A Successive Description property of Monotone-Chain Polar Codes for Slepian-Wolf coding

A Successive Description Property of Monotone-Chain Polar Codes for Slepian–Wolf Coding

Salman Salamatian
EPFL,
Lausanne, Switzerland

Muriel Médard
MIT
Cambdridge, MA, USA

Emre Telatar
EPFL
Lausanne, Switzerland

***Abstract*—We introduce a property that we call Successive Description property for Slepian Wolf coding. We show that Monotone-Chain Polar Codes can be used to construct low-complexity codes that satisfy this property. We discuss applications of this property to network coding problems.**

***Keywords*—Slepian–Wolf Coding, Polar Codes, Separation of Source and Network Coding**

I. Introduction

Consider a memoryless source $(X,Y) \sim P_{X,Y}$ where $P_{X,Y}$ is a joint distribution on $\mathcal{X}\times\mathcal{Y}$ with $\mathcal{X}=\mathcal{Y}=\{0,1\}$. Let (X^n,Y^n) be what the source generates after the first n instances. In this paper we study a version of the Slepian-Wolf problem with 2 sources and 2 receivers. In this setting the first receiver gets nR_1 encoded bits from the first source and nR_2 encoded bits from the second source, and is expected to decode with vanishing probability of error (X^n,Y^n). On the other hand, the second receiver gets respectively $n\tilde{R}_1$ and $n\tilde{R}_2$ encoded bits from each source, and should also decode reliably (X^n,Y^n). The result of Slepian and Wolf in [1] states that if the pairs (R_1,R_2) and $(\tilde{R}_1,\tilde{R}_2)$ are in the region $\mathcal{R}_{SW} = \{(R_x,R_y) : R_x \geq H(X|Y), R_y \geq H(Y|X), R_x+R_y \geq H(X,Y)\}$, then there exist encoders $f_X : \mathcal{X}^n \to \{0,1\}^{nR_1}$ and $f_Y : \mathcal{Y}^n \to \{0.1\}^{nR_2}$, such that receiver 1 can decode reliably x^n and y^n using $(f_X(x^n), f_Y(y^n))$. Similarly, there exist encoders $f_{\tilde{X}}$ and $f_{\tilde{Y}}$ that encode at rates $(\tilde{R}_1,\tilde{R}_2)$ and allow for reliable decoding.

For the encoder to satisfy the rate constraints on the decoders, a naive solution would consist of doing the following encoding:

$$\begin{aligned} F_x &: X^n \mapsto (f_X(X^n), \tilde{f}_X(X^n)) \\ F_y &: Y^n \mapsto (f_Y(Y^n), \tilde{f}_X(Y^n)). \end{aligned} \tag{1}$$

The first and second decoder respectively selects $(f_X(X^n), f_Y(Y^n))$ and $(\tilde{f}_X(X^n), \tilde{f}_Y(Y^n))$ to decode at their desired rates. However, the encoding of (1) is wasteful as we are essentially using two codebooks, and encoding sources X^n and Y^n into respectively $n(R_1+\tilde{R}_1)$ and $n(R_2+\tilde{R}_2)$ bits. In contrast, a code is not *wasteful* in this sense, which we call successive description Slepian–Wolf code, if X^n can be expressed using $n(\max\{R_1,\tilde{R}_1\})$ and Y^n using $n(\max\{R_2,\tilde{R}_2\})$ bits, while still satisfying the rate constraints of each receiver. More precisely this means:

Definition 1. *Let (R_1,R_2) and $(\tilde{R}_1,\tilde{R}_2)$ be 2 pairs of rates on the dominant face of Slepian-Wolf region $\mathcal{R}_{SW}$, that is for which $R_1+R_2=\tilde{R}_1+\tilde{R}_2=H(X,Y)$ and both (R_1,R_2) and $(\tilde{R}_1,\tilde{R}_2)\in\mathcal{R}_{SW}$. Suppose,* w.l.o.g.*, that $R_1\leq\tilde{R}_1$ and $\tilde{R}_2\leq R_2$. A code has a successive description property if its encoding functions:*

$$\begin{aligned} f_X &: \mathcal{X}^n \to \{0,1\}^{nR_1}\times\{0,1\}^{n(\tilde{R}_1-R_1)} \\ &\quad x^n \mapsto (u,a) \end{aligned} \tag{2}$$

$$\begin{aligned} f_Y &: \mathcal{Y}^n \to \{0,1\}^{n\tilde{R}_2}\times\{0,1\}^{n(R_2-\tilde{R}_2)} \\ &\quad y^n \mapsto (v,b) \end{aligned} \tag{3}$$

are such that (x^n,y^n) can be recovered from both (u,a,v) and (u,v,b).

