Keyless Covert Communication via Channel State Information

Hassan ZivariFard, *Student Member, IEEE*, Matthieu R. Bloch, *Senior Member, IEEE*, and Aria Nosratinia, *Fellow, IEEE*

***Abstract*—We consider the problem of covert communication over a state-dependent channel when the Channel State Information (CSI) is available either non-causally, causally, or strictly causally, either at the transmitter alone, or at both transmitter and receiver. Covert communication with respect to an adversary, called "warden," is one in which, despite communication over the channel, the warden's observation remains indistinguishable from an output induced by innocent channel-input symbols. Covert communication involves fooling an adversary in part by a proliferation of codebooks; for reliable decoding at the legitimate receiver, the codebook uncertainty is typically removed via a shared secret key that is unavailable to the warden. In contrast to previous work, we do not assume the availability of a large shared key at the transmitter and legitimate receiver. Instead, we only require a secret key with negligible rate to bootstrap the communication and our scheme extracts shared randomness from the CSI in a manner that keeps it secret from the warden, despite the influence of the CSI on the warden's output. When CSI is available at the transmitter and receiver, we derive the covert capacity region. When CSI is only available at the transmitter, we derive inner and outer bounds on the covert capacity. We also provide examples for which the covert capacity is positive with knowledge of CSI but is zero without it.**

I. Introduction

Covert communication refers to scenarios in which reliable communication over a channel must occur while simultaneously ensuring that a separate channel output at a node, called the warden, has a distribution identical to that induced by an innocent channel symbol [3]–[7]. It is known that in a Discrete Memoryless Channel (DMC) without state, the number of bits that can be reliably and covertly communicated over n channel transmissions scales at most as $O(\sqrt{n})$.[1] This result has motivated the study of other models in which positive rates are achievable [8], [9]. Of particular relevance to the present work, Lee *et al.* [10] have considered the problem of covert communication over a state-dependent channel in which the CSI is known either causally or non-causally to the transmitter but unknown to the receiver and the warden. The authors derived the covert capacity when the transmitter and the receiver share a sufficiently long secret key, as well as a lower bound on the minimum secret key length needed to achieve the covert capacity. Since the presence of CSI provides a natural source of randomness from which to extract secret keys, one may wonder if covert communication with positive rate is possible *without* requiring an external secret key. The present work offers conclusive answers to this question in several scenarios.

The usefulness of exploiting CSI for secrecy has been extensively investigated in the context of state-dependent wiretap channels. A discrete memoryless wiretap channel with random states known non-causally at the transmitter was first studied by Chen and Vinck [11], who established a lower bound on the secrecy capacity based on a combination of wiretap coding with Gel'fand-Pinsker coding. Generally speaking, coding schemes with CSI outperform those without CSI because perfect knowledge of the CSI not only enables the transmitter to align its signal toward the legitimate receiver but also provides a source of common randomness from which to generate a common secret key and enhance secrecy rates. Khisti *et al.* [12] studied the problem of secret key generation from non-causal CSI available at the transmitter and established inner and outer bounds on the secret key capacity. Chia and El Gamal [13] studied a wiretap channel in which the state information is available causally at both transmitter and receiver, proposing a scheme in which the transmitter and the receiver extract a weak secret key from the state and protect the confidential message via a one-time-pad driven with the extracted key (see also [14] and [15]). Han and Sasaki [16] subsequently extended this result to strong secret keys. Goldfeld et al. [17] proposed a superposition coding scheme for the problem of transmitting a semantically secure message over a state-dependent channel with CSI available non-causally at the transmitter. In the context of covert communications, several works have demonstrated the benefits of exploiting common randomness and CSI to generate secret keys. For instance, stealth secret key generation from correlated sources was studied by Lin *et al.* [18], [19] and covert secret key generation was studied by Tahmasbi and Bloch [20], [21]. We note that covert communication over a compound channel was studied by Salehkalaibar *et al.* [22], although the objective therein is to mask the state of the compound channel and not to exploit CSI.

The present paper studies covert communication over a state-dependent discrete memoryless channel with CSI avail-

H. ZivariFard and A. Nosratinia are with the Department of Electrical Engineering, The University of Texas at Dallas, Richardson, TX, USA. M. R. Bloch is with School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. E-mail: hassan@utdallas.edu, matthieu.bloch@ece.gatech.edu, aria@utdallas.edu.

The material in this paper was presented in part at IEEE Information Theory Workshop 2019, Visby, Gotland, Sweden [1], and the IEEE International Symposium on Information Theory 2020, Los Angeles, CA Sweden [2].

The work of H. ZivariFard and A. Nosratinia is supported by National Science Foundation (NSF) grant 1956213. The work of M. R. Bloch is supported by National Science Foundation (NSF) grant 1955401.

[1]In the special case when the output distribution (at the warden) induced by the innocent symbol is a convex combination of the output distributions generated by the other input symbols [5], the scaling is $O(n)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

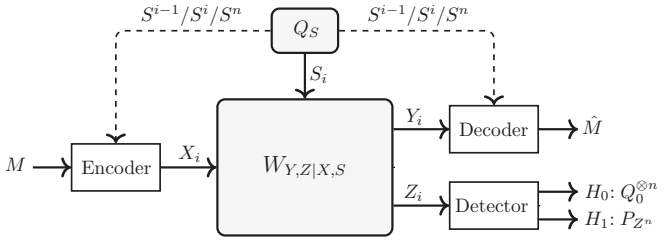

Fig. 1. Model of covert communication over a state-dependent DMC with CSI available at both the transmitter and the receiver

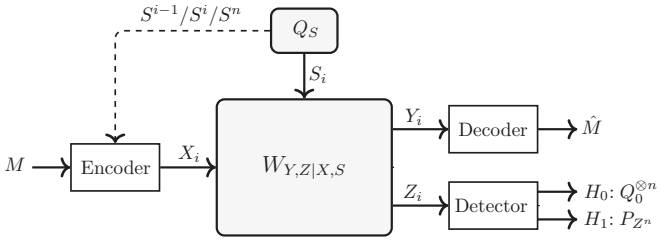

Fig. 2. Model of covert communication over a state-dependent DMC with CSI only available at the transmitter

able either non-causally, causally, or strictly causally, either at both the transmitter and the receiver or at the transmitter alone (see Fig. 1 and Fig. 2). One of the main contributions of the present work is to show that the CSI can be used to simultaneously and efficiently accomplish two necessary tasks: using the CSI for a Shannon strategy or Gel'fand-Pinsker coding, while also extracting a shared secret key at the two legitimate terminals to resolve the multiple codebooks that are necessary for covert communication. Secret key extraction from CSI replaces the external secret key in other models, thus potentially generalizing and expanding the applicability of covert communication. Our scheme requires the transmitter and the receiver to share a secret key with negligible rate to bootstrap the communication. This bootstrapping is common in many security schemes, for instance in all schemes for secret communication based on seeded invertible extractors [23]–[25]. With a slight abuse of terminology, we refer to our model as "keyless" instead of "asymptotically keyless."

Specifically, we characterize the exact covert capacity when CSI is available at both the transmitter and the receiver, and derive inner and outer bounds on the covert capacity when CSI is only available at the transmitter. For some channel models for which the covert capacity is zero without CSI, we show that the covert capacity is positive with CSI. The code constructions behind our proofs combine different coding mechanisms, including channel resolvability for covertness, channel randomness extraction for key generation, and Gel'fand-Pinsker coding for state-dependent channels. The key technical challenge consists in properly combining these mechanisms to ensure the overall covertness of the transmission through block-Markov chaining schemes. In the interest of brevity, the proofs that are standard, or parallel other proofs in this paper, are omitted and made available online [26], [27].

II. Preliminaries

$\mathbb{N}$ is the set of natural numbers, which does not include 0, while $\mathbb{R}$ denotes the set of real numbers. We define $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geqslant 0\}$ and $\mathbb{R}_{++} = \mathbb{R}_+ \backslash \{0\}$. Random variables are denoted by capital letters and their realizations by lower case letters. $\mathbb{E}_X(\cdot)$ is the expectation w.r.t. the random variable X and $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function. The set of $\epsilon-$strongly jointly typical sequences of length n, according to $P_{X,Y}$, is denoted by $\mathcal{T}_\epsilon^{(n)}(P_{X,Y})$. For convenience, typicality will reference the random variables rather than the distribution, e.g., we write $\mathcal{T}_\epsilon^{(n)}(X,Y)$ or $\mathcal{T}_\epsilon^{(n)}(X|Y)$. Superscripts denote the dimension of a vector, e.g., X^n. The integer set $\{1,\dots,M\}$ is denoted by $[\![1:M]\!]$, X_i^j indicates the set $\{X_i, X_{i+1}, \dots, X_j\}$, and $X_{\sim i}^n$ denotes the vector X^n except X_i. The cardinality of a set is denoted by $|\cdot|$. The total variation between Probability Mass Function (PMF) P and PMF Q is defined as, $||P-Q||_1 = \frac{1}{2}\sum_x |P(x)-Q(x)|$ and the Kullback-Leibler (KL) divergence between PMFs is defined as $\mathbb{D}(P||Q) = \sum_x p(x) \log \frac{P(x)}{Q(x)}$. The support of a probability distribution P is denoted by $\text{supp}(P)$. The n-fold product distribution constructed from the same distribution P is denoted $P^{\otimes n}$. Throughout the paper, $\log$ denotes the base 2 logarithm. For a set of random variables $\{X_i\}_{i\in\mathcal{A}}$ indexed over a countable set $\mathcal{A}$, $\mathbb{E}_{\backslash i}(\cdot)$ is the expectation with respect to all the random variables in $\mathcal{A}$ except the one with index $i \in \mathcal{A}$.

Finally, we recall a useful result about the relation between the total variation distance and the KL-divergence.

Lemma 1 (Reverse Pinsker's Inequality [28, eq. (323)])**.** *Pinsker's inequality indicates for two arbitrary distributions P and Q on alphabet $\mathcal{A}$ we have,*

$$||P-Q||_1 \leqslant \sqrt{\frac{1}{2}\mathbb{D}(P||Q)}. \tag{1}$$

A reverse inequality is valid when the alphabet $\mathcal{A}$ is finite. Let P and Q be two arbitrary distributions on a finite alphabet set $\mathcal{A}$ such that P is absolutely continuous with respect to Q. If $\mu \triangleq \min_{a\in\mathcal{Q}:Q(a)>0} Q(a)$, we have,

$$\mathbb{D}(P||Q) \leqslant \log\left(\frac{1}{\mu}\right)||P-Q||_1. \tag{2}$$

III. Channel model

Consider discrete memoryless state-dependent channels as shown in Fig. 1 or Fig. 2. The channel is characterized by input alphabet $\mathcal{X}$, legitimate output alphabet $\mathcal{Y}$, warden output alphabet $\mathcal{Z}$, state alphabet $\mathcal{S}$, and a transition probability $W_{YZ|XS}$. We assume that the CSI is independent and identically distributed (i.i.d.) and drawn according to Q_S and we let $x_0 \in \mathcal{X}$ be an "innocent" symbol corresponding to the absence of communication with the receiver. The distribution induced at the warden in the absence of communication is then

$$Q_0(\cdot) = \sum_{s\in\mathcal{S}} Q_S(s) W_{Z|X,S}(\cdot|x_0,s), \tag{3}$$

and we let $Q_0^{\otimes n} = \prod_{i=1}^n Q_0$. The CSI may be available non-causally, causally, or strictly causally at the transmitter and

may or may not be available at the receiver. Note that the exact causal or non-causal nature of CSI at the receiver is irrelevant because decoding is always done after transmission is completed. The warden is kept ignorant of the CSI.

Formally, a code with CSI available at both the transmitter and the receiver is defined as follows.

Definition 1. *A $(2^{nR}, n)$ code $\mathcal{C}_n$ with CSI available at both the transmitter and the receiver consists of:*

- *a message set $\mathcal{M} = [\![1 : 2^{nR}]\!]$;*
- *when CSI is available non-causally at the transmitter, for each time slot $i \in [\![1:n]\!]$, a deterministic encoder $f_i : \mathcal{M} \times \mathcal{S}^n \mapsto \mathcal{X}_i$ that maps a message m and the entire CSI sequence to a channel input symbol x_i;*
- *when CSI is available causally at the transmitter, for each time slot $i \in [\![1:n]\!]$, a deterministic encoder $f_i : \mathcal{M} \times \mathcal{S}^i \mapsto \mathcal{X}_i$ that maps a message m and the past and current CSI samples to a channel input symbol x_i;*
- *when CSI is available strictly-causally at the transmitter, for each time slot $i \in [\![1:n]\!]$, a deterministic encoder $f_i : \mathcal{M} \times \mathcal{S}^{i-1} \mapsto \mathcal{X}_i$ that maps a message m and the past CSI samples to a channel input symbol x_i;*
- *a decoding function $g : \mathcal{S}^n \times \mathcal{Y}^n \mapsto \mathcal{M} \cup \{?\}$ that maps the channel observations and the CSI sequence to a message $\hat{M} \in \mathcal{M}$ or an error message $?$.*

A code with CSI available only at the transmitter is defined as follows.

Definition 2. *A $(2^{nR}, n)$ code $\mathcal{C}_n$ with CSI available only at the transmitter consists of*

- *a message set $\mathcal{M} = [\![1 : 2^{nR}]\!]$, a local randomness set $\mathcal{J} = [\![1 : 2^{nR_J}]\!]$, and a secret key set $\mathcal{K}$;*
- *when CSI is available non-causally, for each time slot $i \in [\![1:n]\!]$, a stochastic encoder $f_i : \mathcal{M} \times \mathcal{J} \times \mathcal{K} \times \mathcal{S}^n \mapsto \mathcal{X}_i$, that maps a message m, a local randomness j, a secret key k, and the entire CSI sequence to a channel input symbol x_i;*
- *when CSI is available causally, for each time slot $i \in [\![1:n]\!]$, a stochastic encoder $f_i : \mathcal{M} \times \mathcal{J} \times \mathcal{K} \times \mathcal{S}^i \mapsto \mathcal{X}_i$, that maps a message m, a local randomness j, a secret key k, and the past and current CSI samples to a channel input symbol x_i;*
- *when CSI is available strictly-causally, for each time slot $i \in [\![1:n]\!]$, a stochastic encoder $f_i : \mathcal{M} \times \mathcal{J} \times \mathcal{K} \times \mathcal{S}^{i-1} \mapsto \mathcal{X}_i$ that maps a message m, a local randomness j, a secret key k, and the past CSI samples to a channel input symbol x_i;*
- *a decoding function $g : \mathcal{Y}^n \times \mathcal{K} \mapsto \mathcal{M} \cup \{?\}$ that maps the channel observations to a message $\hat{M} \in \mathcal{M}$ or an error message $?$.*

The reason for introducing a stochastic encoder when CSI is only available at the transmitter is that our achievability scheme relies on a likelihood encoder [29]. The stochastic nature of the likelihood encoder greatly simplifies the covertness analysis by providing finer control over the statistics induced by the encoder.

The code is assumed known to all parties and the objective is to design a code that is reliable, covert, and keyless. Reliable means that the probability of error $P_e^{(n)} = \mathbb{P}(\hat{M} \neq M)$ vanishes when $n \to \infty$. Covert means that the warden cannot determine whether communication is happening (hypothesis H_1) or not (hypothesis H_0). Specifically, the probabilities of false alarm α_n (warden deciding H_1 when H_0 is true) and missed detection β_n (warden deciding H_0 when H_1 is true) satisfy $\alpha_n + \beta_n = 1$ for an uninformed warden making random decisions. When the channel carries communication, the warden's channel output distribution is P_{Z^n}, and the optimal hypothesis test by the warden satisfies $\alpha_n + \beta_n \geqslant 1 - \sqrt{\mathbb{D}(P_{Z^n}||Q_0^{\otimes n})}$ [30]. Therefore, we define a code as covert if $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n})$ vanishes when $n \to \infty$. We assume that $\text{supp}(Q_0) = \mathcal{Z}$ for otherwise $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n})$ diverges. Finally, keyless means that $\frac{1}{n}\log|\mathcal{K}|$ vanishes as $n \to \infty$.

A rate R is achievable if there exists a sequence of reliable, covert, and keyless $(2^{nR}, n)$ codes and the covert capacity is the supremum of all achievable covert rates. We denote the covert capacity by $C_{\text{A-B}}$ where A $\in$ {NC,C,SC} indicates the non-causal, causal, or strictly causal nature of the CSI at the transmitter while B $\in$ {T,TR} indicates whether CSI is available only at the transmitter or both the transmitter and receiver. Hence, we are interested in characterizing $C_{\text{NC-TR}}$, $C_{\text{NC-T}}$, $C_{\text{C-TR}}$, $C_{\text{C-T}}$, $C_{\text{SC-TR}}$, and $C_{\text{SC-T}}$.

IV. Channel State Information Available at the Transmitter and the Receiver

Theorem 1. *Let*

$$\mathcal{A} \triangleq \left\{R \geqslant 0 : \exists P_{S,X,Y,Z} \in \mathcal{D} \text{ such that } R \leqslant \mathbb{I}(X;Y|S)\right\}, \tag{4a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_{X|S} W_{Y,Z|S,X} \\ P_Z = Q_0 \\ \mathbb{H}(S|Z) \geqslant \mathbb{I}(X;Z|S) - \mathbb{I}(X;Y|S) \end{array} \right\}. \tag{4b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with non-causal CSI at both the transmitter and the receiver is

$$C_{\text{NC-TR}} = \max\{a : a \in \mathcal{A}\}. \tag{5}$$

Theorem 1 suggests that the key rate $H(S|Z)$ extracted from CSI should exceed the difference between the capacity of the warden and the capacity of the legitimate receiver. The achievability is proved by superposition encoding and the complete proof is available in Appendix A.

Theorem 2. *Let*

$$\mathcal{A} \triangleq \left\{R \geqslant 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that } R \leqslant \mathbb{I}(U;Y|S)\right\}, \tag{6a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_U \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{Y,Z|S,X} \\ P_Z = Q_0 \\ \mathbb{H}(S|Z) \geqslant \mathbb{I}(U;Z|S) - \mathbb{I}(U;Y|S) \\ |\mathcal{U}| \leqslant |\mathcal{X}| + 1 \end{array} \right\}. \tag{6b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with causal CSI at both the transmitter and the receiver is

$$C_{\text{C-TR}} = \max\{a : a \in \mathcal{A}\}. \tag{7}$$

Again, Theorem 2 suggests that the key rate extracted from CSI should exceed the difference between the capacity of the warden and the capacity of the legitimate receiver. The achievability proof is based on block Markov encoding to combine a Shannon strategy for transmitting the message according to CSI with key generation and is available in Appendix B.

Theorem 3. *Let*

$$\mathcal{A} \triangleq \big\{R \geqslant 0 : \exists P_{S,X,Y,Z} \in \mathcal{D} \text{ such that } R \leqslant \mathbb{I}(X;Y|S)\big\}, \tag{8a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_X W_{Y,Z|S,X} \\ P_Z = Q_0 \\ \mathbb{H}(S|Z) \geqslant \mathbb{I}(X;Z|S) - \mathbb{I}(X;Y|S) \end{array} \right\}. \tag{8b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with strictly causal CSI at both the transmitter and the receiver is

$$C_{\text{SC-TR}} = \max\{a : a \in \mathcal{A}\}. \tag{9}$$

Even though *strictly* causal CSI provides limited opportunities to enhance reliability, it is still useful here because it provides shared randomness from which to extract a secret key. The achievability proof merely uses a block Markov encoding scheme for key generation but not for data transmission and is available in Appendix C.

V. Examples of channels with CSI at transmitter and receiver

We provide here two examples of covert communication over state-dependent channels with CSI at the transmitter and receiver. A positive covert capacity is achieved without an external secret key, hence not subject to the square root law. The two examples explore additive and multiplicative CSI, respectively, with the former representing channels in which the channel state can in principle be cancelled and the latter representing fading-like channels.

Binary Additive State: Consider a channel in which X, Y, Z, and S are all binary, Q_S obeys a Bernoulli distribution with parameter $\zeta \in (0 : 0.5)$, and the innocent symbol is $x_0 = 0$. (See Fig. 3). The law of the channel is

$$Y = Z = X \oplus S, \tag{10}$$

so that $Q_0 = Q_S$.

Proposition 1. *The covert capacity of the DMC depicted in Fig. 3 with causal or non-causal CSI available at the transmitter and the receiver is*

$$C_{\text{NC-TR}} = C_{\text{C-TR}} = \mathbb{H}_b(\zeta) = \zeta \log \frac{1}{\zeta} + (1-\zeta) \log \frac{1}{1-\zeta}. \tag{11}$$

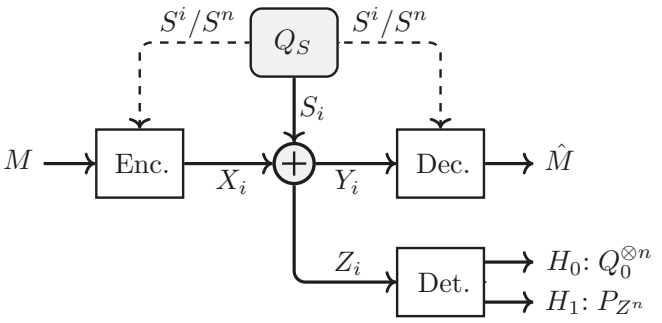

Fig. 3. Binary symmetric channel with additive CSI at the transmitter and the receiver

Table 1
Joint probability distribution of X, S

S \ X	0	1
0	α	β
1	$1-\alpha-\beta-\eta$	η

Intuitively, the encoder perfectly controls the warden's observations because it knows the CSI. By manipulating X, the encoder ensures that Z follows the statistics of S. In part, this means that the symbol $X = 1$ is associated half the time to $S = 0$ and half the time to $S = 1$ to ensure $P_Z = Q_S \sim \text{Bern}(\zeta)$. Further, since the transmitter and receiver share the CSI, the legitimate channel is error-free.

Proof. We first prove Proposition 1 when CSI is available non-causally at both the transmitter and the receiver. Substituting $Y = Z = X \oplus S$ in Theorem 1 results in

$$C_{\text{NC-TR}} = \max_{Q_S P_{X|S}} \mathbb{H}(X|S), \tag{12}$$

with the maximization subject to the constraint $P_Z = Q_0 = Q_S$. Let the joint distribution between X and S be according to Table 1, we have

$$\begin{aligned} P_Z(z=0) &= P_{X,S}(x=0,s=0) + P_{X,S}(x=1,s=1) \\ &= \alpha + \eta, \end{aligned} \tag{13}$$

$$\begin{aligned} Q_S(s=0) &= P_{X,S}(x=0,s=0) + P_{X,S}(x=1,s=0) \\ &= \alpha + \beta. \end{aligned} \tag{14}$$

Therefore $P_Z = Q_S$ implies that

$$Q_Z(z=0) = Q_S(s=0) \Rightarrow \alpha + \eta = \alpha + \beta \Rightarrow \eta = \beta. \tag{15}$$

Therefore,

$$\begin{aligned} &\max_{Q_S P_{X|S}} \mathbb{H}(X|S) = \\ &\max_{\substack{(\alpha,\beta) \\ \zeta = \alpha+\beta}} \Big[-\alpha \log \frac{\alpha}{\alpha+\beta} - (1-\alpha-2\beta) \log \frac{1-\alpha-2\beta}{1-\alpha-\beta} \\ &\qquad - \beta \log \frac{\beta}{\alpha+\beta} - \beta \log \frac{\beta}{1-\alpha-\beta} \Big]. \end{aligned} \tag{16}$$

Considering $Q_S(s=0) = \zeta = \alpha + \beta$ and substituting $\beta = \zeta - \alpha$ in (16) results in

$$\max_{Q_S P_{X|S}} \mathbb{H}(X|S) =$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

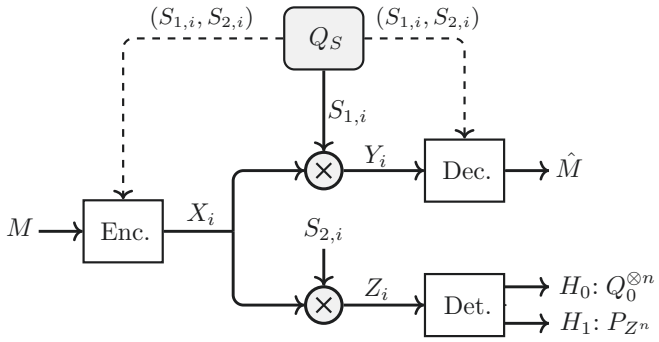

Fig. 4. Binary symmetric channel with multiplicative CSI at the transmitter and the receiver

$$\max_{\alpha}\Big[-\alpha\log\frac{\alpha}{\zeta}-(1+\alpha-2\zeta)\log\frac{1+\alpha-2\zeta}{1-\zeta} -(\zeta-\alpha)\log\frac{\zeta-\alpha}{\zeta}-(\zeta-\alpha)\log\frac{\zeta-\alpha}{1-\zeta}\Big]. \tag{17}$$

Since entropy is a continuous concave function, the maximizer of $\mathbb{H}(X|S)$ is found at the root of the first derivative of (17). This root is $\alpha=\zeta^2$, resulting in $\max\mathbb{H}(X|S)=\mathbb{H}(S)$. Since $Y=Z$, the condition $\mathbb{H}(S|Z)\geqslant\mathbb{I}(X;Z|S)-\mathbb{I}(X;Y|S)$ is automatically satisfied.

We now prove Proposition 1 when CSI is available causally at both the transmitter and the receiver. To prove achievability, we shall substitute specific choices of auxiliary random variables in Theorem 2. We choose U as a Bernoulli random variable with parameter $\eta\in(0:0.5)$ and independent of S, and we set $X=U\oplus S$. Therefore, $Y=Z=U$ and $\mathbb{I}(U;Y|S)=\mathbb{H}(U)$. Since $x_0=0$ we have $Q_0=Q_S$ and the condition $Q_Z=Q_0$ results in $\eta=\zeta$ because

$$Q_S(z=0)=\mathbb{P}(s=0)=\zeta, \tag{18}$$
$$Q_Z(z=0)=\mathbb{P}(u=0)=\eta. \tag{19}$$

Since $Y=Z$, the condition $\mathbb{H}(S|Z)\geqslant\mathbb{I}(U;Z|S)-\mathbb{I}(U;Y|S)$ is automatically satisfied and the covert capacity is lower bounded by $\mathbb{H}_b(\zeta)$. The converse proof follows from the fact $C_{\text{C-TR}}\leqslant C_{\text{NC-TR}}$ by definition. □

Binary Multiplicative State: Consider a channel in which X,Y,Z,S_1, and S_2 are all binary and S_1 and S_2 have a joint distribution with parameters $P(S_1=i,S_2=j)=p_{i,j}$, for $i,j\in\{0,1\}$ and the innocent symbol is $x_0=0$ (See Fig. 4). The law of the channel is

$$Y=X\otimes S_1,\qquad Z=X\otimes S_2. \tag{20}$$

Proposition 2. *The covert capacity of the DMC depicted in Fig. 4 with causal or non-causal CSI available at the transmitter and the receiver is*

$$C_{\text{NC-TR}}=C_{\text{C-TR}}=p_{1,0}. \tag{21}$$

Intuitively, covert communication occurs when the warden's observation is impaired by a bad realization of CSI while the legitimate receiver simultaneously enjoys a good realization of the CSI. Since the receiver knows the CSI, the legitimate channel is effectively noise-free.

Proof. We prove Proposition 2 for the non-causal case, the proof for the causal case is similar and omitted for brevity. Substituting $S=(S_1,S_2)$ in Theorem 1, we obtain

$$\begin{aligned}C_{\text{NC-TR}}&=\max_{\substack{P_{X|S_1,S_2},\\ Q_Z=Q_0}}\big[\mathbb{I}(X;Y|S_1,S_2)\big]\\&=\max_{\substack{P_{X|S_1,S_2},\\ Q_Z=Q_0}}\Big[\sum_{i=0}^{1}\sum_{j=0}^{1}p_{i,j}\mathbb{I}(X;Y|S_1=i,S_2=j)\Big]\\&\overset{(a)}{=}\max_{P_{X|S_1,S_2}}\Big[p_{1,0}\mathbb{I}(X;Y|S_1=1,S_2=0)\Big]\\&=\max_{P_{X|S_1,S_2}}\Big[p_{1,0}\mathbb{H}(X|S_1=1,S_2=0)\Big]\\&=p_{1,0},\end{aligned} \tag{22}$$

where (a) holds because $Y=0$ when $S_1=0$ so that $I(X;Y|S_1=0,S_2=j)=0$ and $Z=X$ when $S_2=1$ so that $P_X=P_Z=Q_0$ imposes $X=0$, and $I(X;Y|S_1=i,S_2=1)=0$. Note that $Q_Z=Q_0$ implies that Z is always equal to zero so that $\mathbb{I}(X;Z|S)\leqslant\mathbb{H}(Z)=0$ and the condition $\mathbb{H}(S|Z)\geqslant\mathbb{I}(X;Z|S)-\mathbb{I}(X;Y|S)$ is automatically satisfied. □

VI. Channel State Information Only Available at the Transmitter

We first recall the definitions of the following classes of broadcast channel, with channel state available only at the transmitter.

