Wiretap Channels With Causal State Information: Strong Secrecy

Te Sun Han, *Fellow, IEEE*, and Masahide Sasaki

***Abstract*— The coding problem for wiretap channels with *causal* channel state information available at the encoder and/or the decoder is studied under the *strong secrecy* criterion. This problem consists of two aspects: one is due to wiretap channel coding and the other is due to one-time pad cipher based on the secret key agreement between Alice and Bob using the channel state information. These two aspects are closely related to each other and give rise to an intriguing tradeoff between exploiting the state to boost secret-message rates versus extracting cryptographic key to improve secrecy capabilities. This issue has yet to be understood how to optimally reconcile the two. We newly devised the "iterative" forward–backward coding scheme, combining wiretap channel coding and secret-key-agreement-based one-time pad cipher. We then established reasonable lower bounds of the secrecy capacity for wiretap channels with causal channel state information available only at the encoder (Theorem 1), which can be easily extended to general cases with various kinds of correlated channel state information at the encoder (Alice), decoder (Bob), and wiretapper (Eve). In particular, for degraded wiretap channels, we give the secret-message (secret-key) capacity bounds (Theorems 2, 4, and 5).**

***Index Terms*— Wiretap channel, channel state information, causal coding, secret key agreement, secrecy capacity, strong secrecy.**

I. INTRODUCTION

IN THIS paper the coding problem for the wiretap channel (WC) with *causal* channel state information (CSI) available at the encoder (Alice) and/or the decoder (Bob) is studied. The concept of WC (*without* CSI) originates in Wyner [1] and was extended to a more general WC by Csiszár and Körner [2]. These landmark papers have been followed by many subsequent extensions and generalizations from the viewpoint of theory and practice. In particular, among others, the WC *with* CSI has also been extensively investigated in the literature. Early works include Luo *et al.* [6], Chen and Vinck [7], and Liu and Chen [8] that have studied the *capacity-equivocation* region for degraded WCs with *non-causal* CSI to establish inner and/or outer bounds on the region, which was motivated by physical-layer security problems to actually intervene in practical fading channel communications. Moreover, subsequent recent developments in this direction with *non-causal* CSI can be found also in Vinck *et al.* [10], Boche and Schaefer [11], Dai and Luo [18], Prabhakaran *et al.* [24], Goldfeld *et al.* [25], Bunin *et al.* [26], etc.

Generally speaking, the coding scheme with causal/non-causal CSI outperforms the one without CSI, because knowledge of the CSI enables us to share a common secret key between Alice and Bob to augment the secrecy capacity. More specifically, then, in addition to the standard WC coding (called the *Wyner's WC coding* [1], [2]) *without* resorting to the CSI, we may incorporate also the cryptographic scheme called the *Shannon's one-time pad (OTP) cipher* (cf. Shannon [4]) based on the *secret key agreement* (cf. Maurer [12], Ahlswede and Csiszár [13]) using the CSI between Alice and Bob. Thus, the problem consists of two aspects: one is due to wiretap channel coding and the other is due to one-time pad cipher based on the secret key agreement. Here is the trade-off between them depending on how to use the state information S.

Recent works taking account of such a secrecy key agreement aspect include Khisti*et al.* [14], Chia and El Gamal [17], Sonee and Hodtani [19], and Fujita [20]. In particular, [14] addresses the problem of key capacity that focuses on the maximum rate of secret key agreement between Alice and Bob rather than on the maximum rate of secure message transmission. However, we cannot say that the secrecy capacity problem in these works with causal CSI has now been fully solved. This is because the problem with causal/non-causal CSI necessarily includes the two separate but closely related coding schemes as mentioned in the above paragraph.

Among others, Chia and El Gamal [17] addresses the case with *causal* common CSI available at both Alice and Bob, whereas Fujita [20] deals with the case with *causal* CSI available only at Alice (given a *physically degraded* WC). Both includes lower bounds on the *weak* secrecy capacity, but with tight secrecy capacity formulas in special cases. The present paper is motivated mainly by these two papers, and the main result to be given in this paper is in nice accordance with their results. In particular, we have newly established the "iterative" forward-backward coding scheme for WCs with causal CSI available at Alice with reasonable lower bounds on secrecy capacity. For degraded channels, we successfully established not only lower/upper bounds, but also several exact secret-massage (secret-key) capacities.

Manuscript received July 30, 2017; revised May 14, 2019; accepted June 23, 2019. Date of publication June 27, 2019; date of current version September 13, 2019. This work was supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI under Grant 17H01281 and in part by the ImPACT Program of the Council for Science, Technology and Innovation (Cabinet Office, Government of Japan).

T. S. Han is with the Quantum ICT Advanced Development Center, National Institute of Information and Communications Technology (NICT), Tokyo 184-8795, Japan (e-mail: han@is.uec.ac.jp).

M. Sasaki is with the Advanced ICT Research Institute, National Institute of Information and Communications Technology, Tokyo 184-8795, Japan (e-mail: psasaki@nict.go.jp).

Communicated by N. Merhav, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2019.2925611

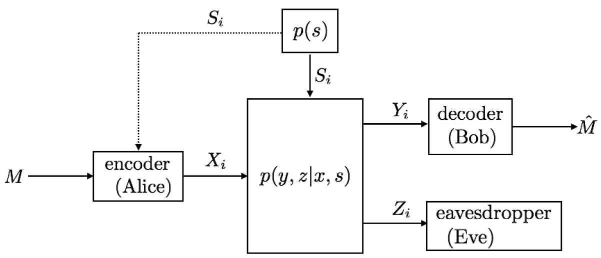

Fig. 1. WC with CSI available only at Alice ($i = 1, 2, \cdots, n$).

The present paper is organized as follows.

In Section II, we give the statement of the problem and the key result (Theorem 1) for the WC with *causal* CSI available only at Alice along with comparison with the work of Chia and El Gamal [17].

In Section III, we give the detailed proof of Theorem 1 to establish lower bounds on the *strong* secrecy capacity. The main ingredients for the proof are Slepian-Wolf coding, Csiszár-Körner's key construction, Gallager's maximum likelihood decoding, and Han-Verdú's resolvability argument, where in the process of these proofs we do *not* invoke the argument of typical sequences at all, which enables us to cope with alphabets that are not necessarily finite (e.g., for Gaussian WCs).

In Section IV, in order to obtain insights into the significance of Theorem 1, we provide specific secrecy capacity bounds (including upper/lower bounds) for degraded WCs with causal/non-causal CSI (Theorems 2, 4, 5 and Corollaries 1, 2, 3).

In Section V, since the present work has partly close bearing with that of Fujita [20], we compare both of them to scrutinize the details of these works.

In Section VI, we conclude the paper with several remarks.

II. Problem Statement and the Result

A *stationary memoryless* WC as illustrated in Fig. 1 is specified by giving the conditional (transition) probability

$$p(y, z|x, s) = P_{YZ|XS}(y, z|x, s) \tag{1}$$

with input random variable X (for Alice), outputs random variables Y (for Bob), Z (for Eve), and CSI random variable S, which are assumed to take values in alphabets $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{S}$, respectively. Alice X (sender), who only has access to *stationary memoryless* CSI S available, wants to send a confidential message $M \in \mathcal{M} = [1 : 2^{nR}]$ (over n channel transmissions) to Bob Y (legitimate receiver) while keeping it secret from Eve Z (eavesdropper), where we use here and hereafter the notation $[i : j] = \{i, i+1, \cdots, j-1, j\}$ for $j \geq i$, and $R \geq 0$ is called the *rate*.

An $(n, 2^{nR})$ code for the WC with *causal* CSI S at the encoder consists of

(i) a message set $\mathcal{M} = [1 : 2^{nR}]$,

(ii) a *stochastic* "causal" encoder $f_i : \mathcal{M} \times \mathcal{S}^i \to \mathcal{X}$ subject to conditional probability $p(x|m, s^i)$ to produce the channel input $X_i(M) = f_i(M, S^i)$ at each time $i \in [1 : n]$, and

(iii) a decoder $g : \mathcal{Y}^n \to \mathcal{M}$ (for Bob) to assign an estimate $\hat{M}$ to each received sequence $\mathbf{Y}$, where we use the notation $a^i = a_1 a_2 \cdots a_i$ (in particular, $\mathbf{a} = a_1 a_2 \cdots a_n$: the bold-faced letters indicate sequences of length n) and assume that the message M is *uniformly* distributed on the message set $\mathcal{M}$.

The probability of error is defined to be $P_e = \Pr\{\hat{M} \neq M\}$. The *information leakage* at Eve with output sequence $\mathbf{Z}$, which measures the amount of information about M that leaks out to Eve, is defined to be $I_E = I(M; \mathbf{Z})$ (the mutual information between M and $\mathbf{Z}$). It should be noted here that this measure is *not* $R_E = \frac{1}{n} I(M; \mathbf{Z})$ (the information leakage *rate*). This means that in this paper we are concerned only with the *strong secrecy* but not the *weak secrecy* as was the case in the literature (e.g., cf. Chia and El Gamal [17], Fujita [20]).

A secrecy rate R is said to be achievable if there exists a sequence of codes $(n, 2^{nR})$ with $P_e \to 0$ and $I_E \to 0$ as $n \to \infty$. The *secrecy capacity* with CSI available only at the encoder (=E), denoted by $C_{\text{CSI-E}}$, is the supremum of all achievable rates.

In order to implement the coding scheme for the WC, it is convenient to introduce its associated channel ω as follows: Let U be an arbitrary auxiliary random variable with values in a set $\mathcal{U}$ that is *independent* of the CSI variable S, and let $h : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ be a *stochastic* mapping subject to conditional probability $p(x|u, s)$. According to the Shannon strategy [5], we define the ω as the WC specified by the conditional probability

$$p(y, z|u, s) = \sum_{x \in \mathcal{X}} p(y, z|x, s) p(x|u, s), \tag{2}$$

which gives the associated WC (called a *test channel*) with input variable U (Alice), outputs variables Y, Z (Bob and Eve) and CSI variable S. Thus, hereafter we may focus solely on the coding problem for the channel ω from the standpoint of achievable rates.

Let us now describe the main result. Set

$$\begin{aligned} &R_{\text{CSI-0}}(p(u), p(x|u, s)) \\ &\quad = I(U; Y) - I(U; Z), \end{aligned} \tag{3}$$

$$\begin{aligned} &R_{\text{CSI-1}}(p(u), p(x|u, s)) \\ &\quad = \min\Big[I(U; Y) - I(U; SZ) \\ &\qquad\qquad + H(S|Z) - H(S|UY), \\ &\qquad\qquad\quad I(U; Y) - H(S|UY) \Big], \end{aligned} \tag{4}$$

$$\begin{aligned} &R_{\text{CSI-2}}(p(u), p(x|u, s)) \\ &\quad = \min\Big[H(S|UZ) - H(S|UY), I(U; Y) - H(S|UY) \Big], \end{aligned} \tag{5}$$

where $I(\cdot;\cdot), I(\cdot;\cdot|\cdot)$ denote the (conditional) mutual informations; and $H(\cdot), H(\cdot|\cdot)$ denote the (conditional) entropies. Moreover, for simplicity we use the notation $A_1 A_2 \cdots A_m$ to denote $(A_1, A_2, \cdots, A_m)$.

Then, we have the following theorem on the *secrecy capacity* $C_{\text{CSI-E}}$ with the understanding that $R_{\text{CSI-1}}(p(u), p(x|u, s)) = 0$ when $I(U; Y) - I(U; SZ) < 0$ or $H(S|Z) - H(S|UY) < 0$:

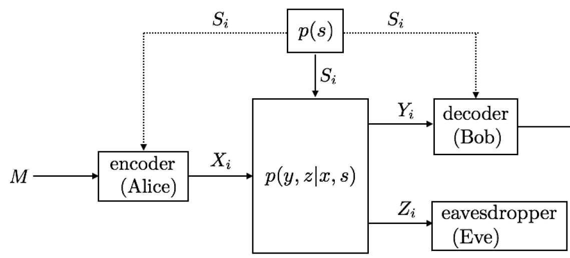

Fig. 2. WC with the same CSI available at Alice and Bob ($i = 1, 2, \cdots, n$).

