"Real" Slepian-Wolf Codes

Bikash Kumar Dey, Sidharth Jaggi, and Michael Langberg

Abstract

We provide a novel achievability proof of the Slepian-Wolf theorem for i.i.d. sources over finite alphabets. We demonstrate that random codes that are linear over the real field achieve the classical Slepian-Wolf rate-region. For finite alphabets we show that typicality decoding is equivalent to solving an integer program. Minimum entropy decoding is also shown to achieve exponentially small probability of error. The techniques used may be of independent interest for code design for a wide class of information theory problems, and for the field of compressed sensing.

I. Introduction

A well-known result by Slepian and Wolf in [2] characterizes the rate-region for near-lossless source coding of distributed sources. The result demonstrates that if two (or more) sources possess correlated data, even independent encoding of the sources' data can still achieve essentially the same performance as when the sources encode jointly. This result has important implications for information theoretic problems as diverse as sensor networks [3], secrecy [4], and low-complexity video encoding [5]. Unfortunately for the distributed source coding problem, codes that are provably both rate-optimal and computationally efficient to implement are hard to come by. Section II gives a partial history of results for the Slepian-Wolf (SW) problem.

In this work we provide novel codes that asymptotically achieve the SW rate-region with vanishing probability of error. Our encoding procedure comprises of random linear operations over the real field $\mathbb{R}$, and are hence called *Real Slepian-Wolf Codes* or RSWCs. In contrast most other codes in the literature operate over appropriate finite fields $\mathbb{F}_q$. We demonstrate that RSWCs can be used in a way that enables the receiver to decode the sources' information by solving a set of integer programs (IPs). Besides being interesting in their own right as a new class of codes achieving the SW rate-region, the relation between RSWCs and IPs has some intriguing implications.

In general IPs are computationally intractable to solve. However, our code design gives us significant flexibility in choosing the particular IPs corresponding to our codes. That is, we show that "almost all" RSWCs result in IPs that have "good" performance for the SW problem. But there are well-studied classes of IPs that are known

[0]The work in this paper was presented in part in ISIT 2008, Toronto, Canada, July 2008 [1].

B. K. Dey is with the Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India, 400 076, email: bikash@ee.iitb.ac.in .

S. Jaggi is with the Department of Information Engineering, Chinese University of Hong Kong, Shatin, N.T., Hong Kong, email: jaggi@ie.cuhk.edu.hk

M. Langberg is with the Computer Science Division, Open University of Israel, 108 Ravutski St., Raanana 43107, Israel, email: mikel@openu.ac.il

to be computationally tractable to solve (for e.g., IPs corresponding to *Totally Unimodular matrices* [6]). It is thus conceivable that suitably chosen RSWCs may be decodable with low computational complexity.

Linear SW codes over finite fields were introduced in [7] and they were shown to achieve the SW rate-region. Decoding such codes is equivalent to finding a vertex of a hypercube satisfying some combinatorial properties. Such problems are computationally intractable. Our SW codes are linear over $\mathbb{R}$. Though decoding our codes may still be difficult, we can use tools from the matured field of convex optimization for decoding our codes.

Also, our work has direct implications for the new field of *Compressed Sensing (CS)*. In the CS setup, N sources each generate a single real number. The resulting length-N sequence is *k-sparse*, i.e., can be written with at most $k \ll N$ non-zero coefficients in a prespecified basis. A typical result [8] in this setup shows that if a receiver gets $\mathcal{O}(k \log(N))$ random linear combinations over $\mathbb{R}$ of the sources' sequence, it can, with high probability, reconstruct the source sequence exactly in a computationally efficient manner by solving a linear program. The CS setup is quite similar to that of the RSWCs we design – the source sequence contains a large amount of redundancy, and a random $\mathbb{R}$-linear mixture of the sequence suffices for exact reconstruction via optimization techniques. There are, however, two major differences. First, RSWCs operate at information-theoretically optimal rates whereas CS codes are bounded away from such performance. Second, CS codes are computationally tractable, whereas we are currently not aware of efficient decoding techniques for RSWCs. We think this tradeoff between computational efficiency and rate-optimality is interesting and worthy of further investigation.

In Section II, we discuss some background and tools to be used in the subsequent sections. In Section III, we present the construction of our RSWCs and the related main results. These results are then proved in Sections IV and V. In Section VI, we present the direct construction of RSWCs for any point on the Slepian-Wolf rate-region without time-sharing between the corner points. The universal minimum-entropy decoding algorithm is shown to work for our RSWCs in Section VII. Section VIII shows that our RSWCs achieve the rate-region of more general *normal source networks without helpers* introduced in [9]. Finally Section IX concludes the paper.

II. Background and Definitions

Shannon's seminal source coding theorem [10] demonstrates that a sequence of discrete random variables can essentially be compressed down to the entropy of the underlying probability distribution generating the sequence. Of the many extensions sparked by this paper, the Slepian-Wolf theorem [2] is the one this paper builds on.

A. Slepian Wolf Theorem for i.i.d. sources [2]

Problem Statement: Two sources named Xavier and Yvonne generate two sequences of discrete random variables, $\mathbf{X} \stackrel{\triangle}{=} X_1, X_2, \ldots, X_n$ over the finite alphabet $\mathcal{X}$, and $\mathbf{Y} \stackrel{\triangle}{=} Y_1, Y_2, \ldots, Y_n$ over the finite alphabet $\mathcal{Y}$, respectively. The sequence $(\mathbf{X}, \mathbf{Y})$ is assumed to be i.i.d. with a joint distribution $p_{X,Y}(x, y)$ that is known in advance to both Xavier and Yvonne. The corresponding marginal distributions over X and Y are denoted by $p_X(x)$ and $p_Y(y)$ respectively. Xavier and Yvonne wish to communicate $(\mathbf{X}, \mathbf{Y})$ to a receiver Zorba. To this end Xavier uses his encoder to transmit a message that is a function only of $\mathbf{X}$ and $p_{X,Y}(x, y)$ to Zorba. Similarly, Yvonne uses her

encoder to transmit a message that is a function only of $\mathbf{Y}$ and $p_{X,Y}(x,y)$ to Zorba. Zorba uses his decoder to attempt to reconstruct $(\mathbf{X},\mathbf{Y})$. Xavier and Yvonne's encoders and Zorba's decoder comprise a SW code $\mathcal{C}$. The SW code $\mathcal{C}$ is said to be *near-lossless* if Zorba's reconstruction of $(\mathbf{X},\mathbf{Y})$ is correct with a probability of error over $p_{XY}(x,y)$ that is asymptotically negligible in the *block-length* n. The *rate-pair* (R_X,R_Y) is said to be *achievable* for the SW problem if for every $\epsilon>0$ there exists a code $\mathcal{C}$ that is near-lossless, and the average (over $p_{X,Y}(x,y)$) number of bits that $\mathcal{C}$ requires Xavier and Yvonne to transmit to Zorba are at most $n(R_X+\epsilon)$ and $n(R_Y+\epsilon)$ respectively. The set of all rate-pairs that are achievable is called the *rate-region*. Slepian and Wolf's characterization of the rate-region is remarkably clean.

Theorem 1: [2] The rate-region for the Slepian-Wolf problem is given by the intersection of

$$\begin{aligned} R_X &\geq H(X|Y), \\ R_Y &\geq H(Y|X), \\ R_X+R_Y &\geq H(X,Y). \end{aligned} \tag{1}$$

Here $H(X|Y)$ and $H(Y|X)$ denote the *conditional entropy* and $H(X,Y)$ denotes the *joint entropy* of (X,Y) (implicitly, over the joint distribution $p_{X,Y}(x,y)$).

B. Linear SW codes over finite fields

The SW codes in [2] have computational complexity that is exponential for both encoding and decoding. An improvement was made in [7], where it was shown that *random linear* encoders suffice. We briefly restate that result here, restricting ourselves to the case when $\mathcal{X}=\mathcal{Y}=\{0,1\}$ for simplicity.

Let $\mathbf{D_X}$ and $\mathbf{D_Y}$ be respectively $\lceil n(R_X+\epsilon)\rceil\times n$ and $\lceil n(R_Y+\epsilon)\rceil\times n$ matrices over the finite field $\mathbb{F}_2$, with each entry of both matrices chosen i.i.d. as either 0 or 1 with probability $1/2$. Here ϵ is an arbitrary positive constant. Abusing notation, let $\mathbf{X}$ and $\mathbf{Y}$ also denote length-n column vectors over $\mathbb{F}_2$. Xavier and Yvonne's encoders are then defined respectively via the matrix multiplications $\mathbf{D_X X}$ and $\mathbf{D_Y Y}$, and their messages to Zorba are respectively the resulting column vectors.

We now define Zorba's decoder. For an arbitrary distribution $p_{X,Y}(x,y)$ over finite alphabets, let the *strongly ϵ-jointly typical set* $A^n_{\epsilon,p_{X,Y}}$ [11] (henceforth simply called the *typical* set) be the set of all length-n sequences $(\mathbf{X},\mathbf{Y})$ such that the empirical distribution induced by $(\mathbf{X},\mathbf{Y})$ differs component-wise from $p_{X,Y}(x,y)$ by at most $\epsilon/(|\mathcal{X}||\mathcal{Y}|)$. That is,

$$A^n_{\epsilon,p_{X,Y}} \triangleq \left\{(\mathbf{x},\mathbf{y}) : \left|\frac{N_{(\mathbf{x},\mathbf{y})}(a,b)}{n} - p_{X,Y}(a,b)\right| < \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|} \text{ for every } (a,b)\in\mathcal{X}\times\mathcal{Y}\right\}$$

where $N_{(\mathbf{x},\mathbf{y})}(a,b)$ denotes the number of component pairs (x_i,y_i) in $(\mathbf{x},\mathbf{y})$ which are equal to (a,b). For simplicity of notation we denote $A^n_{\epsilon,p_{X,Y}}$ as A_ϵ. Zorba checks to see if there exists a unique length-n sequence $(\hat{\mathbf{X}},\hat{\mathbf{Y}})$ satisfying two conditions. First, that $\mathbf{D_X}\hat{\mathbf{X}}$ and $\mathbf{D_Y}\hat{\mathbf{Y}}$ respectively match the messages transmitted by Xavier and Yvonne. Second, whether $(\hat{\mathbf{X}},\hat{\mathbf{Y}})$ lies within A_ϵ. If both conditions are satisfied for exactly one sequence $(\hat{\mathbf{X}},\hat{\mathbf{Y}})$, Zorba outputs $(\hat{\mathbf{X}},\hat{\mathbf{Y}})$, else he declares a decoding error.