Definition 3 (Less Noisy Broadcast Channel With CSI available only at the transmitter)**.** *A discrete memoryless broadcast channel with CSI available only at the transmitter* $\big(\mathcal{X}\times\mathcal{S},W_{Y,Z|X,S},\mathcal{Y}\times\mathcal{Z}\big)$ *is said to be less noisy, if* $\mathbb{I}(U;Y)\geqslant\mathbb{I}(U;Z)$ *for all* $U-(X,S)-(Y,Z)$*. In this case, we say that* Y *is less noisy than* Z*.*

Definition 4 (More Capable Broadcast Channel With CSI available only at the transmitter)**.** *A discrete memoryless broadcast channel with CSI available only at the transmitter* $\big(\mathcal{X}\times\mathcal{S},W_{Y,Z|X,S},\mathcal{Y}\times\mathcal{Z}\big)$ *is said to be more capable, if* $\mathbb{I}(X;Y)\geqslant\mathbb{I}(X;Z)$ *for all* $P_{X,S}$*. In this case, we say that* Y *is more capable than* Z*.*

Theorem 4. *Let*

$$\mathcal{A}\triangleq\left\{\begin{array}{l}R\geqslant 0:\exists P_{U,V,S,X,Y,Z}\in\mathcal{D}\text{ such that}\\ R<\mathbb{I}(U;Y)-\max\big\{\mathbb{I}(U;S),\mathbb{I}(U,V;S)\\ \qquad-\mathbb{I}(V;Y|U)\big\}\end{array}\right\}, \tag{23a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{U,V,S,X,Y,Z}: \\ P_{U,V,S,X,Y,Z} = P_U P_V P_{S|U,V} \mathbb{1}_{\left\{X=X(U,S)\right\}} \\ \qquad \times W_{Y,Z|X,S} \\ Q_S(\cdot) = \sum_{u\in\mathcal{U}} \sum_{v\in\mathcal{V}} P_U(u) P_V(v) \\ \qquad \times P_{S|U,V}(\cdot|u,v) \\ P_Z = Q_0 \\ \mathbb{I}(V;Y|U) > \max\{\mathbb{I}(V;Z), \mathbb{I}(U,V;Z) \\ \qquad -\mathbb{I}(U;Y)\} \\ |\mathcal{U}| \leqslant |\mathcal{X}| + 5 \\ |\mathcal{V}| \leqslant |\mathcal{X}| + 3 \end{array} \right\}. \tag{23b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with non-causal CSI at the transmitter is lower-bounded as

$$C_{\text{NC-T}} \geqslant \sup\{a : a \in \mathcal{A}\}. \tag{24}$$

The proof relies on block-Markov encoding to combine Gel'fand-Pinsker coding, for transmitting the message according to CSI [31], and Wyner-Ziv coding, for secret key generation [32]. The transmitter not only generates a key from S^n, but also selects its codeword according to S^n by using a likelihood encoder [33]–[35]. Instead of directly generating a secret key from the CSI, the transmitter relies on another random variable that is correlated with the CSI to help control the secret key rate. In particular, note that secret keys may not be needed, e.g, when the legitimate receiver's channel is a less noisy version of the warden's channel (see Corollary 1). Proof details are available in Appendix D.

A subset of rated in the region (23a) can be achieved without block-Markov coding or secret key generation. We provide these rates in Theorem 5 for reference. As shown in Section VII, however, secret key generation might be crucial to achieve positive covert rates.

Theorem 5. *Let*

$$\mathcal{A} \triangleq \left\{ \begin{array}{l} R \geqslant 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that} \\ R < \mathbb{I}(U;Y) - \mathbb{I}(U;S) \end{array} \right\}, \tag{25a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z}: \\ P_{S,U,X,Y,Z} = Q_S P_{U|S} \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(U;Y) > \mathbb{I}(U;Z) \\ |\mathcal{U}| \leqslant |\mathcal{X}| + 2 \end{array} \right\}. \tag{25b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with non-causal CSI at the transmitter is lower-bounded as

$$C_{\text{NC-T}} \geqslant \sup\{a : a \in \mathcal{A}\}. \tag{26}$$

Theorem 5 follows from Theorem 4 by choosing S independent of V, so that $P_{S|U,V} = P_{S|U}$. This choice ensures that $\mathbb{I}(V;S) = 0$ and $\mathbb{I}(V;U,Y) = 0$. Alternatively, Theorem 5 can be established with Gel'fand-Pinsker coding with a likelihood encoder but *without* block-Markov encoding or key generation from CSI. Details are omitted for brevity and are available online [26, Appendix E].

Theorem 6. *Let*

$$\mathcal{A} \triangleq \left\{ \begin{array}{l} R \geqslant 0 : \exists P_{S,U,V,X,Y,Z} \in \mathcal{D} \text{ such that} \\ R \leqslant \min\left\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) \right. \\ \qquad \left. -\mathbb{I}(U;S|V)\right\} \end{array} \right\}, \tag{27a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,V,X,Y,Z}: \\ P_{S,U,V,X,Y,Z} = Q_S P_{UV|S} \mathbb{1}_{\left\{X=X(U,S)\right\}} \\ \qquad \times W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \min\left\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \right. \\ \qquad \left. \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\right\} \geqslant \\ \qquad \mathbb{I}(V;Z) - \mathbb{I}(V;S) \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leqslant |\mathcal{X}| + 3 \end{array} \right\}. \tag{27b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with non-causal CSI at the transmitter is upper-bounded as

$$C_{\text{NC-T}} \leqslant \max\{a : a \in \mathcal{A}\}. \tag{28}$$

Proof details are available in Appendix F.

Corollary 1. *Let*

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} R \geqslant 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that} \\ R \leqslant \mathbb{I}(U;Y) - \mathbb{I}(U;S) \end{array} \right\}, \tag{29a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z}: \\ P_{S,U,X,Y,Z} = Q_S P_{U|S} \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ |\mathcal{U}| \leqslant |\mathcal{X}| + 2 \end{array} \right\}. \tag{29b}$$

The covert capacity with CSI available non-causally only at the transmitter when the legitimate receiver's channel is less noisy than the warden's channel, is

$$C_{\text{NC-T}} = \max\{a : a \in \mathcal{A}\}. \tag{30}$$

Proof. The achievability follows from Theorem 5 and the less noisy property of the channel. We can also prove the achievability by using Theorem 4 while generating S independently of V (i.e. $P_{S|U,V} = P_{S|U}$) and the less noisy property of the channel. Furthermore, the converse proof follows from Theorem 6 and the less noisy property of the channel. □

Theorem 7. *Let*

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} R \geqslant 0 : \exists P_{U,V,S,X,Y,Z} \in \mathcal{D} \text{ such that} \\ R < \mathbb{I}(U;Y) + \min\left\{0, \mathbb{I}(V;Y|U) - \mathbb{I}(V;S)\right\} \end{array} \right\}, \tag{31a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{U,V,S,X,Y,Z} : \\ P_{U,V,S,X,Y,Z} = P_U P_V P_{S|V} \mathbb{1}_{\left\{X=X(U,S)\right\}} \\ \qquad \times W_{Y,Z|X,S} \\ Q_S(\cdot) = \sum_{v\in\mathcal{V}} P_V(v) P_{S|V}(\cdot|v) \\ P_Z = Q_0 \\ \mathbb{I}(V;Y|U) > \max\{\mathbb{I}(V;Z), \\ \qquad \mathbb{I}(U,V;Z) - \mathbb{I}(U;Y)\} \\ |\mathcal{U}| \leqslant |\mathcal{X}| + 2 \\ |\mathcal{V}| \leqslant |\mathcal{X}| + 3 \end{array} \right\}. \tag{31b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with causal CSI at the transmitter is lower-bounded as

$$C_{\text{C-T}} \geqslant \sup\{a : a \in \mathcal{A}\}. \tag{32}$$

Theorem 7 is proved using block-Markov encoding to combine a Shannon strategy for sending the message according to CSI and Wyner-Ziv coding for secret key generation. The details of the proof are available in Appendix G.

Theorem 8. *Let*

$$\mathcal{A} \triangleq \left\{R \geqslant 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that } R < \mathbb{I}(U;Y)\right\}, \tag{33a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_U \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(U;Y) > \mathbb{I}(U;Z) \\ |\mathcal{U}| \leqslant |\mathcal{X}| + 1 \end{array} \right\}. \tag{33b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with causal CSI at the transmitter is lower-bounded as

$$C_{\text{C-T}} \geqslant \sup\{a : a \in \mathcal{A}\}. \tag{34}$$

The proof is similar to the proof of Theorem 5, the details are omitted for brevity and are available online; please see [26, Appendix G].

Theorem 9. *Let*

$$\mathcal{A} \triangleq \left\{R \geqslant 0 : \exists P_{S,U,V,X,Y,Z} \in \mathcal{D} \text{ such that } R \leqslant \mathbb{I}(U;Y)\right\}, \tag{35a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,V,X,Y,Z} : \\ P_{S,U,V,X,Y,Z} = Q_S P_V P_{U|V} \mathbb{1}_{\left\{X=X(U,S)\right\}} \\ \qquad \times W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(U;Y) \geqslant \mathbb{I}(V;Z) \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leqslant |\mathcal{X}| \end{array} \right\}. \tag{35b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with causal CSI at the transmitter is upper-bounded as

$$C_{\text{C-T}} \leqslant \max\{a : a \in \mathcal{A}\}. \tag{36}$$

Proof details are available in Appendix H.

Corollary 2. *Let*

$$\mathcal{A} \triangleq \left\{R \geqslant 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D} \text{ such that } R \leqslant \mathbb{I}(U;Y)\right\}, \tag{37a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_S P_U \mathbb{1}_{\left\{X=X(U,S)\right\}} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ |\mathcal{U}| \leqslant |\mathcal{X}| + 1 \end{array} \right\}. \tag{37b}$$

The covert capacity with CSI available causally only at the transmitter when the legitimate receiver's channel is less noisy than the warden's channel is

$$C_{\text{C-T}} = \max\{a : a \in \mathcal{A}\}. \tag{38}$$

Proof. The achievability is proved by using Theorem 8 and the less noisy property of the channel. We can also prove the achievability by using Theorem 7 while generating S independently of V (i.e. $P_{S|V} = Q_S$) and the less noisy property of the channel. Furthermore, the converse proof follows from Theorem 9 and the less noisy property of the channel. □

Theorem 10. *Let*

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} R \geqslant 0 : \exists P_{X,V,S,Y,Z} \in \mathcal{D} \text{ such that} \\ R < \mathbb{I}(X;Y) + \min\left\{0, \mathbb{I}(V;Y|X) - \mathbb{I}(V;S)\right\} \end{array} \right\}, \tag{39a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{X,V,X,Y,Z} : \\ P_{X,V,S,Y,Z} = P_X P_V P_{S|V} W_{Y,Z|X,S} \\ Q_S(\cdot) = \sum_{v\in\mathcal{V}} P_V(v) P_{S|V}(\cdot|v) \\ P_Z = Q_0 \\ \mathbb{I}(V;Y|X) > \max\{\mathbb{I}(V;Z), \\ \qquad \mathbb{I}(X,V;Z) - \mathbb{I}(X;Y)\} \\ |\mathcal{V}| \leqslant |\mathcal{X}| + 3 \end{array} \right\}. \tag{39b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with strictly causal CSI at the transmitter is lower-bounded as

$$C_{\text{SC-T}} \geqslant \sup\{a : a \in \mathcal{A}\}. \tag{40}$$

The proof is similar to the proof of Theorem 7 and we only use the CSI for key generation and not for data transmission. The details are omitted for brevity and are available online; please see [26, Appendix H].

Remark 1. *In the proof of Theorem 4, Theorem 7, and Theorem 10, we assume that there exist a shared secret key for the first two transmission blocks to bootstrap the covert communication between the transmitter and the receiver. The*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

overall rate of this secret key asymptotically amortizes to a negligible value as the number of transmission blocks $B \to \infty$.

Theorem 11. *Let*

$$\mathcal{A} \triangleq \big\{R \geqslant 0 : \exists P_{S,X,Y,Z} \in \mathcal{D} \text{ such that } R < \mathbb{I}(X;Y)\big\}, \tag{41a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_X W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(X;Y) > \mathbb{I}(X;Z) \end{array} \right\}. \tag{41b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with strictly causal CSI at the transmitter is lower-bounded as

$$C_{\text{SC-T}} \geqslant \sup\{a : a \in \mathcal{A}\}. \tag{42}$$

The proof is similar to the proof of Theorem 8, the details of the proof are omitted for brevity and are available online; please see [26, Appendix I]. We now present an upper bound on the covert capacity when the CSI is available strictly causally at the transmitter.

Theorem 12. *Let*

$$\mathcal{A} \triangleq \big\{R \geqslant 0 : \exists P_{S,V,X,Y,Z} \in \mathcal{D} \text{ such that } R \leqslant \mathbb{I}(X;Y)\big\}, \tag{43a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,V,X,Y,Z} : \\ P_{S,V,X,Y,Z} = Q_S P_V P_{X|V} W_{Y,Z|X,S} \\ P_Z = Q_0 \\ \mathbb{I}(X;Y) \geqslant \mathbb{I}(V;Z) \\ |\mathcal{V}| \leqslant |\mathcal{X}| \end{array} \right\}. \tag{43b}$$

The covert capacity of the DMC $W_{Y,Z|S,X}$ with strictly causal CSI at the transmitter is upper-bounded as

$$C_{\text{SC-T}} \leqslant \max\{a : a \in \mathcal{A}\}. \tag{44}$$

Proof details are available in Appendix I.

Corollary 3. *Let*

$$\mathcal{A} \triangleq \big\{R \geqslant 0 : \exists P_{S,X,Y,Z} \in \mathcal{D} \text{ such that } R \leqslant \mathbb{I}(X;Y)\big\}, \tag{45a}$$

where,

$$\mathcal{D} \triangleq \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_X W_{Y,Z|X,S} \\ P_Z = Q_0 \end{array} \right\}. \tag{45b}$$

The covert capacity when CSI is available strictly causally at the transmitter and the legitimate receiver's channel is more capable than the warden's channel is,

$$C_{\text{SC-T}} = \max\{a : a \in \mathcal{A}\}. \tag{46}$$

Proof. The achievability is proved by using Theorem 11 and the more capable property of the channel. We can also prove the achievability by using Theorem 10 while generating S independently of V (i.e. $P_{S|V} = Q_S$) and the more capable property of the channel. Furthermore, the converse is proved by utilizing Theorem 12 and the more capable property of the channel. □

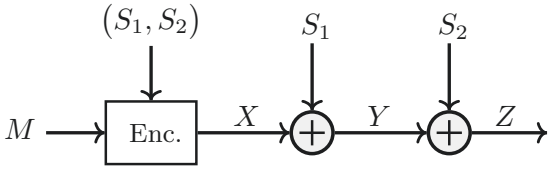

Fig. 5. Degraded channel with binary additive CSI at the transmitter

Remark 2 (Cardinality Bounds)**.** *The cardinality bounds on the auxiliary random variables in Theorems 2 to 10 follows by a standard application of the Eggleston-Fenchel-Carathéodory theorem [36, Theorem 18]. Details are omitted for brevity.*

Remark 3 (Do Stochastic Encoders Improve the Capacity Region?)**.** *We use deterministic encoders when the CSI is available at both of the legitimate terminals, while we use stochastic encoders when the CSI is only available at the transmitter. The use of stochastic encoders is merely motivated by technical convenience in our proof, and we could not conclude whether stochastic encoders outperform deterministic ones.*

VII. Examples of Channels with CSI at Transmitter

We provide two examples of covert communication over state-dependent channels with CSI at the transmitter alone, for which the covert capacity is positive. In both examples, the CSI is additive; however, in the first example the warden's channel is a degraded version of the legitimate receiver's channel while in the second example the legitimate receiver's channel is a degraded version of the warden's channel. The second example shows that our proposed coding scheme with block-Markov encoding and Wyner-Ziv encoding for secret key generation in Theorem 7, can outperform the simple approach for deriving the covert rates in Theorem 8.

Degraded Channel with Binary Additive State: Consider a channel in which X, Y, Z and $S = (S_1, S_2)$ are all binary, and let S_1 and S_2, be independent Bernoulli random variables with parameters $\alpha \in [0 : 0.5]$ and $\beta \in [0 : 0.5]$, respectively, and let $x_0 = 0$ (See Fig. 5). Here, S_1 and S_2 are the CSI of the legitimate receiver's channel and the warden's channel, respectively. The CSI is available causally at the Encoder and the law of the channel is as follows

$$Y = X \oplus S_1, \tag{47}$$

$$Z = Y \oplus S_2. \tag{48}$$

Proposition 3. *The covert capacity of the DMC depicted in Fig. 5 with causal CSI at the transmitter is*

$$C_{\text{C-T}} \overset{(a)}{=} \max_{\substack{P_U, \\ P_Z = Q_0}} \mathbb{H}(U) \overset{(b)}{=} \mathbb{H}_b(\alpha), \tag{49}$$

where $\mathbb{H}_b(\cdot)$ is binary entropy.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

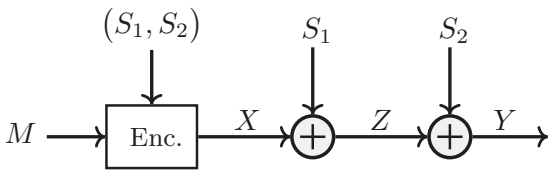

Fig. 6. Reverse degraded channel with binary additive CSI at the transmitter

Proof. The achievability proof for (a) follows from the achievability part of Corollary 2 by considering U, which is the auxiliary random variable that represents the message, as a Bernoulli random variable independent of S_1 and S_2 with parameter $\lambda \in [0 : 0.5]$ and setting $X = U \oplus S_1$. The converse part of (a) follows from the converse part of Corollary 2 and the fact that $\mathbb{I}(U;Y) \leqslant \mathbb{H}(U)$. To prove (b) in Proposition 3, we have

$$\begin{aligned} Q_0(z=0) &= \mathbb{P}(s_1 \oplus s_2 = 0) \\ &= \mathbb{P}(s_1 = 0, s_2 = 0) + \mathbb{P}(s_1 = 1, s_2 = 1) \\ &= (1-\alpha)(1-\beta) + \alpha\beta. \end{aligned} \tag{50}$$

The distribution induced at the output of the warden when transmitting a codeword is

$$\begin{aligned} P_Z(z=0) &= \mathbb{P}(u \oplus s_2 = 0) \\ &= \mathbb{P}(u = 0, s_2 = 0) + \mathbb{P}(u = 1, s_2 = 1) \\ &= (1-\lambda)(1-\beta) + \lambda\beta. \end{aligned} \tag{51}$$

Therefore, the covertness constraint $P_Z = Q_0$ requires $\lambda = \alpha$. □

Reverse Degraded Channel with Binary Additive State: To show the benefits of the proposed scheme, we provide an example in which the region in Theorem 7 strictly improves the region in Theorem 8. Consider a channel in which X, Y, Z and $S = (S_1, S_2)$ are all binary, and let S_1, S_2 and U be independent Bernoulli random variables with parameters $\alpha \in (0 : 0.5]$, $\beta \in (0 : 0.5]$, and $\lambda \in (0 : 0.5]$, respectively, and let $x_0 = 0$ (See Fig. 6). Also, let V be a Bernoulli random variable. Here, S_1 and S_2 are the CSI of the warden's channel and the legitimate receiver's channel, respectively, U is an auxiliary random variable that represents the message, and V is an auxiliary random variable that represents a description of the CSI. The CSI is available causally at the Encoder and the law of the channel is as follows

$$Z = X \oplus S_1, \tag{52}$$
$$Y = Z \oplus S_2. \tag{53}$$

Since for this example $\mathbb{I}(U;Z) \geqslant \mathbb{I}(U;Y)$, the achievable rate region in Theorem 8 results in zero rate but Theorem 7 results in the following region.

Proposition 4. *The covert capacity of the DMC depicted in Fig. 6 with causal CSI at the transmitter is lower bounded as*

$$C_{\text{C-T}} \geqslant \mathbb{H}_b(\eta) - \mathbb{H}_b(\beta), \tag{54}$$

where $\eta = \alpha\beta + (1-\alpha)(1-\beta)$.

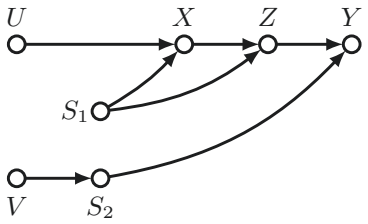

Fig. 7. Chaining between the random variables for the reverse degraded channel with binary additive CSI

Proof. Here we choose $X = U \oplus S_1$ therefore $Z = U$ and $Y = U \oplus S_2$. To prove the region in Proposition 4 by using Theorem 7, we start with covertness constraint $P_Z = Q_0$,

$$Q_0(z=0) = \mathbb{P}(s_1 = 0) = \alpha, \tag{55}$$
$$P_Z(z=0) = \mathbb{P}(u = 0) = \lambda. \tag{56}$$

Therefore, the covertness constraint requires $\lambda = \alpha$. We also choose $V = S_2$ therefore,

$$P_{V|S} = P_{V|S_1,S_2} = P_{V|S_2} = \mathbb{1}_{\{V=S_2\}}. \tag{57}$$

The chaining between the random variables for this example is depicted in Fig. 7. Now we show that the fourth condition in Theorem 7 which includes the following conditions is satisfied,

$$\mathbb{I}(V;Y|U) > \mathbb{I}(V;Z), \tag{58}$$
$$\mathbb{I}(U,V;Y) > \mathbb{I}(U,V;Z). \tag{59}$$

We now have,

$$\begin{aligned} \mathbb{I}(V;Y|U) &= \mathbb{H}(S_2|U) - \mathbb{H}(S_2|U,Y) \\ &= \mathbb{H}(S_2) - \mathbb{H}(S_2|U, U \oplus S_2) \\ &= \mathbb{H}(S_2) = \mathbb{H}_b(\beta), \\ \mathbb{I}(V;Z) &= \mathbb{H}(S_2) - \mathbb{H}(S_2|U) \overset{(a)}{=} 0. \end{aligned}$$

where (a) follows since U and S_2 are independent. Therefore, the condition (58) is satisfied. For the condition in (59) we have,

$$\begin{aligned} \mathbb{I}(U,V;Y) &= \mathbb{H}(U,S_2) - \mathbb{H}(U,S_2|Y) \\ &= \mathbb{H}(U) + \mathbb{H}(S_2) - \mathbb{H}(U,S_2|U \oplus S_2) \\ &= \mathbb{H}(U) + \mathbb{H}(S_2) - \mathbb{H}(S_2|U \oplus S_2), \\ \mathbb{I}(U,V;Z) &= \mathbb{I}(U,S_2;U) = \mathbb{H}(U) \end{aligned}$$

since $\mathbb{H}(S_2) - \mathbb{H}(S_2|U \oplus S_2) > 0$ the condition in (59) is also satisfied. To calculate the covert rate (31a) in Theorem 7 we have,

$$\begin{aligned} \mathbb{I}(V;Y|U) &= \mathbb{H}_b(\beta), \\ \mathbb{I}(V;S) &= \mathbb{I}(S_2;S_1,S_2) \\ &= \mathbb{H}(S_2) = \mathbb{H}_b(\beta), \\ \mathbb{I}(U;Y) &= \mathbb{H}(Y) - \mathbb{H}(Y|U) \\ &= \mathbb{H}(U \oplus S_2) - \mathbb{H}(U \oplus S_2|U) \\ &= \mathbb{H}(U \oplus S_2) - \mathbb{H}(S_2) \\ &= \mathbb{H}_b(\eta) - \mathbb{H}_b(\beta), \end{aligned}$$

where $\eta = \alpha\beta + (1-\alpha)(1-\beta)$.

Remark 4 (Covertness vs. Security)**.** *This example also captures the difference between covertness and security. Here the warden has noiseless access to the transmitted sequence, and therefore it can decode the transmitted message, but since the transmitted sequence has the same statistics as the CSI it cannot prove that communication is happening.*

Remark 5 (Shared Key)**.** *In the examples provided in this section, the codebooks are generated with the same distribution as the CSI S_1 therefore the legitimate terminals need to have access to a shared secret key of negligible rate to discriminate the codewords from the CSI which is consistent with our code definition in Definition 2.*

□

VIII. Conclusion

This paper studies keyless covert communication over state dependent channels, when the CSI is available either at the transmitter alone, or at both the transmitter and receiver, but not to the adversary (warden). Our results show the feasibility of covertly communicating with a positive rate without an externally shared key between the transmitter and the receiver. This is in stark contrast with the known results showing that in the absence of CSI, covert communication without a shared key is impossible at positive rates.

Appendix A
Proof of Theorem 1

Achievability Proof: Fix $P_{X|S}(x|s)$ and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation: For every $s^n \in \mathcal{S}^n$ let $C_n \triangleq \{X^n(s^n, m)\}_{(s^n,m)\in\mathcal{S}^n\times\mathcal{M}}$, where $\mathcal{M} \triangleq [\![1 : 2^{nR}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_{X|S}^{\otimes n}(\cdot|s_i)$. We denote a realization of C_n by $\mathcal{C}_n \triangleq \{x^n(s^n, m)\}_{(s^n,m)\in\mathcal{S}^n\times\mathcal{M}}$.

Encoding: Given the CSI s^n, to send the message m, the transmitter computes $x^n(s^n, m)$ and transmits it over the channel. For a fixed codebook $\mathcal{C}_n$, the induced joint distribution is

$$P^{(\mathcal{C}_n)}_{S^n,M,X^n,Z^n}(s^n, m, \tilde{x}^n, z^n) = Q_S^{\otimes n}(s^n)2^{-nR} \times \mathbb{1}_{\{\tilde{x}^n=x^n(s^n,m)\}}W^{\otimes n}_{Z|S,X}(z^n|s^n, \tilde{x}^n). \tag{60}$$

Covert Analysis: We now show

$$\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \xrightarrow[n\to\infty]{} 0, \tag{61}$$

where

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} Q_S(s)P_{X|S}(x|s)W_{Z|X,S}(\cdot|x, s). \tag{62}$$

Then we choose $P_{X|S}$ such that it satisfies $Q_Z = Q_0$. Henceforth, we denote by $P^{(\mathcal{C}_n)}$ the distributions induced by a fixed codebook $\mathcal{C}_n$, and by $P_{\cdot|C_n}$ the distributions induced by a random codebook C_n. First, consider the following marginal from (60),

$$P_{Z^n|C_n}(z^n) = \sum_{s^n}\sum_{m} Q_S^{\otimes n}(s^n)2^{-nR} \times W^{\otimes n}_{Z|S,X}\big(z^n|s^n, X^n(s^n, m)\big). \tag{63}$$

We now bound $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})]$ as in (67) available at the next page, in which

(a) follows from Jensen's inequality;
(b) follows by taking expectation with respect to $\tilde{s}^n$ and m of the nominator of the second term in the argument of the log function;
(c) follows by defining Ψ_1 and Ψ_2 as in (68) and (70);
(d) follows by defining $\mu_{S,X,Z} = \min_{(s,x,z)\in(\mathcal{S},\mathcal{X},\mathcal{Z})} P_{S,X,Z}(s, x, z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$.

