The Dispersion of Slepian-Wolf Coding

Vincent Y. F. Tan*†, Oliver Kosut‡

* Institute for Infocomm Research, A*STAR (Email: tanyfv@i2r.a-star.edu.sg)
† Department of Electrical and Computer Engineering, National University of Singapore
‡ Stochastic Systems Group (SSG), Massachusetts Institute of Technology, (Email: okosut@mit.edu)

***Abstract*—We characterize second-order coding rates (or dispersions) for distributed lossless source coding (the Slepian-Wolf problem). We introduce a fundamental quantity known as the entropy dispersion matrix, which is analogous to scalar dispersion quantities. We show that if this matrix is positive-definite, the optimal rate region under the constraint of a fixed blocklength and non-zero error probability has a curved boundary compared to being polyhedral for the Slepian-Wolf case. In addition, the entropy dispersion matrix governs the rate of convergence of the non-asymptotic region to the asymptotic one. As a by-product of our analyses, we develop a general universal achievability procedure for dispersion analysis of some other network information theory problems such as the multiple-access channel. Numerical examples show how the region given by Gaussian approximations compares to the Slepian-Wolf region.**

***Index Terms*—Slepian-Wolf, Dispersion, Second-order Rates**

I. Introduction

Distributed lossless source coding consists in *separately* encoding two (or more) correlated sources $(X_1^n, X_2^n) \sim \prod_{k=1}^n p_{X_1,X_2}(x_{1k}, x_{2k})$ into a pair of rate-limited messages (M_1, M_2). Subsequently, given these compressed versions of the sources, a decoder seeks to reconstruct (X_1^n, X_2^n). One of the most remarkable results in information theory, proved by Slepian and Wolf [1], states that the set of achievable rate pairs (R_1, R_2) is equal to that when each of the encoders is given knowledge of the other source, i.e., encoder 1 knows X_2^n and vice versa. The optimal rate region $\mathscr{R}^*$ is the polyhedron

$$\begin{aligned} R_1 &\geq H(X_1|X_2) \\ R_2 &\geq H(X_2|X_1) \\ R_1 + R_2 &\geq H(X_1, X_2). \end{aligned} \tag{1}$$

As with most other statements in information theory, this result is asymptotic in nature. In this paper, we take a step towards non-asymptotic results by analyzing the second-order coding rates of the *Slepian-Wolf* (SW) problem.

An two-sender SW code is characterized by four parameters; the *blocklength* n, the *rates* of the first and second sources (R_1, R_2) and the *probability of error* defined as

$$P_{\mathrm{e}}^{(n)} := \mathsf{P}((\hat{X}_1^n, \hat{X}_2^n) \neq (X_1^n, X_2^n)), \tag{2}$$

where $\hat{X}_1^n$ and $\hat{X}_2^n$ are the reconstructed versions of X_1^n and X_2^n respectively. Traditionally, we require $P_{\mathrm{e}}^{(n)} \to 0$ as $n \to \infty$. In this paper, we fix n and require the code to be such that $P_{\mathrm{e}}^{(n)} \leq \epsilon$. We then ask what the set of achievable pairs of rates as a function of (n, ϵ) is. The main tool that we use is a multidimensional version of the Berry-Essèen theorem [2].

V. Tan is supported by A*STAR, Singapore. O. Kosut is supported by Shell Global Solutions International B.V.

A. Main Contributions

This paper characterizes the (n, ϵ)-optimal rate region for the SW problem $\mathscr{R}^*(n, \epsilon)$ up to an $O(\frac{\log n}{n})$ factor. In the course of doing so, we introduce a fundamental quantity called the *entropy dispersion matrix* of p_{X_1,X_2} and show that if this matrix is non-singular, the boundary of $\mathscr{R}^*(n, \epsilon)$ is, unlike that of SW, a *smooth* curve. We also demonstrate numerically how our region compares to the SW region and to the problem of finite blocklength source coding with side information also at the encoder. While the SW problem is the focus of this paper, our achievability technique is general enough to be applicable to multi-terminal channel coding problems such as the multiple-access, broadcast and interference channels. The results for these other problems are not included this paper. The interested reader may refer to [3] for more details.

B. Related Work

The redundancy of SW coding was discussed in [4]–[6]. However, the authors considered a single source X_1 to be compressed and side information X_2 available only at the decoder. Thus, X_2 is neither coded nor estimated. They showed that a scalar dispersion quantity governs the second-order coding rate. He et al. [5] also analyzed a variable-length variant of the SW problem and showed that the dispersion is smaller than in the fixed-length setting. This dispersion is similar to that for channel coding. Sarvotham et al. [7] considered the SW problem with two sources to be compressed but limited their setting to the case the sources are symmetric. This work generalizes their setting in that we consider all discrete sources. This paper is a network information theory analogue of the works on second-order coding rates [8], [9] and finite blocklength analysis [10]–[13]. We employ the information spectrum method [14] in our converse proof. This was also done in [9].

II. Problem Statement and Main Results

A. Notation

Random variables and the values they take on will be denoted by upper case (e.g., X) and lower case (e.g., x) respectively. Types (empirical distributions) will be denoted by upper case (e.g., P) and distributions by lower case (e.g., p). For a sequence $x^n \in \mathcal{X}^n$, the type is denoted as P_{x^n} and conditional types are denoted similarly. The entropy

and conditional entropy are denoted as $H(X_1) = H(p_{X_1})$ and $H(X_2|X_1) = H(p_{X_2|X_1}|p_{X_1})$ respectively. For a pair of sequences x_1^n, x_2^n, the notations $\hat{H}(x_1^n) := H(P_{x_1^n})$ and $\hat{H}(x_2^n|x_1^n) := H(P_{x_2^n|x_1^n}|P_{x_1^n})$ denote, respectively, the empirical marginal and conditional entropies. For two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$, the notation $\mathbf{u} \leq \mathbf{v}$ means $u_t \leq v_t$ for all $t = 1, \ldots, d$. We also use the notation $[2^{nR}] := \{1, \ldots, \lceil 2^{nR} \rceil\}$.

