that it is part of $\mathcal{R}_{DIR}$, we consider the following joint density for (U_0, U_1, U_2), in Theorem 3. U_2 is the output when X_0 is sent through a BSC with cross over probability P_{02} and U_0 is the output when U_2 is sent through a BSC with cross over probability P_{012} where $P_{02} \bullet P_{012} = P_{01}$. U_1 is set as a constant. It is easy to verify from Theorem 3 that the above rate tuple is achievable and hence is part of $\mathcal{R}_{DIR}$.

Next, we prove that the above point is not in $\mathcal{R}_{HK-I}$. We note that the proof of this claim is, in fact, very similar to the [14, Proof of Theorem 6]. However, the proof does not follow directly from their result. We incorporate some of the required machinery wherever necessary.

Let us say, there exists a joint distribution satisfying (66), such that all conditions in (67) are satisfied for the point (68), i.e.:

$$\begin{aligned} I(X_0; U_0) &\leq 1 - H_b(P_{01}) \\ I(X_0; U_1|U_0) &= 0 \\ H(X_1|U_0, U_1) &\leq H_b(P_1) + \Delta \\ I(X_0; U_0|U_2) &\leq 1 - H_b(P_{01}) \\ I(X_0; U_2|U_0) &\leq H_b(P_{01}) - H_b(P_{02}) \\ I(X_0; U_0, U_2) &\leq 1 - H_b(P_{02}) \\ H(X_2|U_0, U_2) &\leq H_b(P_2) + \Delta \end{aligned} \tag{69}$$

Let us focus on the first three conditions in (69). First, observe that $X_1 \leftrightarrow X_0 \leftrightarrow \{U_0, U_1\}$ form a Markov chain from (66). Therefore, the condition $I(X_0; U_1|U_0) = 0$ implies that the Markov chain $X_1 \leftrightarrow X_0 \leftrightarrow U_0 \leftrightarrow U_1$, must hold. Next, the condition $I(X_0; U_0) \leq 1 - H_b(P_{01})$ is equivalent to $H(X_0|U_0) \geq H_b(P_{01})$ and hence it follows that $H(X_0|U_0, U_1) \geq H_b(P_{01})$. Let $X_1 = X_0 \oplus Z_1$, where $Z_1 \sim Bern(P_1)$ is independent of $\{X_1, U_0, U_1\}$, and $\oplus$ denotes the modulo 2 sum. Then the following series of inequalities follow:

$$\begin{aligned} H_b(P_1 \bullet P_{01}) &= H_b(P_1) + \Delta \\ &\geq H(X_0 \oplus Z_1|U_0, U_1) \\ &\geq^{(a)} H_b(P_1 \bullet H_b^{-1}(H(X_0|U_0, U_1))) \\ &\geq^{(b)} H_b(P_1 \bullet H_b^{-1}(H_b(P_{01}))) \\ &= H_b(P_1 \bullet P_{01}) \end{aligned} \tag{70}$$

where (a) follows from Mrs. Gerber's Lemma [43], (b) follows from the monotonicity of $H_b(\cdot)$ and $H_b^{-1}(\cdot)$ denotes the inverse of the binary entropy function. Here, we assume that the range of the inverse of the binary entropy function is $\{0 - 0.5\}$, i.e., $0 \leq H_b^{-1}(\cdot) \leq 0.5$. As the LHS and the RHS of the above series of inequalities are the same, all the inequalities must be equalities and therefore the following conditions must hold:

$$\begin{aligned} H(X_1|U_0, U_1) &= H(X_0 \oplus Z_1|U_0) = H_b(P_1 \bullet P_{01}) \\ H(X_0|U_0, U_1) &= H(X_0|U_0) = H_b(P_{01}) \end{aligned} \tag{71}$$

From the above arguments, it follows that U_1 can be any random variable that in independent of $\{X_0, X_1\}$ given U_0. In fact, the specific distribution of U_1 does not play a role in the remainder of the proof, as it does not influence the remaining inequalities. Let us next focus on U_0. Let $\mathcal{U}_0$ be the alphabet of U_0 and let $|\mathcal{U}_0| = M$. It follows from standard cardinality bounding techniques that M is finite. Without loss of generality, let $\mathcal{U}_0 = \{1, 2, \ldots, M\}$ and let the distribution over $\mathcal{U}_0$ be $\{q_1, q_2, \ldots, q_M\}$. Let $P(X_0 = 0|U_0 = i) = p_i, \; i \in \{1, \ldots, M\}$. This implies that $P(X_1 = 0|U_0 = i) = p_i(1 - P_1) + (1 - p_i)P_1$. We have:

$$H(X_0|U_0) = \sum_i q_i H_b(p_i) = H_b(P_{01}) \tag{72}$$

$$\begin{aligned} H(X_0 \oplus Z_1|U_0) &= \sum_i q_i H_b(p_i(1 - P_1) + (1 - p_i)P_1) \\ &= \sum_i q_i H_b(p_i \bullet P_1) \\ &= H_b(P_1 \bullet P_{01}) \end{aligned} \tag{73}$$

As $H_b(x)$ is strictly concave in the range $x \in [0, 0.5]$, (72) and (73) imply that M has to be even and $H(X_0|U_0 = i) = H_b(P_{01}), \; \forall i \in \{1, \ldots, M\}$. As X_0 is a binary symmetric random variable, it implies that $\mathcal{U}_0$ can be divided into two sub-groups, such that, $p_i = P_{01} \; \forall i \in \{1, \ldots, \frac{M}{2}\}$, $p_i = 1 - P_{01} \; \forall i \in \{\frac{M}{2} + 1, \ldots, M\}$ and $q_{i+\frac{M}{2}} = q_i \; \forall i \in \{1, \ldots, \frac{M}{2}\}$. As a consequence, U_0 can be split into two independent random variables $\{U_0^b, V\}$, without any loss of optimality, where U_0^b is a binary symmetric random variable and V is a random variable taking values over an alphabet of size $\frac{M}{2}$, with a distribution of $P(V = i) = 2q_i, \; \forall i \in \{1, \ldots, \frac{M}{2}\}$. Note that the conditional distribution $P(X_0|U_0) = P(X_0|U_0^b, V) = P(X_0|U_0^b)$, and is given by the BSC with cross-over probability P_{01}. As V is independent of U_0^b and the Markov chain condition $X_0 \leftrightarrow U_0^b \leftrightarrow V$ holds, V must be independent of (X_0, U_0^b). In fact, if M is 2, U_0 is a standard binary symmetric random variable and V is a constant.

Our objective now is to show that there does not exist a joint density satisfying (66), such that conditions (69) hold, where:

i) $U_0 = \{U_0^b, V\}$, where U_0^b is a binary symmetric random variable and V is a random variable that takes values over an alphabet of size $\frac{M}{2}$
ii) $P(X_0|U_0) = P(X_0|U_0^b)$
iii) $P(X_0|U_0^b)$ is distributed according to a BSC with cross-over probability P_{01} and
iv) V is independent of (X_0, U_0^b).

Let us again suppose that such a joint density exists and let the alphabet of U_2 be $\mathcal{U}_2$. Let us focus on the last four conditions in (69). Note that, as the distribution of (X_0, U_0) is fixed, it is sufficient to consider only the following three conditions:

$$\begin{aligned} I(X_0; U_0|U_2) &\leq 1 - H_b(P_{01}) \\ I(X_0; U_2|U_0) &\leq H_b(P_{01}) - H_b(P_{02}) \\ H(X_2|U_0, U_2) &\leq H_b(P_2) + \Delta \end{aligned} \tag{74}$$

From the second condition above, we have:

$$H(X_0|U_0, U_2) \geq H_b(P_{02}) \tag{75}$$

Next, let $X_2 = X_0 \oplus Z_2$, where $Z_2 \sim Bern(P_2)$ and is independent of (X_0, U_0, U_2). We have the following series of inequalities:

$$\begin{aligned} H_b(P_2 \bullet P_{02}) = H_b(P_2) + \Delta &\geq H(X_2|U_0, U_2) \\ &= H(X_0 \oplus Z_2|U_0, U_2) \\ &\geq^{(a)} H_b(P_2 \bullet H_b^{-1}(H(X_0|U_0, U_2))) \\ &\geq^{(b)} H_b(P_2 \bullet P_{02}) \end{aligned} \tag{76}$$

where (a) follows from Mrs. Gerber's lemma and (b) follows from monotonicity of $H_b(\cdot)$. As the LHS and RHS of the above series of inequalities are the same, each inequality must be an equality and hence, we have:

$$H(X_0|U_0, U_2) = H_b(P_{02})$$
$$H(X_0 \oplus Z_2|U_0, U_2) = H_b(P_2 \bullet P_{02}) \tag{77}$$

