Non-Asymptotic and Second-Order Achievability Bounds for Source Coding With Side-Information

Shun Watanabe[†], Shigeaki Kuzuoka[‡], Vincent Y. F. Tan[⋆]

[†]University of Tokushima and University of Maryland, College Park, Email: shun-wata@is.tokushima-u.ac.jp
[‡]Department of Computer and Communication Sciences, Wakayama University, Email: kuzuoka@ieee.org
[⋆]Institute for Infocomm Research and National University of Singapore, Email: tanyfv@i2r.a-star.edu.sg

***Abstract*—We present a novel achievability bound for the Wyner-Ahlswede-Körner (WAK) problem of lossless source coding with rate-limited side-information. This bound is proved using ideas from channel simulation and channel resolvability. The bound improves on all previous non-asymptotic bounds on the error probability of the WAK problem. We also present achievable second-order coding rates by applying the multidimensional Berry-Essèen theorem to our new non-asymptotic bound.**

***Index Terms*—Source coding, side-information, finite blocklength, non-asymptotic, second-order coding rate**

I. Introduction

We revisit the classical problem of lossless source coding with a helper. In this problem, first tackled by Wyner [1] and Ahlswede-Körner [2] (WAK), a main source X^n is to be reconstructed almost losslessly from rate-limited versions of X^n and Y^n, a correlated random variable regarded as side-information or helper. The compression rates of X^n and Y^n are denoted as R_1 and R_2 respectively. The optimal rate region is the set of rate pairs (R_1, R_2) for which there exists a reliable code. WAK [1], [2] showed that the optimal rate region is

$$R_1 \geq H(X|U), \quad R_2 \geq I(U;Y) \tag{1}$$

for some $P_{U|Y}$. For the direct part, the helper encoder compresses the side-information and transmits a description represented by U. The main encoder then uses binning as in the achievability proof of the Slepian-Wolf theorem to help the decoder recover X given the description U.

With renewed interest in finite blocklength bounds and second-order coding rates [3], [4], in this paper, we derive the tightest finite blocklength bound on the error probability for the WAK problem. The proof makes use of channel resolvability techniques [5, Ch. 6] in the helper's code construction. The use of channel resolvability for lossy (and lossless) source coding is well-recognized in the quantum information theory community. See [6]–[8] for example. The main idea in our proof is that, mixed over the common randomness, the joint distribution of (U, Y) (in the one-shot notation) is close (in the variational distance sense) to $(\hat{U}, Y)$, where $\hat{U}$ designates the chosen auxiliary codeword. As a result, by monotonicity of the variational distance, the joint distribution of (U, Y, X) is also close to $(\hat{U}, Y, X)$. This means that in the asymptotic setting, the triple $(\hat{U}, Y, X)$ is jointly typical with high probability. This circumvents the need to use the so-called piggyback coding lemma (PBL) and the Markov lemma [1].

A. Main Contributions

Our main contribution in this paper is to demonstrate an improved bound for WAK coding using ideas from channel resolvability [5, Ch. 6] and channel simulation [9]. The primary part of the new bound on the error probability is $\Pr(\mathcal{E}_{\rm c} \cup \mathcal{E}_{\rm b})$ where $\mathcal{E}_{\rm c}$ represents the covering error and $\mathcal{E}_{\rm b}$ represents the binning error. Since $\Pr(\mathcal{E}_{\rm c})$ and $\Pr(\mathcal{E}_{\rm b})$ are both information spectrum [5] quantities, they are amenable to be estimated by the Berry-Essèen theorem [10] in the n-fold setting. This yields an achievable second-order coding rate. However, unlike in the point-to-point setting [3], [4], in the multiuser setting, the second-order coding rate is expressed in terms of a so-called dispersion matrix [11].

B. Related Work

Kuzuoka [12] and Miyake and Kanaya [13] used Wyner's PBL to derive a finite blocklength achievability bound and an information spectrum characterization for the WAK problem respectively. Verdú improved on Kuzuoka's bound by using finite blocklength analogues of the packing and covering lemmas in [14]. Kelly and Wagner [15] derived bounds on the error exponent for the WAK problem. We review these results in Section II-C. The second-order coding rate for Slepian-Wolf coding (a related problem) was derived by Tan and Kosut [11].

II. Preliminaries

In this section, we introduce our notation, state the WAK problem and review existing results for this problem.

A. Notations

Random variables (e.g., X) and their realizations (e.g., x) are in capital and lower case respectively. All random variables take values on finite sets which are denoted in calligraphic font (e.g., $\mathcal{X}$). The cardinality of $\mathcal{X}$ is denoted as $|\mathcal{X}|$. Let $X^n := (X_1, \ldots, X_n)$. The set of all distributions supported on alphabet $\mathcal{X}$ is denoted as $\mathcal{P}(\mathcal{X})$. Information-theoretic quantities are denoted in the usual manner [5], [16].