B. Definitions

Let $(\mathcal{X}_1, \mathcal{X}_2, p_{X_1,X_2}(x_1, x_2))$ be a discrete memoryless multiple source (DMMS). This means that $(X_1^n, X_2^n) \sim \prod_{k=1}^n p_{X_1,X_2}(x_{1k}, x_{2k})$. The alphabets $\mathcal{X}_1, \mathcal{X}_2$ are finite.

Definition 1. *An* $(n, 2^{nR_1}, 2^{nR_2}, \epsilon)$-SW code *consists of two encoders* $f_{j,n} : \mathcal{X}_j^n \to \mathcal{M}_j := [2^{nR_j}], j = 1, 2$*, and a decoder* $\varphi_n : \mathcal{M}_1 \times \mathcal{M}_2 \to \mathcal{X}_1^n \times \mathcal{X}_2^n$ *such that the the error probability in* (2) *with* $(\hat{X}_1^n, \hat{X}_2^n) := \varphi_n(f_{1,n}(X_1^n), f_{2,n}(X_2^n))$ *does not exceed* ϵ*. The* rates *are defined as* $R_j := \frac{1}{n} \log |\mathcal{M}_j|$.

Definition 2. *A rate pair* (R_1, R_2) *is* (n, ϵ)-achievable *if there exists an* $(n, 2^{nR_1}, 2^{nR_2}, \epsilon)$*-SW code for the DMMS* $p_{X_1,X_2}(x_1, x_2)$*. The* (n, ϵ)-optimal rate region $\mathscr{R}^*(n, \epsilon) \subset \mathbb{R}^2$ *is the set of all* (n, ϵ)*-achievable rate pairs.*

For a positive-semidefinite symmetric matrix $\mathbf{V} \succeq 0$, let the random vector $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$. Define the set

$$\mathscr{S}(\mathbf{V}, \epsilon) := \{\mathbf{z} \in \mathbb{R}^3 : \mathsf{P}(\mathbf{Z} \leq \mathbf{z}) \geq 1 - \epsilon\}. \tag{3}$$

Note that $\mathscr{S}(\mathbf{V}, \epsilon) \subset \mathbb{R}^3$ and is analogous to the cumulative distribution function of a zero-mean Gaussian with covariance matrix $\mathbf{V}$. If $\epsilon \leq \frac{1}{2}$, $\mathscr{S}(\mathbf{V}, \epsilon)$ is a convex, unbounded set in the positive orthant. The boundary of $\mathscr{S}(\mathbf{V}, \epsilon)$ is a differentiable manifold if $\mathbf{V}$ is positive-definite ($\mathbf{V} \succ 0$).

Definition 3. *The* entropy density vector *is defined as*

$$\mathbf{h}(X_1, X_2) := \begin{bmatrix} -\log p_{X_1|X_2}(X_1|X_2) \\ -\log p_{X_2|X_1}(X_2|X_1) \\ -\log p_{X_1,X_2}(X_1, X_2) \end{bmatrix}. \tag{4}$$

The mean of the entropy density vector is $\mathsf{E}[\mathbf{h}(X_1, X_2)] = \mathbf{H}(p_{X_1,X_2}) := [H(X_1|X_2), H(X_2|X_1), H(X_1, X_2)]^T$.

Definition 4. *The* entropy dispersion matrix $\mathbf{V}(p_{X_1,X_2})$ *is the covariance of the random vector* $\mathbf{h}(X_1, X_2)$.

We abbreviate the deterministic quantities $\mathbf{H}(p_{X_1,X_2})$ and $\mathbf{V}(p_{X_1,X_2})$ as $\mathbf{H}$ and $\mathbf{V}$ respectively. Observe that $\mathbf{V}$ is an analogue of the scalar dispersion quantities that have gained attention in recent years [10]–[13]. We will find it convenient to define the *rate vector* $\mathbf{R} := [R_1, R_2, R_1 + R_2]^T \in \mathbb{R}^3$.

Definition 5. *Define the region* $\mathscr{R}_{\mathrm{in}}(n, \epsilon) \subset \mathbb{R}^2$ *to be the set of rate pairs* (R_1, R_2) *that satisfy*

$$\mathbf{R} \in \mathbf{H} + \frac{1}{\sqrt{n}} \mathscr{S}(\mathbf{V}, \epsilon) + \frac{\nu \log n}{n} \mathbf{1}, \tag{5}$$

where $\nu := |\mathcal{X}_1||\mathcal{X}_2| + 1$ *and* $\mathbf{1} := (1, 1, 1)^T$*. Also let* $\mathscr{R}_{\mathrm{out}}(n, \epsilon) \subset \mathbb{R}^2$ *be the set of rate pairs* (R_1, R_2) *that satisfy*

$$\mathbf{R} \in \mathbf{H} + \frac{1}{\sqrt{n}} \mathscr{S}(\mathbf{V}, \epsilon) - \frac{\log n}{n} \mathbf{1}. \tag{6}$$

An illustration is provided in Fig. 1. Henceforth, $\epsilon \in (0, 1)$.