Then [7] shows the following result.

Theorem 2: [7] For each rate pair (R_X, R_Y) in the region defined by (1) and sufficiently large n, with high probability over choices of $\mathbf{D_X}$ and $\mathbf{D_Y}$ the corresponding SW code is near-lossless.

Many of the SW codes in the literature build on such encoders that are linear over a finite field. Some such codes use iteratively decodable channel codes to attain performance that is empirically "good", but performance guarantees have not been proven (e.g. [12]). Other codes use recent theoretical advances in channel codes to produce near-lossless codes that achieve any point in the SW rate-region, but cannot give guarantees on computational complexity (e.g. [13]).

C. Linear codes over real fields

As mentioned in the introduction, Compressed Sensing codes operate over real (and complex) fields, and are structurally similar to the codes proposed in this work. The primary difference between the two sets of results is that our focus is on achieving information-theoretically optimal performance (at the cost of potentially high decoding complexity), whereas CS codes have lower decoding complexity at the cost of non-optimal rates. Some intriguing results on CS codes can be found in [14], [8].

Concurrently, codes over the real field $\mathbb{R}$ also seem to have applications for the channel coding problem. Using significantly different techniques, Tao et al. [15] obtained channel codes that can be decoded solving a linear program (LP). Also, *lattice codes* have been shown to achieve capacity for the AWGN channel [16].

III. RSWC Model

As is common in the SW literature [11], we focus on just the point $(H(X), H(Y|X))$ in the SW rate-region. Time-sharing between this and the symmetric point $(H(X|Y), H(Y))$ enables us to achieve all points in the rate-region. Thus Xavier encodes his data $\mathbf{X}$ using a classical lossless source code, and Zorba decodes it losslessly. We henceforth discuss only Yvonne's RSWC encoder for $\mathbf{Y}$ and Zorba's corresponding decoder. In Section VI we show how to generalize our proof techniques to get codes that achieve any point in the SW rate-region without time-sharing. We consider only $\mathcal{X}$ and $\mathcal{Y}$ that are ordered finite subsets of $\mathbb{R}$.

RSWC Encoder: We define an $\mathbb{R}^{m\times n}$ *encoding matrix* $\mathbf{D}$. Here m is a code-design parameter to be specified later, and $\mathbf{D}$ is chosen as follows. Each component D_{ij} of $\mathbf{D}$ is chosen randomly from a finite set $\mathcal{D}$. More precisely, each element of $\mathbf{D}$ is chosen i.i.d. from $\mathcal{D}$ according to a distribution p_D. The set $\mathcal{D}$ can be any arbitrary finite subset of $\mathbb{R}$, and the distribution p_D can be chosen arbitrarily on $\mathcal{D}$, as long as the probability of at least two elements of $\mathcal{D}$ is non-zero. For ease of proof, we assume that p_D is zero-mean – the more general case requires only small changes in the proof details. The particular values of $\mathcal{D}$ and p_D can be chosen according to the application. We denote the i-th row of $\mathbf{D}$ by $\mathbf{D}_i$.

For a fixed block-length n, Yvonne's data is arranged as a column vector $\mathbf{Y} \stackrel{\triangle}{=} (Y_1, Y_2, \cdots, Y_n)^T$. To encode, $\mathbf{Y}$ is multiplied by $\mathbf{D}$ to get a length-m real vector $\mathbf{U} \stackrel{\triangle}{=} \mathbf{DY}$. We denote the real interval $(-n^{0.5+\epsilon}, n^{0.5+\epsilon})$ by I_q. Each component U_i of $\mathbf{U}$ is uniformly quantized by dividing I_q into steps of size $\Delta_n = 2n^{-\epsilon}$. Thus $\lceil (0.5+2\epsilon)\log n \rceil$

bits suffice for this quantization. Note that the values outside the range I_q are quantized to the farthest quantization levels from origin. Here and throughout the paper $\log(.)$ denotes the binary logarithm, and ϵ is a code-design parameter that can be used to trade off between the probability of error and the rate of the RSWC. It can be chosen as any arbitrarily small positive real number. The quantized value of U_i is denoted by $\hat{U}_i$ and the corresponding length-m quantized vector is denoted by $\hat{\mathbf{U}}$. We take $m = \lceil (n(H(Y|X) + 3\epsilon))/(0.5 \log n) \rceil$ since then Yvonne's encoder will encode at about $H(Y|X)$ bits per symbol. Thus the total number of bits Yvonne transmits to Zorba equals $m\lceil (0.5 + 2\epsilon) \log n \rceil$, which for all sufficiently large n can be bounded from above by $nH(Y|X) + \rho\epsilon n$ for a universal constant ρ.

RSWC Decoder: Zorba first decodes $\mathbf{X} = \mathbf{x}$. Suppose he received $\hat{\mathbf{U}} = \hat{\mathbf{u}}$ from Y. He finds a vector $\mathbf{y}$ which is strongly ϵ-jointly typical with $\mathbf{x}$, and for which $\widehat{\mathbf{Dy}} = \hat{\mathbf{u}}$. If there is no such $\mathbf{y}$ or there is more than one such $\mathbf{y}$ he declares a decoding error.

The ensemble of RSWC encoder-decoder pairs described above is denoted by $\mathcal{C}(\epsilon, n, p_{X,Y}, p_D)$. The *probability of error of* $\mathcal{C}(\epsilon, n, p_{X,Y}, p_D)$ is defined as the probability over $p_{X,Y}$ and $p_{\mathbf{D}}$ that Zorba makes or declares a decoding error. The *rate of* $\mathcal{C}(\epsilon, n, p_{X,Y}, p_D)$ is defined as the number of bits that Yvonne transmits to Zorba.

We are now in a position to state and prove our main results. The proofs of these results are presented in the next two sections. Theorem 3 shows that our RSWCs achieve the corner point $(H(X), H(Y|X))$ in the Slepian-Wolf rate-region with exponentially small probability of error.

Theorem 3: For all sufficiently large n there are universal positive constants c, ρ, such that the probability of error under typicality decoding and rate of $\mathcal{C}(\epsilon, n, p_{X,Y}, p_D)$ are at most $2^{-cn/\log n}$ and $H(Y|X) + \rho\epsilon$ respectively.

We next show that Yvonne's decoding can be done by solving an IP.

Theorem 4: If Yvonne's source is binary, then the typicality decoding of a RSWC for the point $(H(X), H(Y|X))$ is equivalent to solving an IP.

Further, we show that even for discrete memoryless sources over larger alphabet $\mathcal{Y}$, the encoder can be implemented as a series of RSWC encoders each of which is for a derived binary source. Then the typicality decoder can be implemented as a series of decoders each of which is equivalent to solving an IP.

Theorem 5: For any finite alphabet $\mathcal{Y}$, the real SW encoding can be done using $|\mathcal{Y}| - 1$ RSWC encoders so that the typicality decoder can be implemented by solving $|\mathcal{Y}| - 1$ IPs.

For any rate-pair in the Slepian-Wolf rate-region, a direct construction of the individual RSWC encoders for Xavier and Yvonne without time-sharing between the corner points is presented in Section VI. It is shown that RSWCs constructed this way also achieve the Slepian-Wolf rate-region.

Theorem 6: Any point in the Slepian-Wolf rate-region can be achieved directly by RSWCs without time-sharing.

We also show that RSWCs can be decoded by *minimum entropy decoding*.

Theorem 7: For all sufficiently large n there are universal positive constants c, ρ, such that the probability of error under minimum entropy decoding and rate of $\mathcal{C}(\epsilon, n, p_{X,Y}, p_D)$ are at most $2^{-cn/\log n}$ and $H(Y|X) + \rho\epsilon$

respectively.

It is argued in Section VIII that the achievable rate-region of the more general class of source networks known as *normal source networks without helpers* [9] is also achieved by our RSWCs.

Theorem 8: Random RSWCs achieve the rate region of any normal source network without helpers.

The above results will be proved in the subsequent sections. In the rest of the paper, for simplicity of exposition many different constants, independent of n, will be denoted by the same symbol "c".

IV. Proof of Theorem 3

The probability of decoding error is given by

$$P_e^n \leq P_1 + P_2 \tag{2}$$

where P_1 is the probability that $(\mathbf{X}, \mathbf{Y})$ are not strongly jointly ϵ-typical, and P_2 is the probability that $(\mathbf{X}, \mathbf{Y}) \in A_\epsilon$, but there is another $\mathbf{y}' \neq \mathbf{y}$ such that $(\mathbf{X}, \mathbf{y}') \in A_\epsilon$, and $\widehat{\mathbf{DY}} = \widehat{\mathbf{Dy}'}$.

Bounding P_1: For P_1, note that for any non-typical sequence $(\mathbf{x}, \mathbf{y})$, its type $p_{(\mathbf{x},\mathbf{y})}$ satisfies $|p_{X,Y} - p_{(\mathbf{x},\mathbf{y})}|_1 \geq \epsilon/|\mathcal{X}||\mathcal{Y}|$. So, using $D(p_{X,Y}||p_{(\mathbf{x},\mathbf{y})}) \geq |p_{X,Y} - p_{(\mathbf{x},\mathbf{y})}|_1^2/(2\ln 2)$ [11, Lemma 12.6.1] and Sanov's theorem [11, Theorem 12.4.1], we have

$$\begin{aligned} P_1 &\leq (n+1)^{|\mathcal{X}||\mathcal{Y}|} \exp\left(-n.\frac{\epsilon^2}{2|\mathcal{X}|^2|\mathcal{Y}|^2}\right) \\ &\leq 2^{-cn} \end{aligned} \tag{3}$$

for some positive constant c. The rest of this section focuses on bounding P_2 in (2).