B. Problem Formulation

Let us consider a correlated source (X, Y) taking values in $\mathcal{X} \times \mathcal{Y}$ and having joint distribution P_{XY}. Throughout, X is the main source while Y is the helper or side-information. The WAK problem involves reconstructing X losslessly given rate-limited versions of both X and Y.

Definition 1. *A (possibly stochastic)* source coding with side-information (SSI) code $\Phi = (f, g, \psi)$ *is a triple of mappings that includes two encoders $f : \mathcal{X} \to \mathcal{M}$ and $g : \mathcal{Y} \to \mathcal{L}$ and a decoder $\psi : \mathcal{M} \times \mathcal{L} \to \mathcal{X}$. The* error probability *of the SSI code Φ is defined as*

$$\mathrm{P_e}(\Phi) := \Pr\left\{X \neq \psi(f(X), g(Y))\right\}. \tag{2}$$

In Section IV, we consider n-fold i.i.d. extensions of X and Y, denoted as X^n and Y^n. In this case, we use the subscript n to specify the blocklength, i.e., the code is $\Phi_n = (f_n, g_n, \psi_n)$ and the compression index sets are $\mathcal{M}_n$ and $\mathcal{L}_n$. In this case, we can define the pair of rates of the code Φ_n as

$$R_1(\Phi_n) := \frac{1}{n}\log|\mathcal{M}_n|, \quad R_2(\Phi_n) := \frac{1}{n}\log|\mathcal{L}_n|. \tag{3}$$

Definition 2. *The (n, ε)*-optimal rate region $\mathscr{R}(n, \varepsilon)$ *is defined as the set of all pairs of rates (R_1, R_2) for which there exists a SSI code Φ_n with rates at most (R_1, R_2) and with error probability not exceeding ε. In other words,*

$$\mathscr{R}(n,\varepsilon) := \Bigg\{(R_1,R_2) \in \mathbb{R}_+^2 : \exists\, \Phi_n \text{ s.t. } \frac{1}{n}\log|\mathcal{M}_n| \le R_1, \frac{1}{n}\log|\mathcal{L}_n| \le R_2, \mathrm{P_e}(\Phi_n) \le \varepsilon\Bigg\} \tag{4}$$

We also define the asymptotic rate regions

$$\mathscr{R}(\varepsilon) := \mathrm{cl}\left[\bigcup_{n\ge 1}\mathscr{R}(n,\varepsilon)\right], \quad \mathscr{R} := \bigcap_{0<\varepsilon<1}\mathscr{R}(\varepsilon). \tag{5}$$

where cl *denotes closure in* $\mathbb{R}^2$.

In the following, we will provide an inner bound to $\mathscr{R}(n, \varepsilon)$ that improves on all previous inner bounds [12], [17].

C. Existing Results

In this section, we review some asymptotic and non-asymptotic bounds for the WAK problem. Let $\mathcal{P}(P_{XY})$ be the set of all joint distributions P_{UXY} such that the $\mathcal{X}\times\mathcal{Y}$-marginal is P_{XY}, $U - Y - X$ forms a Markov chain and $|\mathcal{U}| \le |\mathcal{Y}| + 1$. Define the set

$$\mathscr{R}^* := \bigcup_{P_{UXY}\in\mathcal{P}(P_{XY})} \{(R_1,R_2) : R_1 \ge H(X|U), R_2 \ge I(U;Y)\}. \tag{6}$$

Wyner [1] and Ahlswede-Körner [2] proved the following:

Theorem 1 (Wyner [1], Ahlswede-Körner [2])**.** *For every $0 < \varepsilon < 1$, we have*

$$\mathscr{R}(\varepsilon) = \mathscr{R} = \mathscr{R}^*. \tag{7}$$

To prove the direct part, Wyner used the PBL and the Markov lemma [1] while Ahlswede-Körner [2] used the maximal code construction. Ahlswede-Gács-Körner [18] proved the strong converse using entropy and image-size characterizations [16, Ch. 15]. See [16, Thm. 16.4].

The following non-asymptotic version of Wyner's bound was proved by Kuzuoka [12] using the Markov lemma. For fixed alphabet $\mathcal{U}$, joint distribution $P_{UXY} \in \mathcal{P}(P_{XY})$ and arbitrary positive constants γ_b and γ_c, we define

$$\mathcal{T}_\mathrm{b}(\gamma_\mathrm{b}) := \left\{(u,x) \in \mathcal{U}\times\mathcal{X} : -\log P_{X|U}(x|u) < \gamma_\mathrm{b}\right\} \tag{8}$$

$$\mathcal{T}_\mathrm{c}(\gamma_\mathrm{c}) := \left\{(u,y) \in \mathcal{U}\times\mathcal{Y} : \log\frac{P_{Y|U}(y|u)}{P_Y(y)} < \gamma_\mathrm{c}\right\} \tag{9}$$

Theorem 2 (Kuzuoka [12])**.** *For arbitrary $\gamma_\mathrm{b}, \gamma_\mathrm{c} > 0$, there exists an SSI code Φ with error probability satisfying*

$$\mathrm{P_e}(\Phi) \le 2\sqrt{P_{UX}(\mathcal{T}_\mathrm{b}(\gamma_\mathrm{b})^c)} + P_{UY}(\mathcal{T}_\mathrm{c}(\gamma_\mathrm{c})^c) + \frac{2^{\gamma_\mathrm{b}}}{|\mathcal{M}|} + \exp\left\{-\frac{|\mathcal{L}|}{2^{\gamma_\mathrm{c}}}\right\}. \tag{10}$$

The first and second terms are the dominant ones. The second term represents the encoding of Y with U and the first term represents the decoding of X given U. The first term can be large due to the square root resulting from Wyner's PBL. Verdú [17] demonstrated a refined version of Theorem 2 in which the square root in the first term is removed.