Following similar arguments as before, it again follows from strict concavity of $H_b(\cdot)$ that $H(X_0|U_0, U_2) = H_b(P_{02})$, $\forall U_0 \in \mathcal{U}_0$, $U_2 \in \mathcal{U}_2$. Therefore, there exists a function of U_0 and U_2, $\Psi(U_0, U_2) = \hat{X}_0$, such that $P(X_0 \neq \hat{X}_0) \leq P_{02}$. These arguments imply that, if there exists a joint density satisfying (66), such that conditions (69) hold and (X_0, U_0) are distributed such that conditions (i)-(iv) above hold, then there should exist a function $\Psi(U_0, U_2) = \hat{X}_0$, such that $P(X_0 \neq \hat{X}_0) \leq P_{02}$, i.e., the following conditions must hold:

$$I(X_0; U_2|U_0^b, V) \leq H_b(P_{01}) - H_b(P_{02})$$
$$\{U_0^b, V\} \leftrightarrow X_0 \leftrightarrow U_2$$
$$P(X_0 \neq \Psi(U_0^b, V, U_2)) \leq P_{02} \tag{78}$$

However, it is well known that these conditions cannot hold simultaneously if X_0 is a binary symmetric source and $P(U_0^b|X_0)$ is distributed according to a BSC with cross-over probability P_{01} [44].[5] Recall that the same conditions appear in the context of the Wyner-Ziv setup with Hamming distortion, when the source and side information are binary symmetric with the conditional distribution given by BSC (see [11, Sec. 2]). If there existed such a joint density, then there would be no loss in the corresponding Wyner-Ziv setup, compared to a setup where the side information is available to both the encoder and the decoder. However, it is well known that there is a strict loss in the Wyner-Ziv setup with Hamming distortion, when the sources are binary symmetric with the conditional distribution given by BSC. This leads to a contradiction and implies that there cannot exist such a joint density. Hence, for the 2-sink-1-helper setup shown in Fig. 7, $\mathcal{R}_{HK-I} \subset \mathcal{R}_{DIR}$. ■

Remark 8: It is interesting to note that the Wyner-Ziv problem is inherently a lossy setup and the scenarios that we focus on in this paper are all lossless. However, the last condition in (78) mimics the distortion constraint in the Wyner-Ziv framework and hence enables us to incorporate results from the lossy setting. In fact, the source of sub-optimality for the Han and Kobayashi coding scheme when applied to the DIR setup, in exactly same as that in the Wyner-Ziv setting. In the Wyner-Ziv setup, the codeword generated at the source must be independent of the side information sequence given the source sequence. This constraint manifests itself as a Markov chain condition in the Wyner-Ziv characterization and leads to the strict loss over joint encoding, for certain sources and distortion measures. Application of the Han and Kobayashi scheme to the DIR setup requires artificial imposition of independence of certain transmitted messages, which appear as Markov chain conditions in (66), and lead to strict loss over the proposed encoding scheme.

[5]Note that V is independent of (X_0, U_0^b) and hence does not affect the Wyner-Ziv bound.

IX. Conclusion

This paper considers a new routing paradigm called dispersive information routing, wherein each intermediate node is allowed to "split a packet" and forward subsets of the information on individual forward paths. We demonstrated using simple examples the gains of DIR over conventional routing techniques. Under certain assumptions on the cost function, the problem of finding the minimum cost under DIR essentially reduces to characterizing an achievable rate region for a new multi-terminal information theoretic setup. We proposed a new coding scheme for this setup using principles from multiple descriptions encoding and showed that it achieves the complete rate region for certain special cases of the setup. We further showed that this coding scheme strictly improves upon a corresponding variant of the Han and Kobayashi coding scheme, that was proposed for general multi-terminal source networks.

Appendix A
Typicality Definition and Lemmas

We follow the notation and the notion of strong typicality defined in [3]. Let $\gamma = \{1, \ldots, m\}$ and let Z_γ denote a vector of random variables taking values in a finite set $\mathcal{Z}_\gamma = \mathcal{Z}_1 \times \mathcal{Z}_2 \times \cdots \mathcal{Z}_m$.

Definition 1: Let $0 < \epsilon < 1$ be given. Denote by $p(i_\gamma) = Pr(Z_\gamma = i_\gamma)$ $\forall i_\gamma \in \mathcal{Z}_\gamma$. Define the length n $\epsilon-$typical set as:

$$\mathcal{T}_\epsilon^n(Z_\gamma) = \left\{ z_\gamma^n : |N(i_\gamma|z_\gamma^n) - np(i_\gamma)| \leq \frac{\epsilon n p(i_\gamma)}{\log|\mathcal{Z}_\gamma|}\ \forall i_\gamma \in \mathcal{Z}_\gamma \right\} \tag{79}$$

where $N(i_\gamma|z_\gamma^n)$ denotes the number of occurrences (frequency) of i_γ in z_γ^n. Every sequence which belongs to this set is called an $\epsilon-$typical sequence. We denote it by $\mathcal{T}_\epsilon^n$ when the associated variables are clear from the context. We also omit the superscript n when it is obvious from the context. For any subset $\alpha \subseteq \gamma$, we define:

$$\mathcal{T}_\epsilon^n(Z_\alpha) = \left\{ z_\alpha^n : (z_\alpha^n, z_{\alpha^c}^n) \in \mathcal{T}_\epsilon^n(Z_\gamma) \text{ for some } z_{\alpha^c}^n \in \mathcal{Z}_{\alpha^c}^n \right\} \tag{80}$$

where $\alpha^c = \gamma - \alpha$. For two disjoint subsets $\alpha, \beta \subseteq \gamma$, we define the conditional typical set given $z_\beta^n \in \mathcal{Z}_\beta^n$ as:

$$\mathcal{T}_\epsilon^n(Z_\alpha|z_\beta^n) = \left\{ z_\alpha^n : (z_\alpha^n, z_\beta^n) \in \mathcal{T}_\epsilon^n(Z_{\alpha\cup\beta}) \right\} \tag{81}$$

Lemma 1: Let $0 < \epsilon < 1$ be given. For two disjoint subsets $\alpha, \beta \subseteq \gamma$ and for n sufficiently large, the following properties hold:

a)
$$Pr(Z_\alpha^n \in \mathcal{T}_\epsilon^n(Z_\alpha)) \geq 1 - \epsilon \tag{82}$$

b) For any $z_\alpha^n \in \mathcal{Z}_\alpha^n$, we have:
$$2^{-nH(Z_\alpha)-n\epsilon} \leq Pr(Z_\alpha^n = z_\alpha^n) \leq 2^{-nH(Z_\alpha)+n\epsilon} \tag{83}$$

c)
$$(1-\epsilon)2^{nH(Z_\alpha)-n\epsilon} \leq |\mathcal{T}_\epsilon^n(Z_\alpha)| \leq 2^{nH(Z_\alpha)+n\epsilon} \tag{84}$$

d) For any $z_\beta^n \in \mathcal{Z}_\beta^n$, we have:
$$Pr(Z_\alpha^n \in \mathcal{T}_\epsilon^n(Z_\alpha|z_\beta^n)\big|Z_\beta^n = z_\beta^n) \geq 1 - \epsilon \tag{85}$$

e) For any $z_\alpha^n \in \mathcal{Z}_\alpha^n$ and $z_\beta^n \in \mathcal{Z}_\beta^n$, denoting $Pr(Z_\alpha^n = z_\alpha^n | Z_\beta^n = z_\beta^n)$ by $P(z_\alpha^n|z_\beta^n)$, we have:

$$2^{-nH(Z_\alpha|Z_\beta)-2n\epsilon} \leq P(z_\alpha^n|z_\beta^n) \leq 2^{-nH(Z_\alpha|Z_\beta)+2n\epsilon} \quad (86)$$

f) For any $z_\beta^n \in \mathcal{Z}_\beta^n$:

$$2^{-nH(Z_\alpha|Z_\beta)-2n\epsilon} \leq |\mathcal{T}_\epsilon^n(Z_\alpha|z_\beta^n)| \leq 2^{-nH(Z_\alpha|Z_\beta)+2n\epsilon} \quad (87)$$

Proof: Refer to [3]. ■

APPENDIX B
BOUNDING ENCODING/DECODING ERRORS IN THEOREM 1

Probability of Encoding Error: Let us analyze the probability of encoding error at source node E_i. Let $\mathcal{E}$ denote the event of an encoding error. We have:

$$\begin{aligned} P(\mathcal{E}) = P(\mathcal{E}|x_i^n \in \mathcal{T}_\epsilon^n)P(x_i^n \in \mathcal{T}_\epsilon^n) \\ +P(\mathcal{E}|x_i^n \notin \mathcal{T}_\epsilon^n)P(x_i^n \notin \mathcal{T}_\epsilon^n) \end{aligned} \quad (88)$$

From standard typicality arguments, we have $P(x_i^n \notin \mathcal{T}_\epsilon^n) \to 0$ as $n \to \infty$. Hence, it is sufficient to find conditions on the rates to bound $P(\mathcal{E}|x_i^n \in \mathcal{T}_\epsilon^n)$.