C. Main Result and Interpretation

Theorem 1. *The* (n, ϵ)*-optimal rate region* $\mathscr{R}^*(n, \epsilon)$ *satisfies*

$$\mathscr{R}_{\mathrm{in}}(n, \epsilon) \subset \mathscr{R}^*(n, \epsilon) \subset \mathscr{R}_{\mathrm{out}}(n, \epsilon). \tag{7}$$

for all n *sufficiently large.*

This theorem is proved for $\mathbf{V} \succ 0$ in Section III. Sources for which $\mathbf{V}$ is singular include those which are (i) independent, i.e., $I(X_1; X_2) = 0$, (ii) either X_1 or X_2 is uniform over their alphabets. The authors in [7] dealt with the specific case where $X_1, X_2 \in \mathbb{F}_2$, $X_1 = \mathrm{Bern}(\frac{1}{2})$, $X_2 = X_1 \oplus N$ with $N = \mathrm{Bern}(q), q \in (0, \frac{1}{2})$, i.e., a discrete symmetric binary source (DSBS). In Section IV, we comment on how the proof can be adapted to derive $\mathscr{R}^*(n, \epsilon)$ for a DSBS and all $\mathbf{V} \succeq 0$.

The direct part of Theorem 1 is proved using the usual random binning argument together with a multidimensional Berry-Essèen theorem [2]. The converse is proved using an information spectrum technique by Han [14]. Theorem 1 extends to the case where there are more than two senders.

By examining $\mathscr{R}_{\mathrm{in}}(n, \epsilon)$ and $\mathscr{R}_{\mathrm{out}}(n, \epsilon)$, it can be seen that we have characterized the (n, ϵ)-rate region up to an $O(\frac{\log n}{n})$ factor. This residual is a consequence of (i) universal decoding for the direct part and (ii) approximations resulting from using the multidimensional Berry-Essèen theorem [2]. Observe that as $n \to \infty$, the (n, ϵ)-rate region approaches the SW region [1] at a rate of $O(\frac{1}{\sqrt{n}})$. This follows from the multidimensional central limit theorem. However, somewhat unexpectedly, if $\mathbf{V} \succ 0$, the (n, ϵ)-rate region is not-polyhedral [cf. (1)]. Its boundary is a smooth curve in $\mathbb{R}^2$. This curvature, given by $\mathbf{V}$, is due to the fact that the three empirical entropies $\hat{H}(X_1^n|X_2^n)$, $\hat{H}(X_2^n|X_1^n)$ and $\hat{H}(X_1^n, X_2^n)$ have to be *jointly* smaller than some rate vector. By Taylor's theorem, we see that the empirical entropy vector behaves like a multivariate Gaussian with mean $\mathbf{H}$ and covariance $\mathbf{V}$.

III. Proof of Theorem 1

A. Achievability (Inner Bound)

Proof: Let (R_1, R_2) be a rate pair such that $\mathbf{R}$ belongs to the inner bound $\mathscr{R}_{\mathrm{in}}(n, \epsilon)$, defined in (5).

Codebook Generation: For $j = 1, 2$, randomly and independently assign an index $f_{1,n}(x_j^n) \in [2^{nR_j}]$ to each sequence $x_j^n \in \mathcal{X}_j^n$ according to a uniform pmf. The sequences of the same index form a *bin*, i.e., $\mathcal{B}_j(m_j) := \{x_j^n \in \mathcal{X}_j^n : f_{1,n}(x_j^n) = m_j\}$. Note that $\mathcal{B}_j(m_j), m_j \in [2^{nR_j}]$ are random sets. The bin assignments are revealed to all parties. In particular, the decoder knows the bin rates R_j.

Encoding: Given $x_j^n \in \mathcal{X}_j^n$, encoder j transmits the bin index $f_{j,n}(x_j^n)$. Hence, for length-n sequence, the rates of m_1 and m_2 are R_1 and R_2 respectively.

Decoding: The decoder, upon receipt of the bin indices (m_1, m_2) finds the unique sequence pair $(\hat{x}_1^n, \hat{x}_2^n) \in \mathcal{B}_1(m_1) \times \mathcal{B}_2(m_2)$ such that the *empirical entropy vector*

$$\hat{\mathbf{H}}(\hat{x}_1^n, \hat{x}_2^n) := \begin{bmatrix} \hat{H}(\hat{x}_1^n|\hat{x}_2^n) \\ \hat{H}(\hat{x}_2^n|\hat{x}_1^n) \\ \hat{H}(\hat{x}_1^n, \hat{x}_2^n) \end{bmatrix} \leq \mathbf{R} - \delta_n \mathbf{1}, \tag{8}$$

where $\delta_n := (|\mathcal{X}_1||\mathcal{X}_2| + \frac{1}{2})\frac{\log(n+1)}{n}$. Define the *empirical entropy typical set* $\mathscr{T}(\mathbf{R}, \delta_n) := \{\mathbf{z} \in \mathbb{R}^3 : \mathbf{z} \leq \mathbf{R} - \delta_n \mathbf{1}\}$. Then, (8) is equivalent to $\hat{\mathbf{H}}(\hat{x}_1^n, \hat{x}_2^n) \in \mathscr{T}(\mathbf{R}, \delta_n)$. If there is more than one pair or no such pair in $\mathcal{B}_1(m_1) \times \mathcal{B}_2(m_2)$, declare a decoding error. Note that our decoding scheme is *universal* [15], i.e., the decoder does not depend on knowledge of the true distribution p_{X_1,X_2}.