Theorem 3 (Verdú [17])**.** *For arbitrary $\gamma_\mathrm{b}, \gamma_\mathrm{c} > 0$, there exists an SSI code Φ with error probability satisfying*

$$\mathrm{P_e}(\Phi) \le P_{UX}(\mathcal{T}_\mathrm{b}(\gamma_\mathrm{b})^c) + P_{UY}(\mathcal{T}_\mathrm{c}(\gamma_\mathrm{c})^c) + \frac{2^{\gamma_\mathrm{b}}}{|\mathcal{M}|} + \exp\left\{-\frac{|\mathcal{L}|}{2^{\gamma_\mathrm{c}}}\right\}. \tag{11}$$

In Section III, we further improve on Verdú's bound. We show that the two information spectrum terms in (11) (first two terms) can be combined under a single probability.

In another line of work, Kelly and Wagner [15] demonstrated bounds on the error exponent for the WAK problem. Here we present only the direct part (lower bound).

Theorem 4 (Kelly-Wagner [15])**.** *There exists a sequence of SSI codes $\{\Phi_n\}_{n=1}^\infty$ of rates (R_1, R_2) such that*

$$\liminf_{n\to\infty}\frac{1}{n}\log\frac{1}{\mathrm{P_e}(\Phi_n)} \ge \eta_\mathrm{L}(P_{XY}, R_1, R_2) \tag{12}$$

where

$$\eta_\mathrm{L}(P_{XY},R_1,R_2) := \min_{Q_Y}\max_{Q_{U|Y}}\min_{\substack{Q_{X|YU}:\\ H(Q_X)\ge R_1}} D(Q_{XYU}||P_{XY}Q_{U|Y}) + \begin{cases} |R_1 + R_2 - H(Q_{X|U}|Q_U) \\ \quad -I(Q_Y,Q_{U|Y})|^+ & I(Q_Y,Q_{U|Y}) \ge R_2 \\ |R_1 - H(Q_{X|U}|Q_U)|^+ & I(Q_Y,Q_{U|Y}) < R_2 \end{cases} \tag{13}$$

The proof in [15] is based on the method of types [16]. The helper encoder quantizes its observation Y^n using the test channel $Q_{U|Y}$ and the primary encoder uses binning for each source type class. The decoder finds the sequence in the specified bin with the smallest empirical conditional entropy.

III. Novel Non-Asymptotic Bound For WAK

We now present our non-asymptotic bound for WAK coding. Fix $\mathcal{U}$ and $P_{UXY} \in \mathcal{P}(P_{XY})$. Also recall the definitions of the sets $\mathcal{T}_\mathrm{b}(\gamma_\mathrm{b})$ and $\mathcal{T}_\mathrm{c}(\gamma_\mathrm{c})$ in (8) and (9) respectively. The following is our Channel-Simulation-type (CS-type) bound.

Theorem 5 (CS-type bound)**.** *For arbitrary $\gamma_{\rm b}, \gamma_{\rm c}, \delta > 0$, there exists an SSI code Φ with*

$$\begin{aligned} \mathrm{P_e}(\Phi) &\leq P_{UXY}\left[(u,x) \in \mathcal{T}_{\rm b}(\gamma_{\rm b})^c \cup (u,y) \in \mathcal{T}_{\rm c}(\gamma_{\rm c})^c\right] \\ &\quad + \frac{1}{|\mathcal{M}|} \sum_{(u,x') \in \mathcal{T}_{\rm b}(\gamma_{\rm b})} P_U(u) + \sqrt{\frac{\Delta(\gamma_{\rm c}, P_{UY})}{|\mathcal{L}|}} + \delta \end{aligned} \tag{14}$$

where

$$\Delta(\gamma_{\rm c}, P_{UY}) := \sum_{(u,y) \in \mathcal{T}_{\rm c}(\gamma_{\rm c})} P_U(u) \frac{P_{Y|U}(y|u)^2}{P_Y(y)}. \tag{15}$$

See Appendix A for a proof sketch. Using the definitions of $\mathcal{T}_{\rm b}(\gamma_{\rm b})$ and $\mathcal{T}_{\rm c}(\gamma_{\rm c})$, we obtain the following.