Towards finding conditions on the rate to bound $P(\mathcal{E}|x_i^n \in \mathcal{T}_\epsilon^n)$, we define the random variables:

$$\chi(\{j\}_{\mathcal{J}(\Pi)}) = \begin{cases} 1 & \text{if } \left(x_i^n, u_i^n(\{j\}_{\mathcal{J}(\Pi)})\right) \in \mathcal{T}_\epsilon^n \\ 0 & \text{else} \end{cases} \quad (89)$$

We have $P(\mathcal{E}|x_i^n \in \mathcal{T}_\epsilon^n) = P(\Psi = 0)$ where $\Psi = \sum_{\mathcal{J}(\Pi)} \chi(\{j\}_{\mathcal{J}(\Pi)})$. From Chebyshev's inequality, it follows that:

$$P(\Psi = 0) \leq P\left[|\Psi - E[\Psi]| \geq E[\Psi]/2\right] \leq \frac{4\text{Var}(\Psi)}{(E[\Psi])^2} \quad (90)$$

From [3, Lemma 3.1], we can bound $E[\Psi]$ as follows:

$$E[\Psi] \geq 2^{n\sum_{\mathcal{K}\in\mathcal{J}(\Pi)} R'_{\mathcal{K}} - n(\alpha(i,\mathcal{J}(\Pi))+\epsilon)} \quad (91)$$

where

$$\alpha(i, \mathcal{Q}) = -H\left(\{U_i\}_{\mathcal{Q}}|X_i\right) + \sum_{\mathcal{K}\in\mathcal{Q}} H\left(U_{i,\mathcal{K}}|\{U_i\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}\right) \quad (92)$$

$\forall i, \mathcal{Q} \subseteq \mathcal{J}(\Pi)$. We follow the convention $\alpha_W(i, \phi) = 0$. Next consider $\text{Var}(\Psi) = E[\Psi^2] - (E[\Psi])^2$ where,

$$\begin{aligned} E[\Psi^2] &= \sum_{\{j\}_{\mathcal{J}(\Pi)}} \sum_{\{k\}_{\mathcal{J}(\Pi)}} E\left[\chi(\{j\}_{\mathcal{J}(\Pi)})\chi(\{k\}_{\mathcal{J}(\Pi)})\right] \\ &= \sum_{\{j\}_{\mathcal{J}(\Pi)}} \sum_{\{k\}_{\mathcal{J}(\Pi)}} P\left[\chi(\{j\}_{\mathcal{J}(\Pi)}) = 1, \chi(\{k\}_{\mathcal{J}(\Pi)}) = 1\right] \end{aligned} \quad (93)$$

The probability in (93) depends on whether $u_i^n(\{j\}_{\mathcal{J}(\Pi)})$ and $u_i^n(\{k\}_{\mathcal{J}(\Pi)})$ are equal for a subset of indices. Let $\mathcal{Q} \subseteq \mathcal{J}(\Pi)$, $\mathcal{Q} \neq \phi$, such that $\{j\}_{\mathcal{Q}} = \{k\}_{\mathcal{Q}}$. Observe that, due to the hierarchical structure in the conditional codebook generation mechanism, for $u_i^n(\{j\}_{\mathcal{Q}}) = u_i^n(\{k\}_{\mathcal{Q}})$ to hold, $\mathcal{Q}$ must be such that,

$$\text{if } \mathcal{K} \in \mathcal{Q} \Rightarrow \mathcal{I}_{|\mathcal{K}|+}(\mathcal{K}) \subset \mathcal{Q} \quad (94)$$

i.e., $\mathcal{Q} \in \mathcal{Q}^*$ given in (32). It follows from the codebook generation mechanism that given the codeword tuple $\{u_i^n(\{j\}_{\mathcal{Q}})\}$, tuples $\{u_i^n(\{j\}_{\mathcal{J}(\Pi)-\mathcal{Q}})\}$ and $\{u_i^n(\{k\}_{\mathcal{J}(\Pi)-\mathcal{Q}})\}$ are independent and identically distributed. Hence we can rewrite the probability in (93) for some $\mathcal{Q} \subseteq \mathcal{J}(\Pi)$, $\mathcal{Q} \neq \phi$, as:

$$\begin{aligned} P\left[\mathcal{E}(\{j\}_{\mathcal{J}(\Pi)}) \cap \mathcal{E}(\{k\}_{\mathcal{J}(\Pi)})\right] = \left(\frac{P\left[\mathcal{E}(\{j\}_{\mathcal{J}(\Pi)})\right]}{P\left[\mathcal{E}(\{j\}_{\mathcal{Q}})\right]}\right)^2 \\ \times P\left[\mathcal{E}(\{j\}_{\mathcal{Q}})\right] \end{aligned} \quad (95)$$

However, note that if $\mathcal{Q} = \phi$, then:

$$P\left[\mathcal{E}(\{j\}_{\mathcal{J}(\Pi)}) \cap \mathcal{E}(\{k\}_{\mathcal{J}(\Pi)})\right] = \left(P\left[\mathcal{E}(\{j\}_{\mathcal{J}(\Pi)})\right]\right)^2 \quad (96)$$

Next, the total number of ways of choosing $\{j\}_{\mathcal{J}(\Pi)}$ and $\{k\}_{\mathcal{J}(\Pi)}$ such that they overlap in the subset $\mathcal{Q}$ is:

$$\begin{aligned} 2^{n\sum_{\mathcal{K}\in\mathcal{Q}} R'_{i,\mathcal{K}}} \prod_{\mathcal{K}\in\mathcal{J}(\Pi)-\mathcal{Q}} 2^{nR'_{i,\mathcal{K}}}(2^{nR'_{i,\mathcal{K}}} - 1) \\ \leq 2^{n\{\sum_{\mathcal{K}\in\mathcal{J}(\Pi)} R'_{i,\mathcal{K}} + 2\sum_{\mathcal{K}\in\mathcal{J}(\Pi)-\mathcal{Q}} R'_{i,\mathcal{K}}\}} \end{aligned} \quad (97)$$

On substituting (95) and (97) in (93), we bound $\text{Var}(\Psi)$ as:

$$\begin{aligned} \text{Var}(\Psi) \leq \sum \Big\{ 2^{-2n\left(\alpha(i,\mathcal{J}(\Pi)) - \sum_{\mathcal{K}\in\mathcal{J}(\Pi)} R'_{i,\mathcal{K}}\right)} \\ 2^{n\left(\alpha(i,\mathcal{Q}) - \sum_{\mathcal{K}\in\mathcal{Q}} R'_{i,\mathcal{K}}\right)+5n\epsilon} \Big\} \end{aligned} \quad (98)$$

where the summation is over all non-empty $\mathcal{Q}$ such that (94) holds. Observe that the term corresponding to $\mathcal{Q} = \phi$ gets canceled with the '$(\mathcal{E}[\Psi])^2$' term in $\text{Var}(\Psi)$. Inserting, (98) and (91) in (90), we get:

$$P(E|x_i^n \in \mathcal{T}_\epsilon^n) \leq 4\sum 2^{n\left(\alpha(i,\mathcal{Q}) - \sum_{\mathcal{K}\in\mathcal{Q}} R'_{\mathcal{K}}\right)+7n\epsilon} \quad (99)$$

where the summation is over all non-empty $\mathcal{Q}$ satisfying (94). Hence, the probability of encoding error at all the source nodes can be made arbitrarily small if:

$$\sum_{\mathcal{K}\in\mathcal{Q}} R'_{i,\mathcal{K}} \geq \alpha(i, \mathcal{Q}) + 7\epsilon \quad (100)$$

$\forall i, \mathcal{Q}$ satisfying (94).

Probability of Decoding Error: We focus on decoding at sink S_k. We first bound the probability of error for the first stage of decoding. The decoder looks for a unique codeword tuple from $\left\{\{\mathcal{C}_\Sigma\}_{\mathcal{J}(k)}\left(\{m_\Sigma\}_{\mathcal{J}(k)}\right)\right\}$ which are jointly typical. We know that $\{u_\Sigma^*\}_{\mathcal{J}(k)}$ are jointly typical from the Markov Lemma in Appendix C. We have to find conditions on $R''_{i,\mathcal{S}}$ to ensure no other tuple satisfies this property. Denote by $\mathcal{F}$ the event of a decoding error given the encoding is error-free. Due to the symmetry in codebook generation, we can assume that the index tuple of $\{u_\Sigma^*\}_{\mathcal{J}(k)}$ is $(1, \ldots, 1)$. Let $\{j_\Sigma\}_{\mathcal{J}(k)}$ be an index tuple such that:

$$\{j_\Sigma\}_{\mathcal{J}(k)} \neq (1, \ldots, 1) \quad (101)$$

Define the event $\mathcal{F}(\{j_\Sigma\}_{\mathcal{J}(k)})$ as:

$$\mathcal{F}(\{j_\Sigma\}_{\mathcal{J}(k)}) = \left\{\left(u_\Sigma^n(\{j_\Sigma\}_{\mathcal{J}(k)})\right) \in \mathcal{T}_\epsilon^n\right\} \quad (102)$$

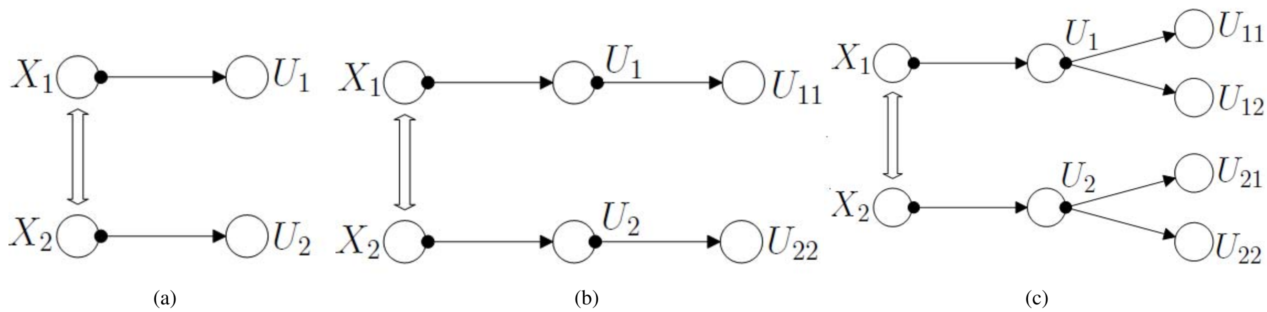

Fig. 8. Depicts the different Markov lemmas. (a) Generalized Markov Lemma [3]. (b) Conditional Markov Lemma [14]. (c) Conditional Markov Lemma - for mutual covering.

