On Distributed Computing for Functions with Certain Structures

Shigeaki Kuzuoka
Wakayama University
E-mail: kuzuoka@ieee.org

Shun Watanabe
Tokyo University of Agriculture and Technology
E-mail: shunwata@cc.tuat.ac.jp

Abstract—The problem of distributed function computation for the class of smooth sources is studied, where functions to be computed are compositions of symbol-wise functions and some outer functions that are not symbol-wise. The optimal rate for computing those functions is characterized in terms of the Slepian-Wolf rate and an equivalence class of sources induced by functions. To prove the result, a new method to derive a converse bound for distributed computing is proposed; the bound is derived by identifying a source that is inevitably conveyed to the decoder and by explicitly constructing a code for reproducing that source. As a byproduct, it provides a conceptually simple proof of the known fact that computing a Boolean function may require as large rate as reproducing the entire source.

I. INTRODUCTION

We study the problem of distributed computation, where the encoder observes $X^n$, the decoder observes $Y^n$, and the function $f_n(X^n, Y^n)$ is to be computed at the decoder based on the message sent from the encoder. A straightforward scheme to compute a function is to use the Slepian-Wolf coding [1]. However, since the decoder does not have to reproduce $X^n$ itself, the Slepian-Wolf rate can be improved in general. Then, our interest is how much improvement we can attain.

The literature of distributed computation can be roughly categorized into two directions, symbol-wise functions and sensitive functions. For symbol-wise functions and the class of i.i.d. sources with positivity condition, Han and Kobayashi derived the converse bound on distributed computation for a variety of sources in a unified manner.

As described above, distributed computation for symbol-wise functions is quite well understood; other than symbol-wise functions, our understanding of distributed computation is limited to the extreme case, i.e., the class of sensitive functions. Our motivation of this paper is to further understand distributed computing for functions that are not symbol-wise nor sensitive. For this purpose, we consider compositions of functions where inner functions are symbol-wise functions and outer functions are the type of a sequence or the modulo sum. Although those functions have simple structures, they exhibit some interesting behaviours; in particular, the Slepian-Wolf rate can be improved in general. For the class of smooth sources, we will characterize the optimal rate for computing those functions in terms of the Slepian-Wolf rate of an equivalence class of sources induced by the function to be computed.

Our main technical contribution of this paper is a new method to derive a converse bound on distributed computation. Unlike the method in [4], [5], our method does not resort to the single-letter characterization argument; instead, we derive a bound by identifying a source that is inevitably conveyed to the decoder and by explicitly constructing a code for reproducing that source. As a byproduct, our method provides a conceptually simple proof of the above mentioned result [6–8] for a subclass of sensitive functions.

The rest of the paper is organized as follows. In Section II, we formulate the function computation problem and introduce the class of smooth sources. In Section III, we revisit distributed computing for symbol-wise functions; the result in this section is a generalization of a result in [5] to smooth sources. In Section IV, we study compositions of functions. Finally, in Section V, we discuss how our approach is extended to sources that are not necessarily smooth.

II. PROBLEM SETTING

Let $(X, Y) = \{(X^n, Y^n)\}_{n=1}^\infty$ be a general correlated source with finite alphabet $\mathcal{X}$ and $\mathcal{Y}$; the source is general in

1Here, we only review papers that are directly related to this work. The problem of distributed function computation (with interactive communication) has been actively studied in the computer science community as well [2], [3].

2For instance, our method applies for the joint type, the Hamming distance, the inner product.
the sense of [9], i.e., it has memory and may not be stationary nor ergodic. Without loss of generality, we assume $\mathcal{X} = \{0, 1, \ldots, |\mathcal{X}| - 1\}$ and $\mathcal{Y} = \{0, 1, \ldots, |\mathcal{Y}| - 1\}$. We consider a sequence $f = \{f_n\}_{n=1}^\infty$ of functions $f_n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}_n$. A code $\Phi_n = (\varphi_n, \psi_n)$ for computing $f_n$ is defined by an encoder $\varphi_n : \mathcal{X}^n \to \mathcal{M}_n$ and a decoder $\psi_n : \mathcal{M}_n \times \mathcal{Y}^n \to \mathcal{Z}_n$. The error probability of the code $\Phi_n$ is given by

$$P_e(\Phi_n | f_n) := \Pr(\psi_n(\varphi_n(X^n), Y^n) \neq f_n(X^n, Y^n)) .$$

**Definition 1:** For a given source $(X, Y)$ and a sequence of functions $f$, a rate $R$ is defined to be achievable if there exists a sequence $\{\Phi_n\}_{n=1}^\infty$ of codes satisfying

$$\lim_{n \to \infty} P_e(\Phi_n | f_n) = 0$$

and

$$\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_n| \leq R .$$

The optimal rate for computing $f$, denoted by $R(X|Y|f)$, is the infimum of all achievable rates.

**Definition 2 (SW Rate):** For a given source $(X, Y)$, the optimal rate $R(X|Y|f)$ for the sequence $f = \{f_n\}_{n=1}^\infty$ of identity functions is called the Slepian-Wolf (SW) rate, and denoted by $R_{SW}(X|Y)$.