When $n \to \infty$ then Ψ_2 in (70) vanishes; and Ψ_1 in (68) vanishes when $n \to \infty$ if,

$$R > \mathbb{I}(S, X; Z) - \mathbb{H}(S). \tag{64}$$

Basic information identities yield:

$$\mathbb{I}(X, S; Z) = \mathbb{I}(X; S, Z) + \mathbb{I}(S; Z) - \mathbb{I}(X; S). \tag{65}$$

Substituting (65) into (64) leads to

$$R > \mathbb{I}(X; Z|S) - \mathbb{H}(S|Z). \tag{66}$$

Decoding and Error Probability Analysis: By access to the CSI s^n, the receiver declares that $\hat{m} = m$ if there exists a unique index $\hat{m}$ such that $(x^n(s^n, \hat{m}), y^n, s^n) \in \mathcal{T}_\epsilon^{(n)}(X, Y, S)$. According to the law of large numbers and the packing lemma the probability of error goes to zero as $n \to \infty$ if [37],

$$R < \mathbb{I}(X; Y|S). \tag{72}$$

The region in Theorem 1 is derived by combining (66) and (72).

Converse Proof: We now develop an upper bound for the non-causal side information. Consider any sequence of length-n codes for a state-dependent channel with CSI available non-causally at both the transmitter and the receiver, such that $P_e^{(n)} \leqslant \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \delta$ with $\lim_{n\to\infty}\epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (4) as follows.

$$\mathcal{A}_\epsilon \triangleq \big\{R \geqslant 0 : \exists P_{S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leqslant \mathbb{I}(X; Y|S) + \epsilon\big\}, \tag{73a}$$

where

$$\mathcal{D}_\epsilon = \left\{\begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_{X|S} W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leqslant \epsilon \\ \mathbb{H}(S|Z) \geqslant \mathbb{I}(X; Z|S) - \mathbb{I}(X; Y|S) - 2\epsilon \end{array}\right\}. \tag{73b}$$

We next show that if a rate R is achievable, then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques,

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leq} \mathbb{H}(M|S^n) - \mathbb{H}(M|Y^n, S^n) + n\epsilon_n \\ &= \mathbb{I}(M; Y^n|S^n) + n\epsilon_n \end{aligned}$$

$$\mathbb{E}_{C_n}\left[\mathbb{D}\big(P_{Z^n|C_n}||Q_Z^{\otimes n}\big)\right] = \mathbb{E}_{C_n}\Big[\sum_{z^n} P_{Z^n|C_n}(z^n)\log\Big(\frac{P_{Z^n|C_n}(z^n)}{Q_Z^{\otimes n}(z^n)}\Big)\Big]$$

$$= \mathbb{E}_{C_n}\left[\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,X^n(s^n,m)\big)\log\left(\frac{\sum_{(\tilde{s}^n,\tilde{m})}Q_S^{\otimes n}(\tilde{s}^n)W_{Z|S,X}^{\otimes n}\big(z^n|\tilde{s}^n,X^n(\tilde{s}^n,\tilde{m})\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right)\right]$$

$$\overset{(a)}{\leq}\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)}P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\log\mathbb{E}_{\setminus(s^n,m)}\left[\frac{\sum_{(\tilde{s}^n,\tilde{m})}Q_S^{\otimes n}(\tilde{s}^n)W_{Z|S,X}^{\otimes n}\big(z^n|\tilde{s}^n,X^n(\tilde{s}^n,\tilde{m})\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right]$$

$$= \sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)}P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)$$
$$\times\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}+\mathbb{E}_{\setminus(s^n,m)}\left[\frac{\sum_{(\tilde{s}^n,\tilde{m})\neq(s^n,m)}Q_S^{\otimes n}(\tilde{s}^n)W_{Z|S,X}^{\otimes n}\big(z^n|\tilde{s}^n,X^n(\tilde{s}^n,\tilde{m})\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right]\right)$$

$$= \sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)}P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)$$
$$\times\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}+\mathbb{E}_{\setminus s^n}\left[\sum_{\tilde{m}\neq m}\frac{\mathbb{E}_{\setminus m}\sum_{\tilde{s}^n}\Big(Q_S^{\otimes n}(\tilde{s}^n)W_{Z|S,X}^{\otimes n}\big(z^n|\tilde{s}^n,X^n(\tilde{s}^n,\tilde{m})\big)\Big)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right]\right)$$

$$\overset{(b)}{=}\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)}P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)$$
$$\times\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}+\mathbb{E}_{\setminus s^n}\left[\sum_{\tilde{m}\neq m}\frac{Q_Z^{\otimes n}(z^n)}{2^{nR}Q_Z^{\otimes n}(z^n)}\right]\right)$$

$$\leqslant\sum_{z^n}\sum_{s^n}\sum_{m}\frac{1}{2^{nR}}\sum_{x^n(s^n,m)}P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}+1\right)$$

$$\overset{(c)}{=}\Psi_1+\Psi_2 \tag{67}$$

$$\Psi_1=\sum_{m}\frac{1}{2^{nR}}\sum_{(s^n,x^n(s^n,m),z^n)\in\mathcal{T}_\epsilon^{(n)}}P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}+1\right) \tag{68}$$

$$\leq\log\left(\frac{2^{-n(1-\epsilon)\left(\mathbb{H}(S)+\mathbb{H}(Z|S,X)\right)}}{2^{nR}2^{-n(1+\epsilon)\mathbb{H}(Z)}}+1\right) \tag{69}$$

$$\Psi_2=\sum_{m}\frac{1}{2^{nR}}\sum_{(s^n,x^n(s^n,m),z^n)\notin\mathcal{T}_\epsilon^{(n)}}P_{S,X,Z}^{\otimes n}\big(s^n,x^n(s^n,m),z^n\big)\log\left(\frac{Q_S^{\otimes n}(s^n)W_{Z|S,X}^{\otimes n}\big(z^n|s^n,x^n(s^n,m)\big)}{2^{nR}Q_Z^{\otimes n}(z^n)}+1\right) \tag{70}$$

$$\overset{(d)}{\leq}2|S||X||Z|e^{-n\epsilon^2\mu_{S,X,Z}}n\log\left(\frac{2}{\mu_Z}+1\right) \tag{71}$$

$$= \sum_{i=1}^{n}\mathbb{I}(M;Y_i|Y^{i-1},S^n)+n\epsilon_n$$
$$= \sum_{i=1}^{n}[\mathbb{H}(Y_i|Y^{i-1},S^n)-\mathbb{H}(Y_i|M,Y^{i-1},S^n)]+n\epsilon_n$$
$$\overset{(b)}{\leq}\sum_{i=1}^{n}[\mathbb{H}(Y_i|S_i)-\mathbb{H}(Y_i|M,Y^{i-1},S^n,X^n)]+n\epsilon_n$$

$$= \sum_{i=1}^{n}[\mathbb{H}(Y_i|S_i)-\mathbb{H}(Y_i|S_i,X_i)]+n\epsilon_n$$
$$= \sum_{i=1}^{n}\mathbb{I}(X_i;Y_i|S_i)+n\epsilon_n$$
$$\overset{(c)}{\leq}n\mathbb{I}(\tilde{X};\tilde{Y}|\tilde{S})+n\epsilon_n$$
$$\overset{(d)}{\leq}n\mathbb{I}(\tilde{X};\tilde{Y}|\tilde{S})+n\epsilon$$

$$\overset{(e)}{=} n\mathbb{I}(X;Y|S) + n\epsilon, \tag{74}$$

where

- (a) follows from Fano's inequality and since M is independent of S^n;
- (b) holds because conditioning does not increase entropy;
- (c) follows from the concavity of mutual information, with the resulting random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ having the following distributions

$$\tilde{P}_{X,S}(x,s) \triangleq \frac{1}{n}\sum_{i=1}^{n} P_{X_i,S_i}(x,s), \tag{75a}$$

$$\tilde{P}_{X,S,Y,Z}(x,s,y,z) \triangleq \tilde{P}_{X,S}(x,s) W_{Y,Z|X,S}(y,z|x,s); \tag{75b}$$

- (d) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \nu\}$, where we choose n large enough such that $\nu \geqslant \frac{\delta}{n}$;
- (e) follows by defining $X \triangleq \tilde{X}$, $Y \triangleq \tilde{Y}$, and $S \triangleq \tilde{S}$.

We also have,

$$\begin{aligned}
nR &= \mathbb{H}(M) \\
&= \mathbb{H}(M|S^n) \\
&\geq \mathbb{I}(M; Z^n|S^n) \\
&\overset{(a)}{=} \mathbb{I}(M, X^n; Z^n|S^n) \\
&\geq \mathbb{I}(X^n; Z^n|S^n) \\
&= \mathbb{I}(X^n, S^n; Z^n) - \mathbb{I}(S^n; Z^n) \\
&= \sum_{x^n}\sum_{s^n}\sum_{z^n} P(x^n,s^n,z^n) \log \frac{W_{Z|X,S}^{\otimes n}(z^n|x^n,s^n)}{P(z^n)} \\
&\quad - \mathbb{H}(S^n) + \mathbb{H}(S^n|Z^n) \\
&\geqslant \sum_{x^n}\sum_{s^n}\sum_{z^n} P(x^n,s^n,z^n) \log \frac{W_{Z|X,S}^{\otimes n}(z^n|x^n,s^n)}{P(z^n)} \\
&\quad + \mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) - \mathbb{H}(S^n) - \delta \\
&\geq \sum_{i=1}^{n}\sum_{x_i}\sum_{s_i}\sum_{z_i} P(x_i,s_i,z_i) \log \frac{W_{Z|X,S}(z_i|x_i,s_i)}{Q_0(z_i)} \\
&\quad - \sum_{i=1}^{n} \mathbb{H}(S_i) - \delta \\
&= \sum_{i=1}^{n} \mathbb{D}(P_{X_i,S_i,Z_i}||P_{X_i,S_i}Q_0) - \sum_{i=1}^{n}\mathbb{H}(S_i) - \delta \\
&\overset{(b)}{\geq} n\mathbb{D}(\tilde{P}_{X,S,Z}||\tilde{P}_{X,S}Q_0) - n\mathbb{H}(\tilde{S}) - \delta \\
&= n\mathbb{D}(\tilde{P}_{X,S,Z}||\tilde{P}_{X,S}\tilde{P}_Z) + \mathbb{D}(\tilde{P}_Z||Q_0) - n\mathbb{H}(\tilde{S}) - \delta \\
&\overset{(c)}{\geq} n\mathbb{I}(\tilde{X},\tilde{S};\tilde{Z}) - n\mathbb{H}(\tilde{S}) - \delta \\
&\overset{(d)}{=} n\mathbb{I}(X,S;Z) - n\mathbb{H}(S) - \delta,
\end{aligned} \tag{76}$$

where

- (a) follows because X^n is a function of (M, S^n);
- (b) follows from Jensen's inequality, the convexity of $\mathbb{D}(\cdot||\cdot)$, and concavity of $\mathbb{H}(\cdot)$;
- (c) follows from the positivity of the KL-divergence and the definition of random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ in (75);
- (d) follows by defining $X \triangleq \tilde{X}$, $Z \triangleq \tilde{Z}$, and $S \triangleq \tilde{S}$.

For any $\nu > 0$, by choosing n large enough and substituting (65) into (76) ensures that

$$\begin{aligned}
R &\geq \mathbb{I}(X;Z|S) - \mathbb{H}(S|Z) - \nu, \\
&\geq \mathbb{I}(X;Z|S) - \mathbb{H}(S|Z) - \epsilon,
\end{aligned} \tag{77}$$

where the last inequality follows from the definition of $\epsilon \triangleq \max\{\epsilon_n, \nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leqslant \epsilon$, note that for n large enough,

$$\begin{aligned}
\mathbb{D}(P_Z||Q_0) &= \mathbb{D}(P_{\tilde{Z}}||Q_0) = \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\
&\leqslant \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leqslant \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \frac{\delta}{n} \leqslant \nu \leqslant \epsilon.
\end{aligned} \tag{78}$$

Combining (74) and (77) shows that $\forall \epsilon_n, \nu > 0$, $R \leqslant \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$C_{\text{NC-TR}} = \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{79}$$

Continuity at Zero: One can prove the continuity at zero of $\mathcal{A}_\epsilon$ by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(X;Y|S)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(X;Z|S) - \mathbb{H}(S|Z)$ in the continuity at zero proof in Appendix F and following the exact same arguments.

Appendix B
Proof of Theorem 2

Achievability Proof: To prove the achievability of Theorem 2 it is convenient to introduce an associated channel $W_{Y,Z|U,S}$ as follows: Let $U \in \mathcal{U}$ be an arbitrary auxiliary random variable which is independent of the state S, and let $x : \mathcal{U} \times \mathcal{S} \mapsto \mathcal{X}$ be a deterministic mapping subject to $\mathbb{1}_{\{x=x(s,u)\}}$. According to the Shannon strategy [38], we define the $W_{Y,Z|U,S}$ as a channel specified by

$$W_{Y,Z|U,S} = \sum_{x\in\mathcal{X}} \mathbb{1}_{\{x=x(s,u)\}} W_{Y,Z|X,S}(y,z|x,s), \tag{80}$$

which results in a channel with input U, outputs Y, Z, and state S. Therefore, we only focus on the coding problem for the channel $W_{Y,Z|U,S}$ for the achievability proof.

We use block-Markov coding in which B independent messages are transmitted over B channel blocks, each of length r, therefore the overall codeword length is $n = rB$ symbols. The warden's observation Z^n can be described in terms of observations in individual block-Markov blocks $Z^n = (Z_1^r, \ldots, Z_B^r)$. The distribution of the warden's observation, induced by the block-Markov coding, is $P_{Z^n} \triangleq P_{Z_1^r,\ldots,Z_B^r}$ and the target output distribution is $Q_0^{\otimes n} = \prod_{j=1}^{B} Q_0^{\otimes r}$. Therefore,

$$\begin{aligned}
&\mathbb{D}\Big(P_{Z^n}||Q_0^{\otimes n}\Big) = \mathbb{D}\Big(P_{Z_1^r\ldots Z_B^r}||Q_0^{\otimes rB}\Big) \\
&= \sum_{j=1}^{B} \mathbb{D}\Big(P_{Z_j^r|Z_{j+1}^{B,r}}||Q_0^{\otimes r}|P_{Z_{j+1}^{B,r}}\Big) \\
&= \sum_{j=1}^{B} \Big[\mathbb{D}\Big(P_{Z_j^r}||Q_0^{\otimes r}\Big) + \mathbb{D}(P_{Z_j^r|Z_{j+1}^{B,r}}||P_{Z_j^r}|P_{Z_{j+1}^{B,r}})\Big]
\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

$$= \sum_{j=1}^{B} \Big[\mathbb{D}\Big(P_{Z_j^r} || Q_0^{\otimes r}\Big) + \mathbb{I}\Big(Z_j^r; Z_{j+1}^{B,r}\Big)\Big], \tag{81}$$

where $Z_{j+1}^{B,r} = \{Z_{j+1}^r, \dots Z_B^r\}$. Hence, $\mathbb{D}\Big(P_{Z^n} || Q_0^{\otimes n}\Big) \underset{n\to\infty}{\longrightarrow} 0$, is equivalent to;

$$\mathbb{D}\Big(P_{Z_j^r} || Q_0^{\otimes r}\Big) \underset{r\to\infty}{\longrightarrow} 0,\ \mathbb{I}\Big(Z_j^r; Z_{j+1}^{B,r}\Big) \underset{r\to\infty}{\longrightarrow} 0, \quad \forall j \in [\![1:B]\!]. \tag{82}$$

This requires constructing a code that approximates $Q_0^{\otimes r}$ in each block, while eliminating the dependencies across blocks created by block-Markov coding. The random code generation is as follows.

Fix $P_U(u)$, $x = x(s,u)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1:B]\!]$, create a function $\Phi : S_j^r \mapsto [\![1:2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(s_j^r)$ independently and uniformly at random for every $s_j^r \in \mathcal{S}^r$. The key $k_j = \Phi(s_j^r)$ obtained in the block $j \in [\![1:B]\!]$ from the state sequence s_j^r is used to assist the encoder in the next block.

Codebook Generation for Messages: For each block $j \in [\![1:B]\!]$, let $C_r \triangleq \{U^r(m_j, k_{j-1})\}_{(m_j,k_{j-1})\in\mathcal{M}\times\mathcal{K}}$, where $\mathcal{M} \triangleq [\![1:2^{rR}]\!]$ and $\mathcal{K} \triangleq [\![1:2^{rR_k}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_U^{\otimes r}$. We denote a realization of C_r by $\mathcal{C}_r \triangleq \{u^r(m_j, k_{j-1})\}_{(m_j,k_{j-1})\in\mathcal{M}\times\mathcal{K}}$.

Encoding: For the first block, we assume that the transmitter and the receiver have access to a shared secret key k_0, in this block to transmit m_1 the encoder computes $u^r(m_1, k_0)$ and transmits codeword x^r, where $x_i = x(u_i(m_1,k_0), s_i)$. At the end of the first block, the encoder generates a key from CSI s_1^r to be used in Block 2.

For block $j \in [\![2:B]\!]$, to send the message m_j according to the generated key k_{j-1} from the previous block, the encoder computes $u^r(m_j, k_{j-1})$ and transmits codeword x^r, where $x_i = x(u_i(m_j,k_{j-1}), s_i)$. Also, at the end of each block $j \in [\![2:B]\!]$, the encoder generates a key from CSI s_j^r to be used in the next block.

Define

$$\begin{aligned}&\Upsilon^{(\mathcal{C}_r)}_{M_j,K_{j-1},U^r,S_j^r,Z_j^r,K_j}(m_j, k_{j-1}, \tilde{u}^r, s_j^r, z_j^r, k_j)\\ &\triangleq 2^{-r(R_k+R)} \mathbb{1}_{\{\tilde{u}^r = u^r(m_j,k_{j-1})\}} Q_S^{\otimes r}(s_j^r)\\ &\quad \times W_{Z|U,S}^{\otimes r}(z_j^r|\tilde{u}^r, s_j^r)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}.\end{aligned} \tag{83}$$

For a fixed codebook $\mathcal{C}_r$, the induced joint distribution by our code design (i.e. $P^{(\mathcal{C}_r)}$) satisfies

$$\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{M_j,K_{j-1},U^r,S_j^r,Z_j^r,K_j} || \Upsilon^{(\mathcal{C}_r)}_{M_j,K_{j-1},U^r,S_j^r,Z_j^r,K_j}\Big) \leqslant \epsilon. \tag{84}$$

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (84) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the key K_{j-1} is assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which is made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (96).

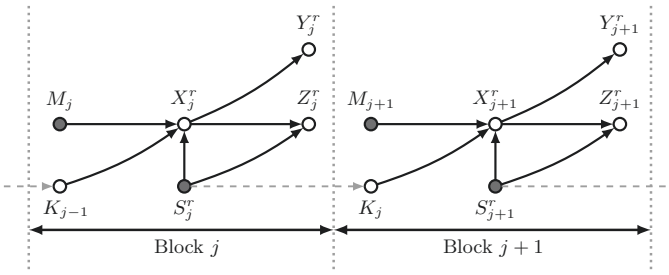

Fig. 8. Functional dependence graph for the block-Markov encoding scheme

Covert Analysis: We now show $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n} || Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$, where C_n is the set of all codebooks from all blocks, and

$$\begin{aligned}Q_Z(\cdot) = \sum_{s\in\mathcal{S}}\sum_{u\in\mathcal{U}}\sum_{x\in\mathcal{X}} Q_S(s)P_U(u)\mathbb{1}_{\{X=X(u,s)\}}\\ \times W_{Z|X,S}(\cdot|x,s).\end{aligned} \tag{85}$$

Then we choose P_U and $X(U,S)$ such that it satisfies $Q_Z = Q_0$. From the expansion in (81), by substituting Q_0 with Q_Z, for every block $j \in [\![2:B]\!]$,

$$\begin{aligned}\mathbb{I}(Z_j^r; Z_{j+1}^{B,r}) &\leqslant \mathbb{I}(Z_j^r; K_j, Z_{j+1}^{B,r})\\ &\overset{(a)}{=} \mathbb{I}(Z_j^r; K_j),\end{aligned} \tag{86}$$

where (a) holds because $Z_j^r - K_j - Z_{j+1}^{B,r}$ forms a Markov chain, as seen in the functional dependence graph depicted in Fig. 8. Also,

$$\begin{aligned}\mathbb{I}(Z_j^r; K_j) &= \mathbb{D}(P^{(\mathcal{C}_n)}_{Z_j^r,K_j} || P^{(\mathcal{C}_n)}_{Z_j^r} P^{(\mathcal{C}_n)}_{K_j})\\ &\overset{(b)}{\leqslant} \mathbb{D}(P^{(\mathcal{C}_n)}_{Z_j^r,K_j} || Q_Z^{\otimes r} Q_{K_j}),\end{aligned} \tag{87}$$

where Q_{K_j} is the uniform distribution over $[\![1:2^{rR_K}]\!]$ and (b) follows from the positivity of relative entropy and

$$\begin{aligned}\mathbb{D}(P_{Z_j^r,K_j} || P_{Z_j^r} P_{K_j}) &= \mathbb{D}(P_{Z_j^r,K_j} || Q_Z^{\otimes r} Q_{K_j})\\ &\quad - \mathbb{D}(P_{K_j} || Q_{K_j}) - \mathbb{D}(P_{Z_j^r} || Q_Z^{\otimes r}).\end{aligned} \tag{88}$$

Therefore by combining (81), (87), and (88),

$$\mathbb{D}(P_{Z^n|C_n} || Q_Z^{\otimes n}) \leqslant 2\sum_{j=1}^{B} \mathbb{D}(P_{Z_j^r,K_j|C_r} || Q_Z^{\otimes r} Q_{K_j}). \tag{89}$$

We now proceed to bound the right-hand side of (89). First, consider the following marginal from (83),

$$\begin{aligned}\Upsilon_{Z_j^r,K_j|C_r}(z_j^r, k_j) &= \sum_{m_j}\sum_{k_{j-1}}\sum_{s_j^r} \frac{1}{2^{r(R+R_k)}}\\ &\times Q_S^{\otimes r}(s_j^r) W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(m_j,k_{j-1}), s_j^r\big)\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}.\end{aligned} \tag{90}$$

From (84) and the monotonicity of KL-divergence we have,

$$\mathbb{D}\big(\Upsilon_{Z_j^r,K_j|C_r} || P_{Z_j^r,K_j|C_r}\big) \leqslant \epsilon. \tag{91}$$

To bound the Right Hand Side (RHS) of (89) by using Lemma 1 and the triangle inequality we have,

$$\mathbb{E}_{C_r} || P_{Z_j^r,K_j|C_r} - Q_Z^{\otimes r} Q_{K_j} ||_1$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

$$\leqslant \mathbb{E}_{C_r}||P_{Z_j^r,K_j|C_r} - \Upsilon_{Z_j^r,K_j|C_r}||_1 + \mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_j|C_r} - Q_Z^{\otimes r}Q_{K_j}||_1. \quad (92)$$

From Lemma 1 and (91) the first term on the RHS of (92) vanishes as r grows. We now bound the second term on the RHS of (92) by using Lemma 1 as in (93) available at the next page, in which

(a) follows from Jensen's inequality;
(b) holds because $\mathbb{1}_{\{\cdot\}} \leqslant 1$;
(c) follows by defining Ψ_1 and Ψ_2 as follows,

$$\begin{aligned}
\Psi_1 &= \sum_{k_j}\sum_{m_j}\sum_{k_{j-1}} \frac{1}{2^{r(R+R_k+R_K)}} \\
&\times \sum_{\left(u^r(m_j,k_{j-1}),s_j^r,z_j^r\right)\in\mathcal{T}_\epsilon^{(n)}} \Upsilon_{U^r,S^r,Z^r}^{\otimes r}\left(u^r(m_j,k_{j-1}),s_j^r,z_j^r\right) \\
&\times \log\Bigg(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\left(z_j^r|u^r(m_j,k_{j-1}),s_j^r\right)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)} \\
&+ \frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)} + \frac{W_{Z|U}^{\otimes r}\left(z_j^r|u^r(m_j,k_{j-1})\right)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)} + 1\Bigg) \\
&\leq \log\Bigg(\frac{2^{rR_K}2^{-r(1-\epsilon)(\mathbb{H}(S)+\mathbb{H}(Z|U,S))}}{2^{r(R+R_k)}2^{-r(1+\epsilon)\mathbb{H}(Z)}} \\
&+ \frac{2^{rR_K}2^{-r(1-\epsilon)\mathbb{H}(S,Z)}}{2^{-r(1+\epsilon)\mathbb{H}(Z)}} + \frac{2^{-r(1-\epsilon)\mathbb{H}(Z|U)}}{2^{r(R+R_k)}2^{-r(1+\epsilon)\mathbb{H}(Z)}} + 1\Bigg)
\end{aligned} \quad (94)$$

$$\begin{aligned}
\Psi_2 &= \sum_{k_j}\sum_{m_j}\sum_{k_{j-1}} \frac{1}{2^{r(R+R_k+R_K)}} \\
&\times \sum_{\left(u^r(m_j,k_{j-1}),s_j^r,z_j^r\right)\notin\mathcal{T}_\epsilon^{(n)}} \Upsilon_{U^r,S^r,Z^r}^{\otimes r}\left(u^r(m_j,k_{j-1}),s_j^r,z_j^r\right) \\
&\times \log\Bigg(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\left(z_j^r|u^r(m_j,k_{j-1}),s_j^r\right)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)} \\
&+ \frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)} + \frac{W_{Z|U}^{\otimes r}\left(z_j^r|u^r(m_j,k_{j-1})\right)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)} + 1\Bigg) \\
&\leq 2|U||S||Z|e^{-r\epsilon^2\mu_{U,S,Z}} r\log\Big(\frac{2}{\mu_Z}+1\Big),
\end{aligned} \quad (95)$$

where $\mu_{U,S,Z} = \min_{(u,s,z)\in(\mathcal{U},\mathcal{S},\mathcal{Z})} P_{U,S,Z}(u,s,z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$. When $r \to \infty$ then $\Psi_2 \to 0$, by choosing $R_K = \mathbb{H}(S|Z) - \epsilon$, Ψ_1 vanishes when r grows if,

$$R + R_k > \mathbb{I}(U;Z|S), \quad (96a)$$
$$R + R_k > \mathbb{I}(U;Z). \quad (96b)$$

Since U and S are independent, (96b) is redundant because of (96a).

Decoding and Error Probability Analysis: At the end of the block $j \in [\![1:B]\!]$, using its knowledge of the CSI s_j^r of the current block and the key k_{j-1} generated from the previous block, the receiver finds a unique $\hat{m}_j$ such that $\left(u^r(\hat{m}_j,k_{j-1}),s_j^r,y_j^r\right) \in \mathcal{T}_\epsilon^{(r)}$. To analyze the probability of error, we define the following error events for $j \in [\![1:B]\!]$

$$\mathcal{E} = \left\{\hat{M} \neq M\right\}, \quad (97a)$$
$$\mathcal{E}_j = \left\{\hat{M}_j \neq M_j\right\}, \quad (97b)$$
$$\mathcal{E}_{1,j} = \left\{\left(U^r(M_j,K_{j-1}),S_j^r\right) \notin \mathcal{T}_{\epsilon_1}^{(r)}(Q_SP_U)\right\}, \quad (97c)$$
$$\mathcal{E}_{2,j} = \left\{\left(U^r(M_j,K_{j-1}),S_j^r,Y_j^r\right) \notin \mathcal{T}_{\epsilon_2}^{(r)}(Q_SP_UW_{Y|U,S})\right\}, \quad (97d)$$
$$\mathcal{E}_{3,j} = \left\{\left(U^r(k_{j-1},\hat{m}_j),S_j^r,Y_j^r\right) \in \mathcal{T}_{\epsilon_2}^{(r)}, \text{for some } \hat{m}_j \neq M_j\right\}, \quad (97e)$$

where $\epsilon_2 > \epsilon_1 > \epsilon > 0$. The probability of error is upper bounded as follows,

$$\mathbb{P}(\mathcal{E}) \leqslant \mathbb{P}\Big\{\bigcup_{j=1}^B \mathcal{E}_j\Big\} \leqslant \sum_{j=1}^B \mathbb{P}(\mathcal{E}_j). \quad (98)$$

Now we bound $\mathbb{P}(\mathcal{E}_j)$ by using union bound

$$\mathbb{P}(\mathcal{E}_j) \leqslant \mathbb{P}(\mathcal{E}_{1,j}) + \mathbb{P}(\mathcal{E}_{1,j}^c \cap \mathcal{E}_{2,j}) + \mathbb{P}(\mathcal{E}_{2,j}^c \cap \mathcal{E}_{3,j}). \quad (99)$$

By the law of large numbers the first and second term on RHS of (99) vanishes when r grows. According to the law of large numbers and the packing lemma, the last term on RHS of (99) vanishes when r grows if [37],

$$R < \mathbb{I}(U;S,Y) = \mathbb{I}(U;Y|S). \quad (100)$$

Furthermore, this scheme requires that

$$R_k \leqslant R_K = \mathbb{H}(S|Z) - \epsilon, \quad (101)$$

The region in Theorem 2 is obtained by applying Fourier-Motzkin to (96a), (100), and (101).