It then follows from union bound that:

$$P(\mathcal{F}) \leq \sum P\big(\mathcal{F}(\{j_\Sigma\}_{\mathcal{J}(k)})\big) \tag{103}$$

where the summation is over all $\{j_\Sigma\}_{\mathcal{J}(k)} \neq (1, \ldots, 1)$. However, a subset of indices of $\{j_\Sigma\}_{\mathcal{J}(k)}$ can still be equal to 1. We expand the above summation over all such possible subsets. Let $\mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N \subseteq \mathcal{J}(k)$ satisfying (94) be such that the following holds[6]:

$$\exists i \in \{1, 2, \ldots, N\} : \mathcal{Q}_i \subset \mathcal{J}(k) \tag{104}$$

i.e., at least one of the $\mathcal{Q}_i$'s is a strict subset of $\mathcal{J}(k)$. Define the set:

$$A_{\mathcal{Q}_1,\mathcal{Q}_2,\ldots,\mathcal{Q}_N} = \{j_\Sigma\}_{\mathcal{J}(k)} : \forall i \begin{cases} j_{i,\mathcal{K}} = 1 & \text{if } \mathcal{K} \in \mathcal{Q}_i \\ j_{i,\mathcal{K}} \neq 1 & \text{otherwise} \end{cases}$$

Then, we can expand (103) as:

$$P(F) \leq \sum\sum P\big(\mathcal{F}(\{j_\Sigma\}_{\mathcal{J}(k)})\big) \tag{105}$$

where the first summation is over all $\{\mathcal{Q}_1, \ldots, \mathcal{Q}_N : \mathcal{Q}_i \subseteq \mathcal{J}(k)$, satisfying (94) and (104)$\}$ and the second summation is over all $\{j_\Sigma\}_{\mathcal{J}(k)} \in A_{\mathcal{Q}_1,\mathcal{Q}_2,\ldots\mathcal{Q}_N}$. We note that, due to the conditional independence of the codewords generated, $P\big(\mathcal{F}(\{j_\Sigma\}_{\mathcal{J}(k)})\big)$ is the same for all $\{j_\Sigma\}_{\mathcal{J}(k)} \in A(\mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N)$, i.e., $P\big(\mathcal{F}(\{j_\Sigma\}_{\mathcal{J}(k)})\big)$ depends only on $\mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N$. We can bound $P(\mathcal{F})$ as:

$$P(\mathcal{F}) \leq \sum\big\{|A_{\mathcal{Q}_1,\mathcal{Q}_2,\ldots\mathcal{Q}_N}| P\big(\mathcal{F}(\mathcal{Q}_i; \forall i \in \Sigma)\big)\big\} \tag{106}$$

where $P(\mathcal{F}(\mathcal{Q}_i; \forall i \in \Sigma)) = P\big(\mathcal{F}(\mathcal{Q}_1, \mathcal{Q}_2, \ldots, \mathcal{Q}_N)\big)$ denotes $P(\mathcal{F}(\{j_\Sigma\}_{\mathcal{J}(k)}))$ for some $\{j_\Sigma\}_{\mathcal{J}(k)} \in A_{\mathcal{Q}_1,\mathcal{Q}_2,\ldots\mathcal{Q}_N}$ and the summation is over all $\{\mathcal{Q}_1, \ldots, \mathcal{Q}_N : \mathcal{Q}_i \subseteq \mathcal{J}(k)$, satisfying (94) and (104)$\}$. We next bound the individual terms in the above product. Recall that each of the bins $\mathcal{C}_{i,\mathcal{S}}(\cdot)$ have $2^{n(R'_{i,\mathcal{S}} - R''_{i,\mathcal{S}})}$ codewords. Using Lemma 3.1 [3], we can bound both the terms in the above product as:

$$P\big(\mathcal{F}(\mathcal{Q}_i; \forall i \in \Sigma)\big) \leq \frac{2^{nH\left(\{U_i\}_{\mathcal{Q}_i^c}\forall i \middle| \{U_i\}_{\mathcal{Q}_i}\forall i\right)}}{2^{n\sum_{i\in\Sigma}\sum_{\mathcal{K}\in\mathcal{Q}_i^c} H\left(U_{i,\mathcal{K}} \middle| \{U_i\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}\right) - 4n\epsilon}}$$

$$|A_{\mathcal{Q}_1,\mathcal{Q}_2,\ldots\mathcal{Q}_N}| \leq 2^{n\sum_{i\in\Sigma}\sum_{\mathcal{K}\in\mathcal{Q}_i^c}(R'_{i,\mathcal{K}} - R''_{i,\mathcal{K}})} \tag{107}$$

where $\mathcal{Q}_i^c = \mathcal{J}(k) - \mathcal{Q}_i$. Substituting (107) in (106), it follows that $P(\mathcal{F})$ can be made arbitrarily small if: $\forall \mathcal{Q}_1, \mathcal{Q}_2, \ldots \mathcal{Q}_N \subseteq \mathcal{J}(k)$ satisfying (94) and (104),

$$\begin{aligned}\sum_{i\in\Sigma}\sum_{\mathcal{K}\in\mathcal{Q}_i^c}(R'_{i,\mathcal{K}} - R''_{i,\mathcal{K}}) \leq \sum_{i\in\Sigma}\sum_{\mathcal{K}\in\mathcal{Q}_i^c} H(U_{i,\mathcal{K}}|\{U_i\}_{\mathcal{I}_{|\mathcal{K}|+}(\mathcal{K})}) \\ - H\left(\{U_i\}_{\mathcal{Q}_i^c}\forall i|\{U_i\}_{\mathcal{Q}_i}\forall i\right) - 4\epsilon\end{aligned} \tag{108}$$

where $\mathcal{Q}_i^c = \mathcal{J}(k) - \mathcal{Q}_i$. On plugging in the bounds for $R'_{i,\mathcal{K}}$ from (100) into (108), we get (40) in Theorem 1.

Appendix C
Conditional Markov Lemma-For Mutual Covering

It was shown in [3][7] that if a codeword of U_1 (denoted by U_1^*) is selected jointly typical with X_1^n and a codeword of U_2 (denoted by U_2^*) is selected jointly typical with X_2^n and if $U_1 \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow U_2$, then $(U_1^*, X_1^n, X_2^n, U_2^*)$ are jointly typical, with probability approaching 1. This is called the generalized Markov lemma and is depicted in Fig. 8. Similarly, Wagner et al. [14] considered the case in which codewords of U_{11} and U_{22} are generated conditioned on codewords of U_1 and U_2, respectively. They showed that if a pair of codewords of (U_1, U_{11}) (denoted by (U_1^*, U_{11}^*)) are jointly typical with X_1^n and a pair of codewords of (U_2, U_{22}) (denoted by (U_2^*, U_{22}^*)) are typical with X_2^n, and if $(U_1, U_{11}) \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow (U_2, U_{22})$, then $(U_1^*, U_{11}^*, X_1^n, X_2^n, U_2^*, U_{22}^*)$ are jointly typical, with probability approaching 1. This is called the

[6]Again observe that it is sufficient for us to consider $\mathcal{Q}_i$s which satisfy (94) due to the hierarchical structure of the conditional codebook generation.

[7]We note that an earlier Markov Lemma proof appeared in [13]. However the proof in [3] is easily extendible to more general settings as it is based on standard typicality arguments.

conditional Markov lemma for obvious reasons and is depicted in Fig. 8. However, these results are not sufficient for our scenario and we require a stronger version of the conditional Markov lemma. In what follows, we will establish a series of lemmas, culminating with the needed variant called the conditional Markov lemma for mutual covering (Lemma 4).