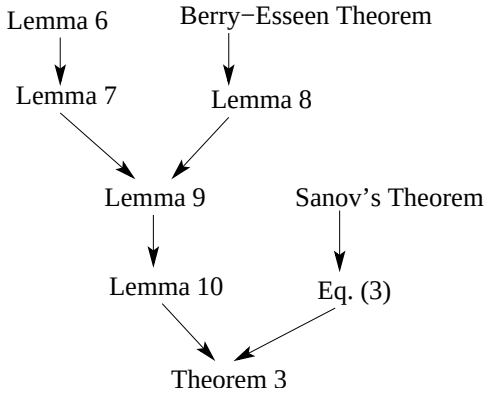

Fig. 1. Dependence structure of Lemmas

Bounding P_2: In the following, we present a sequence of lemmas leading to Lemma 12, which gives a bound on P_2. A dependency "graph" of lemmas is shown in Fig. 1 to ease understanding. We start by a general lemma proved in the Appendix.

Lemma 9: Let $W_1, W_2, \cdots, W_n$ be a sequence of i.i.d. zero-mean random variables taking values from $\mathcal{W}$, and $a \stackrel{\triangle}{=} \max\{|w||w \in \mathcal{W}\}$. Then for any positive constant A,

$$Pr\left\{\left|\sum_{i=1}^{n} W_i\right| > A\right\} \leq 2(n+1)^{|\mathcal{W}|} \exp\left(-\frac{A^2}{2na^2}\right)$$

We now show some properties of our quantization of $U_i = \mathbf{D}_i\mathbf{Y}$.

Lemma 10: There exists a positive constant c so that for any $\mathbf{y} \in \mathcal{Y}^n$,

$$\Pr\{|\mathbf{D}_i\mathbf{y}| > n^{0.5+\epsilon}\} \leq 2^{-cn^{2\epsilon}}.$$

Proof: Let $y_{\max}$ be the element in $\mathcal{Y}$ with maximum absolute value. For any $y \in \mathcal{Y}$, let S_y be the set of indices j such that $y_j = y$, i.e., $S_y \stackrel{\triangle}{=} \{j|y_j = y\}$. If $|\mathbf{D}_i\mathbf{y}| = \left|\sum_{y\in\mathcal{Y}}\left(\sum_{j\in S_y} D_{ij}y_j\right)\right| > n^{0.5+\epsilon}$ then for at least one y, $|\sum_{j\in S_y} D_{ij}y_j| > (1/|\mathcal{Y}|)n^{0.5+\epsilon}$. So,

$$\begin{aligned}
&\Pr\{|\mathbf{D}_i\mathbf{y}| > n^{0.5+\epsilon}\} \\
&\leq \Pr\left\{\left|\sum_{j\in S_y} D_{ij}y_j\right| > \frac{1}{|\mathcal{Y}|}n^{0.5+\epsilon} \text{ for at least one } y\right\} \\
&\leq \sum_{y\in\mathcal{Y}} \Pr\left\{\left|\sum_{j\in S_y} D_{ij}y_j\right| > \frac{1}{|\mathcal{Y}|}n^{0.5+\epsilon}\right\} \\
&= \sum_{y\in\mathcal{Y}} \Pr\left\{\left|\sum_{j\in S_y} D_{ij}\right| > \frac{1}{|\mathcal{Y}||y|}n^{0.5+\epsilon}\right\} \\
&\leq \sum_{y\in\mathcal{Y}} \left\{2(|S_y|+1)^{|\mathcal{D}|} \exp\left(-\frac{1}{2|S_y|\alpha^2}\frac{1}{|\mathcal{Y}|^2|y|^2}n^{1+2\epsilon}\right)\right\} \quad (4) \\
&\leq \sum_{y\in\mathcal{Y}} 2(n+1)^{|\mathcal{D}|} \exp\left(-\frac{n^{1+2\epsilon}}{2n\alpha^2|\mathcal{Y}|^2|y_{\max}|^2}\right) \quad (5) \\
&\leq |\mathcal{Y}|2(n+1)^{|\mathcal{D}|} \exp\left(-\frac{n^{2\epsilon}}{2\alpha^2|\mathcal{Y}|^2|y_{\max}|^2}\right) \\
&\leq 2^{-cn^{2\epsilon}}
\end{aligned}$$

for some constant c, for large enough n, and where $\alpha = \max\{|d||d \in \mathcal{D}\}$. Here (4) follows from Lemma 9, and (5) follows from $|S_y| \leq n$ and $|y_{\max}| \geq |y|\ \forall y \in \mathcal{Y}$. □

The following lemma gives, for two different $\mathbf{y}, \mathbf{y}' \in \mathcal{Y}^n$, an upper bound on the probability that $\widehat{\mathbf{D}_i\mathbf{y}} = \widehat{\mathbf{D}_i\mathbf{y}'}$.

Let $p_\pm$ denote the minimum of $Pr\{D_{ij} > 0\}$ and $Pr\{D_{ij} < 0\}$. Since D_{ij} has zero mean and has at least two symbols with non-zero probability, it follows that $p_\pm \neq 0$.

Lemma 11: If $\mathbf{y} \in \mathcal{Y}^n$ and $\mathbf{y}' \in \mathcal{Y}^n$ differ in t components then

$$Pr\{|\mathbf{D}_i(\mathbf{y} - \mathbf{y}')| < \Delta_n\} \leq \min\left(1 - p_\pm, \frac{c}{\sqrt{t}}\right)$$

for some fixed constant $c \in \mathbb{R}$.

Proof: Let b_y be the *smallest difference in* $\mathcal{Y}$, i.e., $b_y \stackrel{\triangle}{=} \min_{y_1,y_2\in\mathcal{Y},y_1\neq y_2}|y_1-y_2|$. We denote the j-th component $(y_j-y'_j)$ of $\mathbf{y}-\mathbf{y}'$ by α_j. Then there are t nonzero α_j, and w.l.o.g., we assume that $\alpha_1,\alpha_2,\ldots,\alpha_t\neq 0$. Note that $|\{y-y'|y,y'\in\mathcal{Y},y\neq y'\}|\leq|\mathcal{Y}|^2$. So there are at least $\tau\stackrel{\triangle}{=}t/|\mathcal{Y}|^2$ elements among $\alpha_1,\alpha_2,\ldots,\alpha_t$ which are the same. Let us assume, w.l.o.g., that $\alpha_1=\alpha_2=\cdots=\alpha_\tau$. Let σ^2 be the variance of D_{ij}. Then the random variables $V_1=\alpha_1D_{i1},V_2=\alpha_2D_{i2},\ldots,V_\tau=\alpha_\tau D_{i\tau}$ are i.i.d. with zero mean and variance $\sigma'^2=|\alpha_1|^2\sigma^2$. The central limit theorem states that the distribution of the normalized sum $W_\tau=\sum_{j=1}^{\tau}V_j/(\sigma'\sqrt{\tau})$ approaches the normal $\mathcal{N}(0,1)$ distribution as τ increases. The *Berry-Esseen theorem* [17] gives a uniform upper bound on the deviation of the cumulative distribution function (cdf) of W_τ from the cdf of $\mathcal{N}(0,1)$. The Berry-Esseen bound is given by

$$|Pr\{W_\tau<w\}-\Phi(w)|\leq\frac{\beta\gamma}{\sigma'^3\sqrt{\tau}}, \tag{6}$$

for any $w\in\mathbb{R}$. Here $\gamma=E\{|V_1|^3\}$ is the third moment of V_1, and β is a universal constant whose value has been improved over the decades. We use the Berry-Esseen bound to prove the lemma as below.

$$\begin{aligned}
&Pr\{|\mathbf{D}_i(\mathbf{y}-\mathbf{y}')|<\Delta_n\}\\
&\quad=Pr\{-\Delta_n<\mathbf{D}_i(\mathbf{y}-\mathbf{y}')<\Delta_n\}\\
&\quad=Pr\left\{-\frac{\Delta_n}{|\alpha_1|\sigma\sqrt{\tau}}<\frac{\mathbf{D}_i(\mathbf{y}-\mathbf{y}')}{|\alpha_1|\sigma\sqrt{\tau}}<\frac{\Delta_n}{|\alpha_1|\sigma\sqrt{\tau}}\right\}\\
&\quad\leq Pr\left\{-\frac{\Delta_n}{\sigma b_y\sqrt{\tau}}<\frac{\mathbf{D}_i(\mathbf{y}-\mathbf{y}')}{|\alpha_1|\sigma\sqrt{\tau}}<\frac{\Delta_n}{\sigma b_y\sqrt{\tau}}\right\}\\
&\quad=Pr\left\{-\frac{\sum_{j=\tau+1}^{n}D_{ij}(y_j-y'_j)}{|\alpha_1|\sigma\sqrt{\tau}}-\frac{\Delta_n}{\sigma b_y\sqrt{\tau}}<\frac{\sum_{j=1}^{\tau}D_{ij}(y_j-y'_j)}{|\alpha_1|\sigma\sqrt{\tau}}<-\frac{\sum_{j=\tau+1}^{n}D_{ij}(y_j-y'_j)}{|\alpha_1|\sigma\sqrt{\tau}}+\frac{\Delta_n}{\sigma b_y\sqrt{\tau}}\right\}\\
&\quad=Pr\left\{-\frac{\sum_{j=\tau+1}^{n}D_{ij}(y_j-y'_j)}{|\alpha_1|\sigma\sqrt{\tau}}-\frac{\Delta_n}{\sigma b_y\sqrt{\tau}}<W_\tau<-\frac{\sum_{j=\tau+1}^{n}D_{ij}(y_j-y'_j)}{|\alpha_1|\sigma\sqrt{\tau}}+\frac{\Delta_n}{\sigma b_y\sqrt{\tau}}\right\} \qquad (7)\\
&\quad\leq\frac{2\Delta_n}{\sigma b_y\sqrt{\tau}\sqrt{2\pi}}+2\times\frac{\beta\gamma}{\sigma'^3\sqrt{\tau}} \qquad (8)\\
&\quad=\frac{c}{\sqrt{t}}
\end{aligned}$$

Eq. (8) follows by using the Berry-Esseen bound (6) on the normalized sum W_τ. The first term $2\times\frac{1}{\sqrt{2\pi}}\times\frac{\Delta_n}{\sigma b_y\sqrt{\tau}}$ in (8) is an upper bound on the probability of $\mathcal{N}(0,1)$ lying in the interval of length $2\times\frac{\Delta_n}{\sigma b_y\sqrt{\tau}}$ in (7). This bound is obtained by multiplying the maximum value $1/\sqrt{2\pi}$ of the probability density function of $\mathcal{N}(0,1)$ by the length of the interval. The deviation of the cdf of W_τ from that of $\mathcal{N}(0,1)$ at each boundary point of the interval is bounded by the Berry-Esseen bound. The second term in (8) is the sum of this bound at these two boundary points.