Analysis of error probability: Let the sequences sent by the two users be (X_1^n, X_2^n) and let their corresponding bin indices be (M_1, M_2). We bound the probability of error averaged over the random code construction. Clearly, the ensemble probability of error is bounded above by the sum of the probabilities of the following four events:

$$\begin{aligned}
\mathcal{E}_1 &:= \{\hat{\mathbf{H}}(X_1^n, X_2^n) \notin \mathscr{T}(\mathbf{R}, \delta_n)\} \\
\mathcal{E}_2 &:= \{\exists\, \tilde{x}_1^n \in \mathcal{B}_1(M_1) \setminus \{X_1^n\} : \hat{\mathbf{H}}(\tilde{x}_1^n, X_2^n) \in \mathscr{T}(\mathbf{R}, \delta_n)\} \\
\mathcal{E}_3 &:= \{\exists\, \tilde{x}_2^n \in \mathcal{B}_2(M_2) \setminus \{X_2^n\} : \hat{\mathbf{H}}(X_1^n, \tilde{x}_2^n) \in \mathscr{T}(\mathbf{R}, \delta_n)\} \\
\mathcal{E}_4 &:= \{\exists\, \tilde{x}_1^n \in \mathcal{B}_1(M_1) \setminus \{X_1^n\}, \tilde{x}_2^n \in \mathcal{B}_2(M_2) \setminus \{X_2^n\} : \\
&\qquad \hat{\mathbf{H}}(\tilde{x}_1^n, \tilde{x}_2^n) \in \mathscr{T}(\mathbf{R}, \delta_n)\} \qquad (9)
\end{aligned}$$

We bound each of these in turn. Consider

$$\mathsf{P}(\mathcal{E}_1) = 1 - \mathsf{P}(\hat{\mathbf{H}}(P_{X_1^n,X_2^n}) \in \mathscr{T}(\mathbf{R}, \delta_n)) \qquad (10)$$

where we made the dependence of the empirical entropy vector on the type explicit. We now bound the probability in (10). Let $\mathrm{vec}(p_{X_1,X_2}) \in \mathbb{R}_+^{|\mathcal{X}_1||\mathcal{X}_2|}$ be a vectorized version of the joint distribution p_{X_1,X_2}. Consider the Taylor series expansion:

$$\hat{\mathbf{H}}(P_{X_1^n,X_2^n}) = \mathbf{H}(p_{X_1,X_2}) + \mathbf{J}(\mathrm{vec}(P_{X_1^n,X_2^n} - p_{X_1,X_2})) + \mathbf{\Delta}. \qquad (11)$$

where the Jacobian $\mathbf{J} \in \mathbb{R}^{3\times(|\mathcal{X}_1||\mathcal{X}_2|)}$ is defined entry-wise as

$$[\mathbf{J}]_{t,(x_1,x_2)} = \frac{\partial g_t(p_{X_1,X_2})}{\partial p_{X_1,X_2}(x_1,x_2)}\Big|_{p_{X_1,X_2}(x_1,x_2)}, \qquad (12)$$

where $g_1(p_{X_1,X_2}) := H(X_1|X_2)$, $g_2(p_{X_1,X_2}) := H(X_2|X_1)$ and $g_3(p_{X_1,X_2}) := H(X_1, X_2)$. Because the g_t's are twice continuously differentiable, each entry of the second order correction term $\mathbf{\Delta} \in \mathbb{R}^3$ in (11) is bounded above by $C\|\mathrm{vec}(P_{X_1^n,X_2^n} - p_{X_1,X_2})\|^2$ for some constant $C > 0$. Let $[\mathbf{J}]_t$ be the t-th row of the matrix $\mathbf{J}$. Now, note that

$$\begin{aligned}
[\mathbf{J}]_t\, \mathrm{vec}(P_{X_1^n,X_2^n}) &= \sum_{x_1,x_2} P_{X_1^n,X_2^n}(x_1,x_2)[\mathbf{J}]_{t,(x_1,x_2)} \\
&= \frac{1}{n}\sum_{k=1}^{n} [\mathbf{J}]_{t,(X_{1k},X_{2k})} \qquad (13)
\end{aligned}$$

because the joint type $P_{X_1^n,X_2^n}$ places a probability mass $1/n$ on each sample (X_{1k}, X_{2k}). Define the random vector $\mathbf{J}_k := ([\mathbf{J}]_{1,(X_{1k},X_{2k})}, [\mathbf{J}]_{2,(X_{1k},X_{2k})}, [\mathbf{J}]_{3,(X_{1k},X_{2k})})^T$. On account of (10), (11) and (13), we have

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_1^c) &\overset{(a)}{=} \mathsf{P}\left[\mathbf{H} + \frac{1}{n}\sum_{k=1}^{n}(\mathbf{J}_k - \mathsf{E}[\mathbf{J}_k]) + \mathbf{\Delta} \leq \mathbf{R} - \delta_n\mathbf{1}\right] \\
&\overset{(b)}{\geq} \mathsf{P}\left[\mathbf{H} + \frac{1}{n}\sum_{k=1}^{n}(\mathbf{J}_k - \mathsf{E}[\mathbf{J}_k]) \leq \mathbf{R} - (\delta_n + c_n)\mathbf{1}\right] \\
&\qquad - \mathsf{P}[\|\mathbf{\Delta}\|_\infty \geq c_n]. \qquad (14)
\end{aligned}$$

where (a) follows from the definition $\mathscr{T}(\mathbf{R}, \delta_n)$ and (b) follows from the probability relation

$$\mathsf{P}(\mathbf{W} + \mathbf{\Delta} \leq \mathbf{R}') \geq \mathsf{P}(\mathbf{W} \leq \mathbf{R}' - c_n\mathbf{1}) - \mathsf{P}(\|\mathbf{\Delta}\|_\infty \geq c_n).$$

As is shown in [3], $\mathsf{P}(\|\mathbf{\Delta}\|_\infty \geq c_n) \leq 1/n^2$ if $c_n = O(1/n)$. With this choice of c_n,