This property is trivially satisfied by random codes, which we will discuss in Sec. III. The main result of this paper in Sec. III, is to show that Monotone Chain Polar Codes— a family of efficient encoding and decoding codes— satisfy the successive description property. Monotone Chain Polar coding was introduced in [2] and is reviewed in Sec. II. This is a low complexity distributed source coding technique that can operate at any pair of rates on the dominant face. We will also discuss why codes that have a successive description property are important in the study of some network coding problems, where the rate constraints on the receivers is given by the network topology. This will be discussed in Sec. V.

II. Background and Notations

In this paper, we will only analyze the case of two sources X and Y, and two receivers. All of the results can be easily generalized to a arbitrary finite number of sources and receivers. Keeping the discussion to this setting simplifies notation and exposition.

Below we give a quick overview of monotone chain polar coding introduced in [2].

Polar transform: The polar transform is a linear bijective function from $X^n\in\{0,1\}^n$ to $U^n\in\{0,1\}^n$ defined by :

$$U^n = G_n X^n \tag{4}$$

with G_N defined as in [3]. We will not go into details of G_n, but suffice it to say that it follows from a recursive construction

of the $n = 2$ case:

$$U^2 = (X_1 \oplus X_2, X_2) \tag{5}$$

Because the transformation described in (4) is a bijection, it is the case that:

$$H(U^n) = H(X^n) \tag{6}$$

In a similar way, it is possible to do the polar transform on Y^n to obtain the random variable $V^n : V^n = G_n Y^n$.

Monotone Chain Polar Codes: The joint entropy $H(U^n, V^n)$ can be expanded in many ways using the chain rule. Perhaps the simplest expansion is :

$$\begin{aligned} H(U^n, V^n) &= H(U^n) + H(V^n|U^n) \\ &= \sum_j^n H(U_j|U^{j-1}) + \sum_j^n H(V_j|V^{j-1}, U^n) \end{aligned} \tag{7}$$

The main idea of source polarization and monotone chain polar codes is to consider so-called monotone decompositions of $H(U^n, V^n)$ as the one in (7) and to show that each term in that expression converges to either 0 or 1.

Definition 2. *A chain expansion of (U^n, V^n) is called Monotone with respect to U^n, if :*

$$H(U^n, V^n) = \sum_{i=1}^{2n} H(S_i|S^{i-1}) \tag{8}$$

with S^{2n} a permutation of (U^n, V^n) such that the relative order of U^n is preserved. We say a chain expansion of (U^n, V^n) is monotone, if it is monotone with respect to both U^n and V^n.

Note that we will use the notation S^{2n} to represent the permutation in a particular monotone chain.

It is very convenient to represent monotone chains for two sources in diagram form as was done in [2]. The regular diagram for $n = 4$ can be seen in Fig 1: all valid monotone decomposition can be seen by looking at the paths from the lower-left $\emptyset$ to the upper-right U^4V^4. As such, the decomposition in (7) corresponds to the two segment path from $\emptyset$ to U^4, and then to U^4V^4. Finally, monotone decompositions can be represented equivalently by a binary string sequence $b^{2n} = \{0, 1\}^{2n}$, where $b_i = 0$ corresponds to a horizontal move on a regular diagram at step i, and $b_i = 1$ corresponds to a vertical move.

It has been proved in [2] that monotone decompositions that have binary string sequence of the form $b^{2n} \in \mathcal{V} = \{0^i 1^n 0^{n-i}\}$ are sufficient to approximate any pair of rates on the dominant face of $\mathcal{R}_{SW}$. These correspond to the three segments path on the diagram, first i horizontal moves, then n vertical moves, finally the $n - i$ remaining horizontal moves.