For $t>0$, there is at least one j such that $y_j\neq y'_j$. Let us assume, w.l.o.g., that $y_1\neq y'_1$. For large enough n, $\Delta_n<b_y\times\min_{d\in\mathcal{D},d\neq 0}|d|$. So,

$$Pr\{|\mathbf{D}_i(\mathbf{y}-\mathbf{y}')|<\Delta_n\}\leq 1-p_{\pm}.$$

This can be easily checked by considering the change in the value from $\sum_{j=2}^{n}D_{ij}(y_j-y'_j)$ to $\mathbf{D}_i(\mathbf{y}-\mathbf{y}')$. □

Lemma 12: Let $\mathbf{y}$ and $\mathbf{y}'$ be any two vectors differing in t components. Then for some constant c and a constant $\tilde{p} < 1$, both independent of $\mathbf{y}$ and $\mathbf{y}'$, we have

$$\Pr\{\widehat{\mathbf{D}_i\mathbf{y}} = \widehat{\mathbf{D}_i\mathbf{y}'}\} \leq \min\left(\tilde{p}, \frac{c}{\sqrt{t}}\right)$$

for large enough n.

Proof:

$$\begin{aligned}
&\Pr\{\widehat{\mathbf{D}_i\mathbf{y}} = \widehat{\mathbf{D}_i\mathbf{y}'}\} \\
&\quad \leq \Pr\{\widehat{\mathbf{D}_i\mathbf{y}} = \widehat{\mathbf{D}_i\mathbf{y}'} || \mathbf{D}_i\mathbf{y}| \leq n^{0.5+\epsilon}, |\mathbf{D}_i\mathbf{y}'| \leq n^{0.5+\epsilon}\} + \Pr\{|\mathbf{D}_i\mathbf{y}| > n^{0.5+\epsilon}\} + \Pr\{|\mathbf{D}_i\mathbf{y}'| > n^{0.5+\epsilon}\} \\
&\quad \leq \Pr\{|\mathbf{D}_i(\mathbf{y} - \mathbf{y}')| < \Delta_n\} + \Pr\{|\mathbf{D}_i\mathbf{y}| > n^{0.5+\epsilon}\} + \Pr\{|\mathbf{D}_i\mathbf{y}'| > n^{0.5+\epsilon}\} \\
&\quad \leq \min\left(1 - p_{\pm}, \frac{c}{\sqrt{t}}\right) + 2(2^{-cn^{2\epsilon}}) \qquad (9)
\end{aligned}$$

for large enough n. The second term in (9) is obtained by applying Lemma 9 on the last two terms in the previous line. For any constant $c' > c$, we have $\frac{c}{\sqrt{t}} + 2(2^{-cn^{2\epsilon}}) < \frac{c'}{\sqrt{t}}$ for large enough n. Also, for any $\tilde{p} > 1 - p_{\pm}$, $1 - p_{\pm} + 2(2^{-cn^{2\epsilon}}) < \tilde{p}$ for large enough n. So the result follows. □

We are now ready to present an upper bound on P_2.

Lemma 13: For large enough n,

$$P_2 \leq 2^{-cn/\log n}, \qquad (10)$$

where c is a constant.

Proof:

$$\begin{aligned}
P_2 &= \sum_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) Pr\left\{\exists \mathbf{y}' \neq \mathbf{y} \text{ s. t. } \widehat{\mathbf{D}\mathbf{y}'} = \widehat{\mathbf{D}\mathbf{y}}, (\mathbf{x},\mathbf{y}') \in A_\epsilon\right\} \\
&\leq \sum_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{\substack{\mathbf{y}' \neq \mathbf{y} \\ (\mathbf{x},\mathbf{y}') \in A_\epsilon}} Pr\left\{\widehat{\mathbf{D}\mathbf{y}'} = \widehat{\mathbf{D}\mathbf{y}}\right\} \qquad (11) \\
&= \sum_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{t>0} \sum_{\substack{(\mathbf{x},\mathbf{y}') \in A_\epsilon \\ d_H(\mathbf{y},\mathbf{y}')=t}} \left(Pr\{\widehat{\mathbf{D}_1\mathbf{y}'} = \widehat{\mathbf{D}_1\mathbf{y}}\}\right)^m \qquad (12) \\
&\leq \sum_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{t>0} \sum_{\substack{(\mathbf{x},\mathbf{y}') \in A_\epsilon \\ d_H(\mathbf{y},\mathbf{y}')=t}} \left(\min\left(\tilde{p}, \frac{c}{\sqrt{t}}\right)\right)^m \qquad (13) \\
&= \sum_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{t>0} N_{\mathbf{x},\mathbf{y}}(t) \left(\min\left(\tilde{p}, \frac{c}{\sqrt{t}}\right)\right)^m \qquad (14)
\end{aligned}$$

where $N_{\mathbf{x},\mathbf{y}}(t)$ is the number of $\mathbf{y}'$ which are jointly typical with $\mathbf{x}$ and which are at Hamming distance t from $\mathbf{y}$, i.e., $N_{\mathbf{x},\mathbf{y}}(t) \stackrel{\triangle}{=} |\{\mathbf{y}' \in \mathcal{Y}^n | (\mathbf{x},\mathbf{y}') \in A_\epsilon, d_H(\mathbf{y},\mathbf{y}') = t\}|$. Eq. (11) follows by union bound, Eq. (12) follows because the rows of $\mathbf{D}$ are i.i.d., and Eq. (13) follows from Lemma 12. For $t > 0$, let $N(t)$ denote the maximum of $N_{\mathbf{x},\mathbf{y}}(t)$ over all possible typical $(\mathbf{x},\mathbf{y})$ pairs, i.e., $N(t) \stackrel{\triangle}{=} \max_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} N_{\mathbf{x},\mathbf{y}}(t)$. Further, let t_n denote the value of t for which the expression inside the second summation in (14) takes the maximum value for some typical $(\mathbf{x},\mathbf{y})$,

i.e., $t_n \stackrel{\triangle}{=} \arg\max_{t>0}\left(N(t)\left(\min\left(\tilde{p}, c/\sqrt{t}\right)\right)^m\right)$. The subscript in t_n is to emphasize that it is a function of n. Then by substituting in (14),

$$P_2 \leq nN(t_n)\left(\min\left(\tilde{p}, \frac{c}{\sqrt{t_n}}\right)\right)^m.$$

We emphasize here that every appearance of "c" may denote a different constant in the following.

For any $\delta \leq \epsilon/2(H(Y|X)+3\epsilon)$, we consider two regimes: (1) $t_n > n^{1-\delta}$ and (2) $t_n \leq n^{1-\delta}$. In the first regime, we use the bounds $N(t_n) \leq 2^{n(H(Y|X)+2\epsilon)}$ [11, Theorem 14.2.2], $Pr\{\widehat{\mathbf{D}_i\mathbf{y}} \neq \widehat{\mathbf{D}_i\mathbf{y}'}\} \leq c/\sqrt{t}$, and $m = \lceil (n(H(Y|X)+3\epsilon))/(0.5\log n)\rceil$ to get, for large enough n,

$$\begin{aligned}\log(P_2) &\leq \log n + \log N(t_n) - \frac{n(H(Y|X)+3\epsilon)}{0.5\log n}((0.5-0.5\delta)\log n - \log c)\\ &= n(H(Y|X)+2\epsilon) - n(H(Y|X)+3\epsilon)(1-\delta) + n\frac{(H(Y|X)+3\epsilon)}{0.5\log n}c + \log n \qquad (15)\\ &= -n\left(\epsilon - \delta(H(Y|X)+3\epsilon)\right) + n\left(\frac{(H(Y|X)+3\epsilon)}{0.5\log n}c + \frac{\log n}{n}\right).\end{aligned}$$

Now, using $\delta \leq \epsilon/2(H(Y|X)+3\epsilon)$ and $(c(H(Y|X)+3\epsilon)/0.5\log n + (\log n)/n) < \epsilon/4$ for sufficiently large n, we get

$$\begin{aligned}\log(P_2) &\leq -\frac{n\epsilon}{2} + \frac{n\epsilon}{4}\\ &= -\frac{n\epsilon}{4}. \qquad (16)\end{aligned}$$

In the regime $t_n \leq n^{1-\delta}$, we use the bounds $N(t_n) < (|\mathcal{Y}|-1)^{t_n}\binom{n}{t_n} < (|\mathcal{Y}|n)^{t_n}$, and $Pr\{\widehat{\mathbf{D}_i\mathbf{y}} \neq \widehat{\mathbf{D}_i\mathbf{y}'}\} \leq \tilde{p}$ to get

$$\begin{aligned}\log(P_2) &\leq \log n + t_n\log n + t_n\log|\mathcal{Y}| - \frac{n(H(Y|X)+3\epsilon)}{\log n}\log\left(\frac{1}{\tilde{p}}\right)\\ &\leq \log n + n^{1-\delta}\log n + n^{1-\delta}\log|\mathcal{Y}| - \frac{cn}{\log n}, \qquad (17)\end{aligned}$$

where $c = (H(Y|X)+3\epsilon)\log(1/\tilde{p})$. For large enough n, $(\log n)^2 < cn^{(\delta/2)}/3 \Rightarrow \log n < cn^{(\delta/2)}/(3\log(n))$. Also, for large enough n, $n^{-\delta}\log|\mathcal{Y}| < c/(3\log(n))$ for some constant c. So, for some constant c',

$$\begin{aligned}\log(P_2) &\leq \log n - \frac{c'n}{3\log n}\\ &\leq -\frac{cn}{\log n} \qquad (18)\end{aligned}$$

for large enough n and for some constant c.