The following class of sources was introduced in [8].

**Definition 3 (Smooth Source):** A general source $(X, Y)$ is said to be smooth with respect to $Y$ if there exists a constant $0 < q < 1$, which does not depend on $n$, satisfying

$$P_{X^n|Y^n}(x, y) \geq qP_{X^n|Y^n}(x, y)$$

for every $x \in \mathcal{X}^n$ and $y, \tilde{y} \in \mathcal{Y}^n$ with $d_H(y, \tilde{y}) = 1$, where $d_H(\cdot, \cdot)$ is the Hamming distance.

The class of smooth sources is a natural generalization of i.i.d. sources with positivity condition studied in [3, 6], and enables us to study distributed computation for a variety of sources in a unified manner. Indeed this class contains sources with memory, such as Markov sources with positive transition matrices, or non-ergodic sources, such as mixtures of i.i.d. sources with positivity condition; see [8] for the detail.

In this paper, for smooth sources, we are interested in characterizing $R(X|Y|f)$ in terms of $R_{SW}(X|Y)$.

### III. Symbol-wise Functions

In this section, we consider the class of symbol-wise functions, which will be the basis of the next section. For a function $f$ on $\mathcal{X} \times \mathcal{Y}$, a function of the form $f_n(x, y) = (f(x_1, y_1), \ldots, f(x_n, y_n))$ is called a symbol-wise function. The following equivalence class was introduced in [5]. We define equivalence relation $\sim_f$ on $\mathcal{X}$ as

$$x \sim_f \hat{x} \iff f(x, y) = f(\hat{x}, y), \forall y \in \mathcal{Y} .$$

Similarly, we define $\sim_{f_n}$ as

$$x \sim_{f_n} \hat{x} \iff f_n(x, y) = f_n(\hat{x}, y), \forall y \in \mathcal{Y}^n .$$

For a symbol-wise function, it holds that

$$x \sim f \iff x_i \sim_f \hat{x}_i, \forall 1 \leq i \leq n .$$

Let $[x]_f$ be $x$’s equivalence class. Now letting $[x^n]_f := \{[x]_f : x \in \mathcal{X}^n\}$, we can define the function $f_n$ on $[\mathcal{X}^n]_f \times \mathcal{Y}^n$ so that

$$\hat{f}_n([x]_f, y) = f_n(x_f, y), \forall (x_f, y) \in \mathcal{X}^n \times \mathcal{Y}^n .$$

Note also that, for a symbol-wise function, $[\mathcal{X}^n]_f = [\mathcal{X}]_f^n$ and

$$\hat{f}_n([x]_f, y)(i) = \bar{f}([x_1]_f, y_1), \ldots, \bar{f}([x_n]_f, y_n)) .$$

For a given function $f_n$ and a pair $(X^n, Y^n)$ of RVs on $\mathcal{X}^n \times \mathcal{Y}^n$, we can define $(X^n|f_n, Y^n)$ on $[\mathcal{X}^n]_{f_n} \times \mathcal{Y}^n$ such as

$$P_{X^n|f_n,Y^n}(\hat{x}, y) := \sum_{x \in [x|f_n]_f} P_{X^n|Y^n}(x, y) .$$

For a given source $(X, Y)$ and a sequence of functions $f = \{f_n\}_{n=1}^\infty$ and a smooth source $(X, Y)$, it holds that

$$R(X|Y|f) = R([X]|f|Y|f) = R_{SW}([X]|f|Y) .$$

**Proof:** Since

$$R(X|Y|f) \leq R([X]|f|Y|f) \leq R_{SW}([X]|f|Y)$$

holds operationally, it suffices to prove

$$R(X|Y|f) \geq R_{SW}([X]|f|Y) .$$

Let $(\varphi_n, \psi_n)$ be a code satisfying

$$\Pr(\psi_n(\varphi_n(X^n), Y^n) \neq f_n(X^n, Y^n))$$

$$= \sum_{x, y} P_{X^n|Y^n}(x, y)1[\psi_n(\varphi_n(x), y) \neq f_n(x, y)]$$

$$\leq \varepsilon_n .$$

Let $\pi_i : \mathcal{Y}^n \to \mathcal{Y}^n$ be the permutation that shifts only ith symbol of $y$, i.e., $y_i \mapsto y_{i+1} \mod |\mathcal{Y}|$. Then, since $(X, Y)$ is smooth, for every $0 \leq b \leq |\mathcal{Y}| - 1$, we have

$$\Pr(\psi_n(\varphi_n(X^n), \pi_b^n(Y^n)) \neq f_n(X^n, \pi_b^n(Y^n)))$$

$$= \sum_{x, y} P_{X^n|Y^n}(x, y)1[\psi_n(\varphi_n(x), \pi_b^n(y)) \neq f_n(x, \pi_b^n(y))]$$

$$\leq \sum_{x, y} \frac{1}{q} P_{X^n|Y^n}(x, \pi_b^n(y))1[\psi_n(\varphi_n(x), \pi_b^n(y)) \neq f_n(x, \pi_b^n(y))]$$