Corollary 6 (Simplified CS-type bound)**.** *For arbitrary $\gamma_{\rm b}, \gamma_{\rm c}, \delta > 0$, there exists an SSI code Φ with*

$$\begin{aligned} \mathrm{P_e}(\Phi) &\leq P_{UXY}\left[(u,x) \in \mathcal{T}_{\rm b}(\gamma_{\rm b})^c \cup (u,y) \in \mathcal{T}_{\rm c}(\gamma_{\rm c})^c\right] \\ &\quad + \frac{2^{\gamma_{\rm b}}}{|\mathcal{M}|} + \sqrt{\frac{2^{\gamma_{\rm c}}}{|\mathcal{L}|}} + \delta. \end{aligned} \tag{16}$$

If $(X^n, Y^n) \sim P_{XY}^n$, then by designing $\gamma_{\rm b}$ and $\gamma_{\rm c}$ appropriately, we see that the dominating term in (16) is the first one. The other terms vanish with n. In particular, δ stems from the amount of common randomness known to all parties and since the amount of common randomness can be arbitrarily large, δ can be arbitrarily small. In addition, Δ in (15) results from approximating an arbitrary distribution with one that is simulated by a channel [19, Lem. 2]. This is precisely the channel resolvability problem [5, Ch. 6] in which given a channel $W : \mathcal{A} \to \mathcal{B}$ and an input distribution P_A, we would like to approximate the output distribution $P_B(b) = \sum_a P_A(a)W(b|a)$ by using as small an amount of randomness as possible. This is done by means of a deterministic map from a finite set $\mathcal{J}$ to a codebook $\mathcal{C} = \{a_j\}_{j \in \mathcal{J}} \subset \mathcal{A}$.

Notice that the sum of the information spectrum terms (first two terms) in Verdú's bound in (11) is the result upon invoking the union bound on the first term in our simplified bound in (16). We illustrate the differences numerically in Section IV.

IV. Achievable Second-Order Coding Rates

In this section, we derive an inner bound to $\mathscr{R}(n, \varepsilon)$, defined in (4) by using Gaussian approximations. For this purpose, given a Gaussian random vector $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$ where $\mathbf{V} \in \mathbb{R}^{2\times 2}$ is a positive-semidefinite matrix, define

$$\mathscr{S}(\mathbf{V}, \varepsilon) := \{\mathbf{z} \in \mathbb{R}^2 : \Pr(\mathbf{Z} \leq \mathbf{z}) \geq 1 - \varepsilon\}. \tag{17}$$

This set was introduced in [11] and is, roughly speaking, the multidimensional analogue of the Q^{-1} function.

To enlarge our inner bound to $\mathscr{R}(n, \varepsilon)$, we use a "time-sharing" variable T, which is independent of (X, Y) [20]. Note that in the finite blocklength setting, the region $\mathscr{R}(n, \varepsilon)$ does not have to be convex unlike in the asymptotic case; cf. (1). For fixed finite sets $\mathcal{U}$ and $\mathcal{T}$, let $\tilde{\mathcal{P}}(P_{XY})$ be the set of all $P_{UTXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{T} \times \mathcal{X} \times \mathcal{Y})$ such that the $\mathcal{X} \times \mathcal{Y}$-marginal of P_{UTXY} is P_{XY} and $U - (Y, T) - X$.

Definition 3. *The* entropy-information density vector *for the WAK problem for $P_{UTXY} \in \tilde{\mathcal{P}}(P_{XY})$ is defined as*

$$\mathbf{j}(U, X, Y|T) := \begin{bmatrix} -\log P_{X|UT}(X|U,T) \\ \log \frac{P_{Y|UT}(Y|U,T)}{P_Y(Y)} \end{bmatrix}. \tag{18}$$

Note that the mean of the entropy-information density vector is the vector of the entropy and mutual information, i.e.,

$$\mathbb{E}[\mathbf{j}(U, X, Y|T)] = \mathbf{J}(P_{UTXY}) = \begin{bmatrix} H(X|U,T) \\ I(U;Y|T) \end{bmatrix}. \tag{19}$$

Definition 4. *The* entropy-information dispersion matrix *for the WAK problem for $P_{UTXY} \in \tilde{\mathcal{P}}(P_{XY})$ is defined as*

$$\mathbf{V}(P_{UTXY}) := \mathbb{E}_T\left[\mathsf{Cov}(\mathbf{j}(U, X, Y|T))\right]. \tag{20}$$

We abbreviate the deterministic quantities $\mathbf{J}(P_{UTXY}) \in \mathbb{R}_+^2$ and $\mathbf{V}(P_{UTXY}) \succeq 0$ as $\mathbf{J}$ and $\mathbf{V}$ respectively when the distribution P_{UTXY} is obvious from the context.