Converse Proof: We now develop an upper bound when CSI is available causally at both of the legitimate terminals. Consider any sequence of length-n codes for a state-dependent channel with CSI available causally at both the transmitter and the receiver such that $P_e^{(n)} \leqslant \epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \delta$ with $\lim_{n\to\infty}\epsilon_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (6) as follows,

$$\mathcal{A}_\epsilon \triangleq \left\{R \geqslant 0 : \exists P_{S,U,X,Y,Z} \in \mathcal{D}_\epsilon : R \leqslant \mathbb{I}(U;Y|S) + \epsilon\right\}, \quad (102a)$$

where

$$\mathcal{D}_\epsilon = \left\{\begin{array}{l} P_{S,U,X,Y,Z} : \\ P_{S,U,X,Y,Z} = Q_SP_U\mathbb{1}_{\left\{X=X(U,S)\right\}}W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z\|Q_0\right) \leqslant \epsilon \\ \mathbb{H}(S|Z) \geqslant \mathbb{I}(U;Z|S) - \mathbb{I}(U;Y|S) - 2\epsilon \\ |\mathcal{U}| \leqslant |\mathcal{X}| + 1 \end{array}\right\}. \quad (102b)$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques,

$$\begin{aligned}
nR &= \mathbb{H}(M) \\
&\overset{(a)}{\leq} \mathbb{H}(M|S^n) - \mathbb{H}(M|Y^n,S^n) + n\epsilon_n \\
&= \mathbb{I}(M;Y^n|S^n) + n\epsilon_n \\
&= \sum_{i=1}^n \mathbb{I}(M;Y_i|Y^{i-1},S^n) + n\epsilon_n
\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

$$
\begin{aligned}
&\mathbb{E}_{C_r}[\mathbb{D}(\Upsilon_{Z_j^r,K_j|C_r}||Q_Z^{\otimes r}Q_{K_j})] = \mathbb{E}_{C_r}\Big[\sum_{z_j^r,k_j}\Upsilon_{Z_j^r,K_j|C_r}(z_j^r,k_j)\log\Big(\frac{\Upsilon_{Z_j^r,K_j|C_r}(z_j^r,k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_j}(k_j)}\Big)\Big]\\
&= \mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(m_j,k_{j-1}),s_j^r\big)\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}\\
&\qquad\times\log\Bigg(\frac{\sum_{\tilde m_j}\sum_{\tilde k_{j-1}}\sum_{\tilde s_j^r}Q_S^{\otimes r}(\tilde s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(\tilde m_j,\tilde k_{j-1}),\tilde s_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde s_j^r)\}}}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg)\Bigg]\\
&\overset{(a)}{\le}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}\sum_{u^r(m_j,k_{j-1})}\Upsilon_{U^r,S^r,Z^r}^{\otimes r}\big(u^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\mathbb{E}_{\Phi(s_j^r)}\big[\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}\big]\\
&\qquad\times\log\mathbb{E}_{\backslash((m_j,k_{j-1}),\Phi(s_j^r))}\Bigg[\frac{\sum_{\tilde m_j}\sum_{\tilde k_{j-1}}\sum_{\tilde s_j^r}Q_S^{\otimes r}(\tilde s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(\tilde m_j,\tilde k_{j-1}),\tilde s_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde s_j^r)\}}}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg]\\
&\overset{(b)}{\le}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}\sum_{u^r(m_j,k_{j-1})}\Upsilon_{U^r,S^r,Z^r}^{\otimes r}\big(u^r(m_j,k_{j-1},s_j^r,z_j^r)\big)\frac{1}{2^{rR_K}}\\
&\qquad\times\log\frac{1}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg(Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1}),s_j^r\big)\\
&\qquad+\mathbb{E}_{\backslash(m_j,k_{j-1})}\Bigg[\sum_{(\tilde m_j,\tilde k_{j-1})\neq(m_j,k_{j-1})}Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(\tilde m_j,\tilde k_{j-1}),s_j^r\big)\Bigg]\\
&\qquad+\mathbb{E}_{\backslash\Phi(s_j^r)}\Bigg[\sum_{\tilde s_j^r}Q_S^{\otimes r}(\tilde s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1}),\tilde s_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde s_j^r)\}}\Bigg]\\
&\qquad+\mathbb{E}_{\backslash((m_j,k_{j-1}),\Phi(s_j^r))}\Bigg[\sum_{\tilde s_j^r}\sum_{(\tilde m_j,\tilde k_{j-1})\neq(m_j,k_{j-1})}Q_S^{\otimes r}(\tilde s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(\tilde m_j,\tilde k_{j-1}),\tilde s_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde s_j^r)\}}\Bigg]\Bigg)\\
&\le\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{s_j^r}\sum_{u^r(m_j,k_{j-1})}\Upsilon_{U^r,S^r,Z^r}^{\otimes r}\big(u^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\qquad\times\log\Bigg(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}+\sum_{(\tilde k_{j-1},\tilde m_j)\neq(m_j,k_{j-1})}\frac{Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&\qquad+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\le\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{s_j^r}\sum_{u^r(m_j,k_{j-1})}\Upsilon_{U^r,S^r,Z^r}^{\otimes r}\big(u^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\qquad\times\log\Bigg(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|U,S}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\overset{(c)}{=}\Psi_1+\Psi_2,
\end{aligned}
\tag{93}
$$

$$
\begin{aligned}
&=\sum_{i=1}^{n}[\mathbb{H}(Y_i|Y^{i-1},S^n)-\mathbb{H}(Y_i|M,Y^{i-1},S^n)]+n\epsilon_n\\
&\overset{(b)}{\le}\sum_{i=1}^{n}[\mathbb{H}(Y_i|S_i)-\mathbb{H}(Y_i|U_i,S_i)]+n\epsilon_n\\
&=\sum_{i=1}^{n}\mathbb{I}(U_i;Y_i|S_i)+n\epsilon_n\\
&\overset{(c)}{\le}n\mathbb{I}(\tilde U;\tilde Y|\tilde S)+n\epsilon_n\\
&\overset{(d)}{\le}n\mathbb{I}(\tilde U;\tilde Y|\tilde S)+n\epsilon\\
&\overset{(e)}{=}n\mathbb{I}(U;Y|S)+n\epsilon
\end{aligned}
\tag{103}
$$

where

(a) follows from Fano's inequality and since M is indepen-

dent of S^n;

(b) holds because conditioning does not increase entropy and $U_i = (M, Y^{i-1}, S^n_{\sim i})$;

(c) follows from the concavity of mutual information, with the resulting random variables $\tilde{U}$, $\tilde{S}$, and $\tilde{Y}$ having the following distributions

$$\tilde{P}_{U,S,X}(u,s,x) \triangleq \frac{1}{n}\sum_{i=1}^{n} P_{U_i,S_i,X_i}(u,s,x), \tag{104a}$$

$$\tilde{P}_{U,S,X,Y,Z}(u,s,x,y,z) \triangleq \tilde{P}_{U,S,X}(u,s,x) \times W_{Y,Z|X,S}(y,z|x,s); \tag{104b}$$

(d) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \nu\}$, where we choose n large enough such that $\nu \geqslant \frac{\delta}{n}$;

(e) follows by defining $U \triangleq \tilde{U}$, $Y \triangleq \tilde{Y}$, and $S \triangleq \tilde{S}$.

We now have,

$$\begin{aligned} nR &\overset{(a)}{\geq} n\mathbb{I}(\tilde{X},\tilde{S};\tilde{Z}) - n\mathbb{H}(\tilde{S}) - \epsilon \\ &\overset{(b)}{\geq} n\mathbb{I}(\tilde{U},\tilde{S};\tilde{Z}) - n\mathbb{H}(\tilde{S}) - \epsilon \\ &\overset{(c)}{=} \mathbb{I}(U,S;Z) - \mathbb{H}(S) - \epsilon, \end{aligned} \tag{105}$$

where

(a) follows from the exact same steps as in (77);

(b) follows from the Markov chain $U - (X,S) - Z$ and from the definition of random variables $\tilde{U}$, $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ in (104);

(c) follows by defining $U \triangleq \tilde{U}$, $Z \triangleq \tilde{Z}$, and $S \triangleq \tilde{S}$.

Rewriting the bound in (105) by using the basic property in (65) leads to

$$R \geqslant \mathbb{I}(U;Z|S) - \mathbb{H}(S|Z) - \epsilon. \tag{106}$$

To show that $\mathbb{D}(P_Z||Q_0) \leqslant \epsilon$, note that for n large enough,

$$\begin{aligned} \mathbb{D}(P_Z||Q_0) &= \mathbb{D}(P_{\tilde{Z}}||Q_0) = \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leqslant \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leqslant \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \frac{\delta}{n} \leqslant \nu \leqslant \epsilon. \end{aligned} \tag{107}$$

Combining (103) and (106) shows that $\forall \epsilon_n, \nu > 0$, $R \leqslant \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$C_{\text{C-TR}} = \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{108}$$

Continuity at Zero: Continuity at zero for $\mathcal{A}_\epsilon$ is established by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(U;Y|S)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(U;Z|S) - \mathbb{H}(S|Z)$ in the continuity at zero proof in Appendix F and following the same arguments.

Appendix C
Proof of Theorem 3

Achievability Proof: We adopt a block-Markov encoding scheme in which B independent messages are transmitted over B channel blocks each of length r, such that $n = rB$. The warden's observation is $Z^n = (Z_1^r, \ldots, Z_B^r)$, the target output distribution is $Q_0^{\otimes n}$, and Equation (81), describing the distance between the two distributions, continues to hold. The random code generation is as follows.

Fix P_X and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1:B]\!]$, create a function $\Phi : S_j^r \mapsto [\![1:2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(s_j^r)$ independently and uniformly at random for every $s_j^r \in \mathcal{S}^r$. The key $k_j = \Phi(s_j^r)$ obtained in the block $j \in [\![1:B]\!]$ from the state sequence s_j^r is used to assist the encoder in the next block.

Codebook Generation for Messages: For each block $j \in [\![1:B]\!]$, let $C_r \triangleq \{X^r(m_j, k_{j-1})\}_{(m_j,k_{j-1})\in\mathcal{M}\times\mathcal{K}}$, where $\mathcal{M} \triangleq [\![1:2^{rR}]\!]$ and $\mathcal{K} \triangleq [\![1:2^{rR_k}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_X^{\otimes r}$. We denote a realization of C_r by $\mathcal{C}_r \triangleq \{x^r(m_j, k_{j-1})\}_{(m_j,k_{j-1})\in\mathcal{M}\times\mathcal{K}}$.

Encoding: For the first block, we assume that the transmitter and the receiver have access to a shared secret key k_0, in this block to transmit m_1 the encoder computes $x^r(m_1, k_0)$ and transmits it over the channel. At the end of the first block, the encoder generates a key from CSI s_1^r to be used in Block 2.

For block $j \in [\![2:B]\!]$, to send the message m_j according to the generated key k_{j-1} from the previous block, the encoder computes $x^r(m_j, k_{j-1})$ and transmits it over the channel. Also, at the end of the block $j \in [\![2:B]\!]$, the encoder generates a key from CSI s_j^r to be used in the next block.

Define

$$\begin{aligned} &\Upsilon^{(\mathcal{C}_r)}_{M_j,K_{j-1},X^r,S_j^r,Z_j^r,K_j}(m_j,k_{j-1},\tilde{x}^r,s_j^r,z_j^r,k_j) \\ &\triangleq 2^{-r(R+R_k)}\mathbb{1}_{\{\tilde{x}^r = x^r(m_j,k_{j-1})\}} Q_S^{\otimes r}(s_j^r) \\ &\times W_{Z|X,S}^{\otimes r}\big(z_j^r|\tilde{x}^r, s_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}. \end{aligned} \tag{109}$$

For a fixed codebook $\mathcal{C}_r$, the induced joint distribution by our code design (i.e. $P^{(\mathcal{C}_r)}$) satisfies

$$\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{M_j,K_{j-1},X^r,S_j^r,Z_j^r,K_j}||\Upsilon^{(\mathcal{C}_r)}_{M_j,K_{j-1},X^r,S_j^r,Z_j^r,K_j}\Big) \leqslant \epsilon. \tag{110}$$

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (110) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the key K_{j-1} is assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which is made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (118).

Covert Analysis: We now show that this coding scheme guarantees that $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \underset{n\to\infty}{\longrightarrow} 0$, where C_n is the set of all the codebooks for all blocks, and

$$Q_Z(\cdot) = \sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} Q_S(s)P_X(x)W_{Z|X,S}(\cdot|x,s). \tag{111}$$

Then we choose P_X such that it satisfies $Q_Z = Q_0$. Similar to (89), by using the functional dependence graph depicted in Fig. 9,

$$\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n}) \leqslant 2\sum_{j=1}^{B}\mathbb{D}(P_{Z_j^r,K_j|C_r}||Q_Z^{\otimes r}Q_{K_j}). \tag{112}$$

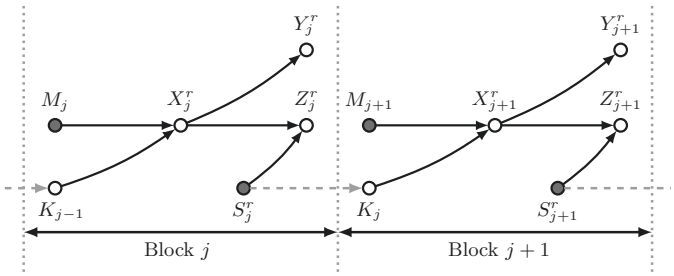

Fig. 9. Functional dependence graph for the block-Markov encoding scheme

We now proceed to bound the RHS of (112). First, consider the following marginal from (109),

$$\Upsilon_{Z_j^r,K_j|C_r}(z_j^r,k_j)=\sum_{m_j}\sum_{k_{j-1}}\sum_{s_j^r}\frac{1}{2^{r(R+R_k)}}\times Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(m_j,k_{j-1}),s_j^r\big)\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}. \tag{113}$$

To bound the RHS of (112) by using Lemma 1 and the triangle inequality we have,

$$\begin{aligned}&\mathbb{E}_{C_r}||P_{Z_j^r,K_j|C_r}-Q_Z^{\otimes r}Q_{K_j}||_1\\&\leqslant\mathbb{E}_{C_r}||P_{Z_j^r,K_j|C_r}-\Upsilon_{Z_j^r,K_j|C_r}||_1\\&\quad+\mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_j|C_r}-Q_Z^{\otimes r}Q_{K_j}||_1.\end{aligned} \tag{114}$$

From Lemma 1 and (110) the first term on the RHS of (114) vanishes as r grows. We now bound the second term on the RHS of (92) by using Lemma 1 as in (115) available at the next page, in which

(a) follows from Jensen's inequality;
(b) holds because $\mathbb{1}_{\{.\}}\leqslant 1$;
(c) follows by defining Ψ_1 and Ψ_2 as follows,

$$\begin{aligned}\Psi_1=&\sum_{k_j}\sum_{k_{j-1}}\sum_{m_j}\frac{1}{2^{r(R+R_k+R_K)}}\\&\times\sum_{\left(x^r(m_j,k_{j-1}),s_j^r,z_j^r\right)\in\mathcal{T}_\epsilon^{(n)}}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\&\times\log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.\\&\left.+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|X}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)\\\leq&\log\left(\frac{2^{rR_K}2^{-r(1-\epsilon)\left(\mathbb{H}(S)+\mathbb{H}(Z|X,S)\right)}}{2^{r(R+R_k)}2^{-r(1+\epsilon)\mathbb{H}(Z)}}\right.\\&\left.+\frac{2^{rR_K}2^{-r(1-\epsilon)\mathbb{H}(S,Z)}}{2^{-r(1+\epsilon)\mathbb{H}(Z)}}+\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|X)}}{2^{r(R+R_k)}2^{-r(1+\epsilon)\mathbb{H}(Z)}}+1\right)\end{aligned} \tag{116}$$

$$\begin{aligned}\Psi_2=&\sum_{k_j}\sum_{k_{j-1}}\sum_{m_j}\frac{1}{2^{r(R+R_k+R_K)}}\\&\times\sum_{\left(x^r(m_j,k_{j-1}),s_j^r,z_j^r\right)\notin\mathcal{T}_\epsilon^{(n)}}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\&\times\log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.\\&\left.+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|X}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)\\\leq&2|X||S||Z|e^{-r\epsilon^2\mu_{X,S,Z}}r\log\Big(\frac{2}{\mu_Z}+1\Big),\end{aligned} \tag{117}$$

where $\mu_{X,S,Z}=\min_{(x,s,z)\in(\mathcal{X},\mathcal{S},\mathcal{Z})}P_{X,S,Z}(x,s,z)$ and $\mu_Z=\min_{z\in\mathcal{Z}}P_Z(z)$. When $r\to\infty$ then $\Psi_2\to 0$, by choosing $R_K=\mathbb{H}(S|Z)-\epsilon$, Ψ_1 vanishes when r grows if,

$$R+R_k>\mathbb{I}(X;Z|S), \tag{118a}$$
$$R+R_k>\mathbb{I}(X;Z). \tag{118b}$$

Since X and S are independent, (118b) is redundant because of (118a).

Decoding and Error Probability Analysis: At the end of the block $j\in[\![1:B]\!]$, using its knowledge of the CSI s_j^r of the current block and the key k_{j-1} generated from the previous block, the receiver finds a unique $\hat{m}_j$ such that $\big(u^r(\hat{m}_j,k_{j-1}),s_j^r,y_j^r\big)\in\mathcal{T}_\epsilon^{(r)}$. To analyze the probability of error, we define the following error events for $j\in[\![1:B]\!]$,

$$\mathcal{E}=\big\{\hat{M}\neq M\big\}, \tag{119a}$$
$$\mathcal{E}_j=\big\{\hat{M}_j\neq M_j\big\}, \tag{119b}$$
$$\mathcal{E}_{1,j}=\big\{\big(X^r(M_j,K_{j-1}),S_j^r\big)\notin\mathcal{T}_{\epsilon_1}^{(r)}(Q_SP_X)\big\}, \tag{119c}$$
$$\mathcal{E}_{2,j}=\big\{\big(X^r(M_j,K_{j-1}),S_j^r,Y_j^r\big)\notin\mathcal{T}_{\epsilon_2}^{(r)}(Q_SP_XW_{Y|X,S})\big\}, \tag{119d}$$
$$\mathcal{E}_{3,j}=\big\{\big(X^r(k_{j-1},\hat{m}_j),S_j^r,Y_j^r\big)\in\mathcal{T}_{\epsilon_2}^{(r)},\ \text{for some }\hat{m}_j\neq M_j\big\}, \tag{119e}$$

where $\epsilon_2>\epsilon_1>\epsilon>0$. The probability of error is upper bounded as follows,

$$\mathbb{P}(\mathcal{E})\leqslant\mathbb{P}\Big\{\bigcup\nolimits_{j=1}^{B}\mathcal{E}_j\Big\}\leqslant\sum_{j=1}^{B}\mathbb{P}(\mathcal{E}_j). \tag{120}$$

Now we bound $\mathbb{P}(\mathcal{E}_j)$ by using union bound

$$\mathbb{P}(\mathcal{E}_j)\leqslant\mathbb{P}(\mathcal{E}_{1,j})+\mathbb{P}(\mathcal{E}_{1,j}^c\cap\mathcal{E}_{2,j})+\mathbb{P}(\mathcal{E}_{2,j}^c\cap\mathcal{E}_{3,j}). \tag{121}$$

By the law of large numbers the first and second term on RHS of (121) vanishes when r grows. According to the law of large numbers and the packing lemma, the last term on RHS of (121) vanishes when r grows if [37],

$$R<\mathbb{I}(X;S,Y)=\mathbb{I}(X;Y|S). \tag{122}$$

Furthermore, this scheme requires that,

$$R_k\leqslant R_K=\mathbb{H}(S|Z)-\epsilon. \tag{123}$$

The region in Theorem 3 is obtained by applying Fourier-Motzkin to (118a), (122), and (123).

Remark 6. *In the achievability proof of Theorem 2 and Theorem 3 we transmit B messages over B blocks. We assume that there exists a shared secret key between the transmitter and the receiver that is used in the first block to bootstrap the covert communication. Consequently, the shared secret key*

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

$$
\begin{aligned}
&\mathbb{E}_{C_r}[\mathbb{D}(\Upsilon_{Z_j^r,K_j|C_r}||Q_Z^{\otimes r}Q_{K_j})] = \mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_j)}\Upsilon_{Z_j^r,K_j|C_r}(z_j^r,k_j)\log\left(\frac{\Upsilon_{Z_j^r,K_j|C_r}(z_j^r,k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_j}(k_j)}\right)\Bigg]\\
&= \mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(m_j,k_{j-1}),s_j^r\big)\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}\\
&\qquad\times\log\left(\frac{\sum\limits_{\tilde{m}_j}\sum\limits_{\tilde{k}_{j-1}}\sum\limits_{\tilde{s}_j^r}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right)\Bigg]\\
&\overset{(a)}{\leq}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}\sum_{x^r(m_j,k_{j-1})}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\times\mathbb{E}_{\Phi(s_j^r)}\big[\mathbb{1}_{\{k_j=\Phi(s_j^r)\}}\big]\\
&\qquad\times\log\mathbb{E}_{\backslash((m_j,k_{j-1}),\Phi(s_j^r))}\left[\frac{\sum\limits_{\tilde{m}_j}\sum\limits_{\tilde{k}_{j-1}}\sum\limits_{\tilde{s}_j^r}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\right]\\
&\overset{(b)}{\leq}\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k)}}\sum_{s_j^r}\sum_{x^r(m_j,k_{j-1})}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\times\frac{1}{2^{rR_K}}\\
&\qquad\times\log\frac{1}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg(Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)\\
&\qquad+\mathbb{E}_{\backslash(m_j,k_{j-1})}\left[\sum_{(\tilde{m}_j,\tilde{k}_{j-1})\neq(m_j,k_{j-1})}Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{k}_{j-1}),s_j^r\big)\right]\\
&\qquad+\mathbb{E}_{\backslash\Phi(s_j^r)}\left[\sum_{\tilde{s}_j^r\neq s_j^r}Q_S^{\otimes r}(\tilde{s}_j^r)\times W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}\right]\\
&\qquad+\mathbb{E}_{\backslash((m_j,k_{j-1}),\Phi(s_j^r))}\left[\sum_{\tilde{s}_j^r\neq s_j^r}\sum_{(\tilde{m}_j,\tilde{k}_{j-1})\neq(m_j,k_{j-1})}Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|X^r(\tilde{m}_j,\tilde{k}_{j-1}),\tilde{s}_j^r\big)\mathbb{1}_{\{k_j=\Phi(\tilde{s}_j^r)\}}\right]\Bigg)\\
&\leq\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{s_j^r}\sum_{x^r(m_j,k_{j-1})}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\qquad\times\log\Bigg(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}+\sum_{(\tilde{m}_j,\tilde{k}_{j-1})\neq(m_j,k_{j-1})}\frac{Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&\qquad+\sum_{\tilde{s}_j^r\neq s_j^r}\frac{Q_S^{\otimes r}(\tilde{s}_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),\tilde{s}_j^r\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\leq\sum_{(z_j^r,k_j)}\sum_{m_j}\sum_{k_{j-1}}\frac{1}{2^{r(R+R_k+R_K)}}\sum_{s_j^r}\sum_{x^r(m_j,k_{j-1})}\Upsilon_{X^r,S^r,Z^r}^{\otimes r}\big(x^r(m_j,k_{j-1}),s_j^r,z_j^r\big)\\
&\times\log\left(\frac{Q_S^{\otimes r}(s_j^r)W_{Z|X,S}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1}),s_j^r\big)}{2^{r(R+R_k-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{2^{rR_K}Q_{S,Z}^{\otimes r}(s_j^r,z_j^r)}{Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|X}^{\otimes r}\big(z_j^r|x^r(m_j,k_{j-1})\big)}{2^{r(R+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)\\
&\overset{(c)}{=}\Psi_1+\Psi_2,
\end{aligned}
\tag{115}
$$

rate is negligible. However, to eliminate the need for this secret key, similar to the block Markov encoding schemes in [39], [40] we can transmit $B-1$ messages over B blocks and remove the decodability condition of the message of the first block, this results in a slight rate loss in the first block, which becomes asymptotically negligible as the number of blocks $B\to\infty$.

Converse Proof: To establish the upper bound, consider any sequence of length-n codes for a state-dependent channel with CSI available strictly causally at both the transmitter and the receiver, such that $P_e^{(n)}\leqslant\epsilon_n$ and $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n})\leqslant\delta$ with $\lim_{n\to\infty}\epsilon_n=0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon>0$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

that expands the region defined in (8) as follows,

$$\mathcal{A}_\epsilon \triangleq \left\{R \geqslant 0 : \exists P_{S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leqslant \mathbb{I}(X;Y|S) + \epsilon\right\}, \tag{124a}$$

where

$$\mathcal{D}_\epsilon = \left\{ \begin{array}{l} P_{S,X,Y,Z} : \\ P_{S,X,Y,Z} = Q_S P_X W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leqslant \epsilon \\ \mathbb{H}(S|Z) \geqslant \mathbb{I}(X;Z|S) - \mathbb{I}(X;Y|S) - 2\epsilon \end{array} \right\}. \tag{124b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques.

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &\overset{(a)}{\leqslant} \mathbb{H}(M|S^n) - \mathbb{H}(M|Y^n, S^n) + n\epsilon_n \\ &= \mathbb{I}(M; Y^n|S^n) + n\epsilon_n \\ &= \sum_{i=1}^n \mathbb{I}(M; Y_i|Y^{i-1}, S^n) + n\epsilon_n \\ &\leqslant \sum_{i=1}^n \left[\mathbb{H}(Y_i|Y^{i-1}, S^n) - \mathbb{H}(Y_i|Y^{i-1}, S^n, X_i, M)\right] + n\epsilon_n \\ &\overset{(b)}{\leqslant} \sum_{i=1}^n \mathbb{I}(X_i; Y_i|S_i) + n\epsilon_n \\ &\overset{(c)}{\leqslant} n\mathbb{I}(\tilde{X}; \tilde{Y}|\tilde{S}) + n\epsilon_n \\ &\overset{(d)}{\leqslant} n\mathbb{I}(\tilde{X}; \tilde{Y}|\tilde{S}) + n\epsilon \\ &\overset{(e)}{=} n\mathbb{I}(X; Y|S) + n\epsilon, \end{aligned} \tag{125}$$

where

- (a) follows from Fano's inequality and since M is independent of S^n;
- (b) holds because conditioning does not increase entropy and $(M, Y^{i-1}, S^n_{\sim i}, X^n_{\sim i}) - (X_i, S_i) - Y_i$ forms a Markov chain;
- (c) follows from concavity of mutual information, with respect to the input distribution, with the random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ having the following distributions

$$\tilde{P}_{X,S}(x,s) \triangleq \frac{1}{n}\sum_{i=1}^n P_{X_i,S_i}(x,s), \tag{126a}$$

$$\tilde{P}_{X,S,Y,Z}(x,s,y,z) \triangleq \tilde{P}_{X,S}(x,s) W_{Y,Z|X,S}(y,z|x,s); \tag{126b}$$

- (d) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \nu\}$, where we choose n large enough such that $\nu \geqslant \frac{\delta}{n}$;
- (e) follows by defining $U \triangleq \tilde{U}$, $Y \triangleq \tilde{Y}$, and $S \triangleq \tilde{S}$.