$$\leq \frac{\varepsilon_n}{q} .$$
where $\pi_i^n$ means applying $\pi_i$ $t$ times. Since $[x]_f$ is uniquely determined from the list $(f(x, y) : y \in \mathcal{Y})$, we can construct a decoder $\hat{\psi}_n : \mathcal{M}_n \times \mathcal{Y}^n \to \{X\}_f^n$ such that $W^n = \hat{\psi}_n(\varphi_n(X^n), Y^n)$ satisfies
\[
\mathbb{E}\left[\frac{1}{n} d_H([X^n]_f, W^n)\right]
\leq \frac{1}{n} \Pr\left([X^n]_f \neq W_i\right)
\leq \frac{1}{n} \Pr\left(\exists 0 \leq b \leq |\mathcal{Y}| - 1, \varphi_n(X^n), \pi_i^n(Y^n) \neq f_n(X^n, \pi_i^n(Y^n))\right)
\leq \frac{|\mathcal{Y}|}{q} \varepsilon_n.
\]
By the Markov inequality, for any $\beta > 0$, we have
\[
\Pr\left(\frac{1}{n} d_H([X^n]_f, W^n) \geq \beta\right) \leq \frac{|\mathcal{Y}|}{q^\beta} \varepsilon_n.
\]
Then, by Lemma 1 in the appendix, there exists a code $(\kappa_n, \tau_n)$ of size $2^{n\delta}$ such that
\[
\Pr\left(\tau_n([X^n]_f), W^n \neq [X^n]_f\right) \leq \frac{|\mathcal{Y}|}{q^\beta} \varepsilon_n + v_n(\beta)2^{-n\delta}.
\]
Since $[X^n]_f$ is a function of $X^n$ and the total code size of $(\varphi_n, \hat{\psi}_n)$ and $(\kappa_n, \tau_n)$ is $|\mathcal{M}_n|2^{n\delta}$, by Lemma 2 in the appendix, we have
\[
\Pr\left(\frac{1}{n} \log \frac{1}{P[X^n]_f | \mathcal{Y} (\varphi_n(X^n), Y^n)} \geq \frac{1}{n} \log |\mathcal{M}_n| + 2\delta\right)
\leq \frac{|\mathcal{Y}|}{q^\beta} \varepsilon_n + (v_n(\beta) + 1)2^{-n\delta}.
\]
Thus, by the standard argument on the Slepian-Wolf coding (cf. Lemma 7.2.1), there exists a code for $([X^n]_f, Y^n)$ with rate $\frac{1}{n} \log |\mathcal{M}_n| + 3\delta$ such that the error probability is less than
\[
\frac{|\mathcal{Y}|}{q^\beta} \varepsilon_n + (v_n(\beta) + 2)2^{-n\delta}.
\]
By taking $\delta > 0$ appropriately compared to $\beta > 0$, the error probability converges to 0, which implies
\[
R_{SW}(X|f) \leq R(X|Y|f) + 3\delta.
\]
Since $\beta > 0$ can be arbitrarily small, and we can make $\delta > 0$ arbitrarily small accordingly, we have (1).

\section{IV. Composition of Functions}

In this section, we consider compositions of functions, where the inner function is symbol-wise and the outer function is the type of a sequence or the modulo-sum. In this section, we assume that the inner symbol-wise function is $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z} = \{0, 1, \ldots, m - 1\}$.

\subsection{A. Type}

For a given symbol-wise function $f_n(x, y) = (f(x_1, y_1), \ldots, f(x_n, y_n))$, we consider its type:
\[
f_1^n(x, y) := P_{f_n(x, y)}.
\]
Then, we denote $f^* = \{f_1^n\}_{n=1}^\infty$.

To characterize the optimal rate for computing $f^*$, we need some preparation. For a given $f$, we introduce $\hat{f}^* : \mathcal{X} \times \mathcal{Y} \to \{0, \ldots, m - 1, -1\}$ as
\[
\hat{f}^*(x, y) := \begin{cases} m & \text{if } f(x, -) \text{ is constant} \\ f(x, y) & \text{else} \end{cases}.
\]
Then, let $\hat{f}^*_n(x, y) = (\hat{f}^*(x_1, y_1), \ldots, \hat{f}^*(x_n, y_n))$ and $f^* = \{f_1^n\}_{n=1}^\infty$.