We begin by illustrating the need for a stronger version. Let us consider random variables $(X_1, X_2, U_1, U_2, U_{11}, U_{12}, U_{21}, U_{22})$, jointly distributed according to a density that satisfies:

$$(U_1, U_{11}, U_{12}) \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow (U_2, U_{21}, U_{22}) \tag{109}$$

Let codebooks of U_1 and U_2 be generated according to their respective marginal densities. Let U_1^* and U_2^* be codewords (selected from the codebooks) that are jointly typical with X_1^n and X_2^n, respectively. From the generalized Markov Lemma, it follows that $(U_1^*, X_1^n, X_2^n, U_2^*) \in \mathcal{T}_\epsilon^n$, with probability 1. Next, let codebooks of U_{11} and U_{12} be generated independently, conditioned on U_1^* (according to $P(U_{11}|U_1)$ and $P(U_{12}|U_2)$, respectively). Similarly, let codebooks of U_{21} and U_{22} be generated independently, conditioned on U_2^*. Let U_{11}^* and U_{12}^* be codewords, selected from the codebooks, that are jointly typical with (X, U_1^*), i.e., $(X_1^n, U_1^*, U_{11}^*, U_{12}^*) \in \mathcal{T}_\epsilon^n$. Similarly, let U_{21}^* and U_{22}^* be selected from the respective codebooks such that $(X_2^n, U_2^*, U_{21}^*, U_{22}^*) \in \mathcal{T}_\epsilon^n$. It follows from the conditional Markov lemma in [14] that the following conditions hold with probability 1:

$$(X_1^n, X_2^n, U_1^*, U_2^*, U_{1i}^*, U_{2j}^*) \in \mathcal{T}_\epsilon^n \quad i, j \in \{1, 2\}$$

However, this does not imply that $(X_1^n, X_2^n, U_1^*, U_2^*, U_{11}^*, U_{12}^*, U_{21}^*, U_{22}^*) \in \mathcal{T}_\epsilon^n$ with probability 1. The primary source of difficulty arises because the codebooks of U_{11} and U_{12} are generated independently, conditioned on U_1^*, and not jointly according to the conditional density $P(U_{11}, U_{12}|U_1)$. Our objective in this appendix is to prove that $(X_1^n, X_2^n, U_1^*, U_2^*, U_{11}^*, U_{12}^*, U_{21}^*, U_{22}^*) \in \mathcal{T}_\epsilon^n$ holds, with probability 1. We note that this setup lies at the heart of the DIR encoding scheme and can be easily extended to more than 2 random variables and layers of encoding, to incorporate the general DIR setup. However, we restrict ourselves to the 2 variable case to keep the notation simple. We also note that the lemmas and proofs here are applicable to more general contexts beyond DIR.

Lemma 2: Let random variables (Y, U, V_1, V_2) be given and let $y^n \in \mathcal{T}_\epsilon^n(Y)$. Let the subset $B_0(y^n) \subset \mathcal{T}_\epsilon^n(U|y^n)$ be such that:

$$2^{n(H(U|Y)-\lambda)} \leq |B_0(y^n)| \leq 2^{n(H(U|Y)+\lambda)} \tag{110}$$

for some $\lambda > 0$. For every $u^n \in B_0(y^n)$, let subset $B_{12}(y^n, u^n) \subset \mathcal{T}_\epsilon^n((V_1, V_2)|u^n)$ be such that:

$$2^{n(H(V_1,V_2|U,Y)-\lambda)} \leq |B_{12}(y^n, u^n)| \leq 2^{n(H(V_1,V_2|U,Y)+\lambda)} \tag{111}$$

and the following hold:

$$\begin{aligned} 2^{n(H(V_1|U,Y)-\lambda)} &\leq |B_1(y^n, u^n)| \leq 2^{n(H(V_1|U,Y)+\lambda)} \\ 2^{n(H(V_2|U,Y)-\lambda)} &\leq |B_2(y^n, u^n)| \leq 2^{n(H(V_2|U,Y)+\lambda)} \\ 2^{n(H(V_1|U,Y,V_2)-\lambda)} &\leq |\hat{B}_1(y^n, u^n, v_2^n)| \leq 2^{n(H(V_1|U,Y,V_2)+\lambda)} \\ 2^{n(H(V_2|U,Y,V_1)-\lambda)} &\leq |\hat{B}_2(y^n, u^n, v_1^n)| \leq 2^{n(H(V_2|U,Y,V_1)+\lambda)} \end{aligned} \tag{112}$$

where $\forall (v_1^n, v_2^n) \in B_{12}(y^n, u^n)$:

$$\begin{aligned} \hat{B}_1(y^n, u^n, v_2^n) &= \{v_1^n : (v_1^n, v_2^n) \in B_{12}(y^n, u^n)\} \\ \hat{B}_2(y^n, u^n, v_1^n) &= \{v_2^n : (v_1^n, v_2^n) \in B_{12}(y^n, u^n)\} \\ B_1(y^n, u^n) &= \{v_1^n : \exists (v_1^n, v_2^n) \in B_{12}(y^n, u^n)\} \\ B_2(y^n, u^n) &= \{v_2^n : \exists (v_1^n, v_2^n) \in B_{12}(y^n, u^n)\} \end{aligned} \tag{113}$$

Let R_0, R_1 and R_2 be given positive rates. Let $\overline{U}_j (j = 1, \ldots, 2^{nR_0})$ be random variables drawn independently and uniformly from $\mathcal{T}_\epsilon^n(U)$. For each $\overline{U}_j$, let $\overline{V}_{jk}^1 (k = 1, \ldots, 2^{nR_1})$ and $\overline{V}_{jk}^2 (k = 1, \ldots, 2^{nR_2})$ be random variables drawn independently and uniformly from $\mathcal{T}_\epsilon^n(V_1|\bar{U}_j)$ and $\mathcal{T}_\epsilon^n(V_2|\bar{U}_j)$, respectively. Then for n sufficiently large,

$$P\big(\nexists j, k_1, k_2 : \overline{U}_j \in B_0(y^n), (\bar{V}_{jk_1}^1, \bar{V}_{jk_2}^2) \in B_{12}(y^n, \overline{U}_j)\big) \leq \delta(\epsilon) \tag{114}$$

where $\delta(\epsilon) \to 0$ as $\epsilon \to 0$, if the rates R_0, R_1 and R_2 satisfy:

$$\begin{aligned} R_0 &\geq I(Y; U) + 7\lambda + 19\epsilon \\ R_0 + R_1 &\geq I(Y; V_1, U) + 8\lambda + 17\epsilon \\ R_0 + R_2 &\geq I(Y; V_2, U) + 8\lambda + 17\epsilon \\ R_0 + R_1 + R_2 &\geq I(Y; V_1, V_2, U) + I(V_1; V_2|U) + 6\lambda + 15\epsilon \end{aligned} \tag{115}$$

Proof: Define the random variable $\mathcal{X}_{j,k_1,k_2}$ as:

$$\mathcal{X}_{j,k_1,k_2} = \begin{cases} 1 & \text{if } \bar{U}_j \in B_0(y^n), (\bar{V}_{k_1}^1, \bar{V}_{k_2}^2) \in B_{12}(y^n, \bar{U}_j) \\ 0 & \text{else} \end{cases} \tag{116}$$

Denote by $\mathcal{X} = \sum_{j,k_1,k_2} \mathcal{X}_{j,k_1,k_2}$. Observe that the probability in (114) is equal to $P(\mathcal{X} = 0)$. From Chebychev's inequality, we have:

$$P(\mathcal{X} = 0) \leq \frac{4Var(\mathcal{X})}{(E[\mathcal{X}])^2} \tag{117}$$

Next we have the following from (110) and (111):

$$\begin{aligned} E[\mathcal{X}] = \sum_{j,k_1,k_2} E[\mathcal{X}_{j,k_1,k_2}] &=^{(a)} 2^{n(R_0+R_1+R_2)} P(\mathcal{X}_{1,1,1}) \\ &\geq 2^{n(R_0+R_1+R_2)} \frac{2^{n(H(V_1,V_2,U|Y))-2\lambda-5\epsilon)}}{2^{n(H(U)+H(V_1|U)+H(V_2|U))}} \end{aligned} \tag{118}$$

where equality in (a) holds because the random variables $\overline{U}_j$, $\overline{V}_{jk}^1$ and $\overline{V}_{jk}^2$ are drawn independently and uniformly from their respective typical sets. Also, using (111) and (112), we can bound $E[\mathcal{X}^2]$ as:

$$\begin{aligned}E[\mathcal{X}^2] &= \sum_{j^1,k_1^1,k_2^1}\sum_{j^2,k_1^2,k_2^2} E[\mathcal{X}_{j^1,k_1^1,k_2^1}\mathcal{X}_{j^2,k_1^2,k_2^2}] \\ &\leq 2^{n(R_0+R_1+R_2)}P_1 + 2^{n(R_0+2R_1+2R_2)}P_2 \\ &\quad +2^{n(R_0+2R_1+R_2)}P_3 + 2^{n(R_0+R_1+2R_2)}P_4 \\ &\quad +2^{2n(R_0+R_1+R_2)}P_1^2 \end{aligned} \tag{119}$$

where

$$\begin{aligned} P_1 &= \frac{2^{n(H(V_1,V_2,U|Y)+2\lambda+5\epsilon)}}{2^{n(H(U)+H(V_1|U)+H(V_2|U))}} \\ P_2 &= \frac{2^{n(H(U|Y)+2H(V_1,V_2|U,Y)+3\lambda+9\epsilon)}}{2^{n(H(U)+2H(V_1|U)+2H(V_2|U))}} \\ P_3 &= \frac{2^{n(H(V_1,U|Y)+2H(V_2|U,Y,V_1)+4\lambda+7\epsilon)}}{2^{n(H(V_1,U)+2H(V_2|U))}} \\ P_4 &= \frac{2^{n(2H(V_1|U,Y,V_2)+H(V_2,U|Y))+4\lambda+7\epsilon)}}{2^{n(2H(V_1|U)+H(V_2,U))}} \end{aligned} \tag{120}$$