Since $cn/\log n < n\epsilon/4$ for large enough n, the result follows by combining (16) and (18). □

From (2), (3), and (10), we have, for large enough n,

$$P_e^n \leq P_1 + P_2 \leq 2P_2 \leq 2^{-cn/\log n},$$

for a constant c, thus completing the proof of Theorem 3.

V. Proof of Theorem 4 and Theorem 5

We first show that for $\mathcal{Y} = \{0, 1\}$ the typicality decoding of our scheme can be done via the solution of an IP. Recall that for a vector $\mathbf{y}$, we defined, for any $y \in \mathcal{Y}$, $S_y = \{i | y_i = y\}$. Similarly with abuse of notation, for any vector $\mathbf{x} = (x_1, \ldots, x_n)$ decoded by Zorba, and $x \in \mathcal{X}$, let us define $S_x = \{i | x_i = x\}$. The constraint $(\mathbf{x}, \mathbf{y}) \in A_\epsilon$ can be written as the linear constraints

$$p(1, x) - \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|} \leq \frac{1}{n} \sum_{i \in S_x} y_i \leq p(1, x) + \frac{\epsilon}{|\mathcal{X}||\mathcal{Y}|}, \quad \forall x \in \mathcal{X}$$

Moreover, the constraints $\widehat{\mathbf{Dy}} = \hat{\mathbf{u}} = (\hat{u}_1, \ldots, \hat{u}_n)$ can be written as

$$\hat{u}_i - \Delta_n/2 \leq \mathbf{D}_i \mathbf{y} \leq \hat{u}_i + \Delta_n/2, \quad \forall i = 1, \ldots m.$$

Finally we add the 'integrality' constraints, namely, that $\mathbf{y} \in \mathcal{Y}^n$.

For arbitrary finite alphabets $\mathcal{Y}$, Yvonne and Zorba perform $|\mathcal{Y}| - 1$ encoding and decoding stages, each of which involves IP decoding of a binary vector. A sketch follows.

Let $y^{(1)}, \ldots, y^{(|\mathcal{Y}|)}$ denote the distinct values of $\mathcal{Y}$. In the first stage, instead of encoding $\mathbf{y}$ directly, Yvonne uses $\mathcal{C}(\epsilon, n, p^1_{X,Y}, p_D)$ to encode the vector $f^1(\mathbf{y})$. Here the vector $f^1(\mathbf{y})$ equals 1 in the locations that $\mathbf{y}$ equals $y^{(1)}$ and equals 0 otherwise, and $p^1_{X,Y}$ is the corresponding induced distribution $p_{X,f^1(Y)}$ defined on $\mathcal{X} \times \{0, 1\}$. Since $f^1(\mathbf{y})$ is a binary vector, Zorba can use the IP decoding described above, and therefore can retrieve the locations where $\mathbf{y}$ equals $y^{(1)}$. Inductively, in the ith stage, Yvonne uses $\mathcal{C}(\epsilon, n(i), p^i_{X,Y}, p_D)$ to encode the vector $f^i(\mathbf{y})$. Here $n(i)$ equals the number of locations whose values are still undetermined before the i^{th} stage, i.e., $n(i)$ equals $|\{j | \mathbf{y}_j \geq y^{(i)}\}|$. The length-$n(i)$ vector $f^i(\mathbf{y})$ is obtained by first throwing away the locations in $f^{i-1}(\mathbf{y})$ that equalled 1, and then marking the remaining locations 1 if and only if the corresponding locations in $\mathbf{y}$ equal $y^{(i)}$. At each stage, Zorba can use the IP decoding described above, and therefore can retrieve the locations where $\mathbf{y}$ equals $y^{(i)}$. Let $f^i(Y)$ denote the corresponding binary random variable s. t. $(X, f^i(Y))$ has the joint distribution given by $p^i_{X,Y}(x, 1) = Pr\{X = x, Y = y^{(i)} | Y \neq y^{(1)}, y^{(2)}, \ldots, y^{(i-1)}\}$, and $p^i_{X,Y}(x, 0) = Pr\{X = x, Y \neq y^{(i)} | Y \neq y^{(1)}, y^{(2)}, \ldots, y^{(i-1)}\}$. Then by a direct extension of the grouping axiom [18, Page 8], we have

$$\begin{aligned} H(Y|X) &= H(f^1(Y)|X) + (1 - p_Y(y^{(1)}))H(f^1(Y)|X) + (1 - p_Y(y^{(1)}) - p_Y(y^{(2)}))H(f^2(Y)|X) + \ldots \\ &\quad + (p_Y(y^{(|\mathcal{Y}|-1)}) + p_Y(y^{(|\mathcal{Y}|)}))H(f^{|\mathcal{Y}|-1}(Y)|X). \end{aligned} \tag{19}$$

Clearly, for a single stage encoding/decoding, the average codelength for Yvonne is bounded by $nH(Y|X) + c\epsilon n$. For a multi-stage encoding/decoding as described above, for a typical $\mathbf{y}$, the block length at the i-th stage is bounded by $n(i) \leq n(1 - Pr\{Y \in \{y^{(1)}, y^{(2)}, \ldots, y^{(i-1)}\}\} + \epsilon)$ and so the codelength is bounded as

$$L_i \leq n(1 - Pr\{Y \in \{y^{(1)}, y^{(2)}, \ldots, y^{(i-1)}\}\} + \epsilon)H(f^i(Y)|X) + c_i \epsilon n$$

for some constants c_i. The average codelength is thus bounded using (19) by

$$L \leq \sum_{i=1}^{|\mathcal{Y}|-1} L_i \leq nH(Y|X) + c\epsilon n \tag{20}$$

for some constant c. If $\mathbf{y}$ is not typical, then in the worst case, the codelength $n(i) = n$ for each i. Then the overall codelength is bounded by $L \leq c'n$ for some constant c'. Since the probability of the non-typical set is exponentially small, the overall average codelength is still bounded by (20) for some constant c. Hence the overall rate of this multistage RSWC differs from $H(Y|X)$ by at most $c\epsilon$, where c is some constant dependent only on $p_{X,Y}$.

The overall probability of error can be bounded as

$$P_e^n \leq P_1 + \sum_{i=1}^{|\mathcal{Y}|} P_{2,i}, \tag{21}$$

where P_1 is the probability that the vector $\mathbf{y}$ is not strongly typical, and $P_{2,i}$ is the conditional probability of error at the i-th stage of decoding given that the vector $\mathbf{y}$ is strongly ϵ-typical and the decoding till the $(i-1)$-th stage is correct. If $\mathbf{y}$ is strongly ϵ-typical, then the codelength at the i-th stage is $n(i) \geq n \times \sum_{j=i}^{|\mathcal{Y}|}(P_Y(y^{(i)}) - \epsilon/|\mathcal{Y}|) \geq n(P_Y(y^{(|\mathcal{Y}|)}) - \epsilon/|\mathcal{Y}|)$. So,

$$\begin{aligned} P_{2,i} &\leq \exp\left(-\frac{c'n(i)}{\log(n(i))}\right) \\ &\leq \exp\left(-\frac{c'n(i)}{\log n}\right) \\ &\leq \exp\left(-\frac{c'(P_Y(|\mathcal{Y}|) - \epsilon/|\mathcal{Y}|)n}{\log n}\right). \end{aligned}$$

Since P_1 is also exponentially small, the overall probability of error for the multistage encoding/decoding is bounded as

$$P_e^n \leq \exp\left(-\frac{cn}{\log n}\right).$$

□

VI. Real SW coding without timesharing

Any rate-pair in the SW rate-region can also be directly achieved by RSWCs without timesharing between the schemes achieving the rate-pairs $(H(X|Y), H(Y))$ and $(H(X), H(Y|X))$. Let (R_1, R_2) be a rate-pair in the SW rate-region. Let $m_1 = \lceil (n(R_1 + 3\epsilon))/(0.5 \log n) \rceil$ and $m_2 = \lceil (n(R_2 + 3\epsilon))/(0.5 \log n) \rceil$. Similar to the encoding scheme of Yvonne described in Section III, Xavier chooses an $m_1 \times n$ encoder matrix $\mathbf{D}_1$ over $\mathcal{D}$ according to a distribution P_D. Similarly Yvonne chooses a random $m_2 \times n$ encoder matrix $\mathbf{D}_2$ over $\mathcal{D}$ according to the distribution P_D [1]. Xavier encodes the length-n vector $\mathbf{X}$ by quantizing each component of $\mathbf{U}_1 \stackrel{\triangle}{=} \mathbf{D}_1\mathbf{X}$ uniformly in the range I_q with step-size $\Delta_n = 2n^{-\epsilon}$ to obtain the vector $\widehat{\mathbf{U}}_1$. Similarly, Yvonne encodes the length-n vector $\mathbf{Y}$ by quantizing each component of $\mathbf{U}_2 \stackrel{\triangle}{=} \mathbf{D}_2\mathbf{Y}$ uniformly in the range I_q with step size $\Delta_n = 2n^{-\epsilon}$ to obtain the vector $\widehat{\mathbf{U}}_2$.

[1] Our arguments go through even if the elements of $\mathbf{D}_1$ and $\mathbf{D}_2$ are chosen from different sets $\mathcal{D}_1$ and $\mathcal{D}_2$ according to some distributions. We restrict to $\mathcal{D}_1 = \mathcal{D}_2$ and the same distribution for the elements of $\mathbf{D}_1$ and $\mathbf{D}_2$ for simplicity.