\begin{theorem}
For a given sequence of symbol-wise function $f$ and a smooth source $(\mathcal{X}, \mathcal{Y})$, it holds that
\[
R(\mathcal{X}|Y|^f) = R_{SW}([X]_{\hat{f}}|Y).
\]
\end{theorem}

\begin{example}
When $f(x, y) = x$, the identity function, then $f_1^n(x, y)$ is the joint type of $(x, y)$. In this case, $R(\mathcal{X}|Y|^f) = R_{SW}(X|Y)$.
\end{example}

\begin{example}
When $f(x, y) = x$, then $f_1^n(x, y)$ is the marginal type of $x$. In this case, $R(\mathcal{X}|Y|^f) = 0$.
\end{example}

\begin{example}
When $f(x, y)$ is given by
\[
x|y| 0 1 2
\]

\begin{tabular} {cccc}
0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 \\
2 & 1 & 0 & 1 \\
3 & 0 & 1 & 1 \\
4 & 1 & 1 & 1 \\
\end{tabular}

then the first row and the fifth row are merged together in $\hat{f}^*(x, y)$, i.e., $\hat{f}^*(0, y) = \hat{f}^*(4, y) = 2$. Thus,
\[
[0]_{f_1} = \{0, 4\}, \quad [1]_{f_1} = \{1, 3\}, \quad [2]_{f_1} = \{2\}.
\]

\begin{proof}
We use the same notation as the proof of Theorem 1.
\end{proof}

Let $(\varphi_n, \psi_n)$ be a code satisfying
\[
\Pr\left(\psi_n(\varphi_n(X^n), Y^n) \neq f_1^n(X^n, \pi_1^n(Y^n))\right) \leq \varepsilon_n.
\]
Then by the same argument as the proof of Theorem 1 we have
\[
\Pr\left(\psi_n(\varphi_n(X^n), \pi_i^n(Y^n)) \neq f_1^n(X^n, \pi_i^n(Y^n))\right) \leq \frac{\varepsilon_n}{q}
\]
for every $0 \leq b \leq |\mathcal{Y}| - 1$.

The key observation here is that, for each $i$, if $\psi_n(\varphi_n(X^n), \pi_i^n(Y^n)) = f_1^n(X^n, \pi_i^n(Y^n))$ for all $0 \leq b \leq |\mathcal{Y}| - 1$ then we can estimate the list of $f_1^n(X^n, \pi_i^n(Y^n))$, for $0 \leq b \leq |\mathcal{Y}| - 1$ correctly (subject $i$ indicates $ith$ component). Indeed we can estimate the list as follows. If $3$Without loss of generality, we identify $\mathcal{X} \times \mathcal{Y}$ with $\mathcal{Z} = \{0, \ldots, |\mathcal{X}||\mathcal{Y}| - 1\}$ in this example.
ψ_n(φ_n(X^n), π^b_n(Y^n)) is unchanged for 0 ≤ b ≤ |Y| − 1, then the estimate of f^n_i(X^n, π^b_i(Y^n)) is m for every 0 ≤ b ≤ |Y| − 1. Otherwise, find z₀ such that (note that the output of the decoder ψ_n is a type on Z)

\[ nψ_n(φ_n(X^n), π^b_n(Y^n))(z₀) < nψ_n(φ_n(X^n), Y^n)(z₀) \]

for some 0 < b ≤ |Y| − 1. Then, the estimate of f^n_i(X^n, Y^n), is z₀. Next, for each 0 < b ≤ |Y| − 1, find z such that

\[ nψ_n(φ_n(X^n), π^b_n(Y^n))(z) > nψ_n(φ_n(X^n), Y^n)(z). \]

Then, the estimate of f^n_i(X^n, π^b_i(Y^n)), is z. If such a z does not exist, then the estimate of f^n_i(X^n, π^b_i(Y^n)), is z₀.

Note that [x]_f is uniquely determined from the list \((\hat{f}^i(x, y) : y ∈ Y)\). Hence, we can construct a decoder ψ_n : M_n × Y^n → |X^n|_f such that W^n = ψ_n(φ_n(X^n), Y^n) satisfies

\[ E \left[ \frac{1}{n}d_H([X^n]_f, W^n) \right] \]

\[ = \sum_{i=1}^{n} \frac{1}{n} \Pr \left( [X_i]_f \neq W_i \right) \]

\[ ≤ \sum_{i=1}^{n} \frac{1}{n} \Pr \left( \exists 0 ≤ b ≤ |Y| − 1, \psi_n(φ_n(X^n), π^b_n(Y^n)) \neq f^n_i(X^n, π^b_n(Y^n)) \right) \]

\[ ≤ \frac{|Y|}{q} \varepsilon_n. \]

Then, by following exactly the same argument as the proof of Theorem 1 we can derive (2).

Next, we show

\[ R(X|Y|f^\oplus) ≤ R_{SW}([X]_{f^\oplus}|Y). \] (3)

In fact, by using a code for \(([X^n]_f, Y^n)\), the decoder can compute the type \(P_{f^n_i(a, y)}\). To compute the type \(P_{f^n_i(a, y)}\), the decoder further needs to know, for each \(z ∈ Z\),

\[ \# \{ i : f(x_i, y_i) = z, \hat{f}^i(x_i, y_i) = m \}. \] (4)

In fact, since \(\hat{f}^i(x_i, y_i) = m\) is equivalent to \(f(x_i, \cdot)\) is constant, the number in (4) can be computed just by the encoder, and these numbers can be sent by using at most \(|Z| \log(n+1)\) bits, which implies (3).