This result can be interpreted using the idea of source spliting introduced in [4] [1]. Source splitting amounts to splitting a source X into two separate sources X_1 and X_2, and then describing the data using $H(X_1) + H(Y|X_1) + H(X_2|Y, X_1)$ bits. Altough the split in [4] is done through a time-sharing variable, we can consider different, possibly deterministic ways of splitting X. In the context of this class of Monotone Chain Polar codes, we split the variable U^n into two variables U^i and U_{i+1}^n. Next, we sequentially compress U^i using $H(U^i)$ bits, then V^n using $H(V^n|U^i)$ bits, and finally U_{i+1}^n using $H(U_{i+1}^n|V^n, U^i)$. Note that as:

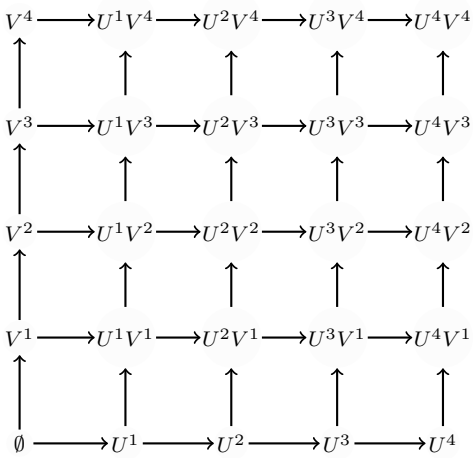

Fig. 1: Regular diagram for Monotone Chain representation for $n = 4$

$$\begin{aligned} H(X^n, Y^n) &= H(U^n, V^n) \\ &= H(U^i) + H(V^n|U^i) + H(U_{i+1}^n|V^n, U^i) \end{aligned} \tag{9}$$

the resulting values

$$R_1 = \frac{1}{n}\left(H(U^i) + H(U_{i+1}^n|V^n, U^i)\right) \tag{10}$$

$$R_2 = \frac{1}{n}\left(H(V^n|U^i)\right) \tag{11}$$

are indeed on the dominant face of $\mathcal{R}_{SW}$. The hope is that by choosing i appropriately, one can tune the (R_1, R_2) as described in eqs. (10) and (11) to approximate any arbitrary pair of rates on the dominant face. Next, using polarization results, one can show that the values (R_1, R_2) correspond to actual operational rates: the ratio of bits that each user describes. The theorem below shows that it is possible to approximate any pair of rates using the simple source spilt above:

Theorem 1. ***from [2]*** *Let $(\hat{R}_1, \hat{R}_2)$ be an arbitrary pair of rates on the dominant face of $\mathcal{R}_{SW}$. For any $\epsilon > 0$, there exist n sufficiently large, such that $|R_1 - \hat{R}_1| \leq \epsilon$ and $|R_2 - \hat{R}_2| < \epsilon$, where R_1 and R_2 are defined in eqs.* (10) *and* (11)

Furthermore, the polarization results of [5] show that when $n \to \infty$, the terms in the chain expansion will tend towards either 0 or 1. This means that the sets of high entropy bits satisfy:

$$\begin{aligned} |A_X(\delta)| &= |\{i : b_i = 0, H(S_i|S^{i-1}) \geq 1 - \delta\}| \approx nR_1 \\ |A_Y(\delta)| &= |\{i : b_i = 1, H(S_i|S^{i-1}) \geq 1 - \delta\}| \approx nR_2. \end{aligned} \tag{12}$$

[1] Relation between source splitting and Monotone Chain Polar codes were pointed out by Arikan himself in the concluding remarks of [2].