On substituting (118), (119) and (120) in (117), we have:

$$\begin{aligned} &P(\mathcal{X}=0) \\ &\leq 4\Big[\delta(\epsilon) + 2^{-n(R_0-I(Y;U)-7\lambda-19\epsilon)} \\ &\quad +2^{-n(R_0+R_1-I(Y;V_1,U)-8\lambda-17\epsilon)} \\ &\quad +2^{-n(R_0+R_2-I(Y;V_2,U)-8\lambda-17\epsilon)} \\ &\quad +2^{-n(R_0+R_1+R_2-I(Y;V_1,V_2,U)-I(V_1;V_2|U)-6\lambda-15\epsilon)}\Big] \end{aligned} \tag{121}$$

which can be made arbitrarily small if the rates satisfy (115). ■

Lemma 3: Let W, Y, U, V_1 and V_2 be random variables with values in finite sets $\mathcal{W}, \mathcal{Y}, \mathcal{U}, \mathcal{V}_1$ and $\mathcal{V}_2$, respectively. Let W^* be a random variable with values in $\mathcal{W}^n$, such that:

$$W^* \leftrightarrow Y^n \leftrightarrow (U^n, V_1^n, V_2^n) \tag{122}$$

Let R_0, R_1 and R_2 be given positive rates. Let $\overline{U}_i|_{i=1}^{2^{nR_0}}$ denote independent random variables chosen uniformly with replacement from $\mathcal{T}_\epsilon^n(U)$. Let $\overline{V}_{i,j}^1(i = 1,\ldots,2^{R_0},\ j = 1,\ldots,2^{nR_1})$ and $\overline{V}_{i,j}^2(i = 1,\ldots,2^{R_0},\ j = 1\ldots 2^{nR_2})$ be random variables drawn independently and uniformly from $\mathcal{T}_\epsilon^n(V_1|\overline{U}_i)$ and $\mathcal{T}_\epsilon^n(V_2|\overline{U}_i)$, respectively $\forall i$. Further, let,

$$P(W^*, Y^n, U^n, V_1^n, V_2^n \in \mathcal{T}_\epsilon^n(W, Y, U, V_1, V_2)) \geq 1-\eta \tag{123}$$

Also, suppose $\forall v_1^n \in \mathcal{T}_\epsilon^n(V_1)$ and $v_2^n \in \mathcal{T}_\epsilon^n(V_2)$:

$$\begin{aligned} P\left((W^*, Y^n, U^n, V_1^n) \in \mathcal{T}_\epsilon^n \middle| V_2^n = v_2^n\right) \geq 1-\eta \\ P\left((W^*, Y^n, U^n, V_2^n) \in \mathcal{T}_\epsilon^n \middle| V_1^n = v_1^n\right) \geq 1-\eta \end{aligned} \tag{124}$$

Then for n sufficiently large, there exists functions $U^*(y^n)$, $V_1^*(y^n, U^*)$ and $V_2^*(y^n, U^*)$, such that:

i) $\overline{U}^*(y^n) = \overline{U}_i$ (for some $i \in \{1,\ldots,2^{R_0}\}$)$\Rightarrow$ $V_1^*(y^n,\overline{U}^*) = \bar{V}_{i,j_1}^1$, $V_2^*(y^n,\overline{U}^*) = \bar{V}_{i,j_2}^2$ for some $j_1 \in \{1,\ldots,2^{nR_1}\}$ and $j_2 \in \{1,\ldots,2^{nR_2}\}$

ii)

$$\begin{aligned} P((W^*, Y^n, \bar{U}^*, V_1^*, V_2^*) \in \mathcal{T}_\epsilon \geq 1-\delta(\epsilon) \\ P((W^*, Y^n, \bar{U}^*, V_1^*) \in \mathcal{T}_\epsilon | V_2^*) \geq 1-\delta(\epsilon) \\ P((W^*, Y^n, \bar{U}^*, V_2^*) \in \mathcal{T}_\epsilon | V_1^*) \geq 1-\delta(\epsilon) \end{aligned} \tag{125}$$

for some $\delta(\epsilon) \to 0$ as $\epsilon \to 0$, if the rates R_0, R_1 and R_2 satisfy:

$$\begin{aligned} R_0 &\geq I(Y;U) + 40\epsilon \\ R_0 + R_1 &\geq I(Y;V_1,U) + 41\epsilon \\ R_0 + R_2 &\geq I(Y;V_2,U) + 41\epsilon \\ R_0+R_1+R_2 &\geq I(Y;V_1,V_2,U)+I(V_1;V_2|U)+33\epsilon \end{aligned} \tag{126}$$

Proof: Let us expand (123) as:

$$\sum_{y^n\in\mathcal{Y}^n}\Big\{P\Big((W^*, U^n, V_1^n, V_2^n) \in \mathcal{T}_\epsilon^n \Big| Y^n = y^n\Big) P(Y^n = y^n)\Big\} \geq 1-\eta \tag{127}$$

Let,

$$A \triangleq \Big\{y^n : P\Big((W^*, U^n, V_1^n, V_2^n) \in \mathcal{T}_\epsilon^n \Big| Y^n = y^n\Big) \geq 1-\sqrt{\eta}\Big\}$$

and

$$A_0 \triangleq A \bigcap \mathcal{T}_\epsilon^n(Y) \tag{128}$$

Then using the reverse Markov inequality, we can show that (similar to [3] and [14]):

$$P(Y^n \in A_0) \geq 1-\delta_1 \tag{129}$$

where $\delta_1 = \sqrt{\eta}+\epsilon$. Then for any $y^n \in A_0$, we have:

$$\sum_{u^n}\Big\{P\Big((W^*, V_1^n, V_2^n) \in \mathcal{T}_\epsilon^n \Big| Y^n = y^n, U^n = u^n\Big) P\Big(U^n = u^n \Big| Y^n = y^n\Big)\Big\} \geq 1-\sqrt{\eta} \tag{130}$$

Let,

$$B(y^n) \triangleq \Big\{u^n : P\Big((W^*, V_1^n, V_2^n) \in \mathcal{T}_\epsilon^n \Big| Y^n = y^n, U^n = u^n\Big) \geq 1-\sqrt[4]{\eta}\Big\}$$

$$B_0 \triangleq B \bigcap \mathcal{T}_\epsilon^n(U|y^n) \tag{131}$$

Using the reverse Markov inequality, we again have:

$$P\Big(U^n \in B_0(y^n)\Big| Y^n = y^n\Big) \geq 1-\delta_2 \tag{132}$$

where $\delta_2 = \sqrt[4]{\eta}+\epsilon$. Hence for any $y^n \in A_0$ and $u^n \in B_0(y^n)$ we have:

$$\sum_{u^n,v_1^n,v_2^n} P\Big(V_1^n = v_1^n, V_2^n = v_2^n \Big| Y^n = y^n, U^n = u^n\Big) Q(y^n, u^n, v_1^n, v_2^n) \geq 1-\sqrt[4]{\eta} \tag{133}$$

where we denote by $Q(\mathcal{S})= P\Big(W^* \in \mathcal{T}_\epsilon^n(W|\mathcal{S})\Big| Y^n = y^n\Big)$ for any set of sequences $\mathcal{S}$. Note that we have used the Markov condition (122) in the above equation. Now define sets $\tilde{B}_{12}(y^n, u^n)$ and $B_{12}(y^n, u^n)$ for any $y^n \in A_0$ and $u^n \in B_0(y^n)$ such that:

$$\begin{aligned} \tilde{B}_{12}(y^n, u^n) &\triangleq \Big\{(v_1^n, v_2^n) : Q(y^n, u^n, v_1^n, v_2^n) \geq 1-\sqrt[8]{\eta}\Big\} \\ B_{12}(y^n) &\triangleq \tilde{B}_{12}(y^n)\bigcap \mathcal{T}_\epsilon^n(U, V_1, V_2|y^n) \end{aligned} \tag{134}$$

Then using the reverse Markov inequality, we can show that:

$$P\Big((V_1^n, V_2^n) \in B_{12}(y^n)\Big| Y^n = y^n, U^n = u^n\Big) \geq 1 - \delta_3 \quad (135)$$

where $\delta_3 = \sqrt[8]{\eta} + \epsilon$. Then from (132), (135) and Lemma 3.1(f) in [3], for n sufficiently large, we have:

$$\begin{aligned} 2^{n(H(U|Y)-3\epsilon)} &\leq |B_0(y^n)| \leq 2^{n(H(U|Y)+\epsilon)} \\ 2^{n(H(V_1,V_2|Y,U)-3\epsilon)} &\leq |B_{12}(y^n, u^n)| \leq 2^{n(H(V_1,V_2|Y,U)+\epsilon)} \end{aligned} \quad (136)$$

Note that we have two of the sets required by Lemma 2. However, we further require bounds on the projections of $B_{12}(y^n, u^n)$ (as in (112)) to invoke Lemma 2. Towards obtaining these bounds, we note that the following inequalities can be shown directly from (123):

$$\begin{aligned} &P\left((W^*, Y^n, U^n, V_1^n) \in \mathcal{T}_\epsilon^n\right) \geq 1 - \eta \\ &P\left((W^*, Y^n, U^n, V_2^n) \in \mathcal{T}_\epsilon^n\right) \geq 1 - \eta \end{aligned} \quad (137)$$