**B. Modulo-Sum**

For a given symbol-wise function f_n(x, y) = (f(x_1, y_1), . . . , f(x_n, y_n)), we consider

\[ f_n^\oplus(x, y) := \sum_{i=1}^{n} f(x_i, y_i) \pmod{m}. \]

Then, let \(f^\oplus = \{f_n^\oplus\}_{n=1}^\infty\).

To characterize the optimal rate for computing \(f^\oplus\), we need some preparation. Let \(f^\oplus : \mathcal{X} × \mathcal{Y}\) be the function defined by

\[ \hat{f}^\oplus(x, y) := f(x, y + 1) - f(x, y) \pmod{m}, \quad y ∈ \mathcal{Y}, \]

where \(f(x, |Y|) = f(x, 0)\). Then, let \(\hat{f}_n^\oplus(x, y) = (\hat{f}^\oplus(x_1, y_1), . . . , \hat{f}^\oplus(x_n, y_n))\) and \(f^\oplus = \{f_n^\oplus\}_{n=1}^\infty\).

**Theorem 3**: For a given symbol-wise function f and smooth source \((X, Y)\), it holds that

\[ R(X|Y|f^\oplus) = R_{SW}([X]_{f^\oplus}|Y). \]

**Example 4**: When \(X = Y = \{0, 1\}\) and \(f(x, y) = x + y\), then \(\hat{f}(x, y) = 1\) for every \((x, y)\). Thus, \(R(X|Y|f^\oplus) = 0\). In fact, the encoder can just send the parity \(\oplus_{i=1}^{n} X_i\). Then, the decoder can reproduce \(f^\oplus(X^n, Y^n) = (\oplus_{i=1}^{n} X_i) \oplus (\oplus_{i=1}^{n} Y_i)\). It is interesting to compare this example with the fact that, for the same function \(f(x, y) = x + y\), \(R(X|Y|f^\oplus) = R_{SW}(X|Y)\).

**Example 5**: When \(X = Y = \{0, 1\}\) and \(f(x, y) = x ∧ y\), then \(\hat{f}(x, y) = x\). Thus,

\[ R(X|Y|f^\oplus) = R_{SW}(X|Y). \]

**Proof**: We use the same notation as the proof of Theorem 1. We first show

\[ R(X|Y|f^\oplus) ≥ R_{SW}([X]_{f^\oplus}|Y). \] (5)

Let \((φ_n, ψ_n)\) be a code satisfying

\[ \Pr(ψ_n(φ_n(X^n), Y^n) ≠ f_n^\oplus(X^n, Y^n)) ≤ \varepsilon_n. \] (6)

Then, by the same argument as Theorem 1 we have

\[ \Pr(ψ_n(φ_n(X^n), π^b_n(Y^n)) ≠ f_n^\oplus(X^n, π^b_n(Y^n))) ≤ \frac{\varepsilon_n}{q} \]

for every 0 ≤ b ≤ |Y| − 1. We construct a decoder \(ψ_n : M_n × Y^n → [X^n]_{f^\oplus}\) as follows. We first get estimates of the list \((\hat{f}^\oplus(x_i, y) : y ∈ Y)\) from estimates of the list

\[ (f^n_i(X^n, π^b_i(Y^n)) = f_n^\oplus(X^n, Y^n) : 1 ≤ b ≤ |Y| − 1). \]

Then, since \([x]_{f^\oplus}\) is uniquely determined from \((\hat{f}^\oplus(x, y) : y ∈ Y)\), we can construct a decoder \(ψ_n\) such that \(W^n = ψ_n(φ_n(X^n), Y^n)\) satisfies

\[ E \left[ \frac{1}{n}d_H([X^n]_{f^\oplus}, W^n) \right] \]

\[ = \sum_{i=1}^{n} \frac{1}{n} \Pr \left( [X_i]_{f^\oplus} ≠ W_i \right) \]

\[ ≤ \sum_{i=1}^{n} \frac{1}{n} \Pr \left( \exists 0 ≤ b ≤ |Y| − 1, \psi_n(φ_n(X^n), π^b_n(Y^n)) ≠ f_n^\oplus(X^n, π^b_n(Y^n)) \right) \]

\[ ≤ \frac{|Y|}{q} \varepsilon_n. \]

Then, by following exactly the same argument as the proof of Theorem 1 we can derive (5).

Next, we prove

\[ R(X|Y|f^\oplus) ≤ R_{SW}([X]_{f^\oplus}|Y). \] (7)
In fact, by using a code for sending $[X^n]_{j=0}^n$ with error $\varepsilon_n$, the decoder can compute $f^n_0(X^n, Y^n)$ with error $\varepsilon_n$ as follows. For each $\tilde{x} \in [X]_{j=0}$, the encoder sends

$$a(\tilde{x}) = \sum_{i=1}^n f(X_i, 0) \mod m$$

to the decoder (by using a constant number of bits). Then, for a given estimate $[\tilde{X}]_{j=0}$ of $[X]_{j=0}$, the decoder computes

$$\sum_{\tilde{x} \in [X]_{j=0}} \left[ a(\tilde{x}) + \sum_{i=1}^n \left( f(\tilde{x}, Y_i) - f(\tilde{x}, 0) \right) \right] \mod m.$$ 

Since

$$f(x_i, y) - f(x_i, 0) = f(\tilde{x}, y) - f(\tilde{x}, 0) \mod m$$

for $[x_i]_{j=0} = \tilde{x}$, the value of (8) coincides with $f^n_0(X^n, Y^n)$ with probability $1 - \varepsilon_n$, which implies (7).