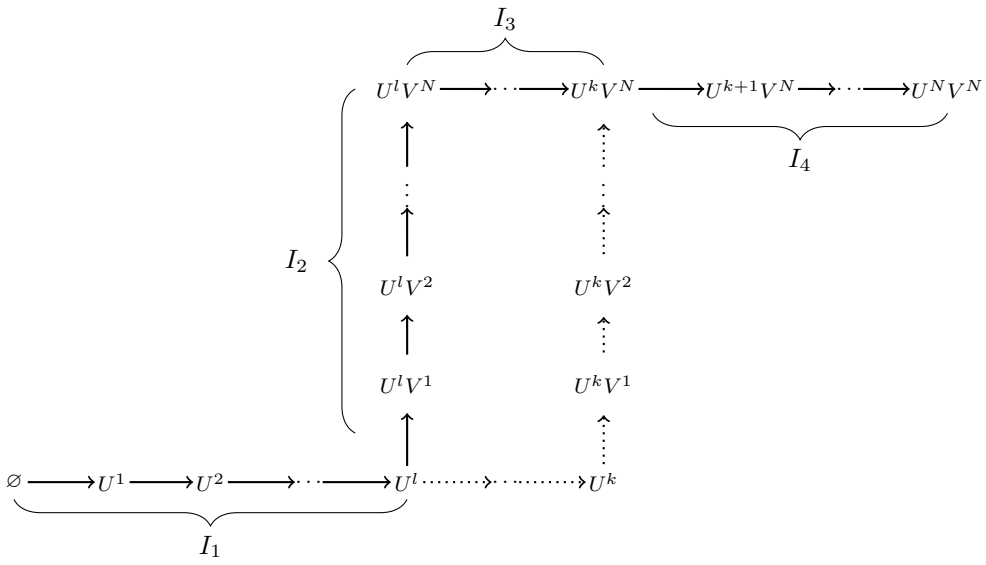

Fig. 2: The paths for two pairs of rates can be decomposed into 4 segments.

This allows a simple encoding technique:

- Given a source realization (x^n, y^n), each sender computes $u^n = G_n x^n$ and $v^n = G_n y^n$ respectively. This defines a realization of the path variables s^{2n}
- The first sender transmits $\{s_i : i \in A_X(\delta)\}$. Similarly the second sender transmits $\{s_i : i \in A_Y(\delta)\}$.
- The decoder has access to $\{s_i : i \in A_X(\delta) \cup A_Y(\delta)\}$. A low complexity successive decoding algorithm given in [2] decodes (x^n, y^n) with high reliability.

In the rest of the paper, we will refer to a Monotone Chain Polar Code, as a (n, b^{2n}, δ) Monotone chain Polar Code, where n is the block length, b^{2n} is the binary sequence of the decomposition, and δ is a parameter that controls the error rate.

III. Successive Description Property

The simplest codes that satisfy the successive description property are perhaps random codes along with joint typicality decoding.

Proposition 1. *Random binning encoding and joint typicality decoding is a successive description coding scheme as defined in Def. 1.*

Proof: Follows from standard techniques. □

This result on random coding is not surprising, as only the number of bits received by the decoder is important in expressing the error probability. Therefore, puncturing some bits from a source can be compensated by using additional bits from the other source, as long as the rates are in the Slepian Wolf region. However, typicality decoding cannot be implemented in practice, and we present now the main result of our paper concerning Monontone Chain Polar Codes.

Theorem 2. *Let (R_1, R_2) and $(\tilde{R}_1, \tilde{R}_2)$, $R_1 \leq \tilde{R}_1$ be two pairs of rates on the dominant face of $\mathcal{R}_{SW}$. Consider two Monotone Chain Polar Codes of same block length n and argument δ at the pair of rates (R_1, R_2) and $(\tilde{R}_1, \tilde{R}_2)$. It is possible to construct encoding functions as defined in Def. 1 from the Monotone Chain Polar codes.*

Proof: Let n be the block size. Consider two pair of rates (R_1, R_2) and $(\tilde{R}_1, \tilde{R}_2)$, and their associated paths b^{2N} and $\tilde{b}^{2N}$. The associated path variables are denoted respectively by S^{2N} and $\tilde{S}^{2N}$. Recall that the paths are in the form $b^{2N} = 0^l 1^N 0^{N-l}$ and $\tilde{b}^{2N} = 0^k 1^N 0^{N-k}$, and suppose without loss of generality that $l \leq k$. This case is shown in Figure 2. Recall that the encoding consist in storing the S_i or $\tilde{S}_i$ that have a high entropy in the chain decomposition. Let $A_X(\delta), A_Y(\delta)$ and $\tilde{A}_X(\delta), \tilde{A}_Y(\delta)$ be the high entropy sets defined in (12) respectively for b^{2n} and $\tilde{b}^{2n}$. Finally, let $\pi : \{1, \ldots, 2N\} \to \{1, \ldots, 2N\}$ be defined such that for all $1 \leq i \leq 2N$, we have $S_i = \tilde{S}_{\pi(i)}$.