Definition 5. *If $\mathbf{V}(P_{UTXY}) \neq \mathbf{0}_{2\times 2}$, $\mathscr{R}_{\rm in}(n, \varepsilon; P_{UTXY})$ is the set of rate pairs (R_1, R_2) such that $\mathbf{R} := [R_1, R_2]^T$ satisfies*

$$\mathbf{R} \in \mathbf{J} + \frac{\mathscr{S}(\mathbf{V}, \varepsilon)}{\sqrt{n}} + \frac{2\log n}{n}\mathbf{1}. \tag{21}$$

If $\mathbf{V}(P_{UTXY}) = \mathbf{0}_{2\times 2}$, $\mathscr{R}_{\rm in}(n, \varepsilon; P_{UTXY})$ is defined as the set of rate pairs (R_1, R_2) such that

$$\mathbf{R} \in \mathbf{J} + \frac{2\log n}{n}\mathbf{1}. \tag{22}$$

From Corollary 6, we can derive the following:

Theorem 7. *For every $0 < \varepsilon < 1$ and all n sufficiently large, the (n, ε)-optimal rate region $\mathscr{R}(n, \varepsilon)$ satisfies*

$$\bigcup_{P_{UTXY} \in \tilde{\mathcal{P}}(P_{XY})} \mathscr{R}_{\rm in}(n, \varepsilon; P_{UTXY}) \subset \mathscr{R}(n, \varepsilon). \tag{23}$$

See Appendix B for a proof sketch. For comparison, for a fixed $P_{UXY} \in \mathcal{P}(P_{XY})$, define $\mathscr{R}_{\rm in}^{\rm V}(n, \varepsilon; P_{UXY})$ to be the set of rate pairs that satisfy

$$R_1 \geq H(X|U) + \sqrt{\frac{V_H(X|U)}{n}} Q^{-1}(\lambda\varepsilon) + \frac{2\log n}{n} \tag{24}$$

$$R_2 \geq I(U;Y) + \sqrt{\frac{V_I(U;Y)}{n}} Q^{-1}((1-\lambda)\varepsilon) + \frac{2\log n}{n} \tag{25}$$

for some $\lambda \in [0, 1]$ where $V_H(X|U) := \mathsf{Var}(\log P_{X|U}(X|U))$ and $V_I(U;Y) := \mathsf{Var}(\log(P_{Y|U}(Y|U)/P_Y(Y)))$. Verdú's bound on the error probability of the WAK problem yields the following inner bound on $\mathscr{R}(n, \varepsilon)$.

$$\bigcup_{P_{UXY} \in \mathcal{P}(P_{XY})} \mathscr{R}_{\rm in}^{\rm V}(n, \varepsilon; P_{UXY}) \subset \mathscr{R}(n, \varepsilon). \tag{26}$$

Example: We now consider the case where (X, Y) is a discrete symmetric binary source DSBS(α) where $\alpha = 0.11$. The optimal rate region in (1) reduces to

$$R_1 \geq h(\beta * \alpha), \qquad R_2 \geq 1 - h(\beta), \tag{27}$$

where $\beta \in [0, 1/2]$ and $h(\,\cdot\,)$ is the binary entropy. The above region is attained by setting the backward test channel from Y

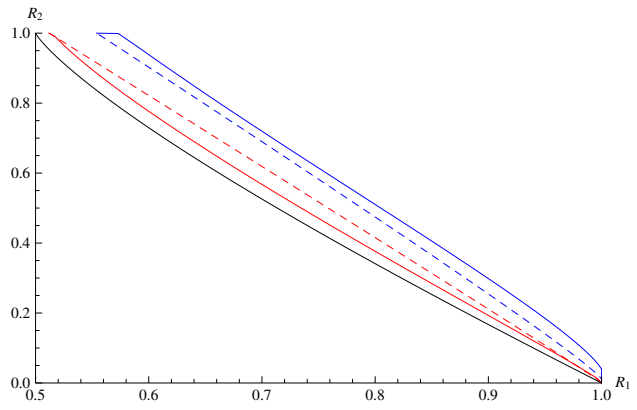

Fig. 1. A comparison between $\tilde{\mathscr{R}}_{\text{in}}(n,\varepsilon)$ without time-sharing (solid line) and the time-sharing region (dashed line) for $\varepsilon = 0.1$. The regions are to the top right of the curves. The blue and red curves are for $n = 500$ and $n = 10,000$ respectively. The black curve is the first-order region (1).

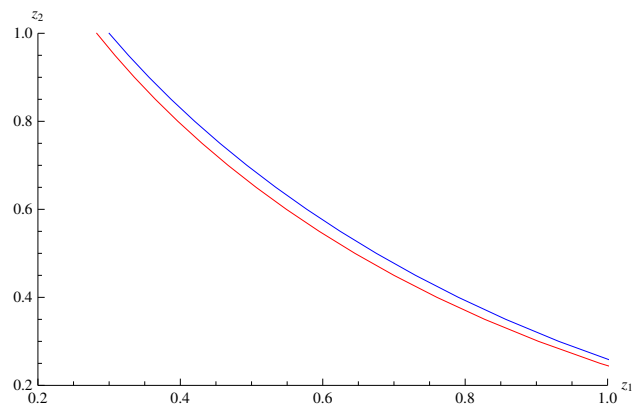

Fig. 2. A comparison between $\mathscr{S}(\mathbf{V}(\beta),\varepsilon)$ (defined in (17)) and $\mathscr{S}^{\text{V}}(\mathbf{V}(\beta),\varepsilon)$ (defined in (29)) for $\beta = h^{-1}(0.5)$ and $\varepsilon = 0.5$. The red and blue curves are the boundaries of $\mathscr{S}(\mathbf{V}(\beta),\varepsilon)$ and $\mathscr{S}^{\text{V}}(\mathbf{V}(\beta),\varepsilon)$ respectively. The regions lie to the top right of the curves.