V. Extension to Restricted Support

The method used to prove Theorems 1, 3 relied on the fact that the source is smooth. In this section, we extend the same approach to sources that are not necessarily smooth.

Definition 4: Let $S \subset X \times Y$ be a subset such that for any $x \in X$ there exists at least one $y \in Y$ such that $(x, y) \in S$. A general source $(X, Y)$ is said to be smooth on support set $S$ if there exists a constant $0 < q < 1$, which does not depend on $n$, satisfying the following two conditions: for any $n$ and any $(x, y) \in X^n \times Y^n$,

1. if $(x_i, y_i) \notin S$ for some $i = 1, \ldots, n$, then $P_{X \times Y}{X^n, Y^n}(x, y) = 0$, and
2. if $(x_i, y_i) \in S$ for all $i = 1, \ldots, n$, then for any $\tilde{y}$ satisfying $d_H(y, \tilde{y}) = 1$,

$$P_{X \times Y}{X^n, Y^n}(x, \tilde{y}) \geq qP_{X \times Y}{X^n, Y^n}(x, y).$$

If $(X, Y)$ is an i.i.d. source, then it is smooth on $S$ if $P_{X \times Y}{X^n, Y^n}(x, y) > 0$ with constant $q = \min_{(x,y)\in S} P_{X \times Y}(x,y).$

In [5], the concept of characteristic graph was used for distributed computing.

Definition 5: For a given function $f$ on $X \times Y$ and $S \subset X \times Y$, the characteristic graph $\Gamma$ is defined as the graph such that the set of nodes is $X$ and there is an edge between two distinct nodes $x$ and $\tilde{x}$ if and only if there exists $y \in Y$ satisfying $(x, y), (\tilde{x}, y) \in S$, and $f(x, y) \neq f(\tilde{x}, y)$.

Let $\Gamma = (G)$ be the maximal independent set of $\Gamma$, and let

$$d_G(x, w) := \begin{cases} 0 & \text{if } x \in w \\ 1 & \text{if } x \notin w \end{cases}$$

be a distortion measure on $X \times \Gamma(G)$. Then, for a sequence $(x, w) \in X^n \times \Gamma(G)^n$, let

$$d^n_\beta(x, w) := \frac{1}{n} \sum_{i=1}^n d_G(x_i, w_i)$$

be the symbol-wise distortion. Now, let us consider the Wyner-Ziv coding problem [11] for $(X, Y)$ with the reproduction alphabet $\Gamma(G)$ and the distortion measure $d^n_\beta$. Let $R_{WZ}^n(X | Y)$ be the optimal coding rate under the average distortion constraint for $D = 0$. Similarly, let us define $R_{WZ}^n(X | Y)$ by replacing the symbol-wise distortion measure with the block-wise distortion measure: $d^n_\beta(x, w) := \max_{1 \leq i \leq n} d_G(x_i, w_i)$.

From the definition, it is apparent that $R_{WZ}^n(X | Y) \leq R_{WZ}^n(X | Y)$. Furthermore, it is not difficult to verify that, for symbol-wise function $f$, $R(X | Y | f) \leq R_{WZ}^n(X | Y)$. By modifying the argument in the proof of Theorem 1 we can show the following.

Theorem 4: For a given symbol-wise function $f$ and a source $(X, Y)$ that is smooth on $S$, it holds that

$$R(X | Y | f) \geq R_{WZ}^n(X | Y).$$

In particular, if there exists, for each $x \in X$, a unique $w \in \Gamma(G)$ satisfying $x \in w$, then

$$R(X | Y | f) = R_{WZ}^n(X | Y) = R_{WZ}^n(X | Y).$$

When $S = X \times Y$, the source $(X, Y)$ is smooth, and Theorem 4 reduces to Theorem 1. The proof of Theorem 4 will be provided in the appendix.

APPENDIX

A. Technical Lemmas

The following lemma says that if there exists a code with small symbol error probability, then, by sending additional message of negligible rate, we can boost that code so that block error probability is small. For given $0 < \beta < 1/2$, let

$$v_n(\beta) := \sum_{i=0}^{[n\beta]-1} (|X| - 1)^i \binom{n}{i} \leq n|X|^{\beta}2^{n\beta}.$$ 

Lemma 1: Suppose that $(X^n, W^n)$ on $X^n \times X^n$ satisfies

$$\Pr \left( \frac{1}{n} d_H(X^n, W^n) \geq \beta \right) \leq \varepsilon_n.$$ 

Then, there exists an encoder $\kappa_n : X^n \to K_n$ with $|K_n| \leq 2^{n\delta}$ and a decoder $\tau_n : K_n \times X^n \to X^n$ such that

$$\Pr(\tau_n(\kappa_n(X^n), W^n) \neq X^n) \leq \varepsilon_n + v_n(\beta)2^{-n\delta}.$$ 