Zorba finds a unique jointly strongly ϵ-typical pair $(\mathbf{x}, \mathbf{y})$ so that $\widehat{\mathbf{D}_1\mathbf{x}} = \widehat{\mathbf{U}}_1$ and $\widehat{\mathbf{D}_2\mathbf{y}} = \widehat{\mathbf{U}}_2$. If there is no such pair, or if there are more than one such pair, then the decoder declares an error. The probability of error can be bounded as

$$P_e^n \leq P_1 + P_{21} + P_{22} + P_{23}, \tag{22}$$

where P_1, as before, is the probability that $(\mathbf{X}, \mathbf{Y})$ is not jointly strongly ϵ-typical, P_{21} is the probability that there is a $\mathbf{x}' \neq \mathbf{X}$ which is also jointly strongly ϵ-typical with $\mathbf{Y}$ and $\widehat{\mathbf{D}_1\mathbf{x}'} = \widehat{\mathbf{U}}_1$, P_{22} is the probability that there is a $\mathbf{y}' \neq \mathbf{Y}$ which is also jointly strongly ϵ-typical with $\mathbf{X}$ and $\widehat{\mathbf{D}_2\mathbf{y}'} = \widehat{\mathbf{U}}_2$, and P_{23} is the probability that there is another jointly typical pair $(\mathbf{x}', \mathbf{y}')$ so that $\mathbf{x}' \neq \mathbf{X}, \mathbf{y}' \neq \mathbf{Y}, \widehat{\mathbf{D}_1\mathbf{x}'} = \widehat{\mathbf{U}}_1$ and $\widehat{\mathbf{D}_2\mathbf{y}'} = \widehat{\mathbf{U}}_2$. We now investigate all the terms in (22).

Let $\mathbf{D}_{1,i}$ and $\mathbf{D}_{2,i}$ denote the i-th rows of the matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ respectively. Similarly as Lemma 12, we have

$$Pr\{\widehat{\mathbf{D}_{1,i}\mathbf{x}} = \widehat{\mathbf{D}_{1,i}\mathbf{x}'}\}, Pr\{\widehat{\mathbf{D}_{2,i}\mathbf{y}} = \widehat{\mathbf{D}_{2,i}\mathbf{y}'}\} \quad \leq \quad \min\left(\tilde{p}, \frac{c_1}{\sqrt{t}}\right),$$

when each pair $\mathbf{x}, \mathbf{x}' \in \mathcal{X}^n$ and $\mathbf{y}, \mathbf{y}' \in \mathcal{Y}^n$ differ in t positions.

We define the following functions.

$$\begin{aligned}
m(R) &\stackrel{\triangle}{=} \left\lceil \frac{n(R+3\epsilon)}{0.5\log n} \right\rceil, \\
\phi_1(h, R, \delta) &\stackrel{\triangle}{=} \log\left(n2^{n(h+2\epsilon)} \left(\frac{c}{\sqrt{n^{1-\delta}}} \right)^{m(R)} \right), \text{ and} \\
\phi_2(L, R, \delta) &\stackrel{\triangle}{=} \log\left(n(Ln)^{n^{1-\delta}} (\tilde{p})^{m(R)} \right).
\end{aligned}$$

Note that in this notation, P_2 in Lemma 13 is given by

$$\log(P_2) \leq \phi_1(H(Y|X), H(Y|X), \delta) \tag{23}$$

for $t_n > n^{1-\delta}$ (See (15)). As shown in (16), this is at most $-n\epsilon/4$ for $\delta \leq \epsilon/2(H(Y|X) + 3\epsilon)$ for large enough n. It can be checked similarly that for $\delta \leq \epsilon/2(R+3\epsilon)$, $R \geq h$, and large enough n, $\phi_1(h, R, \delta) \leq -n((R-h)+\epsilon/4)$. Likewise, for $t_n \leq n^{1-\delta}$, it is shown (See (17)) that

$$\log(P_2) \leq \phi_2(|\mathcal{Y}|, H(Y|X), \delta), \tag{24}$$

which is at most $-cn/\log n$ (See (18)). More generally, it can be similarly proved that for any constants $L > 0$ and $\delta > 0$,

$$\phi_2(L, R, \delta) \leq -\frac{c(R, \epsilon)n}{\log n}$$

for some constant $c(R, \epsilon) > 0$ and for large enough n.

By definition,

$$P_{22} \quad = \quad \sum_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} p_{X,Y}(\mathbf{x}, \mathbf{y}) Pr\left\{ \exists \mathbf{y}' \neq \mathbf{y} \text{ s. t. } \widehat{\mathbf{D}\mathbf{y}'} = \widehat{\mathbf{D}\mathbf{y}}, (\mathbf{x}, \mathbf{y}') \in A_\epsilon \right\}.$$

By similar arguments to those in the proof of Lemma 13, we have $\log(P_{22}) \leq \phi_2(|\mathcal{Y}|, R_2, \delta)$ for $t_n \leq n^{1-\delta}$, and $\log(P_{22}) \leq \phi_1(H(Y|X), R_2, \delta)$ for $t_n > n^{1-\delta}$. Since $R_2 \geq H(Y|X)$, it follows that for large enough n,

$$\log(P_{22}) \leq -\frac{c(R_2,\epsilon)n}{\log n}. \tag{25}$$

Similarly, for large enough n,

$$\log(P_{21}) \leq -\frac{c(R_1,\epsilon)n}{\log n}. \tag{26}$$

As in the proof of Lemma 13, P_{23} can be simplified to (27) below,

$$\begin{aligned}
P_{23} &= \sum_{(\mathbf{x},\mathbf{y})\in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) Pr\left\{\exists(\mathbf{x}',\mathbf{y}') \text{ s. t. } \mathbf{x}'\neq\mathbf{x}, \mathbf{y}'\neq\mathbf{y}, \widehat{\mathbf{D}_1\mathbf{x}'} = \widehat{\mathbf{D}_1\mathbf{x}}, \widehat{\mathbf{D}_2\mathbf{y}'} = \widehat{\mathbf{D}_2\mathbf{y}}, (\mathbf{x}',\mathbf{y}') \in A_\epsilon\right\} \\
&\leq \sum_{(\mathbf{x},\mathbf{y})\in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{\substack{\mathbf{x}'\neq\mathbf{x},\mathbf{y}'\neq\mathbf{y}\\(\mathbf{x}',\mathbf{y}')\in A_\epsilon}} Pr\left\{\widehat{\mathbf{D}_1\mathbf{x}'} = \widehat{\mathbf{D}_1\mathbf{x}},\ \widehat{\mathbf{D}_2\mathbf{y}'} = \widehat{\mathbf{D}_2\mathbf{y}}\right\} \\
&= \sum_{(\mathbf{x},\mathbf{y})\in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{t_1>0,\, t_2>0} \sum_{\substack{(\mathbf{x}',\mathbf{y}')\in A_\epsilon\\ d_H(\mathbf{x},\mathbf{x}')=t_1,\, d_H(\mathbf{y},\mathbf{y}')=t_2}} Pr\left\{\widehat{\mathbf{D}_1\mathbf{x}'} = \widehat{\mathbf{D}_1\mathbf{x}},\ \widehat{\mathbf{D}_2\mathbf{y}'} = \widehat{\mathbf{D}_2\mathbf{y}}\right\} \\
&= \sum_{(\mathbf{x},\mathbf{y})\in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{t_1>0,\, t_2>0} \sum_{\substack{(\mathbf{x}',\mathbf{y}')\in A_\epsilon\\ d_H(\mathbf{x},\mathbf{x}')=t_1,\, d_H(\mathbf{y},\mathbf{y}')=t_2}} \left(Pr\{\widehat{\mathbf{D}_{1,1}\mathbf{x}'} = \widehat{\mathbf{D}_{1,1}\mathbf{x}}\}\right)^{m(R_1)} \left(Pr\{\widehat{\mathbf{D}_{2,1}\mathbf{y}'} = \widehat{\mathbf{D}_{2,1}\mathbf{y}}\}\right)^{m(R_2)} \\
&\leq \sum_{(\mathbf{x},\mathbf{y})\in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{t_1>0,\, t_2>0} \sum_{\substack{(\mathbf{x}',\mathbf{y}')\in A_\epsilon\\ d_H(\mathbf{x},\mathbf{x}')=t_1,\, d_H(\mathbf{y},\mathbf{y}')=t_2}} \left(\min\left(\tilde{p}, \frac{c}{\sqrt{t_1}}\right)\right)^{m(R_1)} \left(\min\left(\tilde{p}, \frac{c}{\sqrt{t_2}}\right)\right)^{m(R_2)} \\
&= \sum_{(\mathbf{x},\mathbf{y})\in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{t_1,\, t_2>0} N_{\mathbf{x},\mathbf{y}}(t_1,t_2) Q_1^{m(R_1)} Q_2^{m(R_2)}.
\end{aligned} \tag{27}$$

In (27), $Q_1 = \min\left(\tilde{p}, c/\sqrt{t_1}\right)$, $Q_2 = \min\left(\tilde{p}, c/\sqrt{t_2}\right)$, and $N_{\mathbf{x},\mathbf{y}}(t_1,t_2)$ is the number of jointly typical $(\mathbf{x}',\mathbf{y}')$ pairs such that $\mathbf{x}'$ differs from $\mathbf{x}$ at t_1 locations and $\mathbf{y}'$ differs from $\mathbf{y}$ at t_2 locations, that is, $N_{\mathbf{x},\mathbf{y}}(t_1,t_2) \stackrel{\triangle}{=} |\{(\mathbf{x}',\mathbf{y}') \in \mathcal{X}^n \times \mathcal{Y}^n | (\mathbf{x}',\mathbf{y}') \in A_\epsilon, d_H(\mathbf{x},\mathbf{x}') = t_1, d_H(\mathbf{y},\mathbf{y}') = t_2\}|$. We define $N(t_1,t_2) \stackrel{\triangle}{=} \max_{\mathbf{x},\mathbf{y}} N_{(\mathbf{x},\mathbf{y})\in A_\epsilon}(t_1,t_2)$, and $(t_{1,n}, t_{2,n})$ as the pair (t_1,t_2) that maximizes $(N(t_1,t_2)Q_1^{m_1}Q_2^{m_2})$, that is, $(t_{1,n},t_{2,n}) \stackrel{\triangle}{=} \arg\max_{t_1,t_2>0}\left(N(t_1,t_2)Q_1^{m_1}Q_2^{m_2}\right)$. Then

$$P_{23} \leq n^2 N(t_{1,n},t_{2,n}) Q_1^{m_1} Q_2^{m_2}.$$

For $\delta < \epsilon/2(R_1+R_2+3\epsilon)$, we consider four cases.

