The Sender-Excited Secret Key Agreement Model: Capacity Theorems

Tzu-Han Chou*, Vincent Y. F. Tan*†, Stark C. Draper*

* Dept. of ECE, University of Wisconsin-Madison, Email: {tchou2@,vtan@,sdraper@ece.}wisc.edu
† LIDS, Massachusetts Institute of Technology, Email: vtan@mit.edu

***Abstract*—We consider the fundamental limits of the secret key generation problem when the sources are excited by the sender. In many practical communication settings, the channel may be influenced by the parties involved. Similar to recent works on probing capacity and channels with action-dependent states, our system model captures such a scenario. We derive single-letter expressions for the secret key capacity. Our coding strategy involves wiretap channel coding and a key generation scheme. By assuming that the eavesdropper receives a degraded version of the legitimate receiver's observation, we also obtain a capacity result that does not involve any auxiliary random variables. Finally, by studying two examples, we show that there is a fundamental tradeoff in between the amount of common randomness (i.e., the secret key rate) and the wiretap secrecy rate.**

***Index Terms*—Secret key capacity, Sender excitation, Probing capacity, Degraded broadcast channel**

I. Introduction

Within the realm of information theoretic secrecy, the foundations of sharing a secret key between two parties in the presence of an eavesdropper were initiated in [1, 2]. Ahlswede and Csiszár [1] studied two models: the *source-type model with wiretapper* (Model SW) and the *channel-type model with wiretapper* (Model CW). In Model SW, the users obtain their observations from a discrete memoryless multiple source (DMMS) and communicate to each other via a noiseless authenticated public channel. The public messages they send can be regarded as compressed versions of the data in a multi-terminal source coding problem. The information that is independent of the public message can be used to generate secret keys. In Model CW, one legitimate user (the sender) controls the input of a discrete memoryless broadcast channel (DMBC), sending information to the legitimate receiver and the eavesdropper. The sender randomly chooses a message and transmits it to receiver. Users may also discuss over a public channel and generate a key based on all the available data sent to them by the other party. It is shown that when one-way public discussion is allowed, the users can adopt wiretap channel coding [3, 4] without the public channel and there is no loss of the secret key rate. These ideas were extended in [5] in which the key generation with helper problem was studied.

However, in many applications, the system is neither a source-type nor a channel-type model. This work explores such a setting. We also derive capacity results for the secret key agreement problem when the sender can control the "state" of the channel in the same spirit as the works on probing capacity and channels with action-dependent states [6–8].

A. Related Work

There are other works dealing with non-source and non-channel models such as [9, 10], where users observe a DMMS and they can also transmit information via a wiretap channel. However, no public discussion is allowed. The key generation scheme is based on the observation that the public message, which assists in generating the key, can be transmitted via the DMBC confidentially, resulting a higher secret key rate.

The authors in [11–14] considered the setting where a wiretap channel is influenced by a random state that is known at the sender (and possibly the receiver) either causally or noncausally and thus can be treated as a correlated source. In [11, 12], the sender transmits a confidential message and the random state is exploited in the coding scheme to confuse the eavesdropper. The lower bound is proved using a combination of Gel'fand-Pinsker coding and wiretap channel coding. In [13], the goal was to generate a secret key when the encoder and decoders have noncausal state information. The authors presented a single-letter expression of the secret key capacity. The resulting key rate consists of two parts; the first is attributed to the rate of the confidential message using wiretap channel coding while treating the state sequence as a time-sharing sequence (multiplexing), while the second key, independent of the first one, is produced by exploiting the common knowledge of the state at the sender and the legitimate receiver. A similar problem with causal state information was studied in [14] and the coding scheme involves block Markov coding, Shannon strategy and wiretap coding.

Another motivation of this paper comes from that fact that in many applications such as storage for computer memories, the system (channel) may be influenced by a *probing* signal that is influenced by some of the users (typically the sender). This problem was first studied in the channel coding context [6] where the channel state of the DMBC depends on the encoder's action sequence, which in turn depends on the message the sender intends to send. As a result, the channel is one whose states are "action-dependent". In [7], the availability of the states at the encoder is controlled by a probing (action) signal, which is subject to a cost constraint. Similar action-dependent ideas were studied in [8] in the

source coding context. However, the models studied in [6], [7] and [8] do not incorporate any secrecy constraints.