Expanding (137) instead of (123) and repeating all steps from (127) through (136), we obtain:

$$\begin{aligned} 2^{n(H(V_1|Y,U)-3\epsilon)} &\leq |B_1(y^n, u^n)| \leq 2^{n(H(V_1|Y,U)+\epsilon)} \\ 2^{n(H(V_2|Y,U)-3\epsilon)} &\leq |B_2(y^n, u^n)| \leq 2^{n(H(V_2|Y,U)+\epsilon)} \end{aligned} \quad (138)$$

where

$$\begin{aligned} B_1(y^n, u^n) &= \{v_1^n : \exists (v_1^n, v_2^n) \in B_{12}(y^n, u^n)\} \\ B_2(y^n, u^n) &= \{v_2^n : \exists (v_1^n, v_2^n) \in B_{12}(y^n, u^n)\} \end{aligned} \quad (139)$$

Similarly, it is easy to show that expanding (124) instead of (123) leads to:

$$\begin{aligned} 2^{n(H(V_1|Y,U,V_2)-3\epsilon)} &\leq |\hat{B}_1(y^n, u^n, v_2^n)| \leq 2^{n(H(V_1|Y,U,V_2)+\epsilon)} \\ 2^{n(H(V_2|Y,U,V_1)-3\epsilon)} &\leq |\hat{B}_2(y^n, u^n, v_1^n)| \leq 2^{n(H(V_2|Y,U,V_1)+\epsilon)} \end{aligned} \quad (140)$$

where $\forall v_1^n \in B_1(y^n, u^n)$ and $v_2^n \in B_2(y^n, u^n)$,

$$\begin{aligned} \hat{B}_1(y^n, u^n, v_2^n) &= \{v_1^n : (v_1^n, v_2^n) \in B_{12}(y^n, u^n)\} \\ \hat{B}_2(y^n, u^n, v_1^n) &= \{v_2^n : (v_1^n, v_2^n) \in B_{12}(y^n, u^n)\} \end{aligned} \quad (141)$$

We now have sets B_0 and B_{12} satisfying all the bounds as required in Lemma 2. Hence, we can define the functions U^*, V_1^* and V_2^* as follows. $U^*(y^n) = \overline{U}_i$ if $\bar{U}_i \in B_0(y^n)$. If no such $\bar{U}_i$ exists, we set $U^*(y^n) = \bar{U}_1$. Next, if there exists a pair $(\overline{V}^1_{i,j_1}, \overline{V}^2_{i,j_2})$ such that $(\overline{V}^1_{i,j_1}, \overline{V}^2_{i,j_2}) \in B_{12}(y^n, \overline{U}_i)$, then define $(V_1^*(y^n, U^*), V_2^*(y^n, U^*)) = (\overline{V}^1_{i,j_1}, \overline{V}^2_{i,j_2})$. If there exists no such pair, define $(V_1^*(y^n, U^*), V_2^*(y^n, U^*)) = (\overline{V}^1_{i,1}, \overline{V}^2_{i,1})$.

It follows from the rate conditions in (126), Lemma 2 with $\lambda = 3\epsilon$ and the bounds on set sizes that:

$$P\Big(U^* \in B_0(Y^n), (V_1^*, V_2^*) \in B_{12}(Y^n, U^*)\Big| Y^n \in A_0\Big) \geq 1 - \delta(\epsilon) \quad (142)$$

for some $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Note that $y^n \in A_0$, $U^* \in B_0(Y^n)$ and $(V_1^*, V_2^*) \in B_{12}(Y^n, U^*)$ imply that $(y^n, U^*, V_1^*, V_2^*) \in \mathcal{T}_\epsilon^n(Y, U, V_1, V_2)$. We then have,

$$P(W^*, Y^n, U^*, V_1^*, V_2^* \in T_\epsilon^n) \geq P(E_1)P(E_2|E_1) \quad (143)$$

where events E_1 and E_2 are defined as:

$$\begin{aligned} E_1 &= \{Y^n \in A_0, U^* \in B_0, (V_1^*, V_2^*) \in B_{12}\} \\ E_2 &= \{W^* \in T_\epsilon^n(W|Y^n, U^*, V_1^*, V_2^*)\} \end{aligned} \quad (144)$$

From (129), (135) and (134), we obtain bounds on $P(E_1)$ and $P(E_2|E_1)$:

$$\begin{aligned} P(E_1) &\geq 1 - \delta_1 - \delta_2 - \delta_3 \\ P(E_2|E_1) &\geq 1 - \sqrt[8]{\eta} \end{aligned} \quad (145)$$

On substituting in (143), we obtain the first bound in (125). The other two bounds in (125) can be shown using similar arguments. ∎

Lemma 4 (Conditional Markov Lemma-for Mutual Covering): Suppose that $(X_1, X_2, U_1, U_2, U_{11}, U_{12}, U_{21}, U_{22})$ *are random variables taking values in arbitrary finite sets* $(\mathcal{X}_1, \mathcal{X}_2, \mathcal{U}_1, \mathcal{U}_2, \mathcal{U}_{11}, \mathcal{U}_{12}, \mathcal{U}_{21}, \mathcal{U}_{22})$*, respectively. Let the random variables satisfy the following Markov condition:*

$$(U_1, U_{11}, U_{12}) \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow (U_2, U_{21}, U_{22}) \quad (146)$$

Let $\overline{U}_{1,i} : i = 1, \ldots, 2^{nR_1}$ *and* $\overline{U}_{2,i} : i = 1, \ldots, 2^{nR_2}$ *be independent codewords of length n each generated using the marginals* $P(U_1)$ *and* $P(U_2)$*, respectively. Let* $2^{nR_{11}}$ *and* $2^{nR_{12}}$ *codewords of* U_{11} *and* U_{12} *(denoted by* $\overline{U}_{11,ij}$ *and* $\overline{U}_{12,ij}$*), respectively, be generated conditioned on each codeword* $\overline{U}_{1,i}$*. Similarly generate codewords of* U_{21} *and* U_{22} *at rates* R_{21} *and* R_{22}*, respectively, conditioned on the codewords of* U_2*. Then for n sufficiently large, there exists functions* $U_1^*(X_1^n)$*,* $U_2^*(X_2^n)$*,* $U_{11}^*(X_1^n, U_1^*)$*,* $U_{12}^*(X_1^n, U_1^*)$*,* $U_{21}^*(X_2^n, U_2^*)$ *and* $U_{22}^*(X_2^n, U_2^*)$ *taking values in* $\mathcal{U}_1^n, \mathcal{U}_2^n, \mathcal{U}_{11}^n, \mathcal{U}_{12}^n, \mathcal{U}_{21}^n$ *and* $\mathcal{U}_{22}^n$*, respectively, such that:*

$$P\left((X_1^n, X_2^n, U_1^*, U_2^*, U_{11}^*, U_{12}^*, U_{21}^*, U_{22}^*) \in \mathcal{T}_\epsilon^n\right) \geq 1 - \delta(\epsilon) \quad (147)$$

where $\delta(\epsilon) \to 0$ *as* $\epsilon \to 0$ *if the rates satisfy:*

$$\begin{aligned} R_1 &> I(X_1; U_1), \\ R_2 &> I(X_2; U_2) \\ R_1 + R_{11} &> I(X_1; U_{11}, U_1), \\ R_1 + R_{12} &> I(X_1; U_{12}, U_1), \\ R_2 + R_{21} &> I(X_2; U_{21}, U_2), \\ R_2 + R_{22} &> I(X_2; U_{22}, U_2), \\ R_1 + R_{11} + R_{12} &> I(X_1; U_{11}U_{12}, U_1) \\ &\quad + I(U_{11}; U_{12}|U_1), \\ R_2 + R_{22} + R_{21} &> I(X_2; U_{21}, U_{22}, U_2) \\ &\quad + I(U_{21}; U_{22}|U_2) \end{aligned} \quad (148)$$

Note that this lemma can be easily extended to the more general case of multiple random variables and multiple layers of encoding using induction (see [3] for the general methodology). While we use the more general version in the proof of Theorem 1 in Appendix B, we restrict to the simpler case here for ease of understanding and to avoid complex notation.