Theorem 1: Let us consider the WC with CSI as in Fig.1 with *causal* CSI available only at Alice. Then, the (strong) secrecy capacity $C_{\text{CSI-E}}$ is lower bounded as

$$C_{\text{CSI-E}} \geq \max\Big[\max_{p(u),p(x|u,s)} R_{\text{CSI-0}}(p(u), p(x|u,s)), \max_{p(u),p(x|u,s)} R_{\text{CSI-1}}(p(u), p(x|u,s)), \max_{p(u),p(x|u,s)} R_{\text{CSI-2}}(p(u), p(x|u,s))\Big], \tag{6}$$

where $p(u), p(x|u,s)$ ranges over all possible (conditional) probability distributions such that $p(u,s) = p(u)p(s)$, and notice here that $p(s)$ is a given distribution and so cannot be varied. □

The term $H(S|UY)$ in (4), (5) specifies the rate of (auxiliary) Slepian-Wolf coding for information reconcillation in secret key agreement (for OPT cipher) between Alice and Bob using the CSI; in (4) the term $I(U;Y) - I(U;SZ)$ specifies the transmission rate of confidential message via WC coding[2]; the term $H(S|Z) - H(S|UY)$ in (4) specifies the key rate to transmit an additional confidential message via OTP cipher with the secret key shared between Alice and Bob using the CSI; the term $I(U;Y) - H(S|UY)$ in (4), (5) specifies the upper bound on total transmission rates for two kinds of confidential messages as above, excluding the Slepian-Wolf auxiliary message.

The achievability of $R_{\text{CSI-0}}(p(u), p(x|u,s))$ is well known, which is attained by the standard WC coding *without* resorting to the OTP cipher using the secret key generated by CSI (cf. Csiszár and Körner [2], El Gamal and Kim [29], Dai and Luo [18]). This is actually attained by employing the "one-time" CSI coding in the sense of Han *et al.* [22].

The achievability proof for $R_{\text{CSI-1}}(p(u), p(x|u,s))$ and $R_{\text{CSI-2}}(p(u), p(x|u,s))$ in Theorem 1 is provided in the next section.

Remark 1: Chia and El Gamal [17] have considered the WC with common CSI available at both Alice and Bob as illustrated in Fig. 2. This channel, however, equivalently reduces to that in Fig. 1 with output $Y_S \equiv SY$ instead of Y. Then, since $H(S|UY_S) = H(S|USY) = 0$, $R_{\text{CSI-1}}(p(u), p(x|u,s))$ and $R_{\text{CSI-2}}(p(u), p(x|u,s))$ in (4), (5) reduce to

$$R_{\text{CSI-1}}(p(u), p(x|u,s)) = \min\Big[I(U;SY) - I(U;SZ) + H(S|Z), I(U;SY)\Big], \tag{7}$$

$$R_{\text{CSI-2}}(p(u), p(x|u,s)) = \min\Big[H(S|UZ), I(U;SY)\Big], \tag{8}$$

where the right-hand side of (7) exactly coincides with the weak secrecy lower bound

$$\min\Big[I(U;SY) - I(U;SZ) + H(S|Z), I(U;SY)\Big] \tag{9}$$

that was given by Chia and El Gamal [17], while the right-hand side of (8) coincides with one more weak secrecy lower bound

$$\min\Big[H(S|UZ), I(U;SY)\Big] \tag{10}$$

that was also given by [17]. Thus, Theorem 1 specialized to the case with "common" CSI available at both Alice and Bob provides the *strong secrecy* version of their results. Specifically, this concludes that Theorems 1, 2 and 3 in [17] all hold with the strong secrecy criterion. □

Remark 2: A basic feature of this paper is that we do *not* invoke the argument of typical sequences at all, so we do not need the finiteness of alphabets $\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$, while the alphabet $\mathcal{S}$ of CSI S needs to be finite. □

III. Proof of Theorem 1

The whole coding scheme involves the transmission of b independent messages over the $b+1$ channel blocks each of length n (b is a sufficiently large fixed positive integer), which are indexed by $j = 0, 1, 2, \cdots, b$. The formal proof is provided in the sequel, where in block j we let $\mathbf{U}_j, \mathbf{S}_j, \mathbf{X}_j, \mathbf{Y}_j, \mathbf{Z}_j$ (correlated i.i.d. sequences of length n subject to joint probability P_{USXYZ}) denote the random variables to indicate channel input sequence, CSI sequence, channel input sequence for Alice, channel output sequences for Bob and Eve, respectively, whereas M_j, M_{0j}, M_{1j}, N_j denote the random variables to indicate uniformly distributed confidential messages to be sent, and auxiliary message, respectively. Their realizations are indicated by the corresponding lower case letters.

Case A): Proof for the Achievability of $R_{\text{CSI-1}}$*:* In what to follow, many kinds of (nonnegative) rates intervene with inequality constraints, which are listed as follows:

$$\overline{R} < I(U;Y), \tag{11}$$

$$R = R_0 + R_1, \tag{12}$$

$$\overline{R} - R_0 > I(U;SZ), \tag{13}$$

$$R_2 > H(S|UY), \tag{14}$$

$$R_0 + R_1 + R_2 < \overline{R}, \tag{15}$$

$$R_1 + R_2 < H(S|Z). \tag{16}$$

Fourier-Motzkin elimination (cf. El Gamal and Kim [29]) claims that the supremum of R over all rates satisfying (11)∼(16) coincides with the right-hand side of (4), so it suffices to show that rates R satisfying (11)∼(16) are indeed

[2]Notice here that the WC ω in this paper is equipped with no public authenticated noiseless channel between Alice and Bob unlike in the standard setting of secret key agreement, but all communications occur inside the WC ω in one-way fashion from Alice to Bob.

achievable, where $\overline{R}$ is used to indicate an achievable rate for usual channel coding (non-WC) between Alice and Bob.

Codebook Generation: For each block $j \in [1:b]$, split message $M_j \in [1:2^{nR}]$ into two independent *uniform* messages $M_{0j} \in [1:2^{nR_0}]$ and $M_{1j} \in [1:2^{nR_1}]$; thus $R = R_0 + R_1$, where, in the process of channel transmission, message M_{0j} is protected by WC coding, and message M_{1j} is protected by OTP cipher with the secret key shared using CSI. The codebook generation consists of the following two parts:

1) Message Codebook Generation: For each block $j \in [0:b]$, randomly and independently generate sequences $\mathbf{u}_j(l), l \in [1:2^{n\overline{R}}]$, each according to probability distribution $\prod_{i=1}^{n} p_U(u_i)$ ($\mathbf{u}_j(l) = u_1 u_2 \cdots u_n$). This is a random code and is denoted by $\mathcal{H}_j$. On the other hand, partition the set $[1:2^{n\overline{R}}]$ of indices into 2^{nR_0} equal-size bins $\mathcal{B}(m_0), m_0 \in [1:2^{nR_0}]$. Moreover, partition the indices within each bin $\mathcal{B}(m_0)$ into 2^{nR_1} equal-size sub-bins $\mathcal{B}(m_0, m_1), m_1 \in [1:2^{nR_1}]$. Furthermore, partition the indices within each bin $\mathcal{B}(m_0, m_1)$ into 2^{nR_2} equal-size sub-sub-bins $\mathcal{B}(m_0, m_1, m_2), m_2 \in [1:2^{nR_2}]$ (cf. Fig. 3 on the next page). These bins are all non-empty because of (15).

2) Key Codebook Generation: In order to construct an efficient key $K_j = \kappa(\mathbf{S}_j)$ of rate R_1 using the CSI $\mathbf{S}_j$, we invoke the following two celebrated lemmas:

Lemma 1 (Slepian and Wolf [3]): Let $\varepsilon > 0$ be an arbitrarily small number and let $R_2 > H(S|UY)$ (cf. (14)). Then, there exists (*deterministic*) functions $\sigma : \mathcal{S}^n \to [1:2^{nR_2}]$ and $\phi : [1:2^{nR_2}] \times \mathcal{U}^n \times \mathcal{Y}^n \to \mathcal{S}^n$ such that

$$\Pr\{\mathbf{S}_j \neq \tilde{\mathbf{S}}_j\} \leq \varepsilon \tag{17}$$

for all sufficiently large n, where $\tilde{\mathbf{S}}_j = \phi(\sigma(\mathbf{S}_j), \mathbf{U}_j, \mathbf{Y}_j)$. □

For simplicity, we use also the notation $N_{j+1} \equiv \sigma(\mathbf{S}_j)$, which is the random variable conveying the auxiliary message used for generating the common secret key between Alice and Bob.

Lemma 2 (Csiszár and Körner [27, Corollary 17.5]): Let $\varepsilon > 0$ be an arbitrarily small number and let $R_1 + R_2 < H(S|Z)$ (cf. (16)). Then, with the same $N_{j+1} \equiv \sigma(\mathbf{S}_j)$ as in Lemma 1, there exists a (*deterministic*) key function $\kappa : \mathcal{S}^n \to [1:2^{nR_1}]$ such that

$$\mathsf{S}\left(\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j)|\mathbf{Z}_j\right) \leq \varepsilon \tag{18}$$

for all sufficiently large n, where we use the notation (called the *security index*):[1]

$$\mathsf{S}(K|F) \overset{\Delta}{=} D(P_{KF}||Q_K \times P_F) \tag{19}$$

with the uniform distribution Q_K on the range of K, the KL divergence $D(\cdot||\cdot)$ and the product distribution $Q_K \times P_F$. □

We use the thus defined *deterministic* function $K_{j-1} \equiv \kappa(\mathbf{S}_{j-1})$ as the key to be used in the next block j.

[1]Specifically, in the proof of this lemma, it suffices to make uniform random hashing $(\kappa, \sigma) : \mathcal{S}^n \to [1:2^{nR_1}] \times [1:2^{nR_2}]$ (and hence uniform random binnning $\sigma : \mathcal{S}^n \to [1:2^{nR_2}]$ simultaneously) to construct a pair of *deterministic* mappings $\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j) \equiv (\kappa(\mathbf{S}_j), \sigma(\mathbf{S}_j))$ satisfying (17) and (18). This is possible owing to rate constraints $R_2 > H(S|UY)$ and $R_1 + R_2 < H(S|Z)$.

Encoding Scheme: We use the block coding scheme as in Fig. 4 on the next page, which is based on the block Markov coding scheme invented by Cover and El Gamal [16] (cf. Fig. 4) and applied to the WC with CSI by Chia and El Gamal [17]. The first block $j = 0$ provides only the CSI sequence $\mathbf{S}_0$ for Alice to be used for encoding in the second block $j = 1$ with $M_0 = N_0 =$ "1" (fixed dummy message). In each block $j \in [1:b]$, given a message triple ($M_{0j} = m_0$, $M_{1j} = m_1$, $N_j = m_2$), Alice first computes $c_j = k_{j-1} \oplus m_1$ (mod 2^{nR_1}) and let $L \overset{\Delta}{=} L(m_0, c_j, m_2)$ be the random index uniformly distributed on the bin $\mathcal{B}(m_0, c_j, m_2)$ with $k_{j-1} = \kappa(\mathbf{s}_{j-1})$ as specified in Lemma 2. Alice then sends out for channel transmission a randomly generated sequence $\mathbf{X}_j$ according to conditional probability $\prod_{i=1}^{n} p_{X|US}(x_i|u_i(L), s_i)$, where

$$\begin{aligned} \mathbf{x}_j &= x_1 x_2 \cdots x_n, \\ \mathbf{u}_j(L) &= u_1(L) u_2(L) \cdots u_n(L), \\ \mathbf{s}_j &= s_1 s_2 \cdots s_n. \end{aligned}$$

and we set $\mathbf{U}_j = \mathbf{u}_j(L)$ for simplicity.