to U to be a BSC with some crossover probability β. All the elements in the entropy-information dispersion matrix $\mathbf{V}(\beta)$ can be evaluated in closed form in terms of β. Define $\mathbf{J}(\beta) := [h(\beta*\alpha), 1-h(\beta)]^T$. In Fig. 1, we plot the second-order region

$$\tilde{\mathscr{R}}_{\text{in}}(n,\varepsilon) := \bigcup_{0\le\beta\le\frac{1}{2}} \left\{ (R_1,R_2) : \mathbf{R} \in \mathbf{J}(\beta) + \frac{\mathscr{S}(\mathbf{V}(\beta),\varepsilon)}{\sqrt{n}} \right\}. \tag{28}$$

The first-order region and the second-order region with time-sharing ($|\mathcal{T}| = 2$) are also shown for comparison. The time-sharing is between $\beta = 0$ and $\beta = 1/2$. As expected, as the blocklength increases, the (n,ε)-optimal rate region tends to the first-order one. Interestingly, at small blocklengths, time-sharing makes the second-order (n,ε)-optimal rate region in (28) larger compared to that without time-sharing.

We also consider the region $\tilde{\mathscr{R}}^{\text{V}}_{\text{in}}(n,\varepsilon)$ which is the analogue of $\tilde{\mathscr{R}}_{\text{in}}(n,\varepsilon)$ but derived from Verdú's bound in Theorem 3. In Fig. 2, we compare the second-order coefficients, namely that derived from our bound $\mathscr{S}(\mathbf{V}(\beta),\varepsilon)$ and

$$\mathscr{S}^{\text{V}}(\mathbf{V}(\beta),\varepsilon) := \bigcup_{0\le\lambda\le 1} \Big\{ (z_1,z_2) : z_1 \ge \sqrt{V_H(\beta)}Q^{-1}(\lambda\varepsilon), \; z_2 \ge \sqrt{V_I(\beta)}Q^{-1}((1-\lambda)\varepsilon) \Big\}. \tag{29}$$

Note that the difference between the two regions is quite small even for $\varepsilon = 0.5$. This is because, for this example, the covariance of the entropy- and information-density (off-diagonal in the dispersion matrix) is negative so the difference between $\Pr(Z_1 \ge z_1 \text{ or } Z_2 \ge z_2)$ and $\Pr(Z_1 \ge z_1) + \Pr(Z_2 \ge z_2)$ is small. The union bound is not very loose in this case.

V. Conclusion

In this paper, we proved a new non-asymptotic bound on the error probability for the WAK problem. The same channel resolvability and channel simulation technique can be used to strengthen finite blocklength bounds for the Wyner-Ziv [14, Thm. 11.3] and Gel'fand-Pinsker [14, Thm. 7.3] problems leading to improved second-order coding rates. These problems are investigated in the full version of this paper [21] (see also independent and concurrent result by Yassaee-Aref-Gohari [22]).

Appendix A
Proof Sketch for Theorem 5

Proof: Fix $P_{UXY} \in \mathcal{P}(P_{XY})$ and $\gamma_{\text{b}}, \gamma_{\text{c}} > 0$. Let $K \in \mathcal{K}$ denote common randomness [6]–[9] where $\mathcal{K}$ is finite. The realization of K is known to all parties. Construct a stochastic encoder $g : \mathcal{K}\times\mathcal{Y} \to \mathcal{L}$ based on a channel resolvability code. Generate $\mathcal{C} := \{u_{11},\ldots,u_{|\mathcal{K}||\mathcal{L}|}\}$ where $u_{kl} \sim P_U$. Define

$$\bar{P}_{Y|U}(y|u) := P_{Y|U}(y|u)\mathbf{1}\{(u,y) \in \mathcal{T}_{\text{c}}(\gamma_{\text{c}})\} \tag{30}$$

$$\bar{P}_{UXY}(u,x,y) := P_U(u)\bar{P}_{Y|U}(y|u)P_{X|Y}(x|y). \tag{31}$$

The marginals $\bar{P}_{UY}$ and $\bar{P}_Y$ are induced by $\bar{P}_{UXY}$. Define

$$\bar{P}_{KL\tilde{U}\tilde{Y}}(k,l,u,y) = \frac{1}{|\mathcal{K}||\mathcal{L}|}\bar{P}_{Y|U}(y|u)\mathbf{1}\{u_{kl} = u\}. \tag{32}$$