By following the same steps as in (77) we also have,

$$nR \geq n\mathbb{I}(\tilde{X}, \tilde{S}; \tilde{Z}) - n\mathbb{H}(\tilde{S}) - \epsilon, \tag{127}$$

where the random variables $\tilde{X}$, $\tilde{S}$, $\tilde{Y}$, and $\tilde{Z}$ have been defined in (126). Substituting (65) into (127) leads to

$$\begin{aligned} R &\geq \mathbb{I}(\tilde{X}; \tilde{Z}|\tilde{S}) - \mathbb{H}(\tilde{S}|\tilde{Z}) - \epsilon \\ &= \mathbb{I}(X; Z|S) - \mathbb{H}(S|Z) - \epsilon, \end{aligned} \tag{128}$$

where the last equality follows by defining $U \triangleq \tilde{U}$, $Z \triangleq \tilde{Z}$, and $S \triangleq \tilde{S}$. To show that $\mathbb{D}(P_Z||Q_0) \leqslant \epsilon$, note that for n large enough,

$$\begin{aligned} \mathbb{D}(P_Z||Q_0) &= \mathbb{D}(P_{\tilde{Z}}||Q_0) = \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^n P_{Z_i}\middle|\middle|Q_0\right) \\ &\leqslant \frac{1}{n}\sum_{i=1}^n \mathbb{D}(P_{Z_i}||Q_0) \leqslant \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \frac{\delta}{n} \leqslant \nu \leqslant \epsilon. \end{aligned} \tag{129}$$

Combining (125) and (128) shows that $\forall \epsilon_n, \nu > 0$, $R \leqslant \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$C_{\text{SC-TR}} = \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{130}$$

Continuity at Zero: One can prove the continuity at zero of $\mathcal{A}_\epsilon$ by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(X;Y|S)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(X;Z|S) - \mathbb{H}(S|Z)$ in the continuity at zero proof in Appendix F and following the exact same arguments.

Appendix D
Proof of Theorem 4

We adopt a block-Markov encoding scheme in which B independent messages are transmitted over B channel blocks each of length r, such that $n = rB$. The warden's observation is $Z^n = (Z_1^r, \ldots, Z_B^r)$, the distribution induced at the output of the warden is P_{Z^n}, the target output distribution is $Q_0^{\otimes n}$, and Equation (81), describing the distance between the two distributions, continues to hold. The random code generation is as follows.

Fix $P_{U|S}(u|s)$, $P_{V|S}(v|s)$, $x(u,s)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1 : B]\!]$, let $C_1^{(r)} \triangleq \left\{V^r(a_j)\right\}_{a_j \in \mathcal{A}}$, where $\mathcal{A} \triangleq [\![1 : 2^{r\tilde{R}}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_V^{\otimes r}$, where $P_V = \sum_{s\in\mathcal{S}} Q_S(s) P_{V|S}(v|s)$. We denote a realization of $C_1^{(r)}$ by $\mathcal{C}_1^{(r)} \triangleq \left\{v^r(a_j)\right\}_{a_j \in \mathcal{A}}$. Partition the set of indices $a_j \in [\![1 : 2^{r\tilde{R}}]\!]$ into bins $\mathcal{B}(t)$, $t \in [\![1 : 2^{rR_T}]\!]$ by using function $\varphi : V^r(a_j) \mapsto [\![1 : 2^{rR_T}]\!]$ through random binning by choosing the value of $\varphi(v^r(a_j))$ independently and uniformly at random for every $v^r(a_j) \in \mathcal{V}^r$. For each block $j \in [\![1 : B]\!]$, create a function $\Phi : V^r(a_j) \mapsto [\![1 : 2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(v^r(a_j))$ independently and uniformly at random for every $v^r(a_j) \in \mathcal{V}^r$. The key $k_j = \Phi(v^r(a_j))$ obtained in block $j \in [\![1 : B]\!]$ from the description of the CSI sequence $v^r(a_j)$ is used to assist the encoder in block $j + 2$.

Codebook Generation for Messages: For each block $j \in [\![1 : B]\!]$, let $C_2^{(r)} \triangleq \left\{U^r(m_j, t_{j-1}, k_{j-2}, \ell_j)\right\}_{(m_j, t_{j-1}, k_{j-2}, \ell_j) \in \mathcal{M}\times\mathcal{T}\times\mathcal{K}\times\mathcal{L}}$, where $\mathcal{M} \triangleq [\![1 : 2^{rR}]\!]$, $\mathcal{T} \triangleq [\![1 : 2^{r\tilde{R}_t}]\!]$, $\mathcal{K} \triangleq [\![1 : 2^{rR_k}]\!]$, and $\mathcal{L} \triangleq [\![1 : 2^{rR'}]\!]$, be a random codebook consisting of independent random sequences each generated

$$\begin{aligned}\Gamma^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},L_j,A_j,U^r,V^r,S_j^r,Z_j^r,K_{j-1},T_j,K_j}&(m_j,t_{j-1},k_{j-2},\ell_j,a_j,\tilde{u}^r,\tilde{v}_j^r,s_j^r,z_j^r,k_{j-1},t_j,k_j)\\ &= 2^{-r(R+R_t+R_k+R'+\tilde{R})}\mathbb{1}_{\{\tilde{u}^r=u^r(m_j,t_{j-1},k_{j-2},\ell_j)\}}\mathbb{1}_{\{\tilde{v}^r=v^r(a_j)\}}P^{\otimes r}_{S|U,V}(s_j^r|\tilde{u}^r,\tilde{v}^r)\\ &\quad\times W^{\otimes r}_{Z|U,S}(z_j^r|\tilde{u}^r,s_j^r)2^{-rR_k}\mathbb{1}_{\{t_j=\sigma(\tilde{v}^r)\}}\mathbb{1}_{\{k_j=\Phi(\tilde{v}^r)\}},\end{aligned}\tag{131}$$

$$f(\ell_j,a_j|s_j^r,m_j,t_{j-1},k_{j-2})=\frac{P^{\otimes r}_{S|U,V}\left(s_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\right)}{\sum\limits_{\ell'_j\in[\![1:2^{rR'}]\!]}\sum\limits_{a'_j\in[\![1:2^{r\tilde{R}}]\!]}P^{\otimes r}_{S|U,V}\left(s_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell'_j),v^r(a'_j)\right)},\tag{132}$$

$$\begin{aligned}\Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j}&(m_j,t_{j-1},k_{j-2},s_j^r,\ell_j,a_j,\tilde{u}^r,\tilde{v}^r,z_j^r,k_{j-1},t_j,k_j)\\ &\triangleq 2^{-r(R+R_t+R_k)}Q_S^{\otimes r}(s_j^r)f(\ell_j,a_j|s_j^r,m_j,t_{j-1},k_{j-2})\mathbb{1}_{\{\tilde{u}^r=u^r(m_j,t_{j-1},k_{j-2},\ell_j)\}}\mathbb{1}_{\{\tilde{v}^r=v^r(a_j)\}}\\ &\quad\times W^{\otimes r}_{Z|U,S}(z_j^r|\tilde{u}^r,s_j^r)2^{-rR_k}\mathbb{1}_{\{t_j=\sigma(\tilde{v}^r)\}}\mathbb{1}_{\{k_j=\Phi(\tilde{v}^r)\}}.\end{aligned}\tag{133}$$

according to $P_U^{\otimes r}$. We denote a realization of $C_2^{(r)}$ by $\mathcal{C}_2^{(r)}\triangleq\left\{u^r(m_j,t_{j-1},k_{j-2},\ell_j)\right\}_{(m_j,t_{j-1},k_{j-2},\ell_j)\in\mathcal{M}\times\mathcal{T}\times\mathcal{K}\times\mathcal{L}}$. Let, $C_r=\left\{C_1^{(r)},C_2^{(r)}\right\}$ and $\mathcal{C}_r=\left\{\mathcal{C}_1^{(r)},\mathcal{C}_2^{(r)}\right\}$. The indices $(m_j,t_{j-1},k_{j-2},\ell_j)$ can be viewed as a three layer binning. We define an ideal PMF for codebook $\mathcal{C}_r$ as in (131) at the top of this page, as an approximate distribution to facilitate the analysis, in (131) $W_{Z|U,S}$ is the marginal distribution $W_{Z|U,S}=\sum_{x\in\mathcal{X}}\mathbb{1}_{\{x=x(u,s)\}}W_{Z|X,S}$ and $P_{S|U,V}$ is defined as follows

$$\begin{aligned}P_{S|U,V}(s|u,v)&\triangleq\frac{P_{S,U,V}(s,u,v)}{P_{U,V}(u,v)}\\ &=\frac{Q_S(s)P_{U|S}(u|s)P_{V|S}(v|s)}{\sum_{s\in\mathcal{S}}Q_S(s)P_{U|S}(u|s)P_{V|S}(v|s)}.\end{aligned}\tag{134}$$

Encoding: We assume that the transmitter and the receiver have access to shared secret keys k_{-1} and k_0 for the first two blocks, but after the first two blocks they use the key that they generate from the CSI.

In the first block, to send the message m_1 according to k_{-1}, the encoder generates the index t_0 uniformly at random and then generates the indices ℓ_1 and a_1 according to the distribution defined in (132) at the top of this page with $j=1$, $P_{S|U,V}$ in (132) is defined in (134). Based on these indices, the encoder computes $u^r(m_1,t_0,k_{-1},\ell_1)$ and $v^r(a_1)$ and transmits codeword x^r, where $x_i=x(u_i(m_1,t_0,k_{-1},\ell_1),s_i)$. Note that, the index t_0 does not convey any useful information. Simultaneously, it uses the description of the CSI $v^r(a_1)$ to generate a reconciliation index t_1 and a key k_1 to be used in the second and the third blocks, respectively.

In the second block, to send the message m_2 and reconciliation index t_1 according to k_0, the encoder generates the indices ℓ_2 and a_2 according to the likelihood encoder described in (132) with $j=2$. Based on these indices, the encoder computes $u^r(m_2,t_1,k_0,\ell_2)$ and $v^r(a_2)$ and transmits codeword x^r, where $x_i=x(u_i(m_2,t_1,k_0,\ell_1),s_i)$. Simultaneously, it uses the description of the CSI $v^r(a_2)$ to generate a reconciliation index t_2 and a key k_2 to be used in the third and the fourth block, respectively.

In block $j\in[\![3:B]\!]$, to send the message m_j and the reconciliation index t_{j-1}, generated in the previous block, according to the key k_{j-2}, generated in the block $j-2$, and the CSI of the current block, the encoder generates indices ℓ_j and a_j from the bin (m_j,t_{j-1},k_{j-2}) according to the likelihood encoder described in (132). The encoder then transmits the codeword x^r, where each coordinate of the transmitted signal is a function of the CSI, as well as the corresponding sample of the transmitter's codeword u_i, i.e., $x_i=x(u_i(m_j,t_{j-1},k_{j-2},\ell_j),s_i)$. Simultaneously, the encoder uses the description of the CSI $v^r(a_j)$ to generate a reconciliation index t_j and a key k_j to be used in the block $j+1$ and the block $j+2$, respectively. The encoding scheme in block $j\in[\![3:B]\!]$ is depicted in Fig. 10.

Considering (133) at the top of this page, for a given codebook $\mathcal{C}_r$, the induced joint distribution over the codebook $\left(\text{i.e., } P^{(\mathcal{C}_r)}\right)$ satisfies

$$\begin{aligned}\mathbb{D}\Big(&P^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j}\\ &||\Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j}\Big)\leqslant\epsilon.\end{aligned}\tag{135}$$

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (135) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the keys K_{j-2}, K_{j-1} and the reconciliation index T_{j-1} are assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which are made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (144) and (150).

Covert Analysis: We now show $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})]\xrightarrow[n\to\infty]{}0$, where C_n is the set of all the codebooks for all blocks and,

$$\begin{aligned}Q_Z(\cdot)=\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}}\sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}}&P_U(u)P_V(v)P_{S|U,V}(s|u,v)\\ &\times\mathbb{1}_{\left\{X=X(u,s)\right\}}W_{Z|X,S}(\cdot|x,s),\end{aligned}\tag{136}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

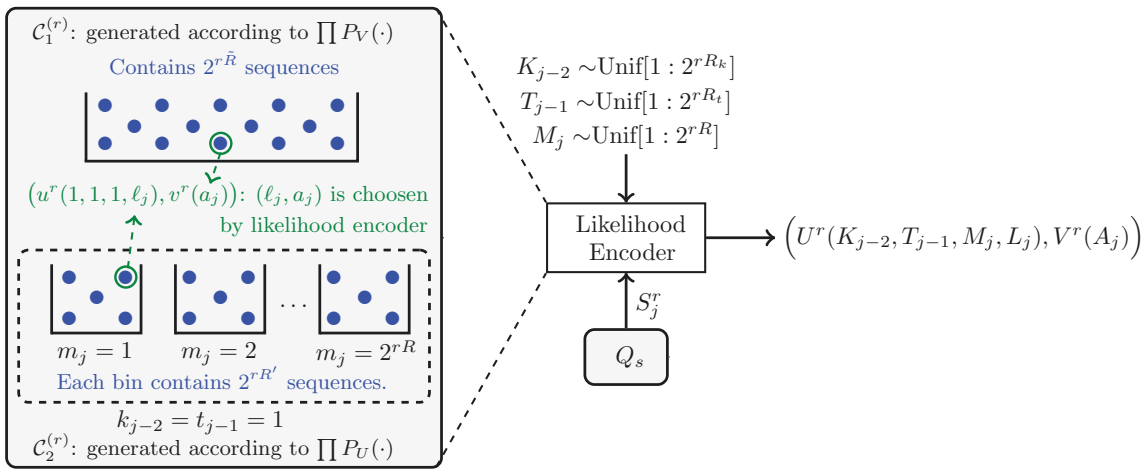

Fig. 10. Proposed coding scheme for the dual use of CSI

such that $\sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}} P_U(u)P_V(v)P_{S|U,V}(\cdot|u,v) = Q_S(\cdot)$. Then we choose P_U, P_V, $P_{S|U,V}$, and $x(u,s)$ such that it satisfies $Q_Z = Q_0$. For every $j \in [\![2:B]\!]$,

$$\begin{aligned}\mathbb{I}(Z_j^r; Z_{j+1}^{B,r}) &\leqslant \mathbb{I}(Z_j^r; K_{j-1}, T_j, K_j, Z_{j+1}^{B,r}) \\ &\overset{(a)}{=} \mathbb{I}(Z_j^r; K_{j-1}, T_j, K_j),\end{aligned} \tag{137}$$

where (a) holds because $Z_j^r - (K_{j-1}, T_j, K_j) - Z_{j+1}^{B,r}$ forms a Markov chain, as seen in the functional dependence graph depicted in Fig. 11. Also,

$$\begin{aligned}\mathbb{I}(Z_j^r; K_{j-1}, T_j, K_j) &= \mathbb{D}\Big(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j} || P_{Z_j^r} P_{K_{j-1}, T_j, K_j}\Big) \\ &\overset{(b)}{\leqslant} \mathbb{D}\Big(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j} || Q_Z^{\otimes r} Q_{K_{j-1}} Q_{T_j} Q_{K_j}\Big),\end{aligned} \tag{138}$$

where $Q_{K_{j-1}} Q_{K_j} Q_{T_j}$ is the uniform distribution over $[\![1:2^{rR_k}]\!] \times [\![1:2^{rR_K}]\!] \times [\![1:2^{rR_T}]\!]$ and (b) follows from

$$\begin{aligned}&\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j} || P^{(\mathcal{C}_r)}_{Z_j^r} P^{(\mathcal{C}_r)}_{K_{j-1}, T_j, K_j}\Big) \\ &= \mathbb{D}\Big(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j} || Q_Z^{\otimes r} Q_{K_{j-1}} Q_{T_j} Q_{K_j}\Big) \\ &- \mathbb{D}\Big(P^{(\mathcal{C}_r)}_{Z_j^r} || Q_Z^{\otimes r}\Big) - \mathbb{D}\Big(P^{(\mathcal{C}_r)}_{K_{j-1}, T_j, K_j} || Q_{K_{j-1}} Q_{T_j} Q_{K_j}\Big).\end{aligned} \tag{139}$$

Therefore, from the expansion in (81), by substituting Q_0 with Q_Z, and also from (138) and (139),

$$\begin{aligned}&\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{Z^n} || Q_Z^{\otimes n}\Big) \\ &\leqslant 2\sum_{j=1}^{B} \mathbb{D}\Big(P^{(\mathcal{C}_r)}_{Z_j^r, K_{j-1}, T_j, K_j} || Q_Z^{\otimes r} Q_{K_{j-1}} Q_{T_j} Q_{K_j}\Big).\end{aligned} \tag{140}$$

To bound the RHS of (140) by using Lemma 1 and the triangle inequality we have,

$$\begin{aligned}&\mathbb{E}_{C_r} || P_{Z_j^r, K_{j-1}, T_j, K_j | C_r} - Q_Z^{\otimes r} Q_{K_{j-1}} Q_{T_j} Q_{K_j} ||_1 \\ &\leqslant \mathbb{E}_{C_r} || P_{Z_j^r, K_{j-1}, T_j, K_j | C_r} - \Gamma_{Z_j^r, K_{j-1}, T_j, K_j | C_r} ||_1 \\ &\quad + \mathbb{E}_{C_r} || \Gamma_{Z_j^r, K_{j-1}, T_j, K_j | C_r} - Q_Z^{\otimes r} Q_{K_{j-1}} Q_{T_j} Q_{K_j} ||_1 \\ &\leqslant \mathbb{E}_{C_r} || P_{Z_j^r, K_{j-1}, T_j, K_j | C_r} - \Upsilon_{Z_j^r, K_{j-1}, T_j, K_j | C_r} ||_1 \\ &\quad + \mathbb{E}_{C_r} || \Upsilon_{Z_j^r, K_{j-1}, T_j, K_j | C_r} - \Gamma_{Z_j^r, K_{j-1}, T_j, K_j | C_r} ||_1 \\ &\quad + \mathbb{E}_{C_r} || \Gamma_{Z_j^r, K_{j-1}, T_j, K_j | C_r} - Q_Z^{\otimes r} Q_{K_{j-1}} Q_{T_j} Q_{K_j} ||_1.\end{aligned} \tag{141}$$

From (135) and the monotonicity of KL-divergence the first term on the RHS of (141) vanishes when r grows. To bound the second term on the RHS of (141) for a fixed codebook $\mathcal{C}_r$, we have (142) at the top of the next page, in which (142b) follows from (132). Hence,

$$\begin{aligned}&\mathbb{E}_{C_r} || \Upsilon_{Z_j^r, K_{j-1}, T_j, K_j | C_r} - \Gamma_{Z_j^r, K_{j-1}, T_j, K_j | C_r} ||_1 \\ &\leqslant \mathbb{E}_{C_r} || \Upsilon_{M_j, T_{j-1}, K_{j-2}, S_j^r, L_j, A_j, U^r, V^r, Z_j^r, K_{j-1}, T_j, K_j | C_r} \\ &\quad - \Gamma_{M_j, T_{j-1}, K_{j-2}, S_j^r, L_j, A_j, U^r, V^r, Z_j^r, K_{j-1}, T_j, K_j | C_r} ||_1 \\ &\overset{(a)}{=} \mathbb{E}_{C_r} || \Upsilon_{M_j, T_{j-1}, K_{j-2}, S_j^r | C_r} - \Gamma_{M_j, T_{j-1}, K_{j-2}, S_j^r | C_r} ||_1 \\ &\overset{(b)}{=} \mathbb{E}_{C_r} || Q_S^{\otimes r} - \Gamma_{S_j^r | M_j = 1, T_{j-1} = 1, K_{j-2} = 1, C_r} ||_1,\end{aligned} \tag{143}$$

where (a) follows from (142b)-(142h) and (b) follows from the symmetry of the codebook construction with respect to M_j, T_{j-1}, and K_{j-2} and (142a). Based on [41] or [42, Theorem 2] the RHS of (143) vanishes if

$$R' > \mathbb{I}(U;S), \tag{144a}$$
$$\tilde{R} > \mathbb{I}(V;S), \tag{144b}$$
$$R' + \tilde{R} > \mathbb{I}(U,V;S). \tag{144c}$$

We now proceed to bound the third term on the RHS of (141). First, consider the following marginal from (131),

$$\begin{aligned}&\Gamma_{Z_j^r, K_{j-1}, T_j, K_j | C_r}(z_j^r, k_{j-1}, t_j, k_j) \\ &= \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\sum_{s_j^r} \frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R})}} \\ &\quad \times P_{S|U,V}^{\otimes r}\big(s_j^r | U^r(m_j, t_{j-1}, k_{j-2}, \ell_j), V^r(a_j)\big) \\ &\quad \times W_{Z|U,S}^{\otimes r}\big(z_j^r | U^r(m_j, t_{j-1}, k_{j-2}, \ell_j), s_j^r\big) \\ &\quad \times \mathbb{1}_{\{t_j = \sigma(V^r(a_j))\}} \mathbb{1}_{\{k_j = \Phi(V^r(a_j))\}} \\ &= \sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j} \frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R})}} \\ &\quad \times W_{Z|U,V}^{\otimes r}\big(z_j^r | U^r(m_j, t_{j-1}, k_{j-2}, \ell_j), V^r(a_j)\big)\end{aligned} \tag{145}$$

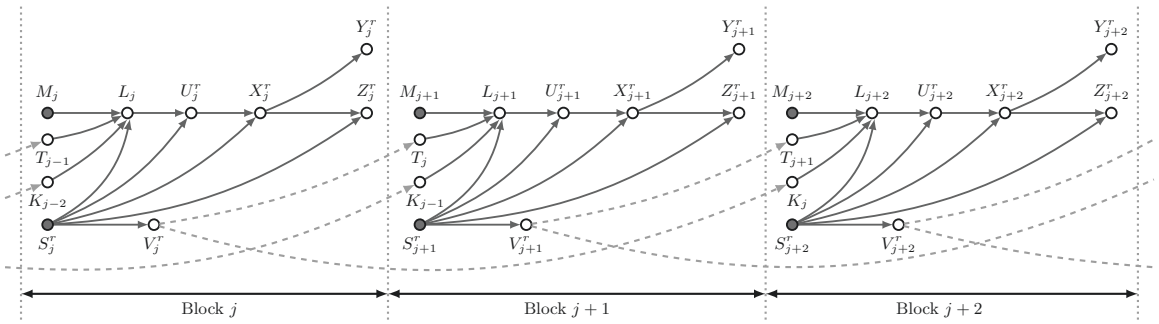

Fig. 11. Functional dependence graph for the block-Markov encoding scheme

$$\Gamma^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2}} = 2^{-r(R+R_t+R_k)} = \Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2}}, \tag{142a}$$

$$\Gamma^{(\mathcal{C}_r)}_{L_j,A_j|M_j,T_{j-1},K_{j-2},S_j^r} = f(\ell_j,a_j|s_j^r,m_j,t_{j-1},k_{j-2}) = \Upsilon^{(\mathcal{C}_r)}_{L_j,A_j|M_j,T_{j-1},K_{j-2},S_j^r}, \tag{142b}$$

$$\Gamma^{(\mathcal{C}_r)}_{U^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j} = \mathbb{1}_{\{\tilde{u}^r=u^r(m_j,t_{j-1},k_{j-2},\ell_j)\}} = \Upsilon^{(\mathcal{C}_r)}_{U^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j}, \tag{142c}$$

$$\Gamma^{(\mathcal{C}_r)}_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r} = \mathbb{1}_{\{\tilde{v}^r=v^r(a_j)\}} = \Upsilon^{(\mathcal{C}_r)}_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r}, \tag{142d}$$

$$\Gamma^{(\mathcal{C}_r)}_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r} = W^{\otimes r}_{Z|U,S} = \Upsilon^{(\mathcal{C}_r)}_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r}, \tag{142e}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r} = 2^{-rR_k} = \Upsilon^{(\mathcal{C}_r)}_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r}, \tag{142f}$$

$$\Gamma^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1}} = \mathbb{1}_{\{t_j=\sigma(v^r)\}} = \Upsilon^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1}}, \tag{142g}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j} = \mathbb{1}_{\{k_j=\Phi(v^r)\}} = \Upsilon^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j}, \tag{142h}$$

$$\times\, \mathbb{1}_{\{t_j=\sigma(V^r(a_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(a_j))\}}, \tag{146}$$

where $W_{Z|U,V}(z|u,v) = \sum_{s\in\mathcal{S}} P_{S|U,V}(s|u,v)W_{Z|U,S}(z|u,s)$.

To bound bound the third term on the RHS of (141), by using Pinsker's inequality in Lemma 1 it is sufficient to bound $\mathbb{E}_{C_r}[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})]$ as in (147) available at the next page, in which

- (a) follows from Jensen's inequality;
- (b) holds because $\mathbb{1}_{\{\cdot\}} \leqslant 1$;
- (c) follows by defining Ψ_1 and Ψ_2 as follows,

$$\Psi_1 = \frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}} \sum_{(k_{j-1},t_j,k_j)} \sum_{m_j} \sum_{t_{j-1}} \sum_{k_{j-2}} \sum_{\ell_j} \sum_{a_j} \sum_{(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r)\in\mathcal{T}_\epsilon^{(n)}}$$
$$\times\, \Gamma^{\otimes r}_{U^r,V^r,Z^r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)$$
$$\times \log\left(\frac{W^{\otimes r}_{Z|U,V}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.$$
$$+\frac{W^{\otimes r}_{Z|V}\big(z_j^r|v^r(a_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}$$
$$\left.+\frac{W^{\otimes r}_{Z|U}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j)\big)}{2^{r(R+R_t+R_k+R')}Q_Z^{\otimes r}(z_j^r)}+1\right)$$
$$\leq \log\left(\frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|U,V)}}{2^{r(R+R_t+R_k+R'+\tilde{R})}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}\right.$$
$$+\frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|V)}}{2^{r\tilde{R}}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}$$
$$\left.+\frac{2^{-r(1-\epsilon)\mathbb{H}(Z|U)}}{2^{r(R+R_t+R_k+R')}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}}+1\right) \tag{148}$$

$$\Psi_2 = \frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}} \sum_{(k_{j-1},t_j,k_j)} \sum_{m_j} \sum_{t_{j-1}} \sum_{k_{j-2}} \sum_{\ell_j} \sum_{a_j} \sum_{(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r)\notin\mathcal{T}_\epsilon^{(n)}}$$
$$\times\, \Gamma^{\otimes r}_{U^r,V^r,Z^r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)$$
$$\times \log\left(\frac{W^{\otimes r}_{Z|U,V}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.$$
$$+\frac{W^{\otimes r}_{Z|V}\big(z_j^r|v^r(a_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}$$
$$\left.+\frac{W^{\otimes r}_{Z|U}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j)\big)}{2^{r(R+R_t+R_k+R')}Q_Z^{\otimes r}(z_j^r)}+1\right)$$
$$\leq 2|V||U||Z|e^{-r\epsilon^2\mu_{V,U,Z}}r\log\Big(\frac{2}{\mu_Z}+1\Big). \tag{149}$$

In (149) $\mu_{V,U,Z} = \min_{(v,u,z)\in(\mathcal{V},\mathcal{U},\mathcal{Z})} P_{V,U,Z}(v,u,z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$. When $r\to\infty$ then $\Psi_2\to 0$ and Ψ_1 goes to zero when r grows if

$$R+R_t+R_k+R'+\tilde{R}-R_T-R_K > \mathbb{I}(U,V;Z), \tag{150a}$$

$$\tilde{R}-R_T-R_K > \mathbb{I}(V;Z), \tag{150b}$$

$$
\begin{aligned}
&\mathbb{E}_{C_r}[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})]\\
&=\mathbb{E}_{C_r}\Big[\sum_{(z_j^r,k_{j-1},t_j,k_j)}\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)\log\Big(\frac{\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_{j-1}}(k_{j-1})Q_{T_j}(t_j)Q_{K_j}(k_j)}\Big)\Big]\\
&=\mathbb{E}_{C_r}\Bigg[\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R})}}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2},\ell_j),V^r(a_j)\big)\\
&\times\mathbb{1}_{\{t_j=\sigma(V^r(a_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(a_j))\}}\\
&\times\log\Bigg(\frac{\sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j}\sum_{\tilde{a}_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j),V^r(\tilde{a}_j)\big)\mathbb{1}_{\{t_j=\sigma(V^r(\tilde{a}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{a}_j))\}}}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg)\Bigg]\\
&\overset{(a)}{\leq}\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R})}}\sum_{(u^r,v^r)}\Gamma_{U^r,V^r,Z^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times\mathbb{E}_{\sigma(v^r(a_j))}\big[\mathbb{1}_{\{t_j=\sigma(v^r(a_j))\}}\big]\times\mathbb{E}_{\Phi(v^r(a_j))}\big[\mathbb{1}_{\{k_j=\Phi(v^r(a_j))\}}\big]\\
&\times\log\mathbb{E}_{\substack{\backslash(m_j,t_{j-1},k_{j-2},\ell_j,a_j),\\ \backslash(\sigma(v^r(a_j)),\Phi(v^r(a_j)))}}\Bigg[\frac{1}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&\times\sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j}\sum_{\tilde{a}_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(,\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j),V^r(\tilde{a}_j)\big)\mathbb{1}_{\{t_j=\sigma(V^r(\tilde{a}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{a}_j))\}}\Bigg]\\
&\overset{(b)}{\leq}\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}}\sum_{(u^r,v^r)}\Gamma_{U^r,V^r,Z^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times\log\frac{1}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\Bigg(W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)\\
&+\mathbb{E}_{\backslash(m_j,t_{j-1},k_{j-2},\ell_j)}\Bigg[\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j)\neq(m_j,t_{j-1},k_{j-2},\ell_j)}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j),v^r(a_j)\big)\Bigg]\\
&+\mathbb{E}_{\backslash(a_j,\sigma(v^r(a_j)),\Phi(v^r(a_j)))}\Bigg[\sum_{\tilde{a}_j\neq a_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),V^r(\tilde{a}_j)\big)\mathbb{1}_{\{t_j=\sigma(V^r(\tilde{a}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{a}_j))\}}\Bigg]\\
&+\mathbb{E}_{\substack{\backslash(m_j,t_{j-1},k_{j-2},\ell_j,a_j),\\ \backslash(\sigma(v^r(a_j)),\Phi(v^r(a_j)))}}\Bigg[\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j)\neq(m_j,t_{j-1},k_{j-2},\ell_j)}\sum_{\tilde{a}_j\neq a_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j),V^r(\tilde{a}_j)\big)\\
&\times\mathbb{1}_{\{t_j=\sigma(V^r(\tilde{a}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{a}_j))\}}\Bigg]\Bigg)\\
&\leq\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}}\sum_{(u^r,v^r)}\Gamma_{U^r,V^r,Z^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times\log\Bigg(\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\\
&+\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2},\tilde{\ell}_j)\neq(m_j,t_{j-1},k_{j-2},\ell_j)}\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\sum_{\tilde{a}_j\neq a_j}\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R})}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\leq\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{a_j}\frac{1}{2^{r(R+R_t+2R_k+R'+\tilde{R}+R_T+R_K)}}\sum_{(u^r,v^r)}\Gamma_{U^r,V^r,Z^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j),z_j^r\big)\\
&\times\log\Bigg(\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j),v^r(a_j)\big)}{2^{r(R+R_t+R_k+R'+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(a_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2},\ell_j)\big)}{2^{r(R+R_t+R_k+R')}Q_Z^{\otimes r}(z_j^r)}+1\Bigg)\\
&\overset{(c)}{=}\Psi_1+\Psi_2,
\end{aligned}
\tag{147}
$$

$$R + R_t + R_k + R' > \mathbb{I}(U; Z). \tag{150c}$$

Decoding and Error Probability Analysis: At the end of the block $j \in [\![1 : B]\!]$, using its knowledge of the key k_{j-2} generated from the block $j-2$, the receiver finds a unique triple $(\hat{m}_j, \hat{t}_{j-1}, \hat{\ell}_j)$ such that $\left(u^r(\hat{m}_j, \hat{t}_{j-1}, k_{j-2}, \hat{\ell}_j), y_j^r\right) \in \mathcal{T}_\epsilon^{(r)}$. To bound the probability of error at the encoder and the decoder, we use the following lemma.