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_1^c) \geq \mathsf{P}\Bigg[&\frac{1}{n}\sum_{k=1}^{n}(\mathbf{J}_k - \mathsf{E}[\mathbf{J}_k]) \leq \\
&\frac{\mathbf{z}}{\sqrt{n}} + \frac{\nu\log n}{n}\mathbf{1} - (\delta_n + c_n)\mathbf{1}\Bigg] - \frac{1}{n^2} \qquad (15)
\end{aligned}$$

because $\mathbf{R} - \mathbf{H} = \frac{\mathbf{z}}{\sqrt{n}} + \frac{\nu\log n}{n}\mathbf{1}$ for some $\mathbf{z}$ such that $\mathsf{P}(\mathbf{Z} \leq \mathbf{z}) \geq 1-\epsilon$ for $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ [cf. definition of $\mathscr{S}(\mathbf{V}, \epsilon)$]. Since $\nu > |\mathcal{X}_1||\mathcal{X}_2| + 1/2$ (the coefficient of δ_n), we have

$$\mathsf{P}(\mathcal{E}_1^c) \geq \mathsf{P}\left[\frac{1}{n}\sum_{k=1}^{n}(\mathbf{J}_k - \mathsf{E}[\mathbf{J}_k]) \leq \frac{\mathbf{z}}{\sqrt{n}} + \psi_n\mathbf{1}\right] - \frac{1}{n^2} \qquad (16)$$

where $\psi_n = O(\frac{\log n}{n})$. Now note that the summands above are i.i.d. random vectors. These random vectors have zero mean, covariance matrix $\mathbf{V} \succ 0$ and finite *third moment* $\xi := \mathsf{E}\|\mathbf{h}(X_1, X_2)\|_2^3$ because $\mathcal{X}_1, \mathcal{X}_2$ are finite sets. Since the set integrated over in (16) is convex, by the multidimensional Berry-Essèen theorem [2] (dimension $d = 3$),

$$\begin{aligned}
\mathsf{P}(\mathcal{E}_1^c) &\geq \mathsf{P}\left(\mathbf{Z} \leq \mathbf{z} + \psi_n\mathbf{1}\right) - \frac{400d^{1/4}\xi}{\lambda_{\min}(\mathbf{V})^{3/2}\sqrt{n}} - \frac{1}{n^2} \\
&\overset{(a)}{\geq} 1 - \epsilon + O\left(\psi_n\right) - \frac{530\xi}{\lambda_{\min}(\mathbf{V})^{3/2}\sqrt{n}} - \frac{1}{n^2}, \qquad (17)
\end{aligned}$$

where (a) follows from Taylor's approximation theorem. Because $\psi_n = O(\frac{\log n}{\sqrt{n}})$ dominates the $O(\frac{1}{\sqrt{n}})$ term resulting from the Berry-Essèen approximation, we conclude that

$$\mathsf{P}(\mathcal{E}_1) \leq \epsilon - O\left(\frac{\log n}{\sqrt{n}}\right). \qquad (18)$$

For the second event, by symmetry and uniformity, $\mathsf{P}(\mathcal{E}_2) = \mathsf{P}(\mathcal{E}_2|X_1^n \in \mathcal{B}_1(1))$. Now consider the chain of inequalities:

$$\begin{aligned}
&\mathsf{P}(\mathcal{E}_2|X_1^n \in \mathcal{B}_1(1)) \\
&\overset{(a)}{=} \sum_{x_1^n,x_2^n} p(x_1^n, x_2^n)\mathsf{P}\Big[\exists\, \tilde{x}_1^n \in \mathcal{B}_1(1) \setminus \{X_1^n\} : \\
&\quad \hat{\mathbf{H}}(\tilde{x}_1^n, x_2^n) \in \mathscr{T}(\mathbf{R}, \delta_n)\Big|(X_1^n, X_2^n) = (x_1^n, x_2^n), X_1^n \in \mathcal{B}_1(1)\Big] \\
&\overset{(b)}{\leq} \sum_{x_1^n,x_2^n} p(x_1^n, x_2^n) \sum_{\tilde{x}_1^n \neq x_1^n : \hat{\mathbf{H}}(\tilde{x}_1^n, x_2^n) \in \mathscr{T}(\mathbf{R}, \delta_n)} \mathsf{P}\left(\tilde{x}_1^n \in \mathcal{B}_1(1)\right) \\
&\overset{(c)}{\leq} \sum_{x_1^n,x_2^n} p(x_1^n, x_2^n) \sum_{\tilde{x}_1^n \neq x_1^n : \hat{H}(\tilde{x}_1^n|x_2^n) \leq R_1 - \delta_n} \mathsf{P}\left(\tilde{x}_1^n \in \mathcal{B}_1(1)\right) \\
&\overset{(d)}{=} \sum_{x_1^n,x_2^n} p(x_1^n, x_2^n) \sum_{\tilde{x}_1^n \neq x_1^n : \hat{H}(\tilde{x}_1^n|x_2^n) \leq R_1 - \delta_n} \frac{1}{\lceil 2^{nR_1}\rceil} \\
&\overset{(e)}{\leq} \sum_{Q} \sum_{(x_1^n,x_2^n) \in \mathcal{T}_Q} p(x_1^n, x_2^n) \sum_{\substack{V \in \mathscr{V}(Q_{\bar{X}_2}): \\ H(V|P_{x_2^n}) \leq R_1 - \delta_n}} \sum_{\tilde{x}_1^n \in \mathcal{T}_V(x_2^n)} 2^{-nR_1}
\end{aligned}$$

$$\overset{(f)}{\leq} \sum_{Q} \sum_{(x_1^n,x_2^n)\in\mathcal{T}_Q} p(x_1^n,x_2^n) \sum_{\substack{V\in\mathscr{V}(Q_{\bar{X}_2}):\\ H(V|P_{x_2^n})\leq R_1-\delta_n}} 2^{nH(V|P_{x_2^n})}2^{-nR_1}$$