Decoding Scheme and Evaluation of Probability of Error: Let $\mathbf{Y}_j$ be the output for Bob due to $\mathbf{U}_j$. Consider the stationary memoryless channel $\omega_n(\mathbf{y}|\mathbf{u}) \equiv P_{\mathbf{Y}_j|\mathbf{U}_j}(\mathbf{y}|\mathbf{u})$ with input $\mathbf{u}$ and output $\mathbf{y}$. For this channel we use the maximum likelihood decoding, that is, we let $\hat{l}$ denote an index such that

$$\omega_n(\mathbf{y}|\mathbf{u}_j(\hat{l})) = \max_{l \in [1:2^{n\overline{R}}]} \omega_n(\mathbf{y}|\mathbf{u}_j(l)), \tag{20}$$

and set $\hat{\mathbf{U}}_j = \mathbf{u}_j(\hat{l})$. Find the $(\hat{m}_0, \hat{c}, \hat{m}_2)$ such that $\hat{l} \in \mathcal{B}(\hat{m}_0, \hat{c}, \hat{m}_2)$. Next, compute $\hat{m}_1 = \hat{c} \ominus \hat{k}_{j-1}$ (mod 2^{nR_1}), where $\hat{k}_{j-1} = \kappa(\hat{\mathbf{s}}_{j-1})$ with $\hat{\mathbf{s}}_{j-1} = \phi(\hat{m}_2, \mathbf{u}_{j-1}(\hat{l}), \mathbf{y}_{j-1})$ and we notice that $\hat{\mathbf{s}}_{j-1} = \tilde{\mathbf{s}}_{j-1}$ if $\hat{m}_2 = m_2$ and $\mathbf{u}_{j-1}(\hat{l}) = \mathbf{u}_{j-1}(L)$ (cf. Lemmas 1 and 2). Finally, declare that the message pair $(\hat{m}_0, \hat{m}_1)$ was sent. In order to evaluate the probability of decoding error

$$Pe(j) \equiv \Pr\{(M_{0j}, M_{1j}) \neq (\hat{M}_{0j}, \hat{M}_{1j})|\mathcal{H}\}, \tag{21}$$

we invoke

Lemma 3 (Gallager [28, Theorem 5.6.2]): Let $\varepsilon > 0$ be an arbitrarily small number and let $\overline{R} < I(U;Y)$ (cf. (11)). Then,

$$\begin{aligned} &\Pr\{(M_{0j}, C_j, N_j \equiv M_{2j}, \mathbf{U}_j) \\ &\qquad \neq (\hat{M}_{0j}, \hat{C}_j, \hat{N}_j \equiv \hat{M}_{2j}, \hat{\mathbf{U}}_j)|\mathcal{H}\} \leq \varepsilon \end{aligned} \tag{22}$$

for all sufficiently large n. □

Then, in view of Lemmas 1 and 3, we have

$$\begin{aligned} Pe(j) &\leq \Pr\{(M_{0j}, C_j, N_j, \mathbf{U}_j) \neq (\hat{M}_{0j}, \hat{C}_j, \hat{N}_j, \hat{\mathbf{U}}_j)|\mathcal{H}\} \\ &\quad + \Pr\{\mathbf{S}_j \neq \tilde{\mathbf{S}}_j\} \\ &\leq 2\varepsilon. \end{aligned} \tag{23}$$

Thus, it is concluded that the total probability of decoding error over all the b blocks is less than or equal to $2b\varepsilon$. It should be remarked here that the total transmission rate averaged over all $b+1$ blocks is $\frac{bR}{b+1}$ because only the b blocks of them are effective for message transmission, which can be made as close to R as desired by letting b large enough.

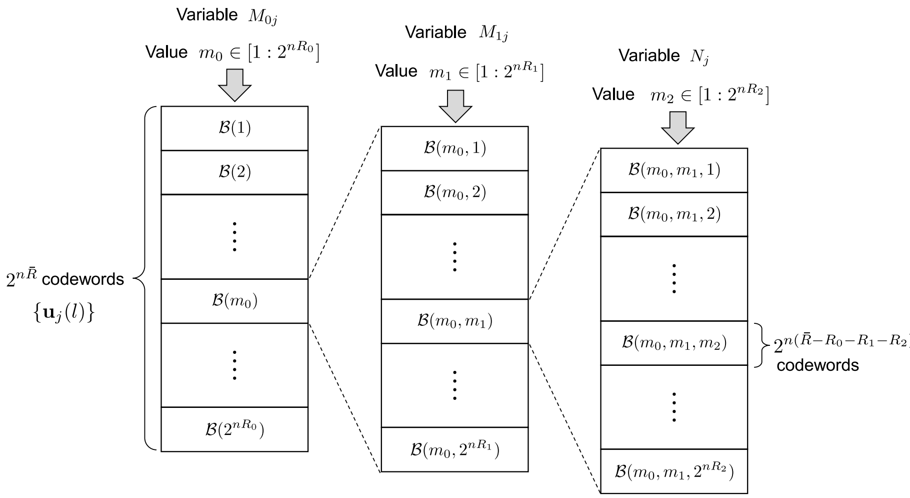

Fig. 3. Bin-partitioning for message codebook generation in each channel block j.

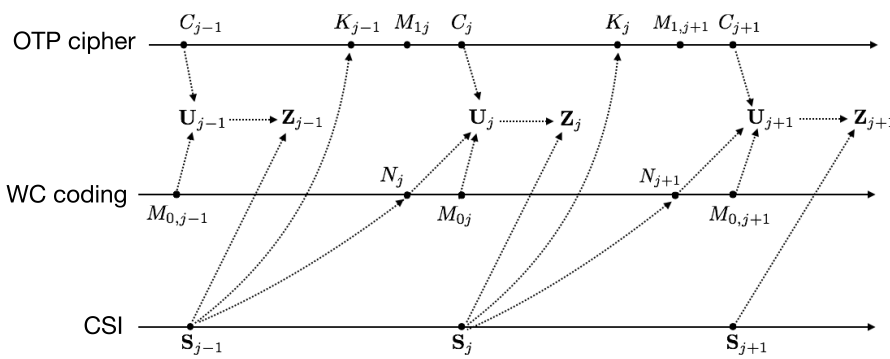

Fig. 4. Sequence diagram of block Markov coding ($C_j = K_{j-1} \oplus M_{1j}$; $j = 1, 2, \cdots, b$).

Evaluation of Information Leakage: We use the following notation: for $j \in [1:b]$,

$$\begin{aligned} \mathcal{H} &= \mathcal{H}_1\mathcal{H}_2\cdots\mathcal{H}_b, \\ M_j &= M_{0j}M_{1j}, \\ M^j &= M_1M_2\cdots M_j, \\ M^{[j]} &= M_jM_{j+1}\cdots M_b, \\ \mathbf{Z}^j &= \mathbf{Z}_1\mathbf{Z}_2\cdots\mathbf{Z}_j, \\ \mathbf{Z}^{[j]} &= \mathbf{Z}_j\mathbf{Z}_{j+1}\cdots\mathbf{Z}_b, \end{aligned}$$

where we notice that $\mathbf{Z}_j$ is the channel output for Eve in block j.

Remark 3: Since K_{j-1} and M_{1j} are independent and M_{1j} is assumed to be uniformly distributed, the OTP cipher claims that K_{j-1} and $C_j = K_{j-1} \oplus M_{1j}$ are independent and C_j is uniformly distributed (cf. Shannon [4]). Notice here that K_{j-1} is not necessarily uniformly distributed, and hence M_{1j} and C_j are *not* necessarily independent. On the other hand, $\mathbf{Z}_{j-1}$ may affect $\mathbf{Z}_j$ only through $K_{j-1}N_j$ and inversely $\mathbf{Z}_{j-1}$ may be affected by $\mathbf{Z}_j$ only through C_jN_j. This property plays the crucial role in evaluating the performance of our coding scheme (cf. Fig.4). □

In the sequel we show that the information leakage to Eve $I_E = I(M^b; \mathbf{Z}^b|\mathcal{H})$ over the whole $b+1$ blocks goes to zero as $n \to \infty$.

To do so, we begin with

$$\begin{aligned} A &\stackrel{\Delta}{=} I(M^b; \mathbf{Z}^b|\mathcal{H}) \\ &= \sum_{j=1}^{b} I(M_j; \mathbf{Z}^b|M^{[j+1]}\mathcal{H}) \\ &\stackrel{(a)}{\leq} \sum_{j=1}^{b} I(M_j; \mathbf{Z}^b|\mathbf{S}_jM^{[j+1]}\mathcal{H}) \\ &\stackrel{(b)}{=} \sum_{j=1}^{b} I(M_j; \mathbf{Z}^j|\mathbf{S}_j\mathcal{H}) \\ &= \sum_{j=1}^{b} I(M_{0j}M_{1j}; \mathbf{Z}^j|\mathbf{S}_j\mathcal{H}) \\ &= \sum_{j=1}^{b} I(M_{0j}M_{1j}; \mathbf{Z}^{j-1}|\mathbf{S}_j\mathcal{H}) \\ &\quad + \sum_{j=1}^{b} I(M_{0j}M_{1j}; \mathbf{Z}_j|\mathbf{Z}^{j-1}\mathbf{S}_j\mathcal{H}) \end{aligned}$$

$$\begin{aligned} &\stackrel{(c)}{=} \sum_{j=1}^{b} I(M_{0j}M_{1j};\mathbf{Z}_j|\mathbf{Z}^{j-1}\mathbf{S}_j\mathcal{H}) \\ &= B + C \end{aligned} \tag{24}$$

with

$$B \stackrel{\Delta}{=} \sum_{j=1}^{b} I(M_{0j};\mathbf{Z}_j|\mathbf{Z}^{j-1}\mathbf{S}_j\mathcal{H}) \tag{25}$$

$$C \stackrel{\Delta}{=} \sum_{j=1}^{b} I(M_{1j};\mathbf{Z}_j|M_{0j}\mathbf{Z}^{j-1}\mathbf{S}_j\mathcal{H}), \tag{26}$$

where (a) follows from the independence of M_j and $\mathbf{S}_j$ given $M^{[j+1]}\mathcal{H}$; (b) follows from the Markov chain property $M_j \to \mathbf{Z}^j\mathbf{S}_j \to \mathbf{Z}^{[j+1]}M^{[j+1]}$ given $\mathcal{H}$; (c) follows from the independence of $M_{0j}M_{1j}\mathbf{S}_j$ and $\mathbf{Z}^{j-1}$ given $\mathcal{H}$.

Let us now separately evaluate B and C in (25) and (26). First,

$$\begin{aligned} B &= \sum_{j=1}^{b} I(M_{0j};\mathbf{Z}_j|\mathbf{Z}^{j-1}\mathbf{S}_j\mathcal{H}) \\ &\le \sum_{j=1}^{b} I(\mathbf{Z}^{j-1}M_{0j};\mathbf{Z}_j|\mathbf{S}_j\mathcal{H}) \\ &= \sum_{j=1}^{b} I(M_{0j};\mathbf{Z}_j|\mathbf{S}_j\mathcal{H}) \\ &\quad + \sum_{j=1}^{b} I(\mathbf{Z}^{j-1};\mathbf{Z}_j|M_{0j}\mathbf{S}_j\mathcal{H}) \\ &\le \sum_{j=1}^{b} I(M_{0j};\mathbf{Z}_j|\mathbf{S}_j\mathcal{H}) \\ &\quad + \sum_{j=1}^{b} I(\mathbf{Z}^{j-1};N_jM_{0j}\mathbf{S}_j\mathbf{Z}_j|\mathcal{H}) \\ &\stackrel{(d)}{=} \sum_{j=1}^{b} I(M_{0j};\mathbf{S}_j\mathbf{Z}_j|\mathcal{H}) \\ &\quad + \sum_{j=1}^{b} I(N_j;\mathbf{Z}^{j-1}|\mathcal{H}), \end{aligned} \tag{27}$$

where (d) follows from the independence of M_{0j} and $\mathbf{S}_j$ and from the Markov chain property $\mathbf{Z}^{j-1} \to N_j \to M_{0j}\mathbf{S}_j\mathbf{Z}_j$ given $\mathcal{H}$.

Next, C can be upper bounded as

$$\begin{aligned} C &= \sum_{j=1}^{b} I(M_{1j};\mathbf{Z}_j|M_{0j}\mathbf{Z}^{j-1}\mathbf{S}_j\mathcal{H}) \\ &\le \sum_{j=1}^{b} I(\mathbf{Z}^{j-1}M_{1j};\mathbf{Z}_j|M_{0j}\mathbf{S}_j\mathcal{H}) \\ &= D + E, \end{aligned} \tag{28}$$

where

$$D \stackrel{\Delta}{=} \sum_{j=1}^{b} I(M_{1j};\mathbf{Z}_j|M_{0j}\mathbf{S}_j\mathcal{H}) \tag{29}$$

$$E \stackrel{\Delta}{=} \sum_{j=1}^{b} I(\mathbf{Z}^{j-1};\mathbf{Z}_j|M_{0j}M_{1j}\mathbf{S}_j\mathcal{H}). \tag{30}$$

Then,

$$\begin{aligned} D &\le \sum_{j=1}^{b} I(M_{1j};C_j\mathbf{Z}_j|M_{0j}\mathbf{S}_j\mathcal{H}) \\ &= F + G, \end{aligned} \tag{31}$$

where

$$F = \sum_{j=1}^{b} I(M_{1j};C_j|M_{0j}\mathbf{S}_j\mathcal{H}),$$

$$G = \sum_{j=1}^{b} I(M_{1j};\mathbf{Z}_j|M_{0j}\mathbf{S}_jC_j\mathcal{H}). \tag{32}$$

Then,

$$\begin{aligned} F &= \sum_{j=1}^{b} I(M_{1j};C_j|M_{0j}\mathbf{S}_j\mathcal{H}) \\ &\stackrel{(f)}{=} \sum_{j=1}^{b} I(M_{1j};C_j) \\ &= H(C_j) - H(C_j|M_{1j}) \\ &\stackrel{(k)}{=} H(C_j) - H(K_{j-1}|M_{1j}) \\ &\stackrel{(g)}{=} H(C_j) - H(K_{j-1}) \\ &\stackrel{(p)}{=} D(P_{K_{j-1}}||Q_{K_{j-1}}), \end{aligned} \tag{33}$$

where (f) follows from the independence of $M_{1j}C_j$ and $M_{0j}\mathbf{S}_j\mathcal{H}$; (k) follows from $K_{j-1} \oplus M_{1j} = C_j$; (g) follows from the independence of K_{j-1} and M_{1j}; (p) follows from that C_j is uniformly distributed on the range of K_{j-1}.