Proof: For an encoder, we use the random binning. Given $k_n \in K_n$ and $w \in X^n$, the decoder finds (if exists) a unique $\tilde{x}$ such that

$$\tilde{x} \in T_\beta^n(w) := \{ x : d_H(x, w) < n\beta \}$$

and $\kappa_n(w) = k_n$. Note that

$$|T_\beta^n(w)| \leq v_n(\beta), \forall w \in X^n.$$
Then, by the standard argument (cf. [10, Lemma 7.2.1]), the error probability averaged over the random binning is bounded as

$$
\mathbb{E}_{\kappa_n} \left[ \Pr \left( \tau_n (\kappa_n (X^n), W^n) \neq X^n \right) \right] \\
\leq \Pr \left( \frac{1}{n} d_H(X^n, W^n) \geq \beta \right) \\
+ \sum_{x, w} P_{X^n|W^n}(x, w) \sum_{\kappa_n(\hat{x}) \neq \kappa_n(x)} \Pr \left( \kappa_n(\hat{x}) \neq \kappa_n(x) \right) \\
\leq \varepsilon_n + \sum_{x, w} P_{X^n|W^n}(x, w) \sum_{\kappa_n(\hat{y}) \neq \kappa_n(y)} \frac{1}{|\mathcal{K}_n|} \\
\leq \varepsilon_n + \nu_n(\beta) 2^{-n\delta}.
$$

The following lemma is also used in the main text; it is a slight modification of the standard converse of the Slepian-Wolf coding (cf. [10, Lemma 7.2.2]), where $X^n$ is replaced by a function value $g_n(X^n)$.

**Lemma 2**: For a given $(X^n, Y^n)$ and a function $g_n : X^n \rightarrow \mathcal{Z}_n$, if a code $(\varphi_n, \psi_n)$ with size $|\mathcal{M}_n|$ satisfies

$$
\Pr \left( \psi_n(\varphi_n(X^n), Y^n) \neq g_n(X^n) \right) \leq \varepsilon_n,
$$

then it holds that

$$
\Pr \left( \frac{1}{n} \log \frac{1}{P_{Z_n|Y^n}(Z_n|Y^n)} > \frac{1}{n} \log |\mathcal{M}_n| + \delta \right) \leq \varepsilon_n + 2^{-n\delta},
$$

where $Z_n = g_n(X^n)$.

**Proof**: By the standard argument, we have

$$
\Pr \left( \frac{1}{n} \log \frac{1}{P_{Z_n|Y^n}(Z_n|Y^n)} > \frac{1}{n} \log |\mathcal{M}_n| + \delta \right) \\
\leq \Pr \left( \psi_n(\varphi_n(X^n), Y^n) \neq g_n(X^n) \right) \\
+ \Pr \left( \frac{1}{n} \log \frac{1}{P_{Z_n|Y^n}(Z_n|Y^n)} > \frac{1}{n} \log |\mathcal{M}_n| + \delta, \psi_n(\varphi_n(X^n), Y^n) = g_n(X^n) \right) \\
\leq \varepsilon_n + \sum_{z_n, m_n} \sum_{y \in g_n^{-1}(z_n) \cap \varphi_n^{-1}(m_n)} P_{Y^n}(y) P_{X^n|Y^n}(x|y) \\
\times \mathbf{1} \left[ P_{Z_n|Y^n}(z_n|y) < \frac{2^{-n\delta}}{|\mathcal{M}_n|}, \psi_n(m_n, y) = z_n \right] \\
\leq \varepsilon_n + \sum_{z_n, m_n, y} P_{Y^n}(y) \frac{2^{-n\delta}}{|\mathcal{M}_n|} \mathbf{1} \left[ \psi_n(m_n, y) = z_n \right] \\
\leq \varepsilon_n + 2^{-n\delta},
$$

where the third inequality follows from

$$
\sum_{x \in g_n^{-1}(z_n) \cap \varphi_n^{-1}(m_n)} P_{X^n|Y^n}(x|y) \leq \sum_{x \in g_n^{-1}(z_n)} P_{Z_n|Y^n}(z_n|y).
$$

**B. Proof of Theorem 2**

We use the same notation as the proof of Theorem 1. Let $(\varphi_n, \psi_n)$ be a code satisfying

$$
\Pr(\psi_n(\varphi_n(X^n), Y^n) \neq f_n(X^n, Y^n)) \\
= \sum_{x,y} P_{X^n=Y^n}(x,y) \mathbf{1}[\psi_n(\varphi_n(x), y) \neq f_n(x, y)] \\
\leq \varepsilon_n.
$$

Then, for every $1 \leq i \leq n$ and $0 \leq b \leq |\mathcal{Y}| - 1$, we have

$$
\Pr \left( \left[ \psi_n(\varphi_n(X^n), \pi^b_i(Y^n)) \neq f_n(X^n, Y^n) \right] \land \left[ (X_i, \pi^b_i(Y_i)) \in S \right] \right) \\
= \sum_{x,y} P_{X^n=Y^n}(x,y) \mathbf{1}[\psi_n(\varphi_n(x), \pi^b_i(y)) \neq f_n(x, y)] \\
\leq \frac{1}{q} \sum_{x,y} P_{X^n=Y^n}(x, \pi^b_i(y)) \mathbf{1}[\psi_n(\varphi_n(x), \pi^b_i(y)) \neq f_n(x, y)] \\
\leq \frac{1}{q} \varepsilon_n,
$$

where the role of the map $\pi^b : \mathcal{Y} \rightarrow \mathcal{Y}$ is the same as $\pi^b_i$ for a single symbol.