$$\overset{(g)}{\leq} \sum_{x_1^n,x_2^n} p(x_1^n,x_2^n)(n+1)^{|\mathcal{X}_1||\mathcal{X}_2|}2^{n(R_1-\delta_n)}2^{-nR_1} \tag{19}$$

where (a) follows because for $\tilde{x}_1^n \neq x_1^n$, the events $\{x_1^n \in \mathcal{B}_1(1)\}$, $\{\tilde{x}_1^n \in \mathcal{B}_1(1)\}$ and $\{(X_1^n,X_2^n)=(x_1^n,x_2^n)\}$ are mutually independent, (b) follows by the union bound and (c) follows from $\{\tilde{x}_1^n : \hat{\mathbf{H}}(\tilde{x}_1^n,x_2^n)\in\mathscr{T}(\mathbf{R},\delta_n)\} \subset \{\tilde{x}_1^n : \hat{H}(\tilde{x}_1^n|x_2^n) \leq R_1-\delta_n\}$. Equality (d) follows from the uniformity in the random binning. In (e), we partitioned the sum over (x_1^n,x_2^n) into type classes indexed by $Q=Q_{\bar{X}_1,\bar{X}_2}$ and $\tilde{x}_1^n\in\mathcal{X}_1^n$ into sums over stochastic matrices $V:\mathcal{X}_2\to\mathcal{X}_1$ for which the V-shell of a sequence of type $Q_{\bar{X}_2}$ in $\mathcal{X}_2^n$ is not empty (denoted as $V\in\mathscr{V}(Q_{\bar{X}_2})$). In (f) we upper bounded the cardinality of the V-shell as $|\mathcal{T}_V(x_2^n)|\leq 2^{nH(V|P_{x_2^n})}$ [15, Lem. 1.2.5]. In (g), we used the Type Counting Lemma [15, Eq. (2.5.1)]. By using the definition of δ_n, (19) gives $\mathsf{P}(\mathcal{E}_2)\leq n^{-1/2}$. Similarly $\mathsf{P}(\mathcal{E}_3)\leq n^{-1/2}$ and $\mathsf{P}(\mathcal{E}_4)\leq n^{-1/2}$.

Combining this with (18), the error probability averaged over the random binning is $\mathsf{P}(\mathcal{E})\leq\epsilon$. Hence, there is a deterministic code whose error probability is no greater than ϵ if the rate pair (R_1,R_2) belongs to $\mathscr{R}_{\text{in}}(n,\epsilon)$. ∎

B. Converse (Outer Bound)

Proof: For the outer bound, [14, Lemma 7.2.2] asserts that every $(n,2^{nR_1},2^{nR_2},P_{\text{e}}^{(n)})$-SW code must satisfy

$$P_{\text{e}}^{(n)} \geq 1-\mathsf{P}\left[\frac{1}{n}\mathbf{h}(X_1^n,X_2^n)\leq\mathbf{R}+\gamma\mathbf{1}\right]-3(2^{-n\gamma}), \tag{20}$$

for all n and for any $\gamma>0$. Recall that $\mathbf{h}(X_1^n,X_2^n)$ is the entropy density vector in (4) evaluated at (X_1^n,X_2^n). Suppose that, to the contrary, there exists a rate pair (R_1,R_2) such that $\mathbf{R}\notin\mathscr{R}_{\text{out}}(n,\epsilon)$ but (R_1,R_2) is (n,ϵ)-achievable. Then, by (6), $\mathbf{z}:=\sqrt{n}(\mathbf{R}-\mathbf{H}+\frac{\log n}{n}\mathbf{1})\notin\mathscr{S}(\mathbf{V},\epsilon)$. By the definition of $\mathscr{S}(\mathbf{V},\epsilon)$ in (3), $\mathbf{z}\in\mathbb{R}^3$ is such that $\mathsf{P}(\mathbf{Z}\leq\mathbf{z})<1-\epsilon$. Now consider the probability in (20), denoted as s_n:

$$\begin{aligned} s_n &\overset{(a)}{=} \mathsf{P}\left[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(\mathbf{h}(X_{1k},X_{2k})-\mathbf{H})\leq\mathbf{z}-\left(\frac{\log n}{\sqrt{n}}-\sqrt{n}\gamma\right)\mathbf{1}\right]\\ &\overset{(b)}{\leq} \mathsf{P}\left[\mathbf{Z}\leq\mathbf{z}-\left(\frac{\log n}{\sqrt{n}}-\sqrt{n}\gamma\right)\mathbf{1}\right]+\frac{530\xi}{\lambda_{\min}(\mathbf{V})^{3/2}\sqrt{n}}\\ &\overset{(c)}{<} 1-\epsilon-O\left(\frac{\log n}{\sqrt{n}}\right)+O\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \tag{21}$$

where (a) follows from the definition of $\mathbf{z}$, (b) follows from the multidimensional Berry-Essèen theorem [2] and (c) follows by taking $\gamma:=\frac{\log n}{2n}$ and using Taylor's approximation theorem. Uniting (20) and (21) yields $P_{\text{e}}^{(n)}>\epsilon$, contradicting the (n,ϵ)-achievability of (R_1,R_2) for all n sufficiently large. ∎

C. Comments on the proof

Instead of the universal decoder in (8), one could use a non-universal one by comparing the entropy density vector with the rate vector. This is likened to maximum-likelihood decoding. Taylor expansion in (11) would not be required. Under this decoding strategy, there is symmetry between the error probabilities in the direct and converse parts. Also see [14, Lem. 7.2.1-2]. The rate penalty of using a universal decoder is of the order $O(\frac{\log n}{n})$. This is insignificant compared to the dispersion term which is of the order $O(\frac{1}{\sqrt{n}})$.