Proof: We note that from standard arguments [2], [15], [33], it follows that if the rates satisfy (148), then there exists functions $U_1^*(X_1^n)$, $U_{11}^*(X_1^n, U_1^*)$ and $U_{12}^*(X_1^n, U_1^*)$ such that:

$$P\left((X_1^n, U_1^*, U_{11}^*, U_{12}^*) \in \mathcal{T}_\epsilon^n\right) \geq 1 - \delta(\epsilon) \quad (149)$$

for some $\delta(\epsilon) \to 0$ as $\epsilon \to 0$. Also, note that X_2^n is drawn according to the right conditional PMF given X_1^n. Hence, we have:

$$P\left((X_1^n, X_2^n, U_1^*, U_{11}^*, U_{12}^*) \in \mathcal{T}_\epsilon^n\right) \geq 1 - \delta(\epsilon) \tag{150}$$

What remains for us to show is that there exists functions $U_2^*(X_2^n)$, $U_{21}^*(X_2^n, U_2^*)$ and $U_{22}^*(X_2^n, U_2^*)$, taking values in $\mathcal{U}_2^n, \mathcal{U}_{21}^n, \mathcal{U}_{22}^n$, jointly typical with $X_1^n, X_2^n, U_1^*, U_{11}^*, U_{12}^*$. We invoke Lemma 3 with $W^* = (X_1^n, U_1^*, U_{11}^*, U_{12}^*)$, $Y^n = X_2^n$, $U = U_2$, $V_1 = U_{21}$ and $V_2 = U_{22}$. Note that given (148) and (150), conditions (122),(123) and (124) are satisfied (for a formal proof of this claim, refer to [33]). Hence, it follows from Lemma 3 that there exist functions $U_2^*(X_2^n)$, $U_{21}^*(X_2^n, U_2^*)$ and $U_{22}^*(X_2^n, U_2^*)$ such that:

$$P\left((X_1^n, X_2^n, U_1^*, U_2^*, U_{11}^*, U_{12}^*, U_{21}^*, U_{22}^*) \in \mathcal{T}_\epsilon^n\right) \geq 1 - \delta(\epsilon) \tag{151}$$

thus proving the lemma. ■

References

[1] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "Networked Slepian–Wolf: Theory, algorithms, and scaling laws," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4057–4073, Dec. 2005.

[2] R. Venkataramani, G. Kramer, and V. K. Goyal, "Multiple description coding with many channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 9, pp. 2106–2114, Sep. 2003.

[3] T. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inf. Theory*, vol. IT-26, no. 3, pp. 277–288, May 1980.

[4] R. Cristescu, B. Beferull-lozano, M. Vetterli, and R. Wattenhofer, "Network correlated data gathering with explicit communication: NP-completeness and algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. 1, pp. 41–54, Feb. 2006.

[5] S. Pattem, B. Krishnamachari, and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," *ACM Trans. Sensor Netw.*, vol. 4, no. 4, p. 24, 2008.

[6] J. Liu, M. Adler, D. Towsley, and C. Zhang, "On optimal communication cost for gathering correlated data through wireless sensor networks," in *Proc. 12th ACM Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2006, pp. 310–321.

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

[8] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 226–228, Mar. 1975.

[9] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. IT-21, no. 3, pp. 294–300, May 1975.

[10] I. Csiszár and J. Körner, "Towards a general theory of source networks," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 155–165, Mar. 1980.

[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[12] R. M. Gray and A. D. Wyner, "Source coding for a simple network," *Bell Syst. Tech. J.*, vol. 53, no. 9, pp. 1681–1721, Nov. 1974.

[13] S. Tung, "Multiterminal source coding," Ph.D. dissertation, Dept. Electr. Eng., Cornell Univ., Ithaca, NY, USA, 1978.

[14] A. B. Wagner, B. G. Kelly, and Y. Altug, "Distributed rate-distortion with common components," *IEEE Trans. Inf. Theory*, vol. 57, no. 7, pp. 4035–4057, Jul. 2011.

[15] A. E. Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 6, pp. 851–857, Nov. 1982.

[16] Z. Zhang and T. Berger, "New results in binary multiple descriptions," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 4, pp. 502–521, Jul. 1987.

[17] H. Luo, Y. Liu, and S. K. Das, "Routing correlated data in wireless sensor networks: A survey," *IEEE Netw.*, vol. 21, no. 6, pp. 40–47, Nov./Dec. 2007.

[18] S. Pattem, B. Krishnamachari, and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," *ACM Trans. Sensor Netw.*, vol. 4, no. 4, p. 24, 2008.

[19] T. Cormen, C. Leiserson, and R. Rivest, *Introduction to Algorithms*. New York, NY, USA: McGraw-Hill, Jul. 2001.

[20] D. S. Lun, M. Médard, T. Ho, and R. Koetter, "Network coding with a cost criterion," in *Proc. Int. Symp. Inform. Theory Appl.*, 2004, pp. 1232–1237.

[21] A. Ramamoorthy, "Minimum cost distributed source coding over a network," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 461–475, Jan. 2011.

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

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

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

[25] C. Tian and S. N. Diggavi, "Side-information scalable source coding," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5591–5608, Dec. 2008.

[26] A. D. Wyner, J. K. Wolf, and F. M. J. Willems, "Communicating via a processing broadcast satellite," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1243–1249, Jun. 2002.

[27] R. Timo, A. J. Grant, and L. Hanlen, "Source coding for a noiseless broadcast channel with partial receiver side information," in *Proc. IEEE Austral. Commun. Theory Workshop*, Feb. 2007.

[28] R. Timo, A. Grant, T. Chan, and G. Kramer, "Source coding for a simple network with receiver side information," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2008, pp. 2307–2311.

[29] T. P. Coleman, E. Martinian, and E. Ordentlich, "Joint source–channel coding for transmitting correlated sources over broadcast networks," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3864–3868, Aug. 2009.

[30] D. Gunduz, E. Erkip, A. Goldsmith, and H. V. Poor, "Source and channel coding for correlated sources over multiuser channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 3927–3944, Sep. 2009.

[31] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[32] K. Viswanatha, E. Akyol, S. Ramaswamy, and K. Rose, "Distributed source coding and dispersive information routing: An integrated approach with networking and database applications," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2010, pp. 1894–1898.

[33] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[34] J. Körner and K. Marton, "How to encode the modulo-two sum of binary sources (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 25, no. 2, pp. 219–221, Mar. 1979.

[35] A. A. Gohari and V. Anantharam, "Evaluation of Marton's inner bound for the general broadcast channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 608–619, Feb. 2012.

[36] E. Akyol, U. Mitra, E. Tuncel, and K. Rose, "Source coding in the presence of exploration-exploitation tradeoff," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jun. 2014.

[37] K. Viswanatha, E. Akyol, and K. Rose, "A strictly improved achievable region for multiple descriptions using combinatorial message sharing," in *Proc. IEEE Inform. Theory Workshop (ITW)*, Oct. 2011, pp. 100–104.

[38] E. Akyol, K. Viswanatha, and K. Rose, "Combinatorial message sharing and random binning for multiple description coding," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jul. 2012, pp. 1371–1375.

[39] E. Akyol, K. Viswanatha, and K. Rose, "On random binning versus conditional codebook methods in multiple descriptions coding," in *Proc. IEEE Inform. Theory Workshop (ITW)*, Sep. 2012, pp. 312–316.

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

[41] E. Akyol, U. Mitra, E. Tuncel, and K. Rose, "On scalable coding in the presence of decoder side information," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jun. 2014.

[42] H. G. Eggleston, *Convexity*. Cambridge, U.K.: Cambridge Univ. Press, 1958.

[43] A. D. Wyner and J. Ziv, "A theorem on the entropy of certain binary sequences and applications—I," *IEEE Trans. Inf. Theory*, vol. 19, no. 6, pp. 769–772, Nov. 1973.

[44] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder—II: General sources," *Inform. Controls*, vol. 38, no. 1, pp. 60–80, Jul. 1978.

[45] K. Viswanatha, E. Akyol, and K. Rose, "Towards optimum cost in multi-hop networks with arbitrary network demands," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jun. 2010, pp. 1833–1837.

[46] K. Viswanatha, E. Akyol, and K. Rose, "On optimum communication cost for joint compression and dispersive information routing," in *Proc. IEEE Inform. Theory Workshop (ITW)*, Aug./Sep. 2010, pp. 1–5.

[47] K. Viswanatha, E. Akyol, and K. Rose, "An achievable rate region for distributed source coding and dispersive information routing," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Jul./Aug. 2011, pp. 776–780.

Kumar B. Viswanatha (S'08–M'14) received his Ph.D. degree in 2013 in electrical and computer engineering from the University of California at Santa Barbara (UCSB), USA. He is currently working for Qualcomm research center in San Diego. Prior to joining Qualcomm, he was an intern associate in the equity volatility desk at Goldman Sachs Co., New York, USA. His research interests include multi-user information theory, wireless communications, joint compression and routing for networks and distributed compression for large scale sensor networks.

Emrah Akyol (S'03–M'12) received the Ph.D. degree in 2011 in electrical and computer engineering from the University of California at Santa Barbara. From 2006 to 2007, he held positions at Hewlett-Packard Laboratories and NTT Docomo Laboratories, both in Palo Alto, where he worked on topics in video compression.

Currently, Dr. Akyol is a postdoctoral researcher in the Electrical Engineering Department, University of Southern California. His current research is on the interplay of networked information theory, communications and control.

Kenneth Rose (S'85–M'91–SM'01–F'03) received the Ph.D. degree in 1991 from the California Institute of Technology, Pasadena.

He then joined the Department of Electrical and Computer Engineering, University of California at Santa Barbara, where he is currently a Professor. His main research activities are in the areas of information theory and signal processing, and include rate-distortion theory, source and source-channel coding, audio and video coding and networking, pattern recognition, and nonconvex optimization. He is interested in the relations between information theory, estimation theory, and statistical physics, and their potential impact on fundamental and practical problems in diverse disciplines.

Dr. Rose was co-recipient of the 1990 William R. Bennett Prize Paper Award of the IEEE Communications Society, as well as the 2004 and 2007 IEEE Signal Processing Society Best Paper Awards.