Now, we construct a decoder $\tilde{\psi}_n$ for the Wyner-Ziv coding problem as follows. For each codeword $m$, side-information $y$, and an index $i = 1, \ldots, n$, let

$$
\omega_i(m, y) := \{ x : f(x, \pi^b_i(y)) = [\psi_n(m, \pi^b_i(y))] \},
$$

for all $b$ s.t. $(x, \pi^b_i(y)) \in S$,

where $[\psi_n(m, \pi^b_i(y))]$ denotes the $i$th symbol of $\psi_n(m, \pi^b_i(y))$. Then, for a given codeword $m$ and $y$, the decoder $\psi_n$ chooses the $i$th output $\tilde{w}_i$ as follows:

1. if there exists at least one $w \in \Gamma(G)$ satisfying $w \supseteq \omega_i(m, y)$, then $\tilde{w}_i = w^\hat{b}$ and
2. otherwise, $\tilde{w}_i$ is chosen arbitrarily.

Let $\hat{W}^n$ be the random variable corresponding to the output of $\tilde{\psi}_n$. Then, by noting $X_i \in \psi_n(\varphi_n(X^n), Y^n) \subseteq \hat{W}_i$ if $\psi_n(\varphi_n(X^n), \pi^b_i(Y^n)) = f_n(X^n, \pi^b_i(Y^n))$ for every $b$ such that $(X_i, \pi^b_i(Y_i)) \in S$, we have

$$
\mathbb{E} \left[ \mathbb{E}_{\tilde{\psi}_n}(X^n, \hat{W}^n) \right] \\
= \frac{1}{n} \sum_{i=1}^n \Pr(X_i \notin \hat{W}_i) \\
\leq \frac{1}{n} \sum_{i=1}^n \Pr \left( \exists b, [\psi_n(\varphi_n(X^n), \pi^b_i(Y^n)) \neq f_n(X^n, \pi^b_i(Y^n))] \land [(X_i, \pi^b_i(Y_i)) \in S] \right) \\
\leq \frac{|\mathcal{Y}|}{q} \varepsilon_n,
$$

5If there exists more than one such $w$, then we pick arbitrary one.
where the last inequality follows from (9) and the union bound. Thus, we have the first claim of the theorem. The latter claim of the theorem can be proved in the same manner as Theorem 1 by using Lemma 1.

ACKNOWLEDGMENT

SW is partially supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 15H06194. SK is partially supported by JSPS KAKENHI Grant Number 26820145.

REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," IEEE Trans. Inf. Theory, vol. IT-19, no. 4, pp. 471–480, Jul. 1973.
[2] A. C. Yao, "Some complexity questions related to distributive computing," Proc. Annual Symposium on Theory of Computing (STOC), pp. 209–213, 1979.
[3] E. Kushilevitz and N. Nisan, Communication Complexity. Cambridge University Press, 1997.
[4] T. S. Han and K. Kobayashi, "A dichotomy of functions $F(X, Y)$ of correlated sources $(X, Y)$ from the viewpoint of the achievable rate region," IEEE Trans. Inf. Theory, vol. 33, no. 1, pp. 69–76, Jan. 1987.
[5] A. Orlitsky and J. R. Roche, "Coding for computing," IEEE Trans. Inf. Theory, vol. 47, no. 3, pp. 903–917, 2001.
[6] R. Ahlswede and I. Csiszár, "To get a bit of information may be as hard as to get full information," IEEE Trans. Inf. Theory, vol. 27, no. 4, pp. 398–408, 1981.
[7] A. A. E. Gamal, "A simple proof of the Ahlswede-Csiszár one-bit theorem," IEEE Trans. Inf. Theory, vol. 29, no. 6, pp. 931–933, 1983.
[8] S. Kuzuoka and S. Watanabe, "A dichotomy of functions in distributed coding: An information-spectrum approach," IEEE Trans. Inf. Theory, vol. 61, no. 9, pp. 5028–5041, Sep. 2015.
[9] T. S. Han and S. Verdú, "Approximation theory of output statistics," IEEE Trans. Inf. Theory, vol. 39, no. 3, pp. 752–772, May 1993.
[10] T. S. Han, Information-spectrum methods in information theory. New York: Springer-Verlag, 2002.
[11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," IEEE Trans. Inf. Theory, vol. IT-22, no. 1, pp. 1–10, Jan. 1976.