Case I: $t_{1,n} > n^{1-\delta}, t_{2,n} > n^{1-\delta}$. In this case, using the bounds $N(t_{1,n},t_{2,n}) \leq 2^{n(H(X,Y)+\epsilon)}, Q_1 \leq c/\sqrt{t_{1,n}}, Q_2 \leq c/\sqrt{t_{2,n}}$, we have

$$\begin{aligned}
\log(P_{23}) &\leq \phi_1(H(X,Y), R_1+R_2, \delta) \\
&\leq -n(R_1+R_2-H(X,Y)+\epsilon/4) \\
&\leq -n\epsilon/4.
\end{aligned} \tag{28}$$

Case II: $t_{1,n} \leq n^{1-\delta}, t_{2,n} \leq n^{1-\delta}$. In this case, using the bounds $N(t_{1,n}, t_{2,n}) \leq (|\mathcal{X}|n)^{t_{1,n}} (|\mathcal{Y}|n)^{t_{2,n}}, Q_1 \leq \tilde{p}, Q_2 \leq \tilde{p}$, we have

$$\begin{aligned} \log(P_{23}) &\leq \phi_2(|\mathcal{X}|, R_1, \delta) + \phi_2(|\mathcal{Y}|, R_2, \delta) \\ &\leq -\frac{c(R_1,\epsilon)n}{\log n} - \frac{c(R_2,\epsilon)n}{\log n} \\ &\leq -\frac{c(R_1,R_2,\epsilon)n}{\log n}, \end{aligned} \tag{29}$$

where $c(R_1, R_2, \epsilon) = c(R_1, \epsilon) + c(R_2, \epsilon)$.

Case III: $t_{1,n} > n^{1-\delta}, t_{2,n} \leq n^{1-\delta}$. In this case, using the bounds $N(t_{1,n}, t_{2,n}) \leq 2^{n(H(X|Y)+2\epsilon)} (|\mathcal{Y}|n)^{t_{2,n}}, Q_1 \leq c/\sqrt{t_{1,n}}, Q_2 \leq \tilde{p}$, we have

$$\begin{aligned} \log(P_{23}) &\leq \phi_1(H(X|Y), R_1, \delta) + \phi_2(|\mathcal{Y}|, R_2, \delta) \\ &\leq -n(R_1 - H(X|Y) + \epsilon/4) - \frac{c(R_2,\epsilon)n}{\log n} \\ &\leq -\frac{c(R_1,R_2,\epsilon)n}{\log n}. \end{aligned} \tag{30}$$

Case IV: $t_{1,n} \leq n^{1-\delta}, t_{2,n} > n^{1-\delta}$: As in Case III, we have

$$\log(P_{23}) \leq -\frac{c(R_1,R_2,\epsilon)n}{\log n}. \tag{31}$$

From (3), (22), (25), (26), (28), (29), (30), and (31), we have,

$$P_e^n \leq 2^{-cn/\log n}$$

for some constant c.

VII. Universal decoding: Proof of Theorem 7

An encoding or decoding operation is said to be universal in a class of sources if the encoding/decoding operation can be chosen without the knowledge of the exact source statistics in the class. The encoding for RSWCs without time-sharing in Section VI results in universal encoding in the class of i.i.d. sources. The two encoders may choose to encode at rates R_1 and R_2 and choose their encoding matrices randomly without the knowledge of the distribution of either source. The joint typicality decoding discussed earlier will be able to recover both the sequences with exponentially small probability of error as long as the rate pair (R_1, R_2) lies in the Slepian-Wolf rate region of the sources. However, though the encoders are universal, the joint typicality decoding is not universal since it requires the decoder to know the joint distribution of the sources.

In this section, we show that the well known universal minimum entropy decoding (MED) [9] which does not need the joint distribution of the sources will also be able to decode our code with exponentially small probability of error provided $m_1 \geq \lceil n(R_1 + 4\epsilon)/(0.5 \log n) \rceil$ and $m_2 \geq \lceil n(R_2 + 4\epsilon)/(0.5 \log n) \rceil$ for some (R_1, R_2) in the Slepian-Wolf rate-region of the sources. Here, the decoder finds the pair $(\mathbf{x}, \mathbf{y})$ with minimum empirical entropy which satisfies the conditions $\widehat{\mathbf{D}_1\mathbf{x}} = \widehat{\mathbf{U}}_1$ and $\widehat{\mathbf{D}_2\mathbf{y}} = \widehat{\mathbf{U}}_2$. If there are more than one such pair then the decoder declares a decoding error.

Before investigating the probability of error under minimum entropy decoding, let us define a weakly ϵ-typical vector $(\mathbf{x}, \mathbf{y})$ as one satisfying

$$\left|\log_2(p^n_{X,Y}(\mathbf{x},\mathbf{y})) + nH(X,Y)\right| \leq n\epsilon,$$
$$|\log_2(p^n_X(\mathbf{x})) + nH(X)| \leq n\epsilon, \text{ and}$$
$$|\log_2(p^n_Y(\mathbf{y})) + nH(Y)| \leq n\epsilon.$$

The set of weakly ϵ-typical vectors will be denoted by $A_{\epsilon,weak}$. A weakly ϵ-typical vector $\mathbf{x}$ (similarly $\mathbf{y}$) is defined as one satisfying

$$|\log_2(p^n_X(\mathbf{x})) + nH(X)| \leq n\epsilon.$$

The properties of the weakly typical set may be found in [11].

Let us denote the joint entropy of the *type* of a pair of vectors $(\mathbf{x}, \mathbf{y})$ as $H(\mathbf{x}, \mathbf{y})$, the corresponding conditional entropies as $H(\mathbf{x}|\mathbf{y})$ and $H(\mathbf{y}|\mathbf{x})$, and the individual entropies of the vectors as $H(\mathbf{x})$ and $H(\mathbf{y})$. The probability of error of a minimum entropy decoder is bounded as

$$P^n_e(MED) \leq P'_1 + P'_{21} + P'_{22} + P'_{23} \tag{32}$$

where P'_1 is the probability that $(\mathbf{X}, \mathbf{Y})$ is not jointly weakly ϵ-typical, P_{21} is the probability that there is a $\mathbf{x}' \neq \mathbf{X}$ such that $H(\mathbf{x}', \mathbf{Y}) \leq H(\mathbf{X}, \mathbf{Y})$ and $\widehat{\mathbf{D}_1\mathbf{x}'} = \widehat{\mathbf{U}}_1$, P_{22} is the probability that there is a $\mathbf{y}' \neq \mathbf{Y}$ such that $H(\mathbf{X}, \mathbf{y}') \leq H(\mathbf{X}, \mathbf{Y})$ and $\widehat{\mathbf{D}_2\mathbf{y}'} = \widehat{\mathbf{U}}_2$, and P_{23} is the probability that there is another pair $(\mathbf{x}', \mathbf{y}')$ so that $\mathbf{x}' \neq \mathbf{X}, \mathbf{y}' \neq \mathbf{Y}, \widehat{\mathbf{D}_1\mathbf{x}'} = \widehat{\mathbf{U}}_1$, $\widehat{\mathbf{D}_2\mathbf{y}'} = \widehat{\mathbf{U}}_2$ and $H(\mathbf{x}', \mathbf{y}') \leq H(\mathbf{X}, \mathbf{Y})$. We will briefly discuss all the terms in (32).

By definition, $P'_1 = Pr\{A^c_{\epsilon,weak}\}$. Since the weakly ϵ-typical set is a superset of the strongly $\epsilon'(\epsilon, p_{X,Y})$-typical set for some $\epsilon'(\epsilon, p_{X,Y})$ [19], P'_1 can be bounded similar to (3) as

$$P'_1 \quad \leq \quad 2^{-cn} \tag{33}$$

where the constant c depends on $p_{X,Y}$.

Following similar steps as the proof of Lemma 13, we have

$$P_{22} \quad = \quad \sum_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} p_{X,Y}(\mathbf{x},\mathbf{y}) \sum_{t>0} N'_{\mathbf{x},\mathbf{y}}(t) \left(\min\left(\tilde{p}, \frac{c}{\sqrt{t}}\right)\right)^{m_2}$$

where $N'_{\mathbf{x},\mathbf{y}}(t) \stackrel{\triangle}{=} |\{\mathbf{y}' \in \mathcal{Y}^n | H(\mathbf{y}'|\mathbf{x}) \leq H(\mathbf{y}|\mathbf{x}), d_H(\mathbf{y}, \mathbf{y}') = t\}|$. Now, let us define $N'(t) \stackrel{\triangle}{=} \max_{(\mathbf{x},\mathbf{y}) \in A_\epsilon} N'_{\mathbf{x},\mathbf{y}}(t)$ for $t > 0$, and $t_n \stackrel{\triangle}{=} \arg\max_{t>0} \left(N'(t)\left(\min\left(\tilde{p}, c/\sqrt{t}\right)\right)^{m_2}\right)$. Then clearly,

$$P_{22} \quad \leq \quad nN'(t_n) \left(\min\left(\tilde{p}, \frac{c}{\sqrt{t_n}}\right)\right)^{m_2}.$$

Note that for a given weakly typical $\mathbf{x}$, the condition $(\mathbf{x}, \mathbf{y}) \in A_{\epsilon,weak}$ implies $H(\mathbf{y}|\mathbf{x}) \leq H(Y|X) + 2\epsilon$. So, $N'_{\mathbf{x},\mathbf{y}}(t) \subseteq |\{\mathbf{y}' \in \mathcal{Y}^n | H(\mathbf{y}'|\mathbf{x}) \leq H(Y|X) + 2\epsilon, d_H(\mathbf{y}, \mathbf{y}') = t\}|$. So, we can use both the bounds $N'(t_n) \leq 2^{n(H(Y|X)+3\epsilon)}$ and $N'(t_n) \leq (|\mathcal{Y}|n)^{t_n}$ for large enough n. Then it can be shown in the same way as in the proof

of Lemma 13 that $P_{22} \leq \exp(-cn/\log n)$ for $m_2 \geq \lceil n(R_2+4\epsilon)/(0.5\log n)\rceil$. Similarly it can be shown that $P_{21}, P_{23} \leq \exp(-cn/\log n)$ for large enough n if $m_1 \geq \lceil n(R_1+4\epsilon)/(0.5\log n)\rceil$ and $m_2 \geq \lceil n(R_2+4\epsilon)/(0.5\log n)\rceil$ for a rate pair (R_1, R_2) in the Slepian-Wolf rate-region. Since P_1' goes to zero exponentially as in (33), it follows that

$$P_e^n(MED) \quad \leq \quad \exp(-cn/\log n)$$

for large enough n for some constant c.