Lemma 2 (Typical With High Probability)**.** *If* $(R', \tilde{R}) \in \mathbb{R}_+^2$ *satisfies* (144)*, then for any* $(m_j, t_{j-1}, k_{j-2}) \in (\mathcal{M}, \mathcal{T}, \mathcal{K})$ *and* $\epsilon > 0$*, we have*

$$\mathbb{E}_{C_r}\mathbb{P}_P\Big(\big(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r\big) \notin \mathcal{T}_\epsilon^{(r)} | C_r\Big) \xrightarrow[r\to\infty]{} 0, \tag{151}$$

where P *is the induced distribution over the codebook defined in* (135)*.*

The proof of Lemma 2 is given in Appendix E.

To analyze the probability of error, we define the following error events for $j \in [\![1 : B]\!]$,

$$\mathcal{E} \triangleq \big\{M \neq \hat{M}\big\}, \tag{152a}$$
$$\mathcal{E}_j \triangleq \big\{M_j \neq \hat{M}_j\big\}, \tag{152b}$$
$$\mathcal{E}_{1,j} \triangleq \big\{\big(U^r(M_j, T_{j-1}, K_{j-2}, L_j), S_j^r\big) \notin \mathcal{T}_{\epsilon_1}^{(r)}(U, S)\big\}, \tag{152c}$$
$$\mathcal{E}_{2,j} \triangleq \big\{\big(U^r(M_j, T_{j-1}, K_{j-2}, L_j), Y_j^r\big) \notin \mathcal{T}_{\epsilon_2}^{(r)}(U, Y)\big\}, \tag{152d}$$
$$\mathcal{E}_{3,j} \triangleq \big\{\big(U^r(M_j, T_{j-1}, K_{j-2}, L_j), Y_j^r\big) \in \mathcal{T}_{\epsilon_2}^{(r)}(U, Y) \text{ for some } m_j \neq M_j \text{ and } \ell_j \in [1 : 2^{rR'}]\big\}. \tag{152e}$$

where $\epsilon_2 > \epsilon_1 > \epsilon > 0$. The probability of error is upper bounded as follows,

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}\Big\{\bigcup\nolimits_{j=1}^{B} \mathcal{E}_j\Big\} \leqslant \sum_{j=1}^{B} \mathbb{P}(\mathcal{E}_j). \tag{153}$$

Now we bound $\mathbb{P}(\mathcal{E}_j)$ by using union bound,

$$\mathbb{P}(\mathcal{E}_j) \leqslant \mathbb{P}(\mathcal{E}_{1,j}) + \mathbb{P}(\mathcal{E}_{1,j}^c \cap \mathcal{E}_{2,j}) + \mathbb{P}(\mathcal{E}_{2,j}^c \cap \mathcal{E}_{3,j}). \tag{154}$$

According to Lemma 2 the first term on the RHS of (154) vanishes when r grows, and by the law of large numbers the second term on the RHS of (154) vanishes when r grows. Also, according to the law of large numbers and the packing lemma, the last term on the RHS of (154) vanishes when r grows if [37],

$$R + R_t + R' \leqslant \mathbb{I}(U; Y). \tag{155}$$

We now analyze the probability of error at the encoder and the decoder for key generation. Let (A_{j-1}, T_{j-1}) denote the chosen indices at the encoder and $\hat{A}_{j-1}$ and $\hat{T}_{j-1}$ be the estimates of the indices A_{j-1} and T_{j-1} at the decoder. At the end of block j, by decoding U_j^r, the decoder knows $\hat{T}_{j-1}$. To find A_{j-1} we define the error event,

$$\mathcal{E}' = \Big\{\Big(V_{j-1}^r(\hat{A}_{j-1}), S_{j-1}^r, U_{j-1}^r, Y_{j-1}^r\Big) \notin \mathcal{T}_\epsilon^{(r)}\Big\}. \tag{156}$$

Also, consider the error events,

$$\mathcal{E}_1' = \Big\{\Big(V_{j-1}^r(a_{j-1}), S_{j-1}^r\Big) \notin \mathcal{T}_{\epsilon'}^{(r)} \text{ for all } a_{j-1} \in [\![1 : 2^{r\tilde{R}}]\!]\Big\}, \tag{157a}$$
$$\mathcal{E}_2' = \Big\{\Big(V_{j-1}^r(A_{j-1}), S_{j-1}^r, U_{j-1}^r, Y_{j-1}^r\Big) \notin \mathcal{T}_\epsilon^{(r)}\Big\}, \tag{157b}$$
$$\mathcal{E}_3' = \Big\{\Big(V_{j-1}^r(\tilde{a}_{j-1}), U_{j-1}^r, Y_{j-1}^r\Big) \in \mathcal{T}_\epsilon^{(r)} \text{ for some } \tilde{a}_{j-1} \in \mathcal{B}(\hat{T}_{j-1}), \tilde{a}_{j-1} \neq A_{j-1}\Big\}, \tag{157c}$$

where $\epsilon > \epsilon' > 0$. By the union bound we have,

$$P(\mathcal{E}') \leqslant P(\mathcal{E}_1') + P(\mathcal{E}_1'^c \cap \mathcal{E}_2') + P(\mathcal{E}_3'). \tag{158}$$

According to Lemma 2 the first term on the RHS of (158) vanishes when r grows if we have (144). Following the steps in [37, Sec. 11.3.1], the last two terms on the RHS of (158) go to zero when r grows if,

$$\tilde{R} > \mathbb{I}(V; S), \tag{159a}$$
$$\tilde{R} - R_t < \mathbb{I}(V; U, Y). \tag{159b}$$

The region in Theorem 4 is derived by remarking that the scheme requires $R_K + R_T \geqslant R_k + R_t$ and applying Fourier-Motzkin to (144) and (150), (155), and (159).

Appendix E
Proof of Lemma 2

For a fix $\epsilon > 0$, consider the PMF Γ defined in (131). For the random experiment described by Γ; since $U^r(m_j, t_{j-1}, k_{j-2}, L_j) \sim P_U^{\otimes r}$, for every $(m_j, t_{j-1}, k_{j-2}) \in \mathcal{M} \times \mathcal{T} \times \mathcal{K}$) and $V^r(A_j) \sim P_V^{\otimes r}$, for every $a_j \in \mathcal{A}$), and S_j^r is derived by passing $\big(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j)\big)$ through the DMC $P_{S|U,V}^{\otimes r}$ by the weak law of large numbers we have

$$\mathbb{E}_{C_r}\mathbb{P}_\Gamma\Big(\big(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r\big) \notin \mathcal{T}_\epsilon^{(r)} | C_r\Big) \xrightarrow[r\to\infty]{} 0. \tag{160}$$

We also have

$$\begin{aligned}&\mathbb{E}_{C_r}||P_{U^r,V^r,S_j^r|C_r} - \Gamma_{U^r,V^r,S_j^r|C_r}||_1 \\ &\leqslant \mathbb{E}_{C_r}||P_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r} \\ &- \Gamma_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,A_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r}||_1 \xrightarrow[r\to\infty]{} 0,\end{aligned} \tag{161}$$

where the RHS of (161) vanishes when r grows because of (143). We now define $g_r : \mathcal{U}^r \times \mathcal{V}^r \times \mathcal{S}_j^r \mapsto \mathbb{R}$ as $g_r(u^r, v^r, s_j^r) \triangleq \mathbb{1}_{\{(u^r,v^r,s_j^r) \notin \mathcal{T}_\epsilon^{(r)}\}}$. We now have,

$$\begin{aligned}&\mathbb{E}_{C_r}\mathbb{P}_P\Big(\big(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r\big) \notin \mathcal{T}_\epsilon^{(r)} | C_r\Big) \\ &= \mathbb{E}_{C_r}\mathbb{E}_P\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r) | C_r\Big] \\ &\leqslant \mathbb{E}_{C_r}\mathbb{E}_\Gamma\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r) | C_r\Big] \\ &+ \mathbb{E}_{C_r}\Big|\mathbb{E}_P\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r) | C_r\Big] \\ &- \mathbb{E}_\Gamma\Big[g_r(U^r(m_j, t_{j-1}, k_{j-2}, L_j), V^r(A_j), S_j^r) | C_r\Big]\Big|\end{aligned}$$

$$\overset{(a)}{\leqslant} \mathbb{E}_{C_r}\mathbb{E}_{\Gamma}\Big[g_r(U^r(m_j,t_{j-1},k_{j-2},L_j),V^r(A_j),S_j^r)|C_r\Big] + \mathbb{E}_{C_r}||P_{U^r,V^r,S_j^r|C_r} - \Gamma_{U^r,V^r,S_j^r|C_r}||_1, \tag{162}$$

where (a) follows from [43, Property 1] for g_r being bounded by 1. From (160) and (161) the RHS of (162) vanishes when r grows.

Appendix F
Proof of Theorem 6

Consider any sequence of length-n codes for a state-dependent channel with CSI available non-causally only at the transmitter such that $P_e^{(n)} \leqslant \epsilon_n$, $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \delta$, and $R_K/n \leqslant \lambda_n$ with $\lim_{n\to\infty}\epsilon_n = \lim_{n\to\infty}\lambda_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish. The following lemma, a version of which with variational distance can be found in [33, Lemma VI.3], will prove useful.

Lemma 3. *If $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \delta$, then $\sum_{i=1}^n \mathbb{I}(Z_i;Z^{i-1}) \leqslant \delta$ and $\sum_{i=1}^n \mathbb{I}(Z_i;Z_{i+1}^n) \leqslant \delta$. In addition, if $T \in [\![1:n]\!]$ is an independent variable uniformly distributed, then $\mathbb{I}(T;Z_T) \leqslant \nu$, where $\nu \triangleq \frac{\delta}{n}$.*

Note that Lemma 3 is slightly different from [33, Lemma VI.3], as the upper bounds are tighter and do not include a factor of n. This is a consequence of using a constraint based on relative entropy instead of total variation. This is crucial in what follows, as we do no necessarily require $\delta \to 0$.

Proof. First note that,

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{I}(Z_i;Z^{i-1}) = \sum_{i=1}^n [\mathbb{H}(Z_i) - \mathbb{H}(Z_i|Z^{i-1})]\\
&= \sum_{i=1}^n \mathbb{H}(Z_i) - \mathbb{H}(Z^n)\\
&= -\sum_{i=1}^n\sum_z P_{Z_i}(z)\log P_{Z_i}(z) + \sum_{z^n} P_{Z^n}(z^n)\log P_{Z^n}(z^n)\\
&= -\sum_{i=1}^n\sum_z P_{Z_i}(z)\log P_{Z_i}(z) + \sum_{i=1}^n\sum_z P_{Z_i}(z)\log Q_0(z)\\
&\quad -\sum_{i=1}^n\sum_z P_{Z_i}(z)\log Q_0(z) + \sum_{z^n} P_{Z^n}(z^n)\log P_{Z^n}(z^n)\\
&= -\sum_{i=1}^n \mathbb{D}(P_{Z_i}||Q_0) - \sum_{z^n} P_{Z^n}(z^n)\log Q_0^{\otimes n}(z^n)\\
&\quad + \sum_{z^n} P_{Z^n}(z^n)\log P_{Z^n}(z^n)\\
&\leqslant \mathbb{D}(P_{Z^n}||Q_0^{\otimes n})\\
&\leqslant \delta.
\end{aligned}$$

Similarly, one can prove $\sum_{i=1}^n \mathbb{I}(Z_i;Z_{i+1}^n) \leqslant \delta$. Next,

$$\begin{aligned}
\mathbb{I}(T;Z_T) &= \mathbb{H}(Z_T) - \mathbb{H}(Z_T|T)\\
&= -\sum_z \frac{1}{n}\sum_{i=1}^n P_{Z_i}(z)\log\frac{1}{n}\sum_{j=1}^n P_{Z_j}(z)\\
&\quad + \frac{1}{n}\sum_{i=1}^n\sum_z P_{Z_i}(z)\log P_{Z_i}(z)\\
&= -\sum_z \frac{1}{n}\sum_{i=1}^n P_{Z_i}(z)\log\frac{1}{n}\sum_{j=1}^n P_{Z_j}(z)\\
&\quad + \sum_z \frac{1}{n}\sum_{i=1}^n P_{Z_i}(z)\log Q_0(z)\\
&\quad - \sum_z \frac{1}{n}\sum_{i=1}^n P_{Z_i}(z)\log Q_0(z)\\
&\quad + \frac{1}{n}\sum_{i=1}^n\sum_z P_{Z_i}(z)\log P_{Z_i}(z)\\
&= -\mathbb{D}\left(\frac{1}{n}\sum_{i=1}^n P_{Z_i}\middle|\middle|Q_0\right) + \frac{1}{n}\sum_{i=1}^n \mathbb{D}(P_{Z_i}||Q_0)\\
&\leqslant \frac{1}{n}\sum_{i=1}^n \mathbb{D}(P_{Z_i}||Q_0)\\
&\leqslant \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n})\\
&\leqslant \frac{\delta}{n}.
\end{aligned} \tag{163}$$

□

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (27) as follows,

$$\mathcal{A}_\epsilon \triangleq \left\{\begin{array}{l} R \geqslant 0 : \exists P_{S,U,V,X,Y,Z} \in \mathcal{D}_\epsilon \text{ such that}\\ R \leqslant \min\big\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y)\\ \qquad -\mathbb{I}(U;S|V)\big\} + \epsilon \end{array}\right\}, \tag{164a}$$

where

$$\mathcal{D}_\epsilon \triangleq \left\{\begin{array}{l} P_{S,U,V,X,Y,Z}:\\ P_{S,U,V,X,Y,Z} = Q_S P_{UV|S}\mathbb{1}_{\{X=X(U,S)\}}\\ \qquad \times W_{Y,Z|X,S}\\ \mathbb{D}\left(P_Z||Q_0\right) \leqslant \epsilon\\ \min\big\{\mathbb{I}(U;Y) - \mathbb{I}(U;S),\\ \qquad \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\big\} \geqslant\\ \qquad \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 4\epsilon\\ \max\{|\mathcal{U}|,|\mathcal{V}|\} \leqslant |\mathcal{X}| + 3 \end{array}\right\}. \tag{164b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques,

$$\begin{aligned}
nR &= \mathbb{H}(M)\\
&\overset{(a)}{\leq} \mathbb{I}(M;Y^n) + n\epsilon_n\\
&= \sum_{i=1}^n \mathbb{I}(M;Y_i|Y^{i-1}) + n\epsilon_n\\
&\leqslant \sum_{i=1}^n \mathbb{I}(M,Y^{i-1};Y_i) + n\epsilon_n\\
&= \sum_{i=1}^n \big[\mathbb{I}(M,Y^{i-1},S_{i+1}^n;Y_i) - \mathbb{I}(S_{i+1}^n;Y_i|M,Y^{i-1})\big] + n\epsilon_n\\
&\overset{(b)}{=} \sum_{i=1}^n \big[\mathbb{I}(M,Y^{i-1},S_{i+1}^n;Y_i) - \mathbb{I}(Y^{i-1};S_i|M,S_{i+1}^n)\big] + n\epsilon_n
\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

$$
\begin{aligned}
&\overset{(c)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M, Y^{i-1}, S_{i+1}^{n}; Y_i) - \mathbb{I}(M, Y^{i-1}, S_{i+1}^{n}; S_i)\big] + n\epsilon_n \\
&\overset{(d)}{=} \sum_{i=1}^{n} [\mathbb{I}(U_i; Y_i) - \mathbb{I}(U_i; S_i)] + n\epsilon_n \\
&= n \sum_{i=1}^{n} \frac{1}{n} \big[\mathbb{I}(U_i; Y_i|T=i) - \mathbb{I}(U_i; S_i|T=i)\big] + n\epsilon_n \\
&= n \sum_{i=1}^{n} \mathbb{P}(T=i) \big[\mathbb{I}(U_i; Y_i|T=i) - \mathbb{I}(U_i; S_i|T=i)\big] + n\epsilon_n \\
&= n\big[\mathbb{I}(U_T; Y_T|T) - \mathbb{I}(U_T; S_T|T)\big] + n\epsilon_n \\
&\overset{(e)}{=} n\big[\mathbb{I}(U_T; Y_T|T) - \mathbb{I}(U_T, T; S_T)\big] + n\epsilon_n \\
&\leq \big[\mathbb{I}(U_T, T; Y_T) - \mathbb{I}(U_T, T; S_T)\big] + n\epsilon_n \\
&\overset{(f)}{=} n\big[\mathbb{I}(U; Y) - \mathbb{I}(U; S)\big] + n\epsilon_n \\
&\overset{(g)}{\leqslant} n\big[\mathbb{I}(U; Y) - \mathbb{I}(U; S)\big] + n\epsilon
\end{aligned} \tag{165}
$$

where

- (a) follows from Fano's inequality for n large enough;
- (b) follows from Csiszár-Körner sum identity [44, Lemma 7];
- (c) follows since S_i is independent of (M, S_{i+1}^n);
- (d) follows by defining $U_i \triangleq (M, Y^{i-1}, S_{i+1}^n)$;
- (e) follows from the independence of S_T and T;
- (f) follows by defining $U \triangleq (U_T, T)$, $Y \triangleq Y_T$, and $S \triangleq S_T$;
- (g) follows by defining $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$, where we choose n large enough such that $\nu \geqslant \frac{\delta}{n}$.

We also have,

$$
\begin{aligned}
nR &= \mathbb{H}(M) \\
&= \mathbb{H}(M|K) \\
&\overset{(a)}{\leq} \mathbb{I}(M; Y^n|K) + n\epsilon_n \\
&= \sum_{i=1}^{n} \mathbb{I}(M; Y_i|Y^{i-1}, K) + n\epsilon_n \\
&\leqslant \sum_{i=1}^{n} \mathbb{I}(M, K, Y^{i-1}, Z^{i-1}; Y_i) + n\epsilon_n \\
&= \sum_{i=1}^{n} \big[\mathbb{I}(M, K, Y^{i-1}, Z^{i-1}, S_{i+1}^{n}; Y_i) \\
&\quad - \mathbb{I}(S_{i+1}^{n}; Y_i|M, K, Y^{i-1}, Z^{i-1})\big] + n\epsilon_n \\
&\overset{(b)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, Y^{i-1}, Z^{i-1}, S_{i+1}^{n}; Y_i) \\
&\quad - \mathbb{I}(Y^{i-1}; S_i|M, K, S_{i+1}^{n}, Z^{i-1})\big] + n\epsilon_n \\
&\overset{(c)}{=} \sum_{i=1}^{n} [\mathbb{I}(U_i, V_i; Y_i) - \mathbb{I}(U_i; S_i|V_i)] + n\epsilon_n \\
&= n \sum_{i=1}^{n} \frac{1}{n} \big[\mathbb{I}(U_T, V_T; Y_T|T=i) \\
&\quad - \mathbb{I}(U_T; S_T|V_T, T=i)\big] + n\epsilon_n \\
&= n \sum_{i=1}^{n} \mathbb{P}(T=i) \big[\mathbb{I}(U_T, V_T; Y_T|T=i) \\
&\quad - \mathbb{I}(U_T; S_T|V_T, T=i)\big] + n\epsilon_n \\
&= n\big[\mathbb{I}(U_T, V_T; Y_T|T) - \mathbb{I}(U_T; S_T|V_T, T)\big] + n\epsilon_n \\
&\leq \big[\mathbb{I}(U_T, V_T, T; Y_T) - \mathbb{I}(U_T; S_T|V_T, T)\big] + n\epsilon_n \\
&\overset{(d)}{=} n\big[\mathbb{I}(U, V; Y) - \mathbb{I}(U; S|V)\big] + n\epsilon_n \\
&\overset{(e)}{\leqslant} n\big[\mathbb{I}(U, V; Y) - \mathbb{I}(U; S|V)\big] + n\epsilon,
\end{aligned} \tag{166}
$$

where

- (a) follows from Fano's inequality for n large enough and the fact that conditioning does not increase entropy;
- (b) follows from Csiszár-Körner sum identity [44, Lemma 7];
- (c) follows by defining $U_i \triangleq (M, Y^{i-1}, S_{i+1}^n)$ and $V_i \triangleq (M, K, Z^{i-1}, S_{i+1}^n)$;
- (d) follows by defining $U \triangleq (U_T, T)$, $V \triangleq (V_T, T)$, $Y \triangleq Y_T$, and $S \triangleq S_T$;
- (e) follows from definition $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$.

Next, we lower bound nR as follows,

$$
\begin{aligned}
&nR + R_K \geqslant \mathbb{H}(M, K) \\
&\geq \mathbb{I}(M, K; Z^n) \\
&= \sum_{i=1}^{n} \mathbb{I}(M, K; Z_i|Z^{i-1}) \\
&= \sum_{i=1}^{n} \big[\mathbb{I}(M, K, S_{i+1}^{n}; Z_i|Z^{i-1}) - \mathbb{I}(S_{i+1}^{n}; Z_i|M, K, Z^{i-1})\big] \\
&\overset{(a)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, S_{i+1}^{n}; Z_i|Z^{i-1}) - \mathbb{I}(Z^{i-1}; S_i|M, K, S_{i+1}^{n})\big] \\
&\overset{(b)}{\geq} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, S_{i+1}^{n}, Z^{i-1}; Z_i) - \mathbb{I}(Z^{i-1}; S_i|M, K, S_{i+1}^{n})\big] - \delta \\
&\overset{(c)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(M, K, S_{i+1}^{n}, Z^{i-1}; Z_i) - \mathbb{I}(M, K, S_{i+1}^{n}, Z^{i-1}; S_i)\big] - \delta \\
&\overset{(d)}{=} \sum_{i=1}^{n} \big[\mathbb{I}(V_i; Z_i) - \mathbb{I}(V_i; S_i)\big] - \delta \\
&= n \sum_{i=1}^{n} \frac{1}{n} \big[\mathbb{I}(V_T; Z_T|T=i) - \mathbb{I}(V_T; S_T|T=i)\big] - \delta \\
&= n \sum_{i=1}^{n} \mathbb{P}(T=i) \big[\mathbb{I}(V_T; Z_T|T=i) - \mathbb{I}(V_T; S_T|T=i)\big] - \delta \\
&= n\big[\mathbb{I}(V_T; Z_T|T) - \mathbb{I}(V_T; S_T|T)\big] - \delta \\
&\overset{(e)}{=} n\big[\mathbb{I}(V_T; Z_T|T) - \mathbb{I}(V_T, T; S_T)\big] - \delta \\
&\overset{(f)}{\geq} n\big[\mathbb{I}(V_T, T; Z_T) - \mathbb{I}(V_T, T; S_T)\big] - 2\delta \\
&\overset{(g)}{=} n\big[\mathbb{I}(V; Z) - \mathbb{I}(V; S)\big] - 2\delta
\end{aligned} \tag{167}
$$

where

- (a) follows from Csiszár-Körner sum identity [44, Lemma 7];
- (b) follows from Lemma 3;
- (c) follows since S_i is independent of (M, K, S_{i+1}^n);
- (d) follows by defining $V_i \triangleq (M, K, S_{i+1}^n, Z^{i-1})$;
- (e) follows from the independence of S_T and T;
- (f) follows from Lemma 3;
- (g) follows by defining $V \triangleq (V_T, T)$, $Z \triangleq Z_T$, and $S \triangleq S_T$.

For any $\nu > 0$, choosing n large enough ensures that,

$$R + \frac{R_K}{n} \geqslant \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 2\nu. \tag{168}$$

Therefore,

$$\begin{aligned} R &\geqslant \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 2\nu - \frac{R_K}{n} \\ &\geqslant \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 2\nu - \lambda_n \\ &\geqslant \mathbb{I}(V;Z) - \mathbb{I}(V;S) - 3\epsilon, \end{aligned} \tag{169}$$

where the last inequality follows since $\epsilon \triangleq \max\{\epsilon_n, \lambda_n, \nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leqslant \epsilon$, note that for n large enough

$$\begin{aligned} &\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) = \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^{n} P_{Z_i}\middle|\middle|Q_0\right) \\ &\leqslant \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leqslant \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \frac{\delta}{n} \leqslant \nu \leqslant \epsilon. \end{aligned} \tag{170}$$

Combining (165), (166), (169), and (170) shows that $\forall \epsilon_n, \lambda_n, \nu > 0$, $R \leqslant \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$R \leqslant \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{171}$$

Continuity at Zero: Our objective is to show that the capacity region is included in the region defined in (27). The challenge, first highlighted in [33, Section VI.D], is that our converse arguments only establish that the capacity region is included in the region $\bigcap_{\epsilon>0}\mathcal{A}_\epsilon$ where $\mathcal{A}_\epsilon$ is defined in (164). In the sequel, the continuity of the slackness in the mutual information inequality (164b) will assume some importance, hence for ease of expression we define and refer to $g(\epsilon) \triangleq 3\epsilon$. As ϵ vanishes, *both* the region $\mathcal{A}_\epsilon$ and the set of distributions $\mathcal{D}_\epsilon$ shrink, so that proving the continuity at $\epsilon = 0$ is not completely straightforward. We carefully lay out the arguments leading to the result in a series of lemmas.