Moreover,

$$\begin{aligned} G &= \sum_{j=1}^{b} I(M_{1j};\mathbf{Z}_j|M_{0j}\mathbf{S}_jC_j\mathcal{H}) \\ &\stackrel{(e)}{=} \sum_{j=1}^{b} I(K_{j-1};\mathbf{Z}_j|M_{0j}\mathbf{S}_jC_j\mathcal{H}) \\ &\stackrel{(j)}{\le} \sum_{j=1}^{b} I(K_{j-1};N_j|M_{0j}\mathbf{S}_jC_j\mathcal{H}) \\ &\stackrel{(m)}{=} \sum_{j=1}^{b} I(K_{j-1};N_j), \end{aligned} \tag{34}$$

where (e) follows from $C_j = K_{j-1} \oplus M_{1j}$; (j) follows from the data processing lemma using the Markov chain property $K_{j-1} \to N_j \to \mathbf{Z}_j$ given $M_{0j}\mathbf{S}_jC_j\mathcal{H}$; (m) follows from the independence of $K_{j-1}N_j$ and $M_{0j}\mathbf{S}_jC_j\mathcal{H}$.

On the other hand,

$$\begin{aligned} E &\le \sum_{j=1}^{b} I(\mathbf{Z}^{j-1};K_{j-1}N_j\mathbf{Z}_j|M_{0j}M_{1j}\mathbf{S}_j\mathcal{H}) \\ &\stackrel{(h)}{=} \sum_{j=1}^{b} I(\mathbf{Z}^{j-1};K_{j-1}N_j|M_{0j}M_{1j}\mathbf{S}_j\mathcal{H}) \\ &\stackrel{(i)}{=} \sum_{j=1}^{b} I(K_{j-1}N_j;\mathbf{Z}^{j-1}|\mathcal{H}), \end{aligned} \tag{35}$$

where (h) follows from the Markov chain property $\mathbf{Z}^{j-1} \to K_{j-1}N_j \to \mathbf{Z}_j$ given $M_{0j}M_{1j}\mathbf{S}_j\mathcal{H}$; (i) follows from the independence of $\mathbf{Z}^{j-1}K_{j-1}N_j\mathcal{H}$ and $M_{0j}M_{1j}\mathbf{S}_j$.

Thus, summarizing up (24)$\sim$ (35), we have the upper bound on the information leakage to Eve $I_E = I(M^b; \mathbf{Z}^b|\mathcal{H})$ as

Lemma 4 (Information leakage bound):

$$I(M^b; \mathbf{Z}^b|\mathcal{H}) \leq \sum_{j=1}^{b} I(M_{0j}; \mathbf{S}_j\mathbf{Z}_j|\mathcal{H}) \tag{36}$$
$$+\sum_{j=1}^{b} I(N_j; \mathbf{Z}^{j-1}|\mathcal{H}). \tag{37}$$
$$+\sum_{j=1}^{b} I(K_{j-1}; N_j) \tag{38}$$
$$+\sum_{j=1}^{b} D(P_{K_{j-1}}||Q_{K_{j-1}}) \tag{39}$$
$$+\sum_{j=1}^{b} I(K_{j-1}N_j; \mathbf{Z}^{j-1}|\mathcal{H}). \tag{40}$$

□

Here, the first term $I(M_{0j}; \mathbf{S}_j\mathbf{Z}_j|\mathcal{H})$ specifies the resolvability performance for Eve; the second term $I(N_j; \mathbf{Z}^{j-1}|\mathcal{H})$ specifies the inter-block interaction effect in the block Markov coding scheme; the third and fourth terms $I(K_{j-1}; N_j)$, $D(P_{K_{j-1}}||Q_{K_{j-1}})$ specify the key performance for Bob; and the fifth term $I(K_{j-1}N_j; \mathbf{Z}^{j-1}|\mathcal{H})$ specifies the key performance for Eve.

The third and fourth ones are evaluated as follows. We can rewrite the security index $\mathsf{S}(\kappa(\mathbf{S}_j)\sigma(sS_j)|\mathbf{Z}_j)$ in (18) of Lemma 2 as

$$\begin{aligned}
&\mathsf{S}(\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j)|\mathbf{Z}_j)\\
&\quad\geq D(P_{\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j)}||Q_{\kappa(\mathbf{S}_j)} \times Q_{\sigma(\mathbf{S}_j)})\\
&\quad= D(P_{\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j)}||P_{\kappa(\mathbf{S}_j)} \times P_{\sigma(\mathbf{S}_j)})\\
&\qquad+D(P_{\kappa(\mathbf{S}_j)}||Q_{\kappa(\mathbf{S}_j)}) + D(P_{\sigma(\mathbf{S}_j)}||Q_{\sigma(\mathbf{S}_j)})\\
&\quad\geq D(P_{\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j)}||P_{\kappa(\mathbf{S}_j)} \times P_{\sigma(\mathbf{S}_j)})\\
&\quad= I(\kappa(\mathbf{S}_j); \sigma(\mathbf{S}_j))\\
&\quad= I(K_j; N_{j+1}).
\end{aligned} \tag{41}$$

Moreover,

$$\begin{aligned}
&\mathsf{S}(\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j)|\mathbf{Z}_j)\\
&\quad\geq \mathsf{S}(\kappa(\mathbf{S}_j)|\mathbf{Z}_j)\\
&\quad= D(P_{K_j}||Q_{K_j}) + I(K_j; \mathbf{Z}_j)\\
&\quad\geq D(P_{K_j}||Q_{K_j}).
\end{aligned} \tag{42}$$

Therefore, Lemma 2 claims that

$$I(K_{j-1}; N_j) \leq \varepsilon, \tag{43}$$
$$D(P_{K_{j-1}}||Q_{K_{j-1}}) \leq \varepsilon. \tag{44}$$

In order to evaluate the second and fifth ones, we use the following lemma, which is the Alice-only CSI counterpart of [17, Proposition 1]:

Lemma 5 (Key secrecy lemma): Let $\varepsilon > 0$ be an arbitrarily small number and let $R_1 + R_2 < H(S|Z)$ (cf. (16)). Then, for $j \in [1 : b]$,

$$\text{i)}\quad I(K_{j-1}N_j; \mathbf{Z}_{j-1}|\mathcal{H}) \leq \varepsilon, \tag{45}$$
$$\text{ii)}\quad I(K_{j-1}N_j; \mathbf{Z}^{j-1}|\mathcal{H}) \leq b\varepsilon \tag{46}$$

for all sufficiently large n. □

Proof: See Appendix A.

From (46) we immediately have

$$I(N_j; \mathbf{Z}^{j-1}|\mathcal{H}) \leq I(K_{j-1}N_j; \mathbf{Z}^{j-1}|\mathcal{H}) \leq b\varepsilon. \tag{47}$$

Now, what remains to be done is to evaluate the first one $I(M_{0j}; \mathbf{S}_j\mathbf{Z}_j|\mathcal{H})$. To do so, we invoke the following resolvability lemma:

Lemma 6 (Resolvability lemma): Let $\varepsilon > 0$ be an arbitrarily small number and let $\overline{R} - R_0 > I(U; SZ)$ (cf. (13)). Then,

$$I(M_{0j}; \mathbf{S}_j\mathbf{Z}_j|\mathcal{H}) \leq \varepsilon \tag{48}$$

for all sufficiently large n. □

Proof: See Appendix B. □

An immediate consequence of Lemma 4 together with (43), (44), (47) and (48) is

$$I(M^b; \mathbf{Z}^b|\mathcal{H}) \leq (3b + 2b^2)\varepsilon, \tag{49}$$

thereby completing the proof for *Case A)*. □

Case B): Proof for the Achievability of $R_{\text{CSI-2}}$: The remainder of Theorem 1 to be proved is the acievability of $R_{\text{CSI-2}}(p(u), p(x|u, s))$ in (5).

The rate constraints in this case are listed as follows ($R_0 = 0$):

$$\overline{R} < I(U; Y), \tag{50}$$
$$R = R_1, \tag{51}$$
$$R_2 > H(S|UY), \tag{52}$$
$$R_1 + R_2 < \overline{R}, \tag{53}$$
$$R_1 + R_2 < H(S|UZ). \tag{54}$$

These constraints are the same as those in *Case A)* with $R_0 = 0$ and $H(S|UZ)$ instead of $H(S|Z)$, where constraint (13) is not necessary here because of $R_0 = 0$. The reason for the replacement of $H(S|Z)$ by $H(S|UZ)$ is that, since $R_0 = 0$, we cannot here leverage the randomization (over input $\mathbf{U}_j$) due to Wyner's WC coding to keep the $\mathbf{U}_j$ secure from the attack by Eve.

Fourier-Motzkin elimination claims that the supremum of R over all rates satisfying (50)$\sim$ (54) coincides with the $R_{\text{CSI-2}}(p(u), p(x|u, s))$, so it suffices to show that rates R satisfying (50)$\sim$ (54) are achievable.

In this case too, the proof argument parallels those as developed in the proof for *Case A)* with $R_0 = 0$, where we notice that $I(M_{0j}; \mathbf{S}_j\mathbf{Z}_j|\mathcal{H}) = 0$ in Lemma 4 and hence Lemma 6 is not needed here, thereby completing the achievability proof for this case. □

IV. Secrecy Capacity Results

Thus far we have developed achievability arguments for WCs with CSI available only at the encoder (Alice) to establish Theorem 1 on lower bounds to the secrecy capacity $C_{\text{CSI-E}}$. In this section, in order to get more insights into this theorem, we address the problem of bounding the secrecy capacities for the case of (*statistically*) degraded WCs, which is an important class of WCs.

Let us first describe the first theorem in this section:

Theorem 2: For any degraded WC (Z is a degraded version of Y) with causal CSI only at Alice, we have

$$C_{\text{CSI-E}} \geq \max_{p(x|s)} \min\big(I(XS;Y) - I(XS;Z), I(XS;Y) - H(S)\big), \tag{55}$$

$$C_{\text{CSI-E}} \leq \max_{p(x|s)} \min\big(I(XS;Y) - I(XS;Z), I(XS;Y) - I(S;Y)\big), \tag{56}$$

$$C^{\mathsf{K}}_{\text{NCSI-E}} \geq \max_{I(XS;Y)\geq H(S)} \big(I(XS;Y) - I(XS;Z)\big), \tag{57}$$

$$C^{\mathsf{K}}_{\text{NCSI-E}} \leq \max_{p(x|s)} \big(I(XS;Y) - I(XS;Z)\big), \tag{58}$$

where $C^{\mathsf{K}}_{\text{NCSI-E}}$ denotes the non-causal *secret-key capacity* (as for the definition, see, e.g., Khisti *et al.* [15], Prabhakaran *et al.* [24], Bunin *et al.* [26]). In contrast with this, $C_{\text{CSI-E}}$ may be called the *secret-message capacity*. The maximization in (57) is taken over all XS such that $I(XS;Y) \geq H(S)$. □

Remark 4: Lower bounds (55) and (57) hold without the assumption of degradedness. It is is easy to check that $I(XS;Y) - I(XS;Z)$ in (55) $\sim$ (58) is nonnegative for degraded WCs, while $I(XS;Y) - H(S)$ in (55) may be negative. □

Proof of (55) (Achievability): Let (X,S) be arbitrarily given, then the functional representation lemma [29] claims that there exist a random variable U and a deterministic function $f : \mathcal{U} \times \mathcal{S} \to \mathcal{X}$ such that U and S are independent and $X = f(U,S)$.

Then, the first term of the achievable rate $R_{\text{CSI-1}}(p(u), p(x|u,s))$ given in Theorem 1 can be rewritten as follows.

$$\begin{aligned}
&I(U;Y) - I(U;SZ) + H(S|Z) - H(S|UY)\\
&\quad= I(U;SY) - I(U;S|Y) - I(U;SZ)\\
&\qquad +H(S|Z) - H(S|UY)\\
&\quad= I(U;Y|S) - I(U;Z|S) + H(S|Z) - H(S|Y)\\
&\quad\overset{(v)}{=} I(XU;Y|S) - I(XU;Z|S) + H(S|Z) - H(S|Y)\\
&\quad\overset{(w)}{=} I(X;Y|S) - I(X;Z|S) + H(S|Z) - H(S|Y)\\
&\quad= I(XS;Y) - I(XS;Z),
\end{aligned} \tag{59}$$

where (v) follows from that X is a deterministic function of (U,S); (w) follows from that $U \to SX \to YZ$ forms a Markov chain.