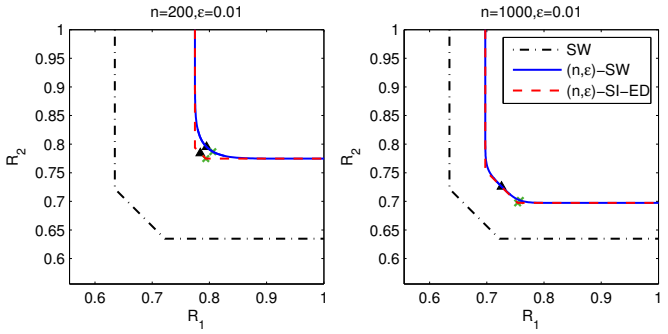

Fig. 1. Plots of the SW boundary, the (n,ϵ)-SI-ED boundary (sharp corners) and the (n,ϵ)-SW boundary (curved) for $\epsilon=0.01$ neglecting the $O(\frac{\log n}{n})$ terms in Theorem 1. The legend applies to both plots. Notice that $\mathscr{R}^*(n,\epsilon)$ and $\mathscr{R}^*_{\text{SI}-\text{ED}}(n,\epsilon)$ are quite different near the equal rate and corner points when n is small. Plots of $\mathscr{R}^*(n,\epsilon)$ and $\mathscr{R}^*_{\text{SI}-\text{ED}}(n,\epsilon)$ as functions of n along the equal rate and corner point slices of $\mathscr{R}^*_{\text{SI}-\text{ED}}(n,\epsilon)$) are given in Figs. 2 and 3. These are indicated by the black ▲ and the green ×.

IV. Singular Entropy Dispersion Matrices

When $\mathbf{V}$ is rank-deficient, consider the set $\mathscr{S}(\mathbf{V},\epsilon)$. Suppose for the moment that $\text{rank}(\mathbf{V})=1$. This is the case considered in [7] where the source is a DSBS(q). For such a DSBS, $\mathbf{V}=v\mathbf{1}_{3\times3}$ for $v=\text{Var}(-\log p_{X_1|X_2}(X_1|X_2))=\text{Var}(-\log p_{X_2|X_1}(X_2|X_1))=\text{Var}(-\log p_{X_1,X_2}(X_1,X_2))$. As such, all the probability mass of the degenerate Gaussian $\mathcal{N}(\mathbf{0},\mathbf{V})$ lies in a subspace of dimension one. Therefore, the set $\mathscr{S}(\mathbf{V},\epsilon)=\{\mathbf{z}\in\mathbb{R}^3:\mathbf{z}\geq\sqrt{v}Q^{-1}(\epsilon)\mathbf{1}\}$ is axis-aligned. The quantity $\sqrt{\frac{v}{n}}Q^{-1}(\epsilon)$ is the *rate redundancy* [4]–[7] for fixed-length SW coding in the finite blocklength regime for a DMMS for which $\text{rank}(\mathbf{V})=1$. In this case, the bounds in (5) and (6) (up to $O(\frac{\log n}{n})$ factors) degenerates to

$$\mathbf{R}\geq\mathbf{H}+\sqrt{(v/n)}\,Q^{-1}(\epsilon)\,\mathbf{1}, \tag{22}$$

where the scalar dispersion $v:=q(1-q)[\log((1-q)/q)]^2$. This reduces to results in previous works [4]–[7]. Our analysis, of course, applies to all sources. Furthermore, we improve on the residual term, which is now of the order $O(\frac{\log n}{n})$. The case where $\text{rank}(\mathbf{V})=2$ follows analogously. All the probability mass of $\mathcal{N}(\mathbf{0},\mathbf{V})$ is concentrated on a two-dimensional subspace in $\mathbb{R}^3$ and the boundary of the set $\mathscr{S}(\mathbf{V},\epsilon)$ are not differentiable. As such only one of the "corners" of $\mathscr{S}(\mathbf{V},\epsilon)$ will be curved and this will be reflected in a result similar to (22). This argument can be formalized and is done in the extended version of this work [3].

V. Numerical Examples

In this section, we present examples to illustrate $\mathscr{R}^*(n,\epsilon)$. We neglect the $O(\frac{\log n}{n})$ terms throughout; thus we are just

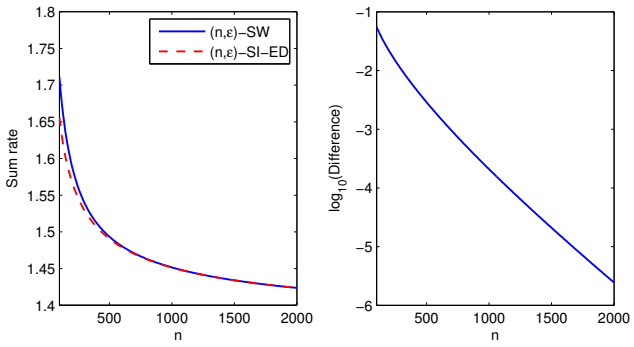

Fig. 2. Comparison between the (n, ϵ)-SW equal rate point and the (n, ϵ)-SI-ED equal rate point and their difference as functions of n. The right plot shows that the difference decays exponentially.