Lemma 4. *For all $\epsilon > 0$, the set $\mathcal{D}_\epsilon$ is closed and bounded, hence compact.*

Proof. We need to check that every constraint defining the set $\mathcal{D}_\epsilon$ defines a closed set of distributions, so that $\mathcal{D}_\epsilon$ is an intersection of closed sets and remains closed. First note that:

- the function that outputs the marginal P_Z of $P_{U,V,S,X,Y,Z}$ is continuous in P_Z;
- Q_0 has support $\mathcal{Z}$ so that the divergence $\mathbb{D}(P_Z \,\|\, Q_0)$ is a continuous function of P_Z;
- mutual information, viewed as a function of the joint distribution of the random variables involved, is continuous;
- all the constraints in the definition of $\mathcal{D}_\epsilon$ are non-strict inequalities.

Consequently, the pre-images of the closed sets defined by the inequalities are pre-images of closed sets through continuous functions, hence closed. $\mathcal{D}_\epsilon$ is bounded because it is a subset of the probability simplex, hence it is compact. □

Lemma 5. *For all $\epsilon > 0$, the set $\mathcal{A}_\epsilon$ is non-empty, closed, and bounded.*

Proof. The set of Pareto optimal points in $\mathcal{A}_\epsilon$ is the image of $\mathcal{D}_\epsilon$ through a continuous function. Since $\mathcal{D}_\epsilon$ is compact, the set of Pareto optimal points is compact. In $\mathbb{R}$, compact sets are closed, hence the set of Pareto optimal points is closed and $\mathcal{A}_\epsilon$ itself is closed by definition. $\mathcal{A}_\epsilon$ is also non-empty because it contains 0. $\mathcal{A}_\epsilon$ is bounded because we can upper bound R by $2\log|\mathcal{X}| + \epsilon$. □

Now define the set

$$\mathcal{A}'_\epsilon \triangleq \left\{\begin{array}{l} R \geqslant 0 : \exists P_{S,U,V,X,Y,Z} \in \mathcal{D}_\epsilon \text{ such that} \\ R \leqslant \min\big\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) \\ \qquad -\mathbb{I}(U;S|V)\big\} \end{array}\right\}. \tag{172}$$

Note that $\mathcal{A}'_\epsilon$ differs from $\mathcal{A}_\epsilon$ in the absence of ϵ in the rate constraint.

Lemma 6. *For all $\epsilon > 0$, the set $\mathcal{A}'_\epsilon$ is closed and bounded.*

Proof. The proof is similar to the proof of Lemma 5. □

Lemma 7.

$$\bigcap_{\epsilon>0}\mathcal{A}'_\epsilon = \bigcap_{\epsilon>0}\mathcal{A}_\epsilon \tag{173}$$

Proof. First, note that $\bigcap_{\epsilon>0}\mathcal{A}'_\epsilon$ is closed, since it is an intersection of closed sets, and bounded, since the sets $\mathcal{A}'_\epsilon$ are nested and bounded. Hence, $\bigcap_{\epsilon>0}\mathcal{A}'_\epsilon$ is compact. Consequently, there exists a maximal element, $r^\dagger$. Consider any $r \in [0, r^\dagger]$. Then $\forall \epsilon > 0$ $\exists P_{U,V,S,X,Y,Z} \in \mathcal{D}_\epsilon$ $r \leqslant r^\dagger \leqslant \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ and $r \in \bigcap_{\epsilon>0}\mathcal{A}'_\epsilon$.

We now want to show that $\bigcap_{\epsilon>0}\mathcal{A}'_\epsilon = \bigcap_{\epsilon>0}\mathcal{A}_\epsilon$. The hard part is showing that, $\bigcap_{\epsilon>0}\mathcal{A}_\epsilon \subset \bigcap_{\epsilon>0}\mathcal{A}'_\epsilon$ since the other direction follows by the definition of $\mathcal{A}_\epsilon$ and $\mathcal{A}'_\epsilon$. We proceed by contradiction. Assume $\exists r^* \in \bigcap_{\epsilon>0}\mathcal{A}_\epsilon$ such that $r^* \notin \bigcap_{\epsilon>0}\mathcal{A}'_\epsilon$. It must be that $r^\dagger < r^*$ for otherwise $r^* \in \bigcap_{\epsilon>0}\mathcal{A}'_\epsilon$ as noted earlier.

Set $r_0 \triangleq \frac{1}{2}(r^\dagger + r^*)$, which is such that $r_0 > r^\dagger$ and therefore $r_0 \notin \bigcap_{\epsilon>0}\mathcal{A}'_\epsilon$. Set $\epsilon' > 0$ such that $\forall \epsilon \leqslant \epsilon'$ $g(\epsilon) < \frac{r^*-r^\dagger}{2}$, which exists by the assumptions on g. Assume that $\forall \epsilon \in (0;\epsilon']$ $r_0 \in \mathcal{A}'_\epsilon$. Then, $r_0 \in \bigcap_{\epsilon>0}\mathcal{A}'_\epsilon$ which contradicts our assumption. Hence, there exists $0 < \epsilon_0 \leqslant \epsilon'$ such that $r_0 \notin \mathcal{A}'_{\epsilon_0}$. Hence $\forall P_{U,V,S,X,Y,Z} \in \mathcal{D}_{\epsilon_0}$ $r_0 > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$. Then $\forall P_{U,V,S,X,Y,Z} \in \mathcal{D}_{\epsilon_0}$

$$r_0 > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} \tag{174}$$

$$\Rightarrow \frac{r^* + r^\dagger}{2} > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} \tag{175}$$

$$\Rightarrow \frac{r^* + r^\dagger}{2} + \frac{r^* - r^\dagger}{2} > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} + \frac{r^* - r^\dagger}{2} \tag{176}$$

$$\Rightarrow r^* > \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} + g(\epsilon_0) \tag{177}$$

Since $r^* \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon$, we have $\forall \epsilon > 0\ \exists P_{U,V,S,X,Y,Z} \in \mathcal{D}_\epsilon$ such that $r^* \leqslant \min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\} + g(\epsilon)$. Hence, there is a contradiction, and we must have $r^* \in \bigcap_{\epsilon>0} \mathcal{A}'_\epsilon$. □

To conclude, one can prove that $\bigcap_{\epsilon>0} \mathcal{A}'_\epsilon = \mathcal{A}_0$, following the exact same arguments as in [33, Section IV.C].

Appendix G
Proof of Theorem 7

We adopt a block-Markov encoding scheme in which B independent messages are transmitted over B channel blocks each of length r, such that $n = rB$. The warden's observation is $Z^n = (Z_1^r, \ldots, Z_B^r)$, the distribution induced at the output of the warden is P_{Z^n}, the target output distribution is $Q_0^{\otimes n}$, and Equation (81), describing the distance between the two distributions, continues to hold. The random code generation is as follows.

Fix $P_U(u)$, $P_{V|S}(v|s)$, $x(u,s)$, and $\epsilon_1 > \epsilon_2 > 0$ such that, $P_Z = Q_0$.

Codebook Generation for Keys: For each block $j \in [\![1:B]\!]$, let $C_1^{(r)} \triangleq \left\{V^r(\ell_j)\right\}_{\ell_j \in \mathcal{L}}$, where $\mathcal{L} \triangleq [\![1:2^{r\tilde{R}}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_V^{\otimes r}$, where $P_V = \sum_{s\in\mathcal{S}} Q_S(s) P_{V|S}(v|s)$. We denote a realization of $C_1^{(r)}$ by $\mathcal{C}_1^{(r)} \triangleq \left\{v^r(\ell_j)\right\}_{\ell_j \in \mathcal{L}}$. Partition the set of indices $\ell_j \in [\![1:2^{r\tilde{R}}]\!]$ into bins $\mathcal{B}(t)$, $t \in [\![1:2^{rR_T}]\!]$ by using function $\varphi : V^r(\ell_j) \mapsto [\![1:2^{rR_T}]\!]$ through random binning by choosing the value of $\varphi(v^r(\ell_j))$ independently and uniformly at random for every $v^r(\ell_j) \in \mathcal{V}^r$. For each block $j \in [\![1:B]\!]$, create a function $\Phi : V^r(\ell_j) \mapsto [\![1:2^{rR_K}]\!]$ through random binning by choosing the value of $\Phi(v^r(\ell_j))$ independently and uniformly at random for every $v^r(\ell_j) \in \mathcal{V}^r$. The key $k_j = \Phi(v^r(\ell_j))$ obtained in block $j \in [\![1:B]\!]$ from the description of the CSI sequence $v^r(\ell_j)$ is used to assist the encoder in block $j+2$.

Codebook Generation for Messages: For each block $j \in [\![1:B]\!]$, let $C_2^{(r)} \triangleq \left\{U^r(m_j, t_{j-1}, k_{j-2})\right\}_{(m_j,t_{j-1},k_{j-2})\in\mathcal{M}\times\mathcal{T}\times\mathcal{K}}$, where $\mathcal{M} \triangleq [\![1:2^{rR}]\!]$, $\mathcal{T} \triangleq [\![1:2^{rR_t}]\!]$, and $\mathcal{K} \triangleq [\![1:2^{rR_k}]\!]$, be a random codebook consisting of independent random sequences each generated according to $P_U^{\otimes r}$. We denote a realization of $C_2^{(r)}$ by $\mathcal{C}_2^{(r)} \triangleq \left\{u^r(m_j, t_{j-1}, k_{j-2})\right\}_{(m_j,t_{j-1},k_{j-2})\in\mathcal{M}\times\mathcal{T}\times\mathcal{K}}$. Also, let $C_r = \left\{C_1^{(r)}, C_2^{(r)}\right\}$ and $\mathcal{C}_r = \left\{\mathcal{C}_1^{(r)}, \mathcal{C}_2^{(r)}\right\}$. The indices (m_j, t_{j-1}, k_{j-2}) can be viewed as a two layer binning. We define an ideal PMF for codebook $\mathcal{C}_r$ as in (178) at the top of the next page, as an approximate distribution to facilitate the analysis, in (178) $W_{Z|U,S}$ is the marginal distribution of $W_{Z|U,S} = \sum_{x\in\mathcal{X}} \mathbb{1}_{\{x=x(u,s)\}} W_{Z|X,S}$ and

$$P_{S|V} = \frac{P_{S,V}(s,v)}{P_V(v)} = \frac{Q_S(s)P_{V|S}(v|s)}{\sum_{s\in\mathcal{S}} Q_S(s)P_{V|S}(v|s)}. \tag{180}$$

Encoding: We assume that the transmitter and the receiver have access to the shared secret keys k_{-1} and k_0 for the first two blocks, but after the first two blocks they use the key that they generate from the CSI.

In the first block, to send the message m_1 according to k_{-1}, the encoder generates the index t_0 uniformly at random and computes $u^r(m_1, t_0, k_{-1})$ and transmits a codeword x^r, where $x_i = x(u_i(m_1, t_0, k_{-1}), s_{1,i})$. Note that, the index t_0 does not convey any useful information. At the end of the first block, to generate a secret key shared between the transmitter and the receiver, the encoder generates the index ℓ_1 according to the following distribution with $j = 1$,

$$f\big(\ell_j|s_j^r\big) = \frac{P_{S|V}^{\otimes r}\big(s_j^r|v^r(\ell_j)\big)}{\sum\limits_{\ell'\in[\![1:2^{r\tilde{R}}]\!]} P_{S|V}^{\otimes r}\big(s_j^r|v^r(\ell'_j)\big)}, \tag{181}$$

where $P_{S|V}$ is defined in (180). Then generates the reconciliation index $t_1 = \varphi(v^n(\ell_1))$; simultaneously, the transmitter generates a key $k_1 = \Phi(v^r(\ell_1))$ from the description of its CSI of the first block $v^r(\ell_1)$ to be used in Block 3.

In the second block, to transmit the message m_2 and the reconciliation index t_1 according to the key k_0, the encoder computes $u^r(m_2, t_1, k_0)$ and transmits a codeword x^r, where $x_i = x(u_i(m_2, t_1, k_0), s_{2,i})$. At the end of the second block, to generate a secret key shared between the transmitter and the receiver, the encoder generates the index ℓ_2 based s_2^r by using the likelihood encoder described in (181) with $j = 2$. Then generates the reconciliation index $t_2 = \varphi(v^n(\ell_2))$; simultaneously, the transmitter generates a key $k_2 = \Phi(v^r(\ell_2))$ from the description of its CSI of the second block $v^r(\ell_1)$ to be used in Block 4.

In block $j \in [\![3:B]\!]$, to send the message m_j and the reconciliation index t_{j-1} according to the generated key k_{j-2} from the previous blocks and the CSI of the current block s_j^r, the encoder computes $u^r(m_j, t_{j-1}, k_{j-2})$ and transmits a codeword x^r, where each coordinate of the transmitted signal is a function of the current state s_j^r as well as the corresponding sample of the transmitter's codeword u_i, i.e., $x_i = x(u_i(m_j, t_{j-1}, k_{j-2}), s_{j,i})$. At the end of this block, the encoder first selects the index ℓ_j based s_j^r by using the likelihood encoder described in (181) and then generates the reconciliation index $t_j = \varphi(v^r(\ell_j))$; simultaneously the encoder generates a key $k_j = \Phi(v^r(\ell_j))$ from the description of its CSI of the block j, $v^r(\ell_j)$, to be used in the Block $j+2$.

Considering (179) at the top of the next page, for a given codebook $\mathcal{C}_r$, the induced joint distribution over the codebook $\big($i.e. $P^{(\mathcal{C}_r)}\big)$ satisfies

$$\mathbb{D}\Big(P^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},U^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j} \\ ||\Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},U^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j}\Big) \leqslant \epsilon. \tag{182}$$

This intermediate distribution $\Upsilon^{(\mathcal{C}_r)}$ approximates the true distribution $P^{(\mathcal{C}_r)}$ and will be used in the sequel for bounding purposes. Expression (182) holds because the main difference between $P^{(\mathcal{C}_r)}$ and $\Upsilon^{(\mathcal{C}_r)}$ is that the keys K_{j-2}, K_{j-1} and the reconciliation index T_{j-1} are assumed to be uniformly distributed in $\Upsilon^{(\mathcal{C}_r)}$, which are made (arbitrarily) nearly uniform in $P^{(\mathcal{C}_r)}$ with appropriate control of rate as in (188) and (193).

Covert Analysis: We now show that this coding scheme guarantees that $\mathbb{E}_{C_n}[\mathbb{D}(P_{Z^n|C_n}||Q_Z^{\otimes n})] \xrightarrow[n\to\infty]{} 0$, where C_n is the set of all the codebooks for all blocks, and

$$Q_Z(\cdot) = \sum_{u\in\mathcal{U}}\sum_{v\in\mathcal{V}}\sum_{s\in\mathcal{S}}\sum_{x\in\mathcal{X}} P_U(u)P_V(v)P_{S|V}(s|v)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

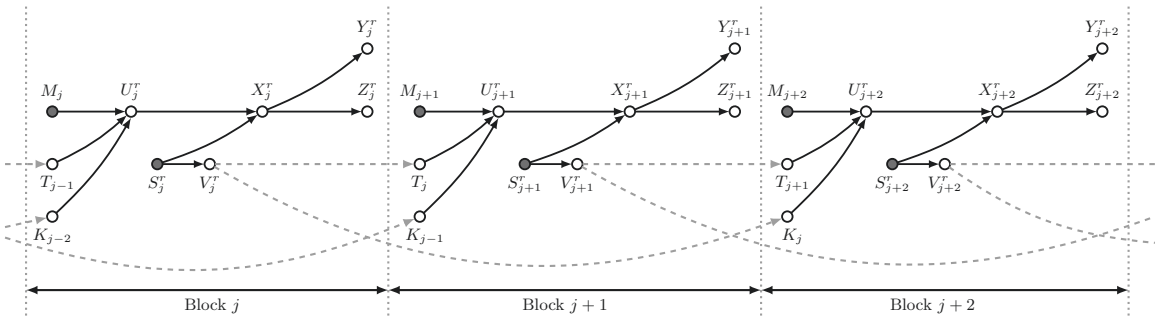

Fig. 12. Functional dependence graph for the block-Markov encoding scheme

$$\Gamma^{(\mathcal{C}_n)}_{M_j,T_{j-1},K_{j-2},L_j,U^r,V^r,S_j^r,Z_j^r,K_{j-1},T_j,K_j}(m_j,t_{j-1},k_{j-2},\ell_j,\tilde{u}^r,\tilde{v}^r,s_j^r,z_j^r,k_{j-1},t_j,k_j)$$
$$\triangleq 2^{-r(R+R_t+R_k+\tilde{R})}\mathbb{1}_{\{\tilde{u}^r=u^r(m_j,t_{j-1},k_{j-2})\}}\mathbb{1}_{\{\tilde{v}^r=v^r(\ell_j)\}}P_{S|V}^{\otimes r}(s_j^r|\tilde{v}^r)W_{Z|U,S}^{\otimes r}(z_j^r|\tilde{u}^r,s_j^r)$$
$$\times 2^{-rR_k}\mathbb{1}_{\{t_j=\varphi(\tilde{v}^r)\}}\mathbb{1}_{\{k_j=\Phi(\tilde{v}^r)\}} \tag{178}$$

$$\Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2},U^r,S_j^r,L_j,V^r,Z_j^r,K_{j-1},T_j,K_j}(m_j,t_{j-1},k_{j-2},\tilde{u}^r,s_j^r,\ell_j,\tilde{v}^r,z_j^r,k_{j-1},t_j,k_j)$$
$$\triangleq 2^{-r(R+R_t+R_k)}\mathbb{1}_{\{\tilde{u}^r=u^r(m_j,t_{j-1},k_{j-2})\}}Q_S^{\otimes r}(s_j^r)f(\ell_j|s_j^r)\mathbb{1}_{\{\tilde{v}^r=v^r(\ell_j)\}}$$
$$\times W_{Z|U,S}^{\otimes r}(z_j^r|\tilde{u}^r,s_j^r)2^{-rR_k}\mathbb{1}_{\{t_j=\varphi(\tilde{v}^r)\}}\mathbb{1}_{\{k_j=\Phi(\tilde{v}^r)\}} \tag{179}$$

$$\times\,\mathbb{1}_{\left\{X=X(u,s)\right\}}W_{Z|X,S}(\cdot|x,s), \tag{183}$$

such that $\sum_{v\in\mathcal{V}}P_V(v)P_{S|V}(\cdot|v)=Q_S(\cdot)$. Then, we choose P_U, P_V, $P_{S|V}$, and $x(u,s)$ such that it satisfies $Q_Z=Q_0$. Similar to (140) using the functional dependence graph depicted in Fig. 12 it follows that,

$$\mathbb{D}\Big(P_{Z^n}^{(\mathcal{C}_r)}||Q_Z^{\otimes n}\Big)\leqslant$$
$$2\sum_{j=1}^{B}\mathbb{D}\Big(P_{Z_j^r,K_{j-1},T_j,K_j}^{(\mathcal{C}_r)}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}\Big). \tag{184}$$

To bound the RHS of (184) by using Lemma 1 and the triangle inequality we have,

$$\begin{aligned}
&\mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r}-Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1\\
&\leqslant\mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r}-\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1\\
&\quad+\mathbb{E}_{C_r}||\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}-Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1\\
&\leqslant\mathbb{E}_{C_r}||P_{Z_j^r,K_{j-1},T_j,K_j|C_r}-\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1\\
&\quad+\mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r}-\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1\\
&\quad+\mathbb{E}_{C_r}||\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}-Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j}||_1.
\end{aligned} \tag{185}$$

From (182) and the monotonicity of KL-divergence the first term on the RHS of (185) goes to zero when r grows. To bound the second term on the RHS of (185) for a fixed codebook $\mathcal{C}_r$, we have (186) at the top of the next page, in which (186c) follows from (181).

Hence,

$$\begin{aligned}
&\mathbb{E}_{C_r}||\Upsilon_{Z_j^r,K_{j-1},T_j,K_j|C_r}-\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||_1\\
&\leqslant\mathbb{E}_{C_r}||\Upsilon_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r}\\
&\quad-\Gamma_{M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1},T_j,K_j|C_r}||_1\\
&\overset{(a)}{=}\mathbb{E}_{C_r}||\Upsilon_{M_j,T_{j-1},K_{j-2},S_j^r|C_r}-\Gamma_{M_j,T_{j-1},K_{j-2},S_j^r|C_r}||_1\\
&\overset{(b)}{=}\mathbb{E}_{C_r}||Q_S^{\otimes r}-\Gamma_{S_j^r|M_j=1,T_{j-1}=1,K_{j-2}=1,C_r}||_1,
\end{aligned} \tag{187}$$

where (a) follows from (186b)-(186h) and (b) follows from the symmetry of the codebook construction with respect to M_j, T_{j-1}, and K_{j-2} and (186a). Based on the soft covering lemma [33, Corollary VII.5] the RHS of (187) vanishes when r grows if

$$\tilde{R}>\mathbb{I}(S;V). \tag{188}$$

We now proceed to bound the third term on the RHS of (185). First, consider the following marginal from (178),

$$\begin{aligned}
&\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)\\
&=\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\sum_{s_j^r}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}}P_{S|V}^{\otimes r}\big(s_j^r|V^r(\ell_j)\big)\\
&\quad\times W_{Z|U,S}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2}),s_j^r\big)\\
&\quad\times\mathbb{1}_{\{t_j=\varphi(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}}\\
&=\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}}\\
&\quad\times W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2}),V^r(\ell_j)\big)\\
&\quad\times\mathbb{1}_{\{t_j=\varphi(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}},
\end{aligned} \tag{189}$$

where $W_{Z|U,V}(z|u,v)=\sum_{s\in\mathcal{S}}P_{S|V}(s|v)W_{Z|U,S}(z|u,s)$. To bound bound the third term on the RHS of (185), by using Pinsker's inequality in Lemma 1 it is sufficient to bound

$$\Gamma^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2}} = 2^{-r(R+R_t+R_k)} = \Upsilon^{(\mathcal{C}_r)}_{M_j,T_{j-1},K_{j-2}}, \tag{186a}$$

$$\Gamma^{(\mathcal{C}_r)}_{U^r|M_j,T_{j-1},K_{j-2},S_j^r} = \mathbb{1}_{\{\tilde{u}^r=u^r(m_j,t_{j-1},k_{j-2})\}} = \Upsilon^{(\mathcal{C}_r)}_{U^r|M_j,T_{j-1},K_{j-2},S_j^r}, \tag{186b}$$

$$\Gamma^{(\mathcal{C}_r)}_{L_j|M_j,T_{j-1},K_{j-2},S_j^r,U^r} = f(\ell_j|s_j^r) = \Upsilon^{(\mathcal{C}_r)}_{L_j|M_j,T_{j-1},K_{j-2},S_j^r,U^r}, \tag{186c}$$

$$\Gamma^{(\mathcal{C}_r)}_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r} = \mathbb{1}_{\{\tilde{v}^r=v^r(\ell_j)\}} = \Upsilon^{(\mathcal{C}_r)}_{V^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r}, \tag{186d}$$

$$\Gamma^{(\mathcal{C}_r)}_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r} = W^{\otimes r}_{Z|U,S} = \Upsilon^{(\mathcal{C}_r)}_{Z_j^r|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r}, \tag{186e}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r} = 2^{-rR_k} = \Upsilon^{(\mathcal{C}_r)}_{K_{j-1}|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r}, \tag{186f}$$

$$\Gamma^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1}} = \mathbb{1}_{\{t_j=\sigma(v^r(\ell_j))\}} = \Upsilon^{(\mathcal{C}_r)}_{T_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1}}, \tag{186g}$$

$$\Gamma^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1},T_j} = \mathbb{1}_{\{k_j=\Phi(v^r(\ell_j))\}} = \Upsilon^{(\mathcal{C}_r)}_{K_j|M_j,T_{j-1},K_{j-2},S_j^r,L_j,U^r,V^r,Z_j^r,K_{j-1},T_j}, \tag{186h}$$

$\mathbb{E}_{C_r}[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})]$ as in (190) available at the next page, in which

(a) follows from Jensen's inequality;
(b) holds because $\mathbb{1}_{\{.\}} \leqslant 1$;
(c) follows by defining Ψ_1 and Ψ_2 as follows,

$$\begin{aligned}
\Psi_1 &= \frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}} \sum_{(k_{j-1},t_j,k_j)} \sum_{m_j} \sum_{t_{j-1}} \sum_{k_{j-2}} \\
&\times \sum_{\ell_j} \sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right)\in\mathcal{T}_\epsilon^{(r)}} \\
&\times \Gamma^{\otimes r}_{U^r,V^r,Z_j^r}\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right) \\
&\times \log\Bigg(\frac{W^{\otimes r}_{Z|U,V}\left(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)} \\
&+ \frac{W^{\otimes r}_{Z|V}\left(z_j^r|v^r(\ell_j)\right)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)} \\
&+ \frac{W^{\otimes r}_{Z|U}\left(z_j^r|u^r(m_j,t_{j-1},k_{j-2})\right)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)} + 1\Bigg) \\
&\leq \log\Bigg(\frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|U,V)}}{2^{r(R+R_t+R_k+\tilde{R})}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}} \\
&+ \frac{2^{r(R_T+R_K)}\times 2^{-r(1-\epsilon)\mathbb{H}(Z|V)}}{2^{r\tilde{R}}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}} \\
&+ \frac{2^{-r(1-\epsilon)\mathbb{H}(Z|U)}}{2^{r(R+R_t+R_k)}\times 2^{-r(1+\epsilon)\mathbb{H}(Z)}} + 1\Bigg)
\end{aligned} \tag{191}$$

$$\begin{aligned}
\Psi_2 &= \frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}} \sum_{(k_{j-1},t_j,k_j)} \sum_{m_j} \sum_{t_{j-1}} \sum_{k_{j-2}} \\
&\times \sum_{\ell_j} \sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right)\notin\mathcal{T}_\epsilon^{(r)}} \\
&\times \Gamma^{\otimes r}_{U^r,V^r,Z_j^r}\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\right) \\
&\times \log\Bigg(\frac{W^{\otimes r}_{Z|U,V}\left(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)} \\
&+ \frac{W^{\otimes r}_{Z|V}\left(z_j^r|v^r(\ell_j)\right)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)} \\
&+ \frac{W^{\otimes r}_{Z|U}\left(z_j^r|u^r(m_j,t_{j-1},k_{j-2})\right)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)} + 1\Bigg) \\
&\leq 2|V||U||Z|e^{-r\epsilon^2\mu_{V,U,Z}} r\log\left(\frac{2}{\mu_Z}+1\right).
\end{aligned} \tag{192}$$

In (192) $\mu_{V,U,Z} = \min_{(v,u,z)\in(\mathcal{V},\mathcal{U},\mathcal{Z})} P_{V,U,Z}(v,u,z)$ and $\mu_Z = \min_{z\in\mathcal{Z}} P_Z(z)$. When $r\to\infty$ then $\Psi_2\to 0$ and Ψ_1 goes to zero when r grows if

$$R + R_t + R_k + \tilde{R} - R_T - R_K > \mathbb{I}(U,V;Z), \tag{193a}$$

$$\tilde{R} - R_T - R_K > \mathbb{I}(V;Z), \tag{193b}$$

$$R + R_t + R_k > \mathbb{I}(U;Z). \tag{193c}$$

Decoding and Error Probability Analysis: At the end of the block $j\in[\![1:B]\!]$, using its knowledge of the key k_{j-2} generated from the block $j-2$, the receiver finds a unique pair $(\hat{m}_j,\hat{t}_{j-1})$ such that $\left(u^r(\hat{m}_j,\hat{t}_{j-1},k_{j-2}),y_j^r\right)\in\mathcal{T}_\epsilon^{(r)}$. According to the law of large numbers and the packing lemma probability of error vanishes when r grows if [37],

$$R + R_t < \mathbb{I}(U;Y). \tag{194}$$

We now analyze the probability of error at the encoder and the decoder for key generation. Let (L_{j-1},T_{j-1}) denote the chosen indices at the encoder and $\hat{L}_{j-1}$ and $\hat{T}_{j-1}$ be the estimate of the index L_{j-1} and T_{j-1} at the decoder. At the end of block j, by decoding U_j^r, the receiver knows T_{j-1} and to find L_{j-1} we define the error event,

$$\mathcal{E} = \left\{\left(V_{j-1}^r(\hat{L}_{j-1}),S_{j-1}^r,U_{j-1}^r,Y_{j-1}^r\right)\notin\mathcal{T}_\epsilon^{(r)}\right\}. \tag{195}$$