On the other hand, the second term of $R_{\text{CSI-1}}(p(u), p(x|u,s))$ can be rewritten as follows.

$$\begin{aligned}
&I(U;Y) - H(S|UY)\\
&\quad= I(U;SY) - I(U;S|Y) - H(S|UY)\\
&\quad= I(U;Y|S) - H(S|Y)\\
&\quad\overset{(y)}{=} I(XU;Y|S) - H(S|Y)\\
&\quad\overset{(z)}{=} I(X;Y|S) - H(S|Y)\\
&\quad= I(XS;Y) - H(S),
\end{aligned} \tag{60}$$

where in $(y),(z)$ we have used the similar argument to $(v),(w)$.

Therefore, in view of Theorem 1, combining (59) and (60) yields (55).

Proof of (56) (Converse): Here, we invoke the following simple but powerful lemma:

Lemma 7 (Chen and Vinck [7]): Let us consider a degraded WC with CSI S such that Z is a degraded version of Y. Then, the secrecy capacity with *non-causal* CSI only at the encoder (=E), denoted by $C_{\text{NCSI-E}}$, is upper bounded as

$$C_{\text{NCSI-E}} \leq \max_{p(u|s)p(x|u,s)} (I(U;Y) - I(U;Z)), \tag{61}$$

where we notice that U and S may be correlated. □

We compute $I(U;Y)$ and $I(U;Z)$ separately with arbitrary USX.

$$\begin{aligned}
&I(U;Y)\\
&\quad= I(USX;Y) - I(SX;Y|U)\\
&\quad= I(S;Y) + I(UX;Y|S) - I(S;Y|U) - I(X;Y|US)\\
&\quad= I(X;Y|S) + I(S;Y) - I(S;Y|U) - I(X;Y|US).
\end{aligned} \tag{62}$$

Similarly,

$$\begin{aligned}
&I(U;Z)\\
&\quad= I(X;Z|S) + I(S;Z) - I(S;Z|U) - I(X;Z|US).
\end{aligned} \tag{63}$$

Hence,

$$\begin{aligned}
&I(U;Y) - (U;Z)\\
&\quad= I(X;Y|S) - I(X;Z|S) + I(S;Y) - I(S;Z)\\
&\qquad -\big(I(S;Y|U) - I(S;Z|U)\big)\\
&\qquad -\big(I(X;Y|US) - I(X;Z|US)\big)\\
&\quad\leq I(X;Y|S) - I(X;Z|S) + I(S;Y) - I(S;Z)\\
&\quad= I(X;Y|S) - I(X;Z|S) + H(S|Z) - H(S|Y)\\
&\quad= I(XS;Y) - I(XS;Z),
\end{aligned} \tag{64}$$

where in the above inequality we have used the propertry $I(S;Y|U) - I(S;Z|U) \geq 0$ and $I(X;Y|US) - I(X;Z|US) \geq 0$, which comes from the assumed degradedness.

Another upper bound $R \leq I(SX;Y) - I(S;Y)$ is derived as follows. For any achievable rate R, Fano inequality yields (with $\varepsilon_n \to 0$ as n tends to ∞):

$$\begin{aligned} nR &= H(M) \\ &\le H(M) - H(M|Y^n) + n\varepsilon_n \\ &= I(M;Y^n) + n\varepsilon_n \\ &= \sum_{i=1}^{n} I(M;Y_i|Y^{i-1}) + n\varepsilon_n \\ &\le \sum_{i=1}^{n} I(MY^{i-1};Y_i) + n\varepsilon_n \\ &\le \sum_{i=1}^{n} I(MY^{i-1};S_iY_i) + n\varepsilon_n \\ &\overset{(p)}{=} \sum_{i=1}^{n} I(MY^{i-1};Y_i|S_i) + n\varepsilon_n \\ &\le \sum_{i=1}^{n} I(X_iMY^{i-1};Y_i|S_i) + n\varepsilon_n \\ &\overset{(q)}{=} \sum_{i=1}^{n} I(X_i;Y_i|S_i) + n\varepsilon_n \\ &\overset{(r)}{=} nI(X_J;Y_J|S_JJ) + n\varepsilon_n \\ &\le nI(JX_J;Y_J|S_J) + n\varepsilon_n \\ &\overset{(s)}{=} nI(X_J;Y_J|S_J) + n\varepsilon_n \\ &\overset{(t)}{=} nI(X;Y|S) + n\varepsilon_n, \end{aligned} \tag{65}$$

where (p) comes from the independence of S_i and MY^{i-1}; (q) follows from the Markov chain property $MY^{i-1} \to X_iS_i \to Y_i$; in (r) J is the random variable such that $\Pr\{J = i)\} = \frac{1}{n}$ $(i = 1, \cdots, n)$; (s) follows from the Markov chain property $J \to X_JS_J \to Y_J$; in (t) we have set $X = X_J, Y = Y_J, S = S_J$.

An immediate consequence (deviding by n and letting $n \to \infty$) of (65) is

$$\begin{aligned} R &\le I(X;Y|S) \\ &= I(XS;Y) - I(S;Y) \end{aligned} \tag{66}$$

with input[2] X. Thus, combining (64) and (66) together with Lemma 7 yields (56). □

Secret-Key Capacity Results: We see that there is a gap between the second terms of (55) and (56), i.e., $H(S) \neq I(S;Y)$. These terms are due to the physical channel capability limitation, which are indispensable when we are concerned with the *secret-message capacity* like in the foregoing. On the other hand, however, as far as we are concerned with the *secret-key capacity*, such terms are not necessarily involved.

Proof of (57) (Achievability): We first invoke the following achievability theorem:

Theorem 3 (Khisti et al. *[15]):* For any WC, the (weak) secret-key capacity with *non-causal* CSI available only at the encoder is lower bounded as

$$C^{\mathsf{K}}_{\text{NCSI-E}} \ge \max_{I(V;Y)\ge I(V;S)} (I(V;Y) - I(V;Z)), \tag{67}$$

where the maximization in (67) is taken over all VS such that $I(V;Y) \ge I(V;S)$ and we notice that V and S may be correlated. □

Remark 5: The "causal" version of formula (67) in Theorem 3 is given by

$$C^{\mathsf{K}}_{\text{CSI-E}} \ge \max_{I(V;Y)\ge I(V;S)} (I(V;Y) - I(V;Z)), \tag{68}$$

where $V = (U,S)$ (U and S are independent) and $C^{\mathsf{K}}_{\text{CSI-E}}$ denotes the (strong) secret-key capacity with *causal* CSI available only at the encoder. Accordingly, $C^{\mathsf{K}}_{\text{NCSI-E}}$ in (57) and (58) can be replaced by $C^{\mathsf{K}}_{\text{CSI-E}}$. The proof of (68) will be given in a forthcoming paper [33] as a special case of more general *causal* WCs. □

Now, let (X,S) be arbitrarily given and let U and f be those as specified by the functional representation lemma [29] as in the proof of (55). We then compute the right-hand side of (67) with $V = (U,S)$ as follows:

$$\begin{aligned} &I(US;Y) - I(US;Z) \\ &\overset{(b)}{=} I(USX;Y) - I(USX;Z) \\ &\overset{(c)}{=} I(XS;Y) - I(XS;Z), \end{aligned} \tag{69}$$

where in (b) we noticed that X is a deterministic function of (U,S); (c) follows from that $U \to XS \to YZ$ forms a Markov chain. On the other hand,

$$\begin{aligned} &I(US;Y) - I(US;S) \\ &= I(SU;Y) - H(S) \\ &\overset{(d)}{=} I(XSU;Y) - H(S) \\ &\overset{(e)}{=} I(XS;Y) - H(S), \end{aligned} \tag{70}$$

where (d) follows since X is a deterministic function of (U,S); (e) follows from the Markov chain property $U \to XS \to Y$. Thus, Theorem 3 together with (69) and (70) yields (57).

Proof of (58) (Converse): To show the converse part, we first observe that Lemma 7 is still valid with $C^{\mathsf{K}}_{\text{NCSI-E}}$ instead of $C_{\text{NCSI-E}}$, which can be ascertained by carefully scrutinizing the proof in [7] (with secret key K instead of secret message M) of Lemma 7. Then, in the entirely same way as above, we have (64), implying the converse here. □

An immediate consequence of Theorem 2 is the following corollaries with degraded WCs, where, hereafter, we denote by $C_{\text{CSI-ED}}, C_{\text{NCS-ED}}, C^{\mathsf{K}}_{\text{NCSI-ED}}$ the (strong) secrecy capacities of WCs with common CSI S available at both the enceder (=E) and decoder (=D):

Corollary 1 (Strengthening of Chia and El Gamal [17]): It holds that

$$\begin{aligned} C_{\text{CSI-ED}} &= C_{\text{NCSI-ED}} \\ &= \max_{p(x|s)} \min\big(I(XS;YS) - I(XS;Z), \\ &\qquad I(XS;YS) - H(S)\big) \\ &= \max_{p(x|s)} \min\big(I(X;Y|S) - I(X;Z|S) + H(S|Z), \\ &\qquad I(X;Y|S)\big) \end{aligned} \tag{71}$$

[2]Actually, in order to conclude (56), we need to show that X in (64) and X (66) can be taken to be the same. However, this can be ascertained by carefully scritinizing the proof of Lemma 7.

Corollary 2: It holds that

$$\begin{aligned} C^{\mathsf{K}}_{\text{CSI-ED}} &= C^{\mathsf{K}}_{\text{NCSI-ED}} \\ &= \max_{p(x|s)} \big(I(XS;YS) - I(XS;Z)\big) \\ &= \max_{p(x|s)} \big(I(X;Y|S) - I(X;Z|S) + H(S|Z)\big). \end{aligned} \quad (72)$$

Proof: It suffices to replace Y by SY in (55) $\sim$ (58), where we have taken account of Remark 5. □.

Remark 6: In fact, Khisti *et al.* [15] has, instead of (72), given the following (weak) formula (*not* assuming the degradedness) as:

$$C^{\mathsf{K}}_{\text{NCSI-ED}} = \max_{p(u,x|s)} \big(I(U;Y|S) - I(U;Z|S) + H(S|Z)\big). \quad (73)$$

However, the proof in [15] for the converse part seems to contain a serious technical flaw. □

Next, following Chia and El Gamal [17], let us consider the following special WC to have

Corollary 3: Let us consider a degraded WC such that Z is a degraded version of Y and $p(y,z|x,s) = p(y,z|x)$, then we have

$$C_{\text{CSI-E}} = C_{\text{NCSI-E}} = \max_{p(x)}(I(X;Y) - I(X;Z)). \quad (74)$$

Remark 7: This result coincides with an intuition that this WC may reduce simply to a WC *without* CSI at Alice and Bob, because CSI S at Alice has no correlation to Bob. In this connection, it will be useful to compare this result with that in [17] with common CSI S available at both the encoder and decoder, the secrecy capacity of which is given as

$$\begin{aligned} &C_{\text{CSI-ED}} = C_{\text{NCSI-ED}} \\ &= \max_{p(x)} \min[I(X;Y) - I(X;Z) + H(S), I(X;Y)]. \end{aligned} \quad (75)$$

Clearly, in (75) the state information S contributes to making achievable rates higher by $H(S)$, whereas in (74) the CSI makes no contribution. This shows that "two-sided" CSI (available both at Alice and Bob) indeed can outperform "one-sided" CSI (available only at Alice). □

Proof of Corollary 3: We first observe that

$$\begin{aligned} R_{\text{CSI-0}}(p(u), p(x|u,s)) &= I(U;Y) - I(U;Z) \\ &= I(X;Y) - I(X;Z) \end{aligned} \quad (76)$$

by setting $X = U$ with S independent of X, which implies the achievability.

In order to show the converse part, we compute as follows:

$$\begin{aligned} &I(U;Y) - I(U;Z) \\ &= I(UX;Y) - I(X;Y|U) \\ &\quad -I(UX;Z) + I(X;Z|U) \\ &= I(X;Y) - I(X;Z) \\ &\quad -I(X;Y|U) + I(X;Z|U). \end{aligned} \quad (77)$$

On the other hand, owing to the assumed degradedness, we have

$$\begin{aligned} &I(X;Y|U) \\ &= I(X;ZY|U) \\ &= I(X;Z|U) + I(X;Y|UZ). \end{aligned} \quad (78)$$

From (77) and (78), it follows that

$$I(U;Y) - I(U;Z) \le I(X;Y) - I(X;Z). \quad (79)$$

Thus, in light of Lemma 7 together with (76) and (79), the corollary is concluded. □

So far, we have studied WCs with *non-binary* alphabets. It would also be interesting to see what happens with *binary* WCs ($\mathcal{U} = \mathcal{X} = \mathcal{Y} = \mathcal{Z} = \mathcal{S} = \{0,1\}$). Letting $\oplus$ denote the exclusive OR, we consider the binary WC defined by

$$Y = X \oplus S \oplus \Psi, \quad (80)$$
$$Z = X \oplus S \oplus \Phi, \quad (81)$$

where X, S, Φ, Ψ are mutually independent. and Φ, Ψ play the role of *external* "additive" noises independent from the CSI S. We assume here that $H(\Phi) > H(\Psi)$ and hence Z is a degraded version of Y in (80) and (81).