The stochastic encoder is

$$g_{\mathcal{C}}(l|k,y) := \frac{\bar{P}_{L\tilde{Y}|K}(l,y|k)}{\bar{P}_{\tilde{Y}|K}(y|k)}. \tag{33}$$

The helper encoder, given k and y, generates $\hat{L} \sim g_{\mathcal{C}}(\,\cdot\,|k,y)$. Define the quantized source $\hat{U} = u_{K\hat{L}}$. The joint distribution of the random variables $K,\hat{L},\hat{U},X,Y$ is given by

$$P_{K\hat{L}\hat{U}XY}(k,l,u,x,y) = \frac{1}{|\mathcal{K}|}g_{\mathcal{C}}(l|k,y)P_{XY}(x,y)\mathbf{1}\{u_{kl} = u\} \tag{34}$$

We also introduce a "smoothed" version of $P_{K\hat{L}\hat{U}XY}$, namely,

$$\bar{P}_{K\hat{L}\hat{U}XY}(k,l,u,x,y) = \frac{1}{|\mathcal{K}|}g_{\mathcal{C}}(l|k,y)\bar{P}_{XY}(x,y)\mathbf{1}\{u_{kl} = u\} \tag{35}$$

The following lemmas form the basis of the bound on the error probability in (14). They can be proved by using monotonicity and the data-processing lemma for the variational distance $d(P,Q) := \frac{1}{2}\sum_a |P(a) - Q(a)|$ as well as [19, Lem. 2].

Lemma 8. *We have*

$$d(P_{\hat{U}XY}, \bar{P}_{UXY}) \leq \frac{P_{UXY}((u,y) \in \mathcal{T}_{\mathrm{c}}(\gamma_{\mathrm{c}})^c)}{2} + d(\bar{P}_{\hat{U}Y}, \bar{P}_{UY}). \tag{36}$$

Lemma 9. *For every $\gamma > 0$, we have*

$$\mathbb{E}_{\mathcal{C}}[d(\bar{P}_{\hat{U}Y}, \bar{P}_{UY})] \leq \frac{1}{2}\sqrt{\frac{\Delta(\gamma_{\mathrm{c}}, P_{UY})}{|\mathcal{L}|}} + \frac{1}{2}\sqrt{\frac{2^{\gamma}}{|\mathcal{K}||\mathcal{L}|}} + P_U(-\log P_U(U) > \gamma). \tag{37}$$

Lemmas 8 and 9 yield a bound on $\mathbb{E}_{\mathcal{C}}[d(P_{\hat{U}XY}, \bar{P}_{UXY})]$ which is an measure of the atypicality of $P_{\hat{U}XY}$ relative to $\bar{P}_{UXY}$ (or equivalently the unsmoothed version P_{UXY}). This is a surrogate for Wyner's PBL and the Markov lemma [1].

As in Slepian-Wolf coding, the main encoder f assigns a bin index $m \in \mathcal{M}$ to each $x \in \mathcal{X}$ uniformly and independently.

Let $\psi' : \mathcal{M} \times \mathcal{U} \to \mathcal{X}$ be defined as follows. ψ' outputs $\hat{x}$ if it is the unique element in bin m (i.e., $f^{-1}(m)$) such that $(u, \hat{x}) \in \mathcal{T}_{\mathrm{b}}(\gamma_{\mathrm{b}})$. Otherwise, ψ' outputs any prescribed constant $x_0 \in \mathcal{X}$. For the common randomness $k \in \mathcal{K}$, messages $m \in \mathcal{M}$ and $l \in \mathcal{L}$, the decoder is defined as $\psi(m, l; k) := \psi'(m, u_{kl})$. For a fixed random binning specified by (random) function f, define

$$\mathcal{E}_f := \{(u, x, y) : \psi'(f(x), u) \neq x\}. \tag{38}$$

Lemma 10. *Averaged over the common randomness, the error probability can be bounded as*

$$\sum_k \frac{1}{|\mathcal{K}|} \mathrm{P_e}(k, f, g_{\mathcal{C}}) \leq \bar{P}_{UXY}(\mathcal{E}_f) + \frac{P_{UXY}((u,y) \in \mathcal{T}_{\mathrm{c}}(\gamma_{\mathrm{c}})^c)}{2} + d(P_{\hat{U}XY}, \bar{P}_{UXY}). \tag{39}$$

Lemma 11. *Averaged over the random binning, the first term in* (39) *can be bounded as*

$$\mathbb{E}_f[\bar{P}_{UXY}(\mathcal{E}_f)] \leq P_{UXY}\left[(u,x) \in \mathcal{T}_{\mathrm{b}}(\gamma_{\mathrm{b}})^c \cap (u,y) \in \mathcal{T}_{\mathrm{c}}(\gamma_{\mathrm{c}})\right] + \frac{1}{|\mathcal{M}|} \sum_{(u,x') \in \mathcal{T}_{\mathrm{b}}(\gamma_{\mathrm{b}})} P_U(u). \tag{40}$$