We investigate all the segments :

- $\mathbf{I_1}$: let $1 \leq i \leq l$. In this range, $\pi(i) = i$, which means that $H(S_i|S^{i-1}) = H(\tilde{S}_\pi(i)|\tilde{S}^{\pi(i)-1})$, therefore $i \in A_X(\delta) \iff \pi(i) \in \tilde{A}_X(\delta)$.
- $\mathbf{I_4}$: let $i > N + k$, similarly we have $\pi(i) = i$, so $H(S_i|S^{i-1}) = H(\tilde{S}_\pi(i)|\tilde{S}^{\pi(i)-1})$. Therefore $i \in A_X(\delta) \iff \pi(i) \in \tilde{A}_X(\delta)$
- $\mathbf{I_3}$: let $N + l < i \leq N + k$. Because $S_i = \tilde{S}_{i-N} = U_{i-N}$, we have $\pi(i) = i - N$ and $H(S_i|S^{i-1}) = H(U_{i-N}|U^{i-N-1}, V^N) \leq H(U_{i-N}|U^{i-N-1}) = H(\tilde{S}_{\pi(i)}|\tilde{S}^{\pi(i)-1})$. Therefore $i \in A_X(\delta) \implies \pi(i) \in \tilde{A}_X(\delta)$.
- $\mathbf{I_2}$: let $l < i \leq N + l$. Because $S_i = \tilde{S}_{i+k-l} = V_{i-l}$, we have $\pi(i) = i + k - l$ and $H(S_i|S^{i-1}) = H(V_{i-l}|V^{i-l-1}, U^l) \geq H(V_{i-l}|V^{i-l-1}, U^k) =$

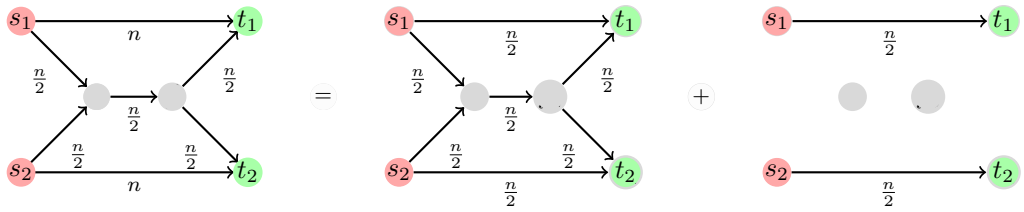

Fig. 3: Decomposition of the network G into G_1 and G_2

$H(\tilde{S}_{\pi(i)}|\tilde{S}^{\pi(i)-1})$. Therefore, $\pi(i) \in \tilde{A}_Y(\delta) \implies i \in A_Y(\delta)$.

This means that the set of high entropy bits is nested: If $i \in A_X(\delta)$, then $\pi(i) \in \tilde{A}_X(\delta)$. Conversely, if $\pi(i) \in \tilde{A}_Y(\delta)$ then $i \in A_Y(\delta)$. Therefore, we define encoding functions F_X and F_Y as follows:

$$F_X : x^n \mapsto (\{s_i : i \in A_{X\cap\tilde{X}}\}, \{s_i : i \in A_X \backslash A_{X\cap\tilde{X}}\}) \quad (13)$$
$$F_Y : y^n \mapsto (\{s_i : i \in A_{Y\cap\tilde{Y}}\}, \{s_i : i \in A_{\tilde{Y}} \backslash A_{Y\cap\tilde{Y}}\}) \quad (14)$$

where $A_{X\cap\tilde{X}} = \{i : i \in A_X \text{ and } \pi(i) \in \tilde{A}_X(\delta)\}$ and $A_{Y\cap\tilde{Y}} = \{i : \pi(i) \in \tilde{A}_Y(\delta) \text{ and } i \in A_Y(\delta)\}$. The previous observations imply that $\{s_i : i \in A_{X\cap\tilde{X}}\} = \{s_i : i \in \tilde{A}_X(\delta)\}$, and $\{s_i : i \in A_{Y\cap\tilde{Y}}\} = \{s_i : i \in A_Y(\delta)\}$. Using (12) concludes the proof. □

IV. Separation in Network Coding

In this section we look at an application of successive description codes to solve network coding problems at low complexity. The problem of multicast network coding [6] consists in transmitting information from a set of sources nodes, to a set of terminal nodes, where all terminals are interested in all sources. This setting has been well studied, and most fundamental limits in this simple case are well-known and understood. When the sources are independent, there exist low complexity algorithms, most notably random linear network codes in a large enough field [7] that allow for a low complexity encoding and decoding of the sources. However, the case of correlated sources is more delicate, and all general methods rely on a joint typicality decoder, or a maximum likelihood or minimum entropy decoder, neither of which can be implemented in practice.