Theorem 4: For the thus defined binary degraded WC, we have

$$C_{\text{CSI-E}} = C_{\text{NCSI-E}} = H(\Phi) - H(\Psi). \quad (82)$$

Remark 8: For comparison, let us consider the case where the encoder is *not* provided the CSI S. In this case, it is natural to regard S as an additive noise to the channel, then we have the secrecy capacity C_{M} without CSI:

$$C_{\text{M}} = H(S \oplus \Phi) - H(S \oplus \Psi). \quad (83)$$

It is obvious that

$$H(\Phi) - H(\Psi) > H(S \oplus \Phi) - H(S \oplus \Psi), \quad (84)$$

which implies that the existence of CSI S can indeed outperform the channel without CSI. Formula (82) means that the secrecy capacity for this WC does not depend on S, which is a consequence of elimination of "noise" S by making use of the CSI and is in nice accordance with the formula of Costa [32] on writing on (Gaussian) dirty paper. A Gaussian counterpart is discussed also in Khisti *et al.* [15]. □

Proof of Theorem 4: Set $X = U \oplus S$ where U and X, S, Φ, Ψ are independent, then

$$Y = U \oplus \Psi, \quad (85)$$
$$Z = U \oplus \Phi. \quad (86)$$

To show the achievability part, it suffices only to consider

$$\begin{aligned} &R_{\text{CSI-0}}(p(u), p(x|u,s)) \\ &= I(U;Y) - I(U;Z) \\ &= H(U) - H(U|Y) - (H(U) - H(U|Z)) \\ &= H(U|Z) - H(U|Y) \\ &= H(U|U \oplus \Phi) - H(U|U \oplus \Psi) \\ &= H(\Phi|U \oplus \Phi) - H(\Psi|U \oplus \Psi) \\ &= H(\Phi) - H(\Psi), \end{aligned} \quad (87)$$

where the last step follows by setting $U \sim (1/2,\ 1/2)$, which implies the achievability.

On the other hand, in order to show the converse part, we invoke (61) of Lemma 7. Let us evaluate the right-hand side of (61) as follows:

$$\begin{aligned} I(U;Y) &= I(U; X \oplus S \oplus \Psi) \\ &\stackrel{(a)}{=} I(U \oplus S; X \oplus \Psi) \\ &= I(X, U \oplus S; X \oplus \Psi) \\ &\quad - I(X; X \oplus \Psi | U \oplus S) \\ &= I(X; X \oplus \Psi) \\ &\quad - I(X; X \oplus \Psi | U \oplus S), \end{aligned} \tag{88}$$

where in (a) we noticed that $(U, X \oplus S \oplus \Psi)$ and $(U \oplus S, X \oplus \Psi)$ are in one-to-one correspondence under operation $\oplus S$.

Similarly, we have

$$\begin{aligned} I(U;Z) &= I(X; X \oplus \Phi) \\ &\quad - I(X; X \oplus \Phi | U \oplus S). \end{aligned} \tag{89}$$

Hence,

$$\begin{aligned} I(U;Y) &- I(U;Z) \\ &= I(X; X \oplus \Psi) - I(X; X \oplus \Phi) \\ &\quad -(I(X; X \oplus \Psi | U \oplus S) - I(X; X \oplus \Phi | U \oplus S)) \end{aligned} \tag{90}$$

We now notice that $X \oplus \Phi$ is a degraded version of $X \oplus \Psi$ to obtain

$$I(X; X \oplus \Psi | U \oplus S) \geq I(X; X \oplus \Phi) | U \oplus S), \tag{91}$$

from which together with (90) it follows that

$$I(U;Y) - I(U;Z) \leq I(X; X \oplus \Psi) - I(X; X \oplus \Phi). \tag{92}$$

It is easy also to see that

$$\max_{p(x)} (I(X; X \oplus \Psi) - I(X; X \oplus \Phi)) = H(\Phi) - H(\Psi), \tag{93}$$

where max can be attained with $X \sim (1/2.1/2)$, which implies the converse. □

In passing this section, let us look back at Theorem 2 to scrutinize more the significance. We first notice that the achievability of $R_{\text{CSI-1}}(p(u), p(x|u,s))$ in Theorem 1 (and hence the achievability (55) in Theorem 2) is based on one-time pad cipher that is attained by reproducing CSI S^n at Alice as $\hat{S}^n$ at Bob. Furthermore, the achievability in Theorem 3 with $V = (U, S)$ (and hence the achievability (57) in Theorem 2) is also based on the reproduction of CSI S^n at Alice as $\hat{S}^n$ at Bob as well.

In view of these observations along with Remark 5, we are now interested in what happens if we confine ourselves to within those coding schemes that the CSI S^n at Alice is required to be reproduced as $\hat{S}^n$ at Bob (this kind of coding schemes are said to be *state-reproducing*). To see this, let the corresponding secret-message capacity and secret-key capacity be denoted by the overlined quantities as $\overline{C}$, then we have the following theorem:

Theorem 5: For any degraded WC (Z is a degraded version of Y) with causal CSI only at Alice, we have

$$\begin{aligned} \overline{C}_{\text{CSI-E}} &= \overline{C}_{\text{NCSI-E}} \\ &= \max_{p(x|s)} \min\big(I(XS;Y) - I(XS;Z), \\ &\qquad\qquad I(XS;Y) - H(S)\big), \end{aligned} \tag{94}$$

$$\begin{aligned} \overline{C}^{\mathsf{K}}_{\text{CSI-E}} &= \overline{C}^{\mathsf{K}}_{\text{NCSI-E}} \\ &= \max_{I(XS;Y) \geq H(S)} \big(I(XS;Y) - I(XS;Z)\big). \end{aligned} \tag{95}$$

Remark 9: It is easy to check that the right-hand side of (94) is not greater than the right-hand side of (95). □

Proof of (94): It suffices to prove only the converse. Since message M and CSI S^n are independent and S^n is reproducible at Bob, Fano inequality with achievable rates R and $\varepsilon_n \to 0$ claims that

$$\begin{aligned} nR &= H(M) \\ &\leq H(M) - H(MS^n|Y^n) + n\varepsilon_n \\ &\leq H(MS^n) - H(S^n) - H(MS^n|Y^n) + n\varepsilon_n \\ &\leq I(MS^n;Y^n) - nH(S) + n\varepsilon_n \\ &= \sum_{i=1}^n I(MS^n;Y_i|Y^{i-1}) - nH(S) + n\varepsilon_n \\ &\leq \sum_{i=1}^n I(MS^nY^{i-1};Y_i) - nH(S) + n\varepsilon_n \\ &\leq \sum_{i=1}^n I(X_iMS^nY^{i-1};Y_i) - nH(S) + n\varepsilon_n \\ &= \sum_{i=1}^n I(X_iS_iMS^{i-1}S_{i+1}^nY^{i-1};Y_i) - nH(S) + n\varepsilon_n \\ &\stackrel{(u)}{=} \sum_{i=1}^n I(X_iS_i;Y_i) - nH(S) + n\varepsilon_n \\ &= \sum_{i=1}^n (I(X_iS_i;Y_i) - H(S_i)) + n\varepsilon_n \\ &\stackrel{(y)}{=} n(I(XS;Y) - H(S)) + n\varepsilon_n, \end{aligned} \tag{96}$$

where (u) follows from the Markov chain property $MS^{i-1}S_{i+1}^nY^{i-1} \to X_iS_i \to Y_i$; in (y) we have used the argument similar to that in (65). Thus, $R \leq I(XS;Y) - H(S)$, which together with the proof of (56) implies the converse here.

Proof of (95): It suffices to prove only the converse. Since S^n is reproducible at Bob, similarly to the derivation in (96) we have

$$\begin{aligned} nH(S) &\leq H(S^n) - H(S^n|Y^n) + \varepsilon_n \\ &= I(S^n;Y^n) + \varepsilon_n \\ &= \sum_{i=1}^n I(S^n;Y_i|Y^{i-1}) + \varepsilon_n \\ &\leq \sum_{i=1}^n I(S^nY^{i-1};Y_i) + \varepsilon_n \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^{n} I(X_i S_i S^{i-1} S_{i+1}^{n} Y^{i-1}; Y_i) + n\varepsilon_n \\
&= \sum_{i=1}^{n} I(X_i S_i; Y_i) + n\varepsilon_n \\
&= nI(XS; Y) + n\varepsilon_n. \qquad (97)
\end{aligned}$$

Thus, $H(S) \leq I(XS; Y)$, which together with the proof of (58) implies the converse here. □

V. Comparison With the Previous Result

We have so far studied the problem of how to convey confidential message over WCs with *causal* CSI available only at Alice under the information leakage $I_E = I(M^b; \mathbf{Z}^b) \to 0$. In this connection, we notice that this kind of problem with *causal* CSI has not yet been brought to enough attention of the researcher, although the problem for WCs with *non-causal* CSI has extensively been investigated in the literature. On the other hand, to the best of our knowledge, Fujita [20] is supposed to be the first who has significantly addressed the problem of WCs with *causal* CSI available only at Alice (used for key agreement with Bob), although its *non-causal* counterpart had been studied by Khisti *et al.* [15]. In this section, we develop the comparison with our results.

In order to describe the main result of [20] in our terminology, define

$$\begin{aligned}
&F_{\text{CSI-1}}(p(u), p(x|u,s)) \\
&\quad = \min\Big[I(U; SY) - I(U; SZ) \\
&\qquad + H(S|Z) - H(S|Y), \\
&\qquad I(U; SY) - H(S|Y)\Big], \qquad (98)
\end{aligned}$$

and let $C^{\mathrm{W}}_{\text{CSI-E}}$ denote the secrecy capacity under the weak secrecy criterion $\frac{1}{n}I(M^b; \mathbf{Z}^b) \to 0$ instead of $C_{\text{CSI-E}}$. Then,

Theorem 6 (Fujita [20, Lemma 1]): Let us consider a degraded WC where Z is a physically degraded version of Y, then

$$C^{\mathrm{W}}_{\text{CSI-E}} \geq \max_{p(u), p(x|u,s)} F_{\text{CSI-1}}(p(u), p(x|u,s)) \qquad (99)$$

holds. □

For comparison, we rewrite $F_{\text{CSI-1}}(p(u), p(x|u,s))$ in (98) as follows.

$$\begin{aligned}
&F_{\text{CSI-1}}(p(u), p(x|u,s)) \\
&\quad = \min\Big[I(U; Y) - I(U; SZ) \\
&\qquad + H(S|Z) - H(S|UY), \\
&\qquad I(U; Y) - H(S|UY)\Big], \qquad (100)
\end{aligned}$$

which is justified because

$$\begin{aligned}
I(U; SY) &= I(U; Y) + I(U; S|Y), \qquad (101)\\
H(S|Y) &= H(S|UY) + I(U; S|Y). \qquad (102)
\end{aligned}$$

Recalling that the lower bound $R_{\text{CSI-1}}(p(u), p(x|u,s))$ in Theorem 1 is

$$\begin{aligned}
&R_{\text{CSI-1}}(p(u), p(x|u,s)) \\
&\quad = \min\Big[I(U; Y) - I(U; SZ) \\
&\qquad + H(S|Z) - H(S|UY), \\
&\qquad I(U; Y) - H(S|UY)\Big] \qquad (103)
\end{aligned}$$

and comparing it with (100), it turns out that $R_{\text{CSI-1}}(p(u), p(x|u,s))$ exactly coincides with $F_{\text{CSI-1}}(p(u), p(x|u,s))$. Hence, the two largest lower bounds in Theorems 1 and 6 coincide with one another:

$$\max_{p(u), p(x|u,s)} R_{\text{CSI-1}}(p(u), p(x|u,s)) = \max_{p(u), p(x|u,s)} F_{\text{CSI-1}}(p(u), p(x|u,s)). \qquad (104)$$

On the other hand, the other largest lower bound in Theorem 1:

$$\max_{p(u), p(x|u,s)} R_{\text{CSI-2}}(p(u), p(x|u,s)) \qquad (105)$$

can be shown to be strictly larger than the left-hand side of (104) for an approximately selected WC in which Y is a degraded version of Z (e.g., see [17, Example 2]), that is

$$\max_{p(u), p(x|u,s)} R_{\text{CSI-2}}(p(u), p(x|u,s)) > \max_{p(u), p(x|u,s)} R_{\text{CSI-1}}(p(u), p(x|u,s)), \qquad (106)$$

which together with (104) implies that for this WC the lower bound in Theorem 1 is strictly larger than the lower bound in Theorem 6.

Now, we are in a position to point out further crucial differences between [20] and this paper, which is due to the completely different approaches taken to the problem. These are summarized as follows.