The proof of the CS-type bound is completed by uniting Lemmas 8–11 and choosing γ and $|\mathcal{K}|$ to be sufficiently large so that the last two terms in (37) are arbitrarily small. ∎

Appendix B
Proof Sketch for Theorem 7

Proof: Fix a $P_{UTXY} \in \tilde{\mathcal{P}}(P_{XY})$. We only consider the case $\mathbf{V}(P_{UTXY}) \succ 0$ here. See [11] to deal with the singular case. Suppose that (R_1, R_2) are such that $\mathbf{R} \in \mathscr{R}_{\mathrm{in}}(n, \varepsilon; P_{UTXY})$, defined in (21). Then the vector

$$\tilde{\mathbf{z}} := \sqrt{n}\left(\mathbf{R} - \mathbf{J} - \frac{2\log n}{n}\mathbf{1}\right) \in \mathscr{S}(\mathbf{V}, \varepsilon). \tag{41}$$

Also fix a sequence $t^n \in \mathcal{T}^n$ whose type is $O(1/n)$-close to P_T. Then, consider the test channel $P_{U^n|Y^n}(u^n|y^n) := P^n_{U|TY}(u^n|t^n, y^n) \in \tilde{\mathcal{P}}(P^n_{XY})$. Set $\gamma_{\mathrm{c}} := \log|\mathcal{L}_n| - \log n$, $\gamma_{\mathrm{b}} := \log|\mathcal{M}_n| - \log n$ and $\delta := 1/n$. Then, by using Corollary 6, there exists an SSI code Φ_n satisfying

$$1 - \mathrm{P_e}(\Phi_n) \geq \Pr\left\{\frac{1}{n}\sum_{i=1}^n \mathbf{j}(U_i, X_i, Y_i|t_i) \leq \mathbf{R} - \frac{\log n}{n}\mathbf{1}\right\} - \frac{3}{\sqrt{n}}. \tag{42}$$

The multi-dimensional Berry-Essèen theorem [10] yields

$$1 - \mathrm{P_e}(\Phi_n) \geq \Pr\left\{\mathbf{Z} \leq \tilde{\mathbf{z}} + \frac{\log n}{\sqrt{n}}\mathbf{1}\right\} - O\left(\frac{1}{\sqrt{n}}\right). \tag{43}$$

Since $\tilde{\mathbf{z}}$ satisfies (41), by using Taylor's approximation theorem, we can assert that the probability in (43) is $\geq 1 - \varepsilon - O(1/\sqrt{n})$ and so $\mathrm{P_e}(\Phi_n) \leq \varepsilon$ for all n sufficiently large. ∎

References

[1] A. D. Wyner, "On source coding with side information at the decoder," *IEEE Trans. on Inf. Th.*, vol. 21, no. 3, pp. 294–300, 1975.

[2] R. Ahlswede and J. Körner, "Source coding with side information and a converse for the degraded broadcast channel," *IEEE Trans. on Inf. Th.*, vol. 21, no. 6, pp. 629–637, 1975.

[3] 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.

[4] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. on Inf. Th.*, vol. 55, pp. 4947–66, Nov 2009.

[5] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer Berlin Heidelberg, Feb 2003.

[6] Z. Luo and I. Devetak, "Channel simulation with quantum side information," *IEEE Trans. on Inf. Th.*, vol. 55, no. 3, pp. 1331–1342, 2009.

[7] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. on Inf. Th.*, vol. 48, no. 10, pp. 2637–2655, Oct 2002.

[8] A. Winter, "Compression of sources of probability distributions and density operators," *arXiv:quant-ph/0208131*, 2002.

[9] P. Cuff, "Distributed channel synthesis," *arXiv:1208.4415*, 2012.

[10] F. Göetze, "On the rate of convergence in the multivariate CLT," *The Annals of Probability*, vol. 19, no. 2, pp. 721–739, 1991.

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

[12] S. Kuzuoka, "A simple technique for bounding the redundancy of source coding with side information," in *Int. Symp. Inf. Th.*, Boston, MA, 2012.

[13] S. Miyake and F. Kanaya, "Coding theorems on correlated general sources," *IEICE Trans. on Fundamentals of Electronics, Communications and Computer*, vol. E78-A, no. 9, pp. 1063–70, 1995.

[14] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge University Press, 2012.

[15] B. Kelly and A. Wagner, "Reliability in source coding with side information," *IEEE Trans. on Inf. Th.*, vol. 58, no. 8, pp. 5086–5111, Aug 2012.

[16] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[17] S. Verdú, "Non-asymptotic achievability bounds in multiuser information theory," in *Allerton Conference*, 2012.

[18] R. Ahlswede and P. Gács and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 34, no. 3, pp. 157–177, 1976.

[19] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. on Inf. Th.*, vol. 52, no. 4, pp. 1562–75, Apr 2006.

[20] Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *Int. Symp. Inf. Th.*, 2012.

[21] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Non-asymptotic and second-order achievability bounds for coding with side-information," *arXiv:1301.6467*, Jan 2013.

[22] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," *arXiv:1303.0696*, Mar 2013.