Separation of source and network coding has been studied in [8], where a notion of separation of source and network coding is proposed. Although it is not discussed in the paper itself, the use of successive description codes is often necessary in this context. We illustrate this with the following example: Consider the butterfly network in Fig. 3 . On this network, the source X with $H(X) = 1$ generated at s_1, Y with $H(Y) = 1$ generated at s_2, are to be transmitted to two terminals t_1 and t_2. Further suppose that the sources are correlated such that $H(X;Y) = \frac{3}{2}$. We propose a low complexity separation approach in the following steps:

- Use a successive description Slepian-Wolf Code with $(R_1, R_2) = (1, \frac{1}{2}) = (\tilde{R}_2, \tilde{R}_1)$ to encode x^n into (u, a), and y^n into (v, b).
- Observe that the network can be decomposed into networks G_1, and G_2 as shown in Figure 3.
- Send (u, v) to both terminals using the network G_1. This is possible by the max-flow theorem of [6]. A possible linear code is shown in Fig. 4 where the summation is a bit to bit summation.
- Send sequence a from S_1 to T_1, and sequence b from S_2 to T_2 using network G_2 (this is not multicast).
- Terminal t_1 receives (u, a, v) and terminal t_2 receives (u, v, b). Since the code is a successive description code, it is possible to decode reliably at both terminals.

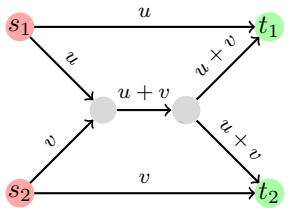

Fig. 4: Linear code to multicast (u, v) to t_1 and t_2

The use of a successive description code in this example allows each terminal to decode at different points on $\mathcal{R}_{SW}$, here on each corner point. Having this flexibility is necessary. Indeed, a naive approach where one encodes x^n and y^n using a Slepian-Wolf codes that does not allow for a successive description cannot succeed due to the constraints on the rates for each decoder given by the topology of the network. This simple example illustrates how a successive description code can be used in a practical setting. Note that using monotone chain polar codes would allow for an overall low-complexity scheme for encoding and decoding of many network coding problems with correlated sources, as shown in [8].

V. Final Remarks

Successive description is a property that is trivially satisfied by random coding, since the only parameter that matters is the number of bits received by the decoder. When going from random codes to structured codes, it is not clear that puncturing bits from one source in exchange for additional bits from another would yield a low complexity decoding. Monotone Chain Polar Codes happen to be a family of codes in which puncturing bits preserves the structure.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *Information Theory, IEEE Transactions on*, vol. 19, no. 4, pp. 471–480, 1973.

[2] E. Arikan, "Polar coding for the slepian-wolf problem based on monotone chain rules," in *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, July 2012, pp. 566–570.

[3] ——, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *Information Theory, IEEE Transactions on*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[4] B. Rimoldi and R. Urbanke, "Asynchronous slepian-wolf coding via source-splitting," in *Information Theory. 1997. Proceedings., 1997 IEEE International Symposium on*, Jun 1997, pp. 271–.

[5] E. Arikan, "Source polarization," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, June 2010, pp. 899–903.

[6] L. Song and R. Yeung, "Network information flow-multiple sources," in *Information Theory, 2001. Proceedings. 2001 IEEE International Symposium on*, 2001, pp. 102–.

[7] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A Random Linear Network Coding Approach to Multicast," *Information Theory, IEEE Transactions on*, vol. 52, no. 10, pp. 4413–4430, 2006.

[8] A. Ramamoorthy, K. Jain, P. Chou, and M. Effros, "Separating distributed source coding from network coding," *Information Theory, IEEE Transactions on*, vol. 52, no. 6, pp. 2785–2795, June 2006.