- [20] heavily depends on the assumption that the WC treated needs to be physically degraded, whereas this paper makes no such assumption.
- [20] confines itself to within the weak secrecy criterion problem ($\frac{1}{n}I(M; \mathbf{Z}) \to 0$), whereas this paper employs the strong secrecy criterion approach ($I(M; \mathbf{Z}) \to 0$). As a consequence, all the results in [20] (and [17]) are guaranteed to hold as they are under the strong secrecy criterion too.
- In [20] all alphabets such as $\mathcal{U}, \mathcal{S}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are required to be finite, whereas in this paper $\mathcal{U}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$ except for $\mathcal{S}$ may be arbitrary (including continuous alphabet cases), so that Theorem 1 as stated in Section II is directly applicable also to, e.g., Gaussian WCs with causal CSI available at Alice.
- The fundamental mathematical tool in [20] to deal with the problem is the typical sequence argument (of course, well established), whereas in this paper the fundamental ingredients consist of Slepian-Wolf coding,

Csiszár-Körner's key construction, Gallager's maximum likelihood decoding, and Han-Verdú's resolvability argument (of course, well established). This methodological difference brings about a new look from the viewpoint of information theoretic perspective and applicability. One of the consequences is that the way of proving the main theorem here is significantly different from that in [20]. This, for example, enabled us to naturally establish the strong secrecy property, which, as is well known, would not be quite easy to be attained by the usual typical sequence arguments.

- Most importantly, we see that there exists a crucial difference between [20] and this paper from the coding theoretic standpoint. Seemingly, both invoke the block Markov coding scheme as devised in [16], which is furnished with a kind of *forward-backward* coding procedure.

  However, in [20], the "recursive" forward-backward coding procedure is employed in the sense that the j-th encoding in each block j is carried out (which is carried over to the next block $j+1$) according to the order $j = 1, 2, \cdots, b$. During this procedure over the total b blocks no decoding is carried out. When the encoding reaches the final block $j = b$ the decoding procedure gets started, which is carried back to block $j = b - 1$. This decoding procedure is repeated backward according to the order $j = b, b-1, \cdots, 1$, which causes at worst "$2b$ block decoding delay" in the whole process.

  On the other hand, this paper employs the "iterative" forward-backward coding procedure in the sense that not only the j-th encoding in each block j (which is carried over to the next block $j+1$) but also the decoding for the previous block $j-1$ are carried out according to the order $j = 1, 2, \cdots, b$. This one-way coding scheme causes only "one block decoding delay."

  Why is this difference? The reason for this is that in [20] the decoding operation in block j is to be made upon receiving the information $\mathbf{S}_j\mathbf{Y}_j$ but the decoding operation for $\mathbf{S}_j$ is postponed to the next block $j+1$ and it is in turn postponed to block $j+2$, and recursively so on to reach the final block $j = b$. Thus, actually, $\mathbf{S}_j$ is decoded according to the order $j = b, b-1, \cdots, 1$. In contrast with this, in this paper the decoding operation in block j is made upon receiving the information $\mathbf{Y}_j$, based on which $\mathbf{U}_j$ is decoded and used to decode $\mathbf{S}_{j-1}$ in block $j-1$, and then proceed to the next block $j+1$. This means that only one block decoding delay and hence low complexities are needed.

VI. Concluding Remarks

In this last section, let us get started with quoting a paragraph from Chia and El Gamal [17], which addressed an interesting non-trivial problem:

"We used key generation from state information to improve the message transmission rate. It may be possible to extend this idea to the case when the state information is available only at the encoder. This case, however, is not straightforward to analyze since it would be necessary for the encoder to reveal some state information to the decoder (and, hence, partially to the eavesdropper) in order to agree on a secret key, which would reduce the wiretap coding part of the rate."

Motivated by it, we have investigated the coding problem for WCs with causal CSI at Alice and/or Bob, and established reasonable lower bounds on the secrecy capacity, which are summarized as Theorems 1 (one of the key results in this paper). Although Theorem 1 treats the WC with CSI available only at Alice, it can actually be useful enough for investigating general WCs with three correlated causal CSIs available at Alice, Bob and Eve, respectively. We would like to remind that this seemingly "general" WCs can actually be reduced to our WCs with CSI available only at Alice. In this connection, the reader may refer, for example, to Khisti *et al.* [15], and Goldfeld *et al.* [25].

As was pointed out in Section V, the main ingredients thereby to establish Theorems 1 actually consist of the well-established information-theoretic lemmas such as Slepian-Wolf coding, Csiszár-Körner's key construction, Gallager's maximum likelihood decoding, and Han-Verdú's resolvability argument, while not invoking the celebrated argument of typical sequences, which enabled us to well handle also the case with alphabets not necessarily finite, for example, including possibly the case of Gaussian WCs with CSI. Actually, this approach enabled us to derive some interesting results for degraded WCs as follows. Theorem 2 gives lower and upper bounds for the secret (-message) capacity, while, fortunately, the exact formula for the secret-key capacity has been determined. Corollary 3 shows a causal secrecy capacity with one-sided CSI, which has nice correspondence with the interesting result of Chia and El Gamal with two-sided CSI [17], while Theorem 4 gives the secrecy capacity for binary WCs with one-sided CSI to establish a counterpart of Gaussian WCs studied by Costa [32] as "Writing on dirty paper." Thus, these results together would provide a basic basis for further investigation of WCs with *causal* CSI.

Appendix A
Proof of Lemma 5

Proof of i)*:* We can rewrite the security index $\mathsf{S}(\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j)|\mathbf{Z}_j)$ in (18) of Lemma 2 as

$$\begin{aligned}&\mathsf{S}(\kappa(\mathbf{S}_j)\sigma(\mathbf{S}_j)|\mathbf{Z}_j)\\ &\quad= \mathsf{S}(K_jN_{j+1}|\mathbf{Z}_j)\\ &\quad= D(P_{K_jN_{j+1}}||Q_{K_jN_{j+1}}) + I(K_jN_{j+1};\mathbf{Z}_j)\\ &\quad\geq I(K_jN_{j+1};\mathbf{Z}_j),\end{aligned} \tag{107}$$

which together with Lemma 2 gives i).

Proof of ii)*:* Here we use the following recurrence relation:

$$\begin{aligned}&I(K_{j-1}N_j;\mathbf{Z}^{j-1}|\mathcal{H})\\ &\quad= I(K_{j-1}N_j;\mathbf{Z}_{j-1}|\mathcal{H})\\ &\qquad+I(K_{j-1}N_j;\mathbf{Z}^{j-2}|\mathbf{Z}_{j-1}\mathcal{H})\\ &\quad\leq I(K_{j-1}N_j;\mathbf{Z}_{j-1}|\mathcal{H})\\ &\qquad+I(K_{j-2}N_{j-1}K_{j-1}N_j;\mathbf{Z}^{j-2}|\mathbf{Z}_{j-1}\mathcal{H})\end{aligned}$$

$$\begin{aligned}&\overset{(j)}{=} I(K_{j-1}N_j;\mathbf{Z}_{j-1}|\mathcal{H})\\&\quad +I(K_{j-2}N_{j-1};\mathbf{Z}^{j-2}|\mathbf{Z}_{j-1}\mathcal{H})\\&\overset{(k)}{\le} I(K_{j-1}N_j;\mathbf{Z}_{j-1}|\mathcal{H})\\&\quad +I(K_{j-2}N_{j-1};\mathbf{Z}^{j-2}|\mathcal{H}),\end{aligned} \tag{108}$$

where (j) follows from the Markov chain property $\mathbf{Z}^{j-2} \to K_{j-2}N_{j-1} \to K_{j-1}N_j$ given $\mathbf{Z}_{j-1}\mathcal{H}$; (k) follows from the Markov chain property $\mathbf{Z}^{j-2} \to K_{j-2}N_{j-1} \to \mathbf{Z}_{j-1}$ given $\mathcal{H}$. Then, taking the summation of both sides in (108) over $j \in [1:l]$ $(1 \le l \le b)$ we have

$$\begin{aligned}I(K_{l-1}N_l;\mathbf{Z}^{l-1}|\mathcal{H}) &\le \sum_{j=1}^{l} I(K_{j-1}N_j;\mathbf{Z}_{j-1}|\mathcal{H})\\&\le \sum_{j=1}^{b} I(K_{j-1}N_j;\mathbf{Z}_{j-1}|\mathcal{H})\\&\overset{(m)}{=} b\varepsilon,\end{aligned} \tag{109}$$

where we have noticed that $I(K_{j-2}N_{j-1};\mathbf{Z}^{j-2}|\mathcal{H}) = 0$ for $j = 1$ and (m) follows from i) of Lemma 5, thereby completing the proof. □

Appendix B
Proof of Lemma 6

The proof is carried out basically along the line of Han and Verdú [31, (8.3)] and Hayashi [21, Theorem 3]). We evaluate here the resolvability in terms of $I(M_{0j};\mathbf{S}_j\mathbf{Z}_j|\mathcal{H})$ under rate constraint

$$\overline{R} - R_0 > I(U;SZ), \tag{110}$$

which is developed as follows.

For each $m_0 \in [1 : 2^{nR_0}]$, let $\mathbf{U}(m_0)$ denote the random variable $\mathbf{u}_j(L(m_0))$ where $L(m_0)$ is distributed uniformly on the bin $\mathcal{B}(m_0)$ with rate constraint (110), and define the channel

$$W(\mathbf{t}|\mathbf{u}) \overset{\Delta}{=} P_{\mathbf{T}(m_0)|\mathbf{U}(m_0)},$$

where $\mathbf{T}(m_0) \overset{\Delta}{=} (\mathbf{S}(m_0), \mathbf{Z}(m_0))$, $\mathbf{t} \overset{\Delta}{=} (\mathbf{s}, \mathbf{z})$ and we notice that $P_{\mathbf{S}(m_0)\mathbf{Z}(m_0)|\mathbf{U}(m_0)}$ does not depend on m_0, so that we can write $P_{\mathbf{USZ}}$ instead of $P_{\mathbf{U}(m_0)\mathbf{S}(m_0)\mathbf{Z}(m_0)}$. Now, set

$$L_n = 2^{n(\overline{R}-R_0)} \tag{111}$$

and

$$i_{\mathbf{U}W}(\mathbf{u},\mathbf{t}) = \log\frac{W(\mathbf{t}|\mathbf{u})}{P_{\mathbf{T}}(\mathbf{t})}. \tag{112}$$

Then,

$$\begin{aligned}&I(M_{0j};\mathbf{S}_j\mathbf{Z}_j|\mathcal{H})\\&= \frac{1}{2^{nR_0}}\sum_{m_0=1}^{2^{nR_0}} \mathrm{E}_{\mathcal{H}} D(P_{\mathbf{T}(m_0)|\mathbf{U}(m_0)}||P_{\mathbf{T}(m_0)})\\&\overset{(a)}{=} \mathrm{E}_{\mathcal{H}} D(P_{\mathbf{T}|\mathbf{U}}||P_{\mathbf{T}})\\&= \sum_{\mathbf{t}\in\mathcal{S}^n\times\mathcal{Z}^n}\sum_{\mathbf{c}_1\in\mathcal{U}^n}\cdots\sum_{\mathbf{c}_{L_n}\in\mathcal{U}^n} P_{\mathbf{U}}(\mathbf{c}_1)\cdots P_{\mathbf{U}}(\mathbf{c}_{L_n})\\&\quad\cdot\frac{1}{L_n}\sum_{j=1}^{L_n} W(\mathbf{t}|\mathbf{c}_j)\log\left(\frac{1}{L_n}\sum_{k=1}^{L_n}\exp i_{\mathbf{U}W}(\mathbf{c}_k,\mathbf{t})\right)\\&= \sum_{\mathbf{c}_1\in\mathcal{U}^n}\cdots\sum_{\mathbf{c}_{L_n}\in\mathcal{U}^n} P_{\mathbf{U}}(\mathbf{c}_1)\cdots P_{\mathbf{U}}(\mathbf{c}_{L_n})\\&\quad\cdot\sum_{\mathbf{t}\in\mathcal{S}^n\times\mathcal{Z}^n} W(\mathbf{t}|\mathbf{c}_1)\log\left(\frac{1}{L_n}\sum_{k=1}^{L_n}\exp i_{\mathbf{U}W}(\mathbf{c}_k,\mathbf{t})\right)\\&\overset{(b)}{\le} \sum_{\mathbf{c}_1\in\mathcal{U}^n}\sum_{\mathbf{t}\in\mathcal{S}^n\times\mathcal{Z}^n} W(\mathbf{t}|\mathbf{c}_1)P_{\mathbf{U}}(\mathbf{c}_1)\\&\quad\cdot\log\left(\frac{1}{L_n}\exp i_{\mathbf{U}W}(\mathbf{c}_1,\mathbf{t})+\frac{1}{L_n}\sum_{k=2}^{L_n}\mathrm{E}\exp i_{\mathbf{U}W}(\mathbf{C}_k,\mathbf{t})\right)\\&\overset{(c)}{\le} \mathrm{E}\left[\log\left(1+\frac{1}{L_n}\exp i_{\mathbf{U}W}(\mathbf{U},\mathbf{T})\right)\right],\end{aligned} \tag{113}$$