Also, consider the error events,

$$\mathcal{E}_1 = \left\{\left(V_{j-1}^r(\ell_{j-1}),S_{j-1}^r\right)\notin\mathcal{T}_{\epsilon'}^{(r)} \text{ for all } \ell_{j-1}\in[\![1:2^{r\tilde{R}}]\!]\right\}, \tag{196a}$$

$$\mathcal{E}_2 = \left\{\left(V_{j-1}^r(L_{j-1}),S_{j-1}^r,U_{j-1}^r,Y_{j-1}^r\right)\notin\mathcal{T}_\epsilon^{(r)}\right\}, \tag{196b}$$

$$\mathcal{E}_3 = \Big\{\left(V_{j-1}^r(\tilde{\ell}_{j-1}),U_{j-1}^r,Y_{j-1}^r\right)\in\mathcal{T}_\epsilon^{(r)} \text{ for some } \ell_{j-1}\in\mathcal{B}(T_{j-1}),\tilde{\ell}_{j-1}\neq\ell_{j-1}\Big\}, \tag{196c}$$

where $\epsilon>\epsilon'>0$. By the union bound we have,

$$P(\mathcal{E}) \leqslant P(\mathcal{E}_1) + P(\mathcal{E}_1^c\cap\mathcal{E}_2) + P(\mathcal{E}_3). \tag{197}$$

$$
\begin{aligned}
&\mathbb{E}_{C_r}\left[\mathbb{D}(\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}||Q_Z^{\otimes r}Q_{K_{j-1}}Q_{T_j}Q_{K_j})\right]\\
&=\mathbb{E}_{C_r}\left[\sum_{z_j^r,k_{j-1},t_j,k_j}\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)\log\left(\frac{\Gamma_{Z_j^r,K_{j-1},T_j,K_j|C_r}(z_j^r,k_{j-1},t_j,k_j)}{Q_Z^{\otimes r}(z_j^r)Q_{K_{j-1}}(k_{j-1})Q_{T_j}(t_j)Q_{K_j}(k_j)}\right)\right]\\
&=\mathbb{E}_{C_r}\left[\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(m_j,t_{j-1},k_{j-2}),V^r(\ell_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\ell_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\ell_j))\}}}{2^{r(R+R_t+2R_k+\tilde{R})}}\right.\\
&\left.\times\log\left(\frac{\sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right)\right]\\
&\overset{(a)}{\leq}\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R})}}\sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\mathbb{E}_{\varphi(v^r(\ell_j))}\left[\mathbb{1}_{\{t_j=\varphi(v^r(\ell_j))\}}\right]\times\mathbb{E}_{\Phi(v^r(\ell_j))}\left[\mathbb{1}_{\{k_j=\Phi(v^r(\ell_j))\}}\right]\\
&\times\log\mathbb{E}_{\substack{\setminus(m_j,t_{j-1},k_{j-2},\ell_j),\\ \setminus(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\left[\frac{\sum_{\tilde{m}_j}\sum_{\tilde{t}_{j-1}}\sum_{\tilde{k}_{j-2}}\sum_{\tilde{\ell}_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right]\\
&\overset{(b)}{\leq}\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}}\sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\log\frac{1}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\times\left(W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)\right.\\
&+\mathbb{E}_{\setminus(m_j,t_{j-1},k_{j-2})}\left[\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),v^r(\ell_j)\big)\right]\\
&+\mathbb{E}_{\substack{\setminus\ell_j,\\ \setminus(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\left[\sum_{\tilde{\ell}_j\neq\ell_j}W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}\right]\\
&\left.+\mathbb{E}_{\substack{\setminus(m_j,t_{j-1},k_{j-2},\ell_j),\\ \setminus(\varphi(v^r(\ell_j)),\Phi(v^r(\ell_j)))}}\left[\sum_{\tilde{\ell}_j\neq\ell_j}\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})}W_{Z|U,V}^{\otimes r}\big(z_j^r|U^r(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2}),V^r(\tilde{\ell}_j)\big)\mathbb{1}_{\{t_j=\varphi(V^r(\tilde{\ell}_j))\}}\mathbb{1}_{\{k_j=\Phi(V^r(\tilde{\ell}_j))\}}\right]\right)\\
&\leq\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}}\sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\log\left(\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\sum_{(\tilde{m}_j,\tilde{t}_{j-1},\tilde{k}_{j-2})\neq(m_j,t_{j-1},k_{j-2})}\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}\right.\\
&\left.+\sum_{\tilde{\ell}_j\neq\ell_j}\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2})\big)}{2^{r(R+R_t+R_k+\tilde{R})}Q_Z^{\otimes r}(z_j^r)}+1\right)\\
&\leq\sum_{(z_j^r,k_{j-1},t_j,k_j)}\sum_{m_j}\sum_{t_{j-1}}\sum_{k_{j-2}}\sum_{\ell_j}\frac{1}{2^{r(R+R_t+2R_k+\tilde{R}+R_T+R_K)}}\sum_{\left(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\right)}\Gamma_{U^r,V^r,Z_j^r}^{\otimes r}\big(u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j),z_j^r\big)\\
&\times\log\left(\frac{W_{Z|U,V}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2}),v^r(\ell_j)\big)}{2^{r(R+R_t+R_k+\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|V}^{\otimes r}\big(z_j^r|v^r(\ell_j)\big)}{2^{r(\tilde{R}-R_T-R_K)}Q_Z^{\otimes r}(z_j^r)}+\frac{W_{Z|U}^{\otimes r}\big(z_j^r|u^r(m_j,t_{j-1},k_{j-2})\big)}{2^{r(R+R_t+R_k)}Q_Z^{\otimes r}(z_j^r)}+1\right)\\
&\overset{(c)}{\triangleq}\Psi_1+\Psi_2,
\end{aligned}
\tag{190}
$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

Similar to the proof of Lemma 2 one can show that the first term on the RHS of (197) vanishes when r grows if we have (188). Following the steps in [37, Sec. 11.3.1], the last two terms on the RHS of (197) go to zero when r grows if,

$$\tilde{R} > \mathbb{I}(S;V), \tag{198a}$$

$$\tilde{R} - R_t < \mathbb{I}(V;U,Y). \tag{198b}$$

Applying Fourier-Motzkin to (188), (193), (194), and (198) and remarking that the scheme requires $R_t + R_k \leqslant R_T + R_K$ results in the achievable region in Theorem 7.

Appendix H
Proof of Theorem 9

Consider any sequence of length-n codes for a state-dependent channel with CSI available causally only at the transmitter such that $P_e^{(n)} \leqslant \epsilon_n$, $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \delta$, and $R_K/n \leqslant \lambda_n$ with $\lim_{n\to\infty}\epsilon_n = \lim_{n\to\infty}\lambda_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (35) as follows,

$$\mathcal{A}_\epsilon \triangleq \left\{R \geqslant 0 : \exists P_{U,V,S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leqslant \mathbb{I}(U;Y) + \epsilon\right\}, \tag{199a}$$

where

$$\mathcal{D}_\epsilon = \left\{\begin{array}{l} P_{U,V,S,X,Y,Z} : \\ P_{U,V,S,X,Y,Z} = Q_S P_V P_{U|V} \mathbb{1}_{\left\{X = X(U,S)\right\}} W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leqslant \epsilon \\ \mathbb{I}(U;Y) \geqslant \mathbb{I}(V;Z) - 4\epsilon \\ \max\{|\mathcal{U}|, |\mathcal{V}|\} \leqslant |\mathcal{X}| \end{array}\right\}. \tag{199b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques,

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &= \mathbb{H}(M|K) \\ &\overset{(a)}{\leq} \mathbb{I}(M;Y^n|K) + n\epsilon_n \\ &= \sum_{i=1}^n \mathbb{I}(M;Y_i|Y^{i-1},K) + n\epsilon_n \\ &\leqslant \sum_{i=1}^n \mathbb{I}(M,K,Y^{i-1};Y_i) + n\epsilon_n \\ &\overset{(b)}{\leqslant} \sum_{i=1}^n \mathbb{I}(M,K,S^{i-1};Y_i) + n\epsilon_n \\ &\overset{(c)}{=} \sum_{i=1}^n \mathbb{I}(U_i;Y_i) + n\epsilon_n \\ &= n\sum_{i=1}^n \frac{1}{n}\mathbb{I}(U_i;Y_i) + n\epsilon_n \\ &= n\sum_{i=1}^n \mathbb{P}(T=i)\mathbb{I}(U_T;Y_T|T=i) + n\epsilon_n \\ &= n\mathbb{I}(U_T;Y_T|T) + n\epsilon_n \\ &\leqslant n\mathbb{I}(U_T,T;Y_T) + n\epsilon_n \\ &\overset{(d)}{=} n\mathbb{I}(U;Y) + n\epsilon_n \\ &\overset{(e)}{\leqslant} n\mathbb{I}(U;Y) + n\epsilon, \end{aligned} \tag{200}$$

where

- (a) follows from Fano's inequality;
- (b) follows since $(M,K,Y^{i-1}) - (M,K,S^{i-1}) - Y_i$, note that we also have $V_i - (M,K,S^{i-1}) - Y_i$, where $V_i \triangleq (M,K,Z^{i-1})$;
- (c) follows by defining $U_i \triangleq (M,K,S^{i-1})$;
- (d) follows by defining $U \triangleq (U_T,T)$ and $Y \triangleq Y_T$;
- (e) follows by defining $\epsilon \triangleq \max\{\epsilon_n,\lambda_n,\nu\}$, where we choose n large enough such that $\nu \geqslant \frac{\delta}{n}$.

Next, we lower bound nR as follows,

$$\begin{aligned} nR + R_K &\geqslant \mathbb{H}(M,K) \\ &\geq \mathbb{I}(M,K;Z^n) \\ &= \sum_{i=1}^n \mathbb{I}(M,K;Z_i|Z^{i-1}) \\ &\overset{(a)}{\geq} \sum_{i=1}^n \mathbb{I}(M,K,Z^{i-1};Z_i) - \delta \\ &\overset{(b)}{=} \sum_{i=1}^n \mathbb{I}(V_i;Z_i) - \delta \\ &= n\sum_{i=1}^n \mathbb{P}(T=i)\mathbb{I}(V_T;Z_T|T=i) - \delta \\ &= n\mathbb{I}(V_T;Z_T|T) - \delta \\ &\overset{(c)}{\geq} n\mathbb{I}(V_T,T;Z_T) - 2\delta \\ &\overset{(d)}{=} n\mathbb{I}(V;Z) - 2\delta \end{aligned} \tag{201}$$

where

- (a) and (c) follow from Lemma 3;
- (b) follows from the definition of $V_i \triangleq (M,K,Z^{i-1})$, which is defined in the process of deriving (200);
- (d) follows by defining $V \triangleq (V_T,T)$ and $Z \triangleq Z_T$.

For any $\nu > 0$, choosing n large enough ensures that,

$$R + \frac{R_K}{n} \geqslant \mathbb{I}(V;Z) - 2\nu. \tag{202}$$

Therefore,

$$\begin{aligned} R &\geqslant \mathbb{I}(V;Z) - 2\nu - \frac{R_K}{n} \\ &\geqslant \mathbb{I}(V;Z) - 2\nu - \lambda_n \\ &\geqslant \mathbb{I}(V;Z) - 3\epsilon, \end{aligned} \tag{203}$$

where the last inequality follows since $\epsilon \triangleq \max\{\epsilon_n,\lambda_n,\nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leqslant \epsilon$, note that for n large enough

$$\begin{aligned} \mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) &= \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^n P_{Z_i}\middle\|Q_0\right) \\ \leqslant \frac{1}{n}\sum_{i=1}^n \mathbb{D}(P_{Z_i}||Q_0) &\leqslant \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \frac{\delta}{n} \leqslant \nu \leqslant \epsilon. \end{aligned} \tag{204}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3135291, IEEE Transactions on Information Theory

Combining (200) and (203), and (204) shows that $\forall \epsilon_n, \lambda_n, \nu > 0$, $R \leqslant \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$R \leqslant \max\left\{a : a \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon\right\}. \tag{205}$$

Continuity at Zero: One can prove the continuity at zero of $\mathcal{A}_\epsilon$ by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(U;Y)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(V;Z)$ in the continuity at zero proof in Appendix F and following the exact same arguments.

Appendix I
Proof of Theorem 12

Consider any sequence of length-n codes for a state-dependent channel with CSI available strictly causally only at the transmitter such that $P_e^{(n)} \leqslant \epsilon_n$, $\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \delta$, and $R_K/n \leqslant \lambda_n$ with $\lim_{n\to\infty} \epsilon_n = \lim_{n\to\infty} \lambda_n = 0$. Note that the converse is consistent with the model and does *not* require δ to vanish.

Epsilon Rate Region: We first define a region $\mathcal{A}_\epsilon$ for $\epsilon > 0$ that expands the region defined in (43) as follows,

$$\mathcal{A}_\epsilon \triangleq \left\{R \geqslant 0 : \exists P_{V,S,X,Y,Z} \in \mathcal{D}_\epsilon : R \leqslant \mathbb{I}(X;Y) + \epsilon\right\}, \tag{206a}$$

where

$$\mathcal{D}_\epsilon = \left\{\begin{array}{l} P_{V,S,X,Y,Z} : \\ P_{V,S,X,Y,Z} = Q_S P_V P_{X|V} W_{Y,Z|X,S} \\ \mathbb{D}\left(P_Z \| Q_0\right) \leqslant \epsilon \\ \mathbb{I}(X;Y) \geqslant \mathbb{I}(V;Z) - 4\epsilon \\ |\mathcal{V}| \leqslant |\mathcal{X}| \end{array}\right\}. \tag{206b}$$

We next show that if a rate R is achievable then $R \in \mathcal{A}_\epsilon$ for any $\epsilon > 0$. For any $\epsilon_n > 0$ and $\nu > 0$, we start by upper bounding nR using standard techniques.

$$\begin{aligned} nR &= \mathbb{H}(M) \\ &= \mathbb{H}(M|K) \\ &\overset{(a)}{\leq} \mathbb{I}(M;Y^n|K) + n\epsilon_n \\ &= \sum_{i=1}^n \mathbb{I}(M;Y_i|Y^{i-1},K) + n\epsilon_n \\ &\leqslant \sum_{i=1}^n \mathbb{I}(M,K,Y^{i-1};Y_i) + n\epsilon_n \\ &\overset{(b)}{\leq} \sum_{i=1}^n \mathbb{I}(M,K,S^{i-1};Y_i) + n\epsilon_n \\ &\leqslant \sum_{i=1}^n \mathbb{I}(M,K,S^{i-1},Z^{i-1};Y_i) + n\epsilon_n \\ &\overset{(c)}{\leq} \sum_{i=1}^n \mathbb{I}(X_i;Y_i) + n\epsilon_n \\ &= n\sum_{i=1}^n \frac{1}{n}\mathbb{I}(X_i;Y_i) + n\epsilon_n \\ &= n\sum_{i=1}^n \mathbb{P}(T=i)\mathbb{I}(X_i;Y_i|T=i) + n\epsilon_n \\ &= n\mathbb{I}(X_T;Y_T|T) + n\epsilon_n \\ &\leqslant n\mathbb{I}(X_T,T;Y_T) + n\epsilon_n \\ &\overset{(d)}{=} n\mathbb{I}(X;Y) + n\epsilon_n \\ &\overset{(e)}{\leqslant} n\mathbb{I}(X;Y) + n\epsilon, \end{aligned} \tag{207}$$

where

- (a) follows from Fano's inequality;
- (b) follows from the Markov chain $(M,K,Y^{i-1}) - (M,K,S^{i-1}) - Y_i$;
- (c) follows since S_i is independent of $\left(M,K,S^{i-1},Z^{i-1},X_i(M,S^{i-1})\right)$ and therefore

$$\begin{aligned} &\mathbb{I}(M,K,S^{i-1},Z^{i-1};Y_i|X_i) \\ &\leqslant \mathbb{I}(M,K,S^{i-1},Z^{i-1};Y_i|X_i,S_i) = 0, \end{aligned} \tag{208}$$

  that is $(M,K,S^{i-1},Z^{i-1}) - X_i - Y_i$ forms a Markov chain, which implies $V_i - X_i - Y_i$, where $V_i \triangleq (M,K,Z^{i-1})$, also forms a Markov chain;
- (d) follows by defining $X \triangleq (X_T,T)$ and $Y \triangleq Y_T$.
- (e) follows by defining $\epsilon \triangleq \max\{\epsilon_n,\lambda_n,\nu\}$, where we choose n large enough such that $\nu \geqslant \frac{\delta}{n}$.

Next, we lower bound nR as follows,

$$\begin{aligned} nR + R_K &\geqslant \mathbb{H}(M,K) \\ &\geq \mathbb{I}(M,K;Z^n) \\ &= \sum_{i=1}^n \mathbb{I}(M,K;Z_i|Z^{i-1}) \\ &\overset{(a)}{\geq} \sum_{i=1}^n \mathbb{I}(M,K,Z^{i-1};Z_i) - \delta \\ &\overset{(b)}{=} \sum_{i=1}^n \left[\mathbb{I}(V_i;Z_i)\right] - \delta \\ &= n\mathbb{I}(V_T;Z_T|T) - \delta \\ &\overset{(c)}{\geq} n\mathbb{I}(V_T,T;Z_T) - 2\delta \\ &\overset{(d)}{=} n\mathbb{I}(V;Z) - 2\delta \end{aligned} \tag{209}$$

where

- (a) and (c) follow from Lemma 3;
- (b) follows by defining $V_i \triangleq (M,K,Z^{i-1})$ which is defined in the process of deriving (207);
- (d) follows by defining $V \triangleq (V_T,T)$ and $Z \triangleq Z_T$.

For any $\nu > 0$, choosing n large enough ensures that,

$$R + \frac{R_K}{n} \geqslant \mathbb{I}(V;Z) - 2\nu. \tag{210}$$

Therefore,

$$\begin{aligned} R &\geqslant \mathbb{I}(V;Z) - 2\nu - \frac{R_K}{n} \\ &\geqslant \mathbb{I}(V;Z) - 2\nu - \lambda_n \\ &\geqslant \mathbb{I}(V;Z) - 3\epsilon, \end{aligned} \tag{211}$$

where the last inequality follows since $\epsilon \triangleq \max\{\epsilon_n,\lambda_n,\nu\}$. To show that $\mathbb{D}(P_Z||Q_0) \leqslant \epsilon$, note that for n large enough

$$\mathbb{D}(P_Z||Q_0) = \mathbb{D}(P_{Z_T}||Q_0) = \mathbb{D}\left(\frac{1}{n}\sum_{i=1}^n P_{Z_i}\middle\|Q_0\right)$$

$$\leqslant \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}(P_{Z_i}||Q_0) \leqslant \frac{1}{n}\mathbb{D}(P_{Z^n}||Q_0^{\otimes n}) \leqslant \frac{\delta}{n} \leqslant \nu \leqslant \epsilon. \tag{212}$$

Combining (207) and (211), and (212) shows that $\forall \epsilon_n, \nu > 0$, $R \leqslant \max\{a : a \in \mathcal{A}_\epsilon\}$. Therefore,

$$R \leqslant \max\left\{a : a \in \bigcap_{\epsilon>0}\mathcal{A}_\epsilon\right\}. \tag{213}$$

Continuity at Zero: One can prove the continuity at zero of $\mathcal{A}_\epsilon$ by substituting $\min\{\mathbb{I}(U;Y) - \mathbb{I}(U;S), \mathbb{I}(U,V;Y) - \mathbb{I}(U;S|V)\}$ with $\mathbb{I}(X;Y)$ and $\mathbb{I}(V;Z) - \mathbb{I}(V;S)$ with $\mathbb{I}(V;Z)$ in the continuity at zero proof in Appendix F and following the exact same arguments.

REFERENCES

[1] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Keyless covert communication in the presence of non-causal channel state information," in *Proc. IEEE Info. Theory Workshop (ITW)*, Visby, Sweden, Aug. 2019, pp. 1–5.

[2] ——, "Keyless covert communication in the presence of channel state information," in *IEEE Int. Symp. on Info. Theory (ISIT)*, Los Angeles, CA, USA, Jun. 2020, pp. 834–839.

[3] A. B. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[4] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: Hiding messages in noise," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2945–2949.

[5] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[6] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[7] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.

[8] T. V. Sobers, A. B. Bash, S. Guha, D. Towsley, and D. Goeckel, "Covert communication in the presence of an uninformed jammer," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6193–6206, Sep. 2017.

[9] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. IEEE Info. Theory Workshop (ITW)*, Hobart, TAS, Australia, Nov. 2014, pp. 30–34.

[10] S.-H. Lee, L. Wang, A. Khisti, and G. W. Wornell, "Covert communication with channel-state information at the transmitter," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2310–2319, Sep. 2018.

[11] Y. Chen and A. J. Han Vinck, "Wiretap channel with side information," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 395–402, Jan. 2008.

[12] A. Khisti, S. N. Diggavi, and G. W. Wornell, "Secret key agreement using asymmetry in channel state knowledge," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Seoul, South Korea, Jul. 2009, pp. 2286–2290.

[13] Y.-K. Chia and A. El Gamal, "Wiretap channel with causal state information," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2838–2849, May 2012.

[14] 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.

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

[16] T. S. Han and M. Sasaki, "Wiretap channels with causal state information: Strong secrecy," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6750–6765, Oct. 2019.

[17] Z. Goldfeld, P. Cuff, and H. H. Permuter, "Wiretap channels with random states non-causally available at the encoder," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1497–1519, Mar. 2020.

[18] P.-H. Lin, C. R. Janda, and E. A. Jorswieck, "Stealthy secret key generation," in *Proc. IEEE Global Conf. on Signal and Info. Processing (GlobalSIP)*, Montreal, QC, Canada, Mar. 2017, pp. 492–496.

[19] P.-H. Lin, C. R. Janda, E. A. Jorswieck, and R. F. Schaefer, "Stealthy keyless secret key generation from degraded sources," in *51st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Apr. 2018, pp. 14–18.

[20] M. Tahmasbi and M. R. Bloch, "Covert secret key generation," in *Proc. IEEE Conf. on Commun. and Network Security (CNS)*, Las Vegas, NV, USA, Dec. 2017, pp. 540–544.

[21] ——, "Framework for covert and secret key expansion over classical-quantum channels," *Phys. Rev. A*, vol. 99, p. 052329, May 2019.

[22] S. Salehkalaibar, M. H. Yassaee, V. Y. F. Tan, and M. Ahmadipour, "State masking over a two-state compound channel," *IEEE Trans. Inf. Theory*, vol. 67, no. 9, pp. 5651–5673, Sep. 2021.

[23] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 1254–1274, Feb. 2012.

[24] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Advances in Cryptology & CRYPTO 2012*, ser. Lecture Notes in Computer Science, R. Safavi-Naini and R. Canetti, Eds., vol. 7417. Springer Berlin Heidelberg, 2012, pp. 294–311, hard-copy.

[25] M. R. Bloch, M. Hayashi, and A. Thangaraj, "Error-control coding for physical-layer secrecy," *Proceedings of IEEE*, vol. 103, no. 10, pp. 1725–1746, Oct. 2015.

[26] H. ZivariFard, M. R. Bloch, and A. Nosratinia, "Keyless covert communication via channel state information," *available at https://arxiv.org/abs/2003.03308*, Mar. 2020.

[27] H. ZivariFard, "Secrecy and covertness in the presence of multi-casting, channel state information, and cooperative jamming," *Dept. Elect. Eng., Univ. Texas at Dallas, Dallas, TX, USA*, Dec. 2021.

[28] I. Sason and S. Verdú, "f-divergence inequalities," *IEEE Trans. Inf. Theory*, vol. 62, no. 11, pp. 5973–6006, Nov. 2016.

[29] C. S. Song, P. Cuff, and H. V. Poor, "The likelihood encoder for lossy compression," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1836–1849, Apr. 2016.

[30] E. L. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. New York, NY, USA: Springer-Verlag, 2005.

[31] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Problem Control Inf. Theory*, vol. 9, no. 1, pp. 19–31, Jan. 1980.

[32] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[33] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[34] M. H. Yassaee, M. R. Aref, and A. A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *Proc. IEEE Int. Symp. on Info. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 1287–1291.

[35] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1574–1605, Apr. 2015.

[36] H. G. Eggleston, *Convexity*, 6th ed. Cambridge, U.K: Cambridge University Press, 1958.

[37] A. El Gamal and Y.-H. Kim, *Network Information Theory*, 1st ed. Cambridge, U.K: Cambridge University Press, 2012.

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

[39] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channels," *IEEE Trans. Inf. Theory*, vol. 25, no. 6, pp. 572–584, Sep. 1979.

[40] F. M. J. Willems and E. C. van der Meulen, "The discrete memoryless multiple-access channel with cribbing encoders," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 313–327, May 1985.

[41] Y. Steinberg, "Resolvability theory for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 472–487, Mar. 1998.

[42] M. H. Yassaee and M. R. Aref, "Multiple access wiretap channels with strong secrecy," in *Proc. IEEE Info. Theory Workshop (ITW)*, Dublin, Ireland, Sep. 2010, pp. 1–5.

[43] E. C. Song, P. Cuff, and H. V. Poor, "A rate-distortion based secrecy system with side information at the decoders," in *Proc. 52th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, Sep. 2014, pp. 755–762.

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

Hassan ZivariFard is a Ph.D. candidate in electrical engineering department at the University of Texas at Dallas. Prior to this, he received a M.Sc. degree from K. N. Toosi University of Technology in Electrical Engineering, Iran, in 2012. His research interests include information theory and physical layer security.

Matthieu R. Bloch is a Professor in the School of Electrical and Computer Engineering. He received the Engineering degree from Supélec, Gif-sur-Yvette, France, the M.S. degree in Electrical Engineering from the Georgia Institute of Technology, Atlanta, in 2003, the Ph.D. degree in Engineering Science from the Université de Franche-Comté, Besançon, France, in 2006, and the Ph.D. degree in Electrical Engineering from the Georgia Institute of Technology in 2008. In 2008-2009, he was a postdoctoral research associate at the University of Notre Dame, South Bend, IN. Since July 2009, Dr. Bloch has been on the faculty of the School of Electrical and Computer Engineering, and from 2009 to 2013 Dr. Bloch was based at Georgia Tech Lorraine. His research interests are in the areas of information theory, error-control coding, wireless communications, and cryptography. Dr. Bloch has served on the organizing committee of several international conferences; he was the chair of the Online Committee of the IEEE Information Theory Society from 2011 to 2014, an Associate Editor for the IEEE Transactions on Information Theory from 2016 to 2019, and he has been on the Board of Governors of the IEEE Information Theory Society since 2016 and currently serves as the 2nd Vice-President. He has been an Associate Editor for the IEEE Transactions on Information Forensics and Security since 2019. He is the co-recipient of the IEEE Communications Society and IEEE Information Theory Society 2011 Joint Paper Award and the co-author of the textbook Physical-Layer Security: From Information Theory to Security Engineering, published by Cambridge University Press.

Aria Nosratinia is Erik Jonsson Distinguished Professor and associate head of the electrical engineering department at the University of Texas at Dallas. He received his Ph.D. in Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign in 1996. He has held visiting appointments at Princeton University, Rice University, and UCLA. His interests lie in the broad area of information theory and signal processing, with applications in wireless communications, machine learning, and data security and privacy. Dr. Nosratinia is a fellow of IEEE for contributions to multimedia and wireless communications. He has served as editor and area editor for the IEEE Transactions on Wireless Communications, and editor for the IEEE Transactions on Information Theory, IEEE Transactions on Image Processing, IEEE Signal Processing Letters, IEEE Wireless Communications (Magazine), and Journal of Circuits, Systems, and Computers. He has received the National Science Foundation career award, and the outstanding service award from the IEEE Signal Processing Society, Dallas Chapter. He has served as the secretary of the IEEE information theory society, treasurer for ISIT, publications chair for the IEEE Signal Processing Workshop, as well as member of the technical committee for a number of conferences. He was the general co-chair of IEEE Information Theory Workshop in 2018. Dr. Nosratinia is a registered professional engineer in the state of Texas and a highly cited researcher.