VIII. Generalization to other source networks: Proof of Theorem 8

The most simple generalization of the Slepian-Wolf source network is to multiple sources as shown in Fig. 2. The same proof technique can be used to show that the decoder can recover all the sources with exponentially small probability of error if the encoders do random real encoding at rates satisfying

$$\sum_{i\in\mathcal{L}} R_i \geq H(X_{\mathcal{L}}|X_{\mathcal{L}^c})$$

for each $\mathcal{L} \subseteq \{1, 2, \cdots, k\}$. Here $\mathcal{L}^c$ denotes the complement of $\mathcal{L}$. Using the same proof technique as outlined in Sec. VII, one can show that the decoder can also do minimum entropy decoding to attain vanishing probability of error.

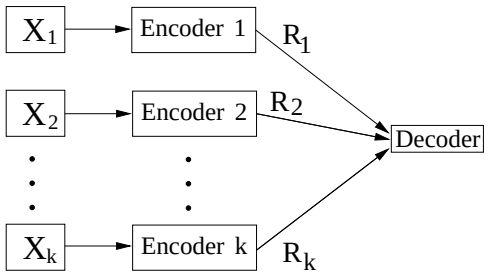

Fig. 2. A simple multi-source network

Csiszar and Korner [9] extended the result of Slepian and Wolf to more general source networks called *normal source networks (NSN) without helpers*. In the following, we briefly discuss their source network and argue that our coding technique can achieve the achievable rate-region of NSN without helpers.

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ denote the set of sources, encoders and decoders respectively in the network. For any $c \in \mathcal{C}$, let $\mathcal{S}_c$ denote the set of source nodes from which information is received at the decoder node c. Let $\mathcal{D}_c$ denote the set of sources which are to be reproduced at c.

An NSN, as defined in [9] and an example of which is shown in Fig. 3, is a source network where

(i) there are no direct edges from the sources to the decoders,

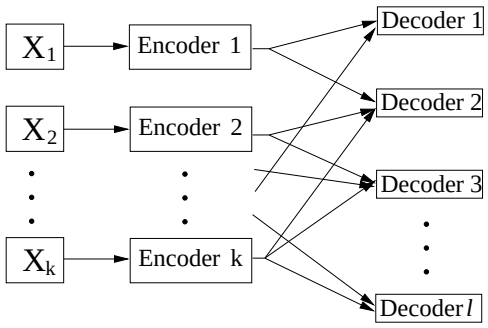

Fig. 3. Normal Source Network

(ii) $|\mathcal{A}| = |\mathcal{B}|$ and the edges from $\mathcal{A}$ to $\mathcal{B}$ define a one-to-one correspondence between the sources and encoders,

(iii) all the sets $\mathcal{S}_c, c \in \mathcal{C}$ are different, and

(iv) for each pair of output vertices c' and c'', the inclusion $\mathcal{S}_{c'} \subseteq \mathcal{S}_{c''}$ implies $\mathcal{D}_{c'} \subseteq \mathcal{D}_{c''}$.

For a source $a \in \mathcal{A}$, let X_a denote the i.i.d. data generated by the source. Similarly, for a subset $\mathcal{L} \subseteq \mathcal{A}$, let $X_{\mathcal{L}}$ denote the vector $(X_a)_{a\in\mathcal{L}}$. A source a in an NSN is called a helper if for some $c \in \mathcal{C}$, $a \in \mathcal{S}_c \setminus \mathcal{D}_c$. Clearly, a source network without helpers satisfy $\mathcal{S}_c = \mathcal{D}_c$ for all $c \in \mathcal{C}$. For any encoder $b \in \mathcal{B}$, let R_b denote its encoding rate. For a source network without helpers, Csiszar and Korner characterized the rate-region.

Theorem 14: [9] The achievable rate-region of an NSN without helpers equals the set of those vectors $\bar{\mathbf{R}} = \{R_b\}_{b\in\mathcal{B}}$ which satisfy the inequalities

$$\sum_{b\in\mathcal{L}} R_b \geq H\left(X_{\mathcal{L}}|X_{\mathcal{S}_c\setminus\mathcal{L}}\right) \tag{34}$$

for every output $c \in \mathcal{C}$ and set $\mathcal{L} \subset \mathcal{S}_c$.

The achievability proof of this rate-region reduces to the achievability proof of the corresponding rate-region for each of the networks obtained by taking all the sources and one decoder. In other words, if the encoders encode at rates satisfying the conditions in Theorem 14, the probability of error for each decoder is negligible. So the proof reduces to the proof for the multiple source network as shown in Fig. 2. It thus follows that the rate-region of any NSN without helpers is achievable by random real encoding at each encoder. Moreover, the rate-region is also achievable with minimum entropy decoders.

IX. Conclusion

The Real Slepian-Wolf Codes analyzed here provide a novel achievability proof of the Slepian-Wolf theorem. Perhaps just as importantly, they demonstrate the intriguing possibility of design of information-theoretic codes via convex optimization techniques. For instance, since decoding RSWCs is equivalent to solving an optimization problem, it is natural to consider similar "real" codes for problems where some function of the code simultaneously needs to be optimized. We are currently investigating the performance of RSWCs under more structured choices of encoding matrices, with the hope of obtaining codes for which IP decoding is equivalent to LP decoding, and is therefore computationally tractable.

APPENDIX

PROOF OF LEMMA 9

First consider $Pr\left\{\sum_{i=1}^{n} W_i > A\right\}$. We define $E \stackrel{\triangle}{=} \{(w_1, w_2, \cdots, w_n) | \sum_{i=1}^{n} w_i > A\}$. Let p_w denote the probability mass distribution of W_i. Then

$$\begin{aligned} Pr\left\{\sum_{i=1}^{n} W_i > A\right\} &= Pr\{E\} \\ &= Pr\left\{p_n | \mu_{p_n} > \frac{A}{n}\right\}. \end{aligned}$$

Here p_n denotes the type of $(w_1, w_2, \cdots, w_n)$ and μ_{p_n} denotes the mean of p_n. By Sanov's Theorem [11, Theorem 12.4.1], we have

$$Pr\left\{\sum_{i=1}^{n} W_i > A\right\} = p_w^n(E) \leq (n+1)^{|\mathcal{W}|} 2^{-nD(p_n^* || p_w)},$$

where $p_n^* = \arg\min_{p_n : \mu_{p_n} > A/n} D(p_n || p_w)$. Since p_w has zero mean, the "nearest" distribution to p_w that has mean greater than A/n in absolute value would differ from p_w in the largest absolute component by at least $A/(an)$. So, $\mu_{p_n^*} > A/n$ implies $|p_n^* - p_w|_1 > A/(an)$. We then have $D(p_n^* || p_w) \geq (1/2\ln 2)|p_n^* - p_w|_1^2 > A^2/(2(na)^2 \ln 2)$ by [11, Lemma 12.6.1]. So,

$$Pr\left\{\sum_{i=1}^{n} W_i > A\right\} \leq (n+1)^{|\mathcal{W}|} \exp\left(-\frac{A^2}{2na^2}\right).$$

Similarly one can show that $Pr\left\{\sum_{i=1}^{n} W_i < -A\right\} \leq (n+1)^{|\mathcal{W}|} \exp\left(-A^2/(2na^2)\right)$. So the result follows. □

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the CUHK direct grant, the CU-MS-JL grant, and a grant from the Bharti Centre for Communication. We would like to thank S. Shenvi for his interest and involvement in several stages of this work. We would also like to thank D. Manjunath for fruitful discussions.

REFERENCES

[1] S. Shenvi, B. K. Dey, S. Jaggi, and M. Langberg, ""Real" slepian-wolf codes," in *IEEE International Symposium on Information Theory (ISIT)*, (Toronto, Canada), July 2008.

[2] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, pp. 471–480, July 1973.

[3] S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, pp. 51–60, March 2002.

[4] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Transactions on Information Theory*, vol. 46, pp. 344–366, Mar. 2000.

[5] R. Puri and K. Ramchandran, "Prism: a new robust video coding architecture based on distributed compression principles," in *Proceedings of the Allerton Conference on Communications, Control, and Computing*, October 2002.

[6] A. J. Hoffmann, "The role of unimodularity in applying linear inequalities to combinatorial theorems," *Annals of Discrete Mathematics*, vol. 4, pp. 73–84, 1979.

[7] I. Csiszar, "Linear codes for sources and source networks: error exponents, universal coding," *IEEE Transactions on Information Theory*, vol. 28, no. 4, pp. 585–592, 1982.

[8] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, February 2006.

[9] I. Csiszar and J. Korner, “Towards a general theory of source networks,” *IEEE Transactions on Information Theory*, vol. 26, no. 2, pp. 155–165, 1980.

[10] C. E. Shannon, “A mathematical theory of communication,” *Bell Systems Technical Journal*, vol. 27, pp. 379–423,623–656, 1948.

[11] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.

[12] J. Garcia-Frias and Y. Zhao, “Compression of correlated binary sources using turbo codes,” *IEEE Communication Letters*, pp. 417–419, October 2001.

[13] T. P. Coleman, A. H. Lee, M. Médard, and M. Effros, “On some new approaches to practical slepian-wolf compression inspired by channel coding,” in *Proceedings of the Conference on Data Compression*, p. 282, March 2004.

[14] D. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, April 2006.

[15] E. Candès and T. Tao, “Decoding by linear programming,” *IEEE Transactions on Information Theory*, vol. 51, pp. 4203–4215, December 2005.

[16] R. Urbanke and B. Rimoldi, “Lattice codes can achieve capacity on the AWGN channel,” *IEEE Transactions on Information Theory*, vol. 44, no. 1, pp. 273–278, 1998.

[17] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume II (2nd ed.)*. New York: John Wiley & Sons, 1972.

[18] R. B. Ash, *Information Theory*. New York: Dover Publications, Inc., 1965.

[19] R. W. Yeung, *Information Theory and Network Coding*. Available at http://www.springerlink.com/content/978-0-387-79233-0: Springer.