where (a) follows from the symmetry of the random code $\mathcal{H}$; (b) follows from the concavity of the logarithm; (c) is the result of

$$\mathrm{E}[\exp i_{\mathbf{U}W}(\mathbf{C}_k,\mathbf{t})] = 1$$

for all $\mathbf{t} \in \mathcal{S}^n \times \mathcal{Z}^n$ and $k = 1, 2, \cdots, L_n$. Now, with $Q(\mathbf{u}) = P_{\mathbf{U}}(\mathbf{u})$, apply a simple inequality with $0 < \rho < 1$ and $x \ge 0$:

$$\log(1+x) = \frac{\log(1+x)^\rho}{\rho} \le \frac{\log(1+x^\rho)}{\rho} \le \frac{x^\rho}{\rho}$$

to (113) to eventually obtain

$$\begin{aligned}&I(M_{0j};\mathbf{S}_j\mathbf{Z}_j|\mathcal{H})\\&\le \frac{1}{\rho L_n^\rho}\mathrm{E}\left(\frac{W(\mathbf{T}|\mathbf{U})}{P_{\mathbf{T}}(\mathbf{T})}\right)^\rho\\&= \frac{1}{\rho L_n^\rho}\sum_{\mathbf{t}\in\mathcal{S}^n\times\mathcal{Z}^n}\sum_{\mathbf{u}\in\mathcal{U}^n} Q(\mathbf{u})W(\mathbf{t}|\mathbf{u})\left(\frac{W(\mathbf{t}|\mathbf{u})}{P_{\mathbf{T}}(\mathbf{t})}\right)^\rho\\&= \frac{1}{\rho L_n^\rho}\sum_{\mathbf{t}\in\mathcal{S}^n\times\mathcal{Z}^n}\sum_{\mathbf{u}\in\mathcal{U}^n} Q(\mathbf{u})W(\mathbf{t}|\mathbf{u})^{1+\rho}P_{\mathbf{T}}(\mathbf{t})^{-\rho}.\end{aligned} \tag{114}$$

On the other hand, by virtue of Hölder's inequality,

$$\begin{aligned}&\left(\sum_{\mathbf{u}\in\mathcal{U}^n} Q(\mathbf{u})W(\mathbf{t}|\mathbf{u})^{1+\rho}\right)P_{\mathbf{T}}(\mathbf{t})^{-\rho}\\&= \left(\sum_{\mathbf{u}\in\mathcal{U}^n} Q(\mathbf{u})W(\mathbf{t}|\mathbf{u})^{1+\rho}\right)\left(\sum_{\mathbf{u}\in\mathcal{U}^n} Q(\mathbf{u})W(\mathbf{t}|\mathbf{u})\right)^{-\rho}\\&\le \left(\sum_{\mathbf{u}\in\mathcal{U}^n} Q(\mathbf{u})W(\mathbf{t}|\mathbf{u})^{\frac{1}{1-\rho}}\right)^{1-\rho}\end{aligned} \tag{115}$$

for $0 < \rho < 1$. Therefore, it follows from (111) that

$$\begin{aligned}&I(M_{0j};\mathbf{S}_j\mathbf{Z}_j|\mathcal{H})\\&\le \frac{1}{\rho L_n^\rho}\sum_{\mathbf{t}\in\mathcal{S}^n\times\mathcal{Z}^n}\left(\sum_{\mathbf{u}\in\mathcal{U}^n} Q(\mathbf{u})W(\mathbf{t}|\mathbf{u})^{\frac{1}{1-\rho}}\right)^{1-\rho}\\&= \frac{1}{\rho}\exp\left[-[n\rho(\overline{R}-R_0)+E_0(\rho,Q)]\right],\end{aligned} \tag{116}$$

where

$$E_0(\rho, Q) = -\log\left[\sum_{\mathbf{t}\in\mathcal{S}^n\times\mathcal{Z}^n}\left(\sum_{\mathbf{u}\in\mathcal{U}^n} Q(\mathbf{u})W(\mathbf{t}|\mathbf{u})^{\frac{1}{1-\rho}}\right)^{1-\rho}\right]. \tag{117}$$

Then, by means of Gallager [28, Theorem 5.6.3], we have $E_0(\rho, Q)|_{\rho=0} = 0$ and

$$\left.\frac{\partial E_0(\rho, Q)}{\partial \rho}\right|_{\rho=0} = -I(Q, W) = -I(\mathbf{U}; \mathbf{SZ}) \overset{(d)}{=} -nIU; SZ), \tag{118}$$

where (d) follows because $(\mathbf{U}, \mathbf{SZ})$ is a correlated i.i.d. sequence with generic variable (U, SZ). Thus, for any small constant $\tau > 0$ there exists a $\rho_0 > 0$ such that, for all $0 < \rho \le \rho_0$,

$$E_0(\rho, Q) \ge -n\rho(1+\tau)I(U; SZ) \tag{119}$$

which is substituted into (116) to obtain

$$I(M_{0j}; \mathbf{S}_j\mathbf{Z}_j|\mathcal{H}) \le \frac{1}{\rho}\exp\left[-n\rho(\overline{R} - R_0 - (1+\tau)I(U; SZ))\right]. \tag{120}$$

On the other hand, in view of (110), with some $\delta > 0$ we can write

$$\overline{R} - R_0 = I(U; SZ) + 2\delta, \tag{121}$$

which leads to

$$\overline{R} - R_0 - (1+\tau)I(U; SZ) = I(U; SZ) + 2\delta - I(U; SZ) - \tau I(U; SZ) = 2\delta - \tau I(U; SZ)). \tag{122}$$

We notice here that $\tau > 0$ can be arbitrarily small, so that the last term on the right-hand side of (122) can be made larger than $\delta > 0$. Then, (120) yields

$$I(M_{0j}; \mathbf{S}_j\mathbf{Z}_j|\mathcal{H}) \le \frac{1}{\rho}\exp[-n\rho\delta], \tag{123}$$

which implies that, for any small $\varepsilon > 0$,

$$I(M_{0j}; \mathbf{S}_j\mathbf{Z}_j|\mathcal{H}) \le \varepsilon \tag{124}$$

for all sufficiently large n, completing the proof of Lemma 6. □

Acknowledgments

The authors are grateful to Hiroyuki Endo for useful discussions. Special thanks go to Alex Bunin for useful comments, which occasioned to improve Theorem 1. Especially, the authors greatly appreciate the excellent editorial leadership of Matthieu Bloch who thoroughly read the earlier version to provide insightful comments. □

References

[1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. IT-54, no. 8, pp. 1355–1387, 1975.

[2] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.

[3] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[4] C. E. Shannon, "Communication theory of secrecy systems," *Bell Labs Tech. J.*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[5] C. E. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.

[6] C. Mitrpant, A. J. H. Vinck, and Y. Luo, "An achievable region for the Gaussian wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2181–2190, May 2006.

[7] Y. Chen and A. J. H. Vinck, "Wiretap channel with side information," *IEEE Int. Symp. Inf. Theory*, vol. 54, no. 1, pp. 395–402, Jun. 2008.

[8] W. Liu and B. Chen, "Wiretap channel with two-sided channel state information," in *Proc. 41st Asilomar Conf. Signals, Syst. Comput.*, Nov. 2007, pp. 893–897.

[9] M. Bloch and J. N. Laneman, "Information-spectrum methods for information-theoretic security," in *Proc. Inf. Theory Appl. Workshop*, Feb. 2009, pp. 23–28.

[10] B. Dai, Z. Zhuang, and A. J. H. Vinck, "Some new results on the wiretap channel with causal side information," in *Proc. IEEE ICCT*, Chengdu, China, Nov. 2012, pp. 609–614.

[11] H. Boche and R. F. Schaefer, "Wiretap channels with side information—Strong secrecy capacity and optimal transceiver design," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 8, pp. 1397–1408, Aug. 2013.

[12] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 733–742, May 1993.

[13] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography. I. Secret sharing," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul. 1993.

[14] A. Khisti, S. Diggavi, and G. Wornell, "Secret key agreement using asymmetry in channel state knowledge," in *Proc. IEEE Int. Symp. Inf. Theory*, Seoul, South Korea, Jun./Jul. 2009, pp. 2286–2290.

[15] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret-key agreement with channel state information at the transmitter," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 672–681, Sep. 2011.

[16] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. IT-25, no. 5, pp. 572–584, Sep. 1979.

[17] Y. K. Chia and A. El Gamal, "Wiretap channel with causal state information," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 2838–2849, May 2012.

[18] B. Dai and Y. Luo, "Some new results on the wiretap channel with side information," *Entropy*, vol. 14, no. 9, pp. 1671–1702, 2012.

[19] A. Sonee and G. A. Hodtani, "Wiretap channel with strictly causal side information at encoder," in *Proc. Iran Workshop Commun. Inf. Theory (IWCIT)*, May 2014, pp. 1–6.

[20] H. Fujita, "On the secrecy capacity of wiretap channels with side information at the transmitter," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 11, pp. 2441–2452, Nov. 2016.

[21] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3989–4001, Jun. 2011.

[22] T. S. Han, H. Endo, and M. Sasaki, "Wiretap channels with one-time state information: Strong secrecy," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 1, pp. 224–236, Jan. 2018.

[23] G. Kramer, "Topics in multi-user information theory," *Found. Trends Commun. Inf. Theory*, vol. 4, nos. 4–5, pp. 265–444, 2008.

[24] V. M. Prabhakaran, K. Eswaran, and K. Ramchandran, "Secrecy via sources and channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6747–6765, Nov. 2012.

[25] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Wiretap channels with random states non-causally available at the encoder," 2016, *arXiv:1608.00743*. [Online]. Available: https://arxiv.org/abs/1608.00743

[26] A. Bunin, Z. Goldfeld, H. Permuter, S. Shamai (Shitz), P. Cuff, and P. Piantanida, "Semantically-secured message-key trade-off over wiretap channels with random parameters," in *Proc. 2nd Workshop Commun. Secur.*, 2018, pp. 33–48.

[27] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[28] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ, USA: Wiley, 1968.

[29] A. El Gamal and Y. H. Kim, *Network Information Theory*. New York, NY, USA: Cambridge Univ. Press, 2011.

[30] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, Apr. 2006.

[31] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 399, no. 3, pp. 752–772, May 1993.

[32] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. IT-29, no. 3, pp. 439–441, May 1983.

[33] T. S. Han and M. Sasaki, "Wiretap channels with causal state information: Revisited," Nat. Inst. Inf. Commun. Technol., Tokyo, Japan, Tech. Rep., 2020.

Te Sun Han (M'80–SM'88–F'90–LF'11) received the B.Eng., M.Eng., and D.Eng. degrees in mathematical engineering from the University of Tokyo, Japan, in 1964, 1966, and 1971, respectively. Since 1993, he has been a Professor with the University of Electro-Communications, where he has been a Professor Emeritus since 2007. He has published papers on information theory, most of which appeared in the IEEE TRANSACTIONS ON INFORMATION THEORY. Also, he has published two books: one of them is *Information-Spectrum Methods in Information Theory* (Springer Verlag, 2003). Especially, this book was written to try to demonstrate part of the general logics latent in information theory. His research interests include basic problems in Shannon theory, multi-user source/channel coding systems, multiterminal hypothesis-testing and parameter estimation under data compression, large-deviation approach to information-theoretic problems, and especially, information-spectrum theory. He is a member of the Board of Governors for the IEEE Information Theory Society. He has been an IEICE Fellow and an Honorary Member since 2011. He was a recipient of the 2010 Shannon Award. From 1994 to 1997, he was the Associate Editor for Shannon Theory of the IEEE TRANSACTIONS ON INFORMATION THEORY.

Masahide Sasaki received the B.S., M.S., and Ph.D. degrees in physics from Tohoku University, Sendai Japan, in 1986, 1988 and 1992, respectively. During 1992–1996, he worked on the development of semiconductor devices in Nippon-Kokan Company (currently JFE Holdings). In 1996, he joined the Communications Research Laboratory, Ministry of Posts and Telecommunications (since 2004, National Institute of Information and Communications Technology (NICT), Ministry of Internal Affairs and Communications). He has been working on quantum optics, quantum communication and quantum cryptography. He is presently Distinguished Researcher of Advanced ICT Research Institute, and NICT Fellow. Dr. Sasaki is a member of Japanese Society of Physics, and the Institute of Electronics, Information and Communication Engineers of Japan.