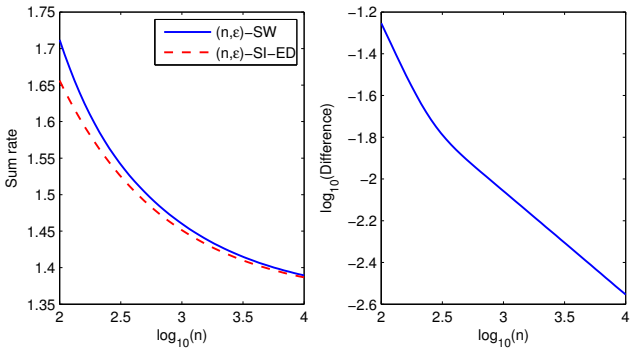

Fig. 3. Comparison between the corner rates and their difference. Note that the x-axis is $\log_{10}(n)$ and the difference decays as $\Theta(n^{-1/2})$.

concerned about *Gaussian approximations*. The source is taken to be $p_{X_1,X_2} = [1-3a, a; a, a]$ where $a = 0.1$. This source has a positive-definite dispersion. In Fig. 1, we plot the boundaries of the SW region [1] and the boundary of $\mathscr{R}^*(n, \epsilon)$ for $\epsilon = 0.01$. We also plot the boundary of the (n, ϵ)-region for coding with side information at encoders and decoder (SI-ED). This region $\mathscr{R}^*_{\mathrm{SI-ED}}(n, \epsilon) \subset \mathbb{R}^2$ is the set of (R_1, R_2) satisfying

$$\mathbf{R} \geq \mathbf{H} + \sqrt{\frac{\mathrm{diag}(\mathbf{V}(p_{X_1,X_2}))}{n}}\, Q^{-1}(\epsilon). \tag{23}$$

From Fig. 1, we see that $\mathscr{R}^*(n, \epsilon)$ has a curved boundary, reflecting the correlations among the entropy densities. Also, it approaches the SW boundary as n grows. The boundaries of $\mathscr{R}^*(n, \epsilon)$ and $\mathscr{R}^*_{\mathrm{SI-ED}}(n, \epsilon)$ coincide if R_2 meets the condition in (23) with equality and R_1 is large (and vice versa).

There are two interesting "slices" of the plots in Fig. 1. These are the equal rate slice (along the 45° line) and the slice passing through the origin and a *corner point* $(R^*_{1,n}, R^*_{2,n})$ of $\mathscr{R}^*_{\mathrm{SI-ED}}(n, \epsilon)$, defined as follows:

$$\begin{aligned} R^*_{2,n} &:= \inf\{R_2 : (R_1, R_2) \in \mathscr{R}^*_{\mathrm{SI-ED}}(n, \epsilon) \text{ for some } R_1\} \\ R^*_{1,n} &:= \inf\{R_1 : (R_1, R^*_{2,n}) \in \mathscr{R}^*_{\mathrm{SI-ED}}(n, \epsilon)\}. \end{aligned} \tag{24}$$

These two slices are indicated by the markers $(\times, \blacktriangle)$ in Fig. 1. The sum rates along both slices are plotted as functions of n in Figs. 2 and 3 respectively. We observe from Fig. 2 that the two sum rates on the 45° equal rate line approach each other as n grows. Moreover, empirically we observe (and can prove) that their difference decays as $\exp(-\Theta(n))$, which is subsumed by the $O(\frac{\log n}{n})$ term, i.e., the dispersions are the same. Thus, when $n \geq 10^3$, there is essentially no loss in performing SW coding versus cooperative encoding if we wish to optimize the sum rate. On the other hand, from Fig. 3, we see that the corresponding difference in corner points decays at a much slower rate of $\Theta(n^{-1/2})$. Thus, the corner rate *dispersions are different* and if we wish to operate at this point, SW loses second-order coding rate relative to the cooperative scenario. See [3] for further analysis of this point.

VI. Conclusions and Future Work

In this paper, we quantified the second-order coding rates of the Slepian-Wolf problem. We showed that these rates are governed by a so-called entropy dispersion matrix. Admittedly, our results cannot be described as being *finite blocklength*. We seek to work towards such results in the future and to compare the accuracy of the Gaussian approximation in Theorem 1 to upper and lower bounds on the blocklength required to achieve a target error probability.

References

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inf. Th.*, vol. 19, pp. 471–80, 1973.

[2] V. Bentkus, "On the dependence of the Berry-Esseen bound on dimension," *J. Stat. Planning and Inference*, vol. 113, pp. 385 – 402, 2003.

[3] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *arXiv:1201.3901*, Feb 2012, [Online].

[4] D. Baron, M. A. Khojastepour, and R. G. Baraniuk, "Redundancy rates of Slepian-Wolf coding," in *Allerton Conf.*, 2004.

[5] D.-K. He, L. A. Lastras-Montaño, E.-H. Yang, A. Jagmohan, and J. Chen, "On the redundancy of Slepian-Wolf coding," *IEEE Trans. on Inf. Th.*, vol. 55, no. 12, pp. 5607–27, Dec 2009.

[6] S. Watanabe, R. Matsumoto, and T. Uyematsu, "Strongly secure privacy amplification cannot be obtained by encoder of Slepian-Wolf code," *IEICE Trans. on Fund. Elec., Comms. and Comp. Sciences*, vol. E93.A, no. 9, pp. 1650–1659, 2010.

[7] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Non-asymptotic performance of symmetric Slepian-Wolf coding," in *Conference on Information Sciences and Systems*, 2005.

[8] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," in *Trans. Third. Prague Conf. Inf. Th.*, 1962, pp. 689–723.

[9] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Trans. on Inf. Th.*, vol. 54, pp. 4619–37, Oct 2008.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding in the finite blocklength regime," *IEEE Trans. on Inf. Th.*, vol. 56, pp. 2307 – 59, May 2010.

[11] V. Kostina and S. Verdú, "Fixed-length lossy compression in the finite blocklength regime: Discrete memoryless sources," in *Int. Symp. Inf. Th.*, 2011.

[12] A. Ingber and Y. Kochman, "The dispersion of lossy source coding," in *Data Compression Conference (DCC)*, 2011.

[13] D. Wang, A. Ingber, and Y. Kochman, "The dispersion of joint source-channel coding," in *Allerton Conference*, 2011, arXiv:1109.6310.

[14] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, Feb 2010.

[15] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Akademiai Kiado, 1981.