In another line of research, Chou et al. [15] studied the problem of secret key generation from a DMMS with one-way public discussion. The DMMS studied is influenced by the inherent randomness of a DMBC that is excited by a deterministic external sounding signal. Capacity and reliability results were derived.

B. Our Contributions

In this paper, we consider a system with the model shown in Fig. 1. The users obtain correlated sources via the outputs of a DMBC $p(x, y, z|s)$, where the input S, controlled by the sender, is subject to an input cost constraint. This generalizes the model of [15] in the sense that the input sounding signal can be randomly selected by the sender based on her private source of randoness and the information of the chosen sounding sequence is protected by using wiretap channel coding. This allows us to optimize over the distribution of the sounding signal to maximize the secrecy key rate. We give a single-letter expression for the secret key capacity of this system. The capacity-achieving coding scheme is one in which the optimal tradeoff between two coding strategies has to be found: (1) Treat the DMBC $p(y, z|s)$ as a wiretap channel [3, 4] and apply wiretap channel coding, and (2) Treat the channel outputs (X, Y, Z) as excited correlated sources and use key generation scheme (as [1]) to extract the key. We demonstrate this tradeoff by using an example in which the channel is degraded in favor of the legitimate receiver.

C. Paper Organization

This paper is organized as follows: In Section II, we describe the system model and define the notion of degradedness (of the eavesdropper). Our main results pertaining to the secret key capacity are provided in Section III. We also prove a (looser) upper bound for the secret key capacity that does not contain any auxiliary random variables. We show that this upper bound is in fact tight for degraded channels. In Section IV, we present two examples to demonstrate how the preceding theorems can be applied to channels of interest. We show that there is an inherent tradeoff between the wiretap rate and the key rate. We conclude our discussion by suggesting avenues for future research in Section V. The proofs of the main results are provided in Section VI.

II. Problem Setup

We adopt the notational conventions employed in the lecture notes by El Gamal and Kim [16]. The setting is depicted in Fig. 1. Consider a DMBC $(\mathcal{S}, p(x, y, z|s), \mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ consisting of four finite sets $\mathcal{S}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$ and a collection of conditional pmfs $p(x, y, z|s)$ on $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. The sender, Alice at terminal $\mathcal{X}$, controls the channel input *sounding signal* s^n via n uses of the channel. Alice has a private source of randomness used to select an index m, which influences s^n. The legitimate receiver at terminal $\mathcal{Y}$ is known as Bob and the eavesdropper at terminal $\mathcal{Z}$ is known as Eve. There is also a noiseless public discussion channel which allows Alice to

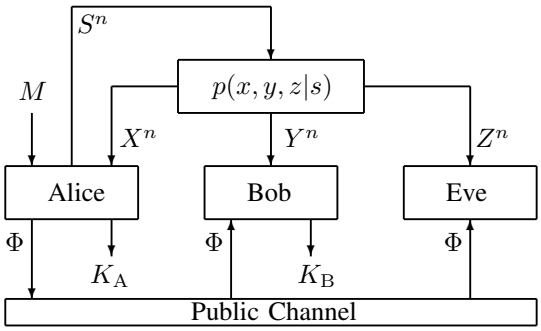

Fig. 1. Our problem setup: Based on her private source of randomness M, Alice excites the channel via the sounding signal $S^n(M)$. She generates a public message $\Phi(M, X^n)$, which is transmitted through the noiseless public channel and hence known to all parties. Alice and Bob generate keys $K_{\mathrm{A}}(M, X^n)$ and $K_{\mathrm{B}}(\Phi, Y^n)$ respectively. The keys should agree, while at the same time, they should be kept secret from Eve.

transmit a message Φ to Bob and Eve. A $(2^{nR_M}, 2^{nR_\Phi}, n)$ *code* for the secret key generation protocol consists of:

1) *Channel Excitation*: Alice first selects an message $m \in [1 : 2^{nR_M}]$. Then, she chooses a message-dependent input sequence $s^n = s^n(m)$ such that every input codeword to the channel satisfies the cost constraint
$$\Lambda(s^n(m)) = \frac{1}{n} \sum_{i=1}^{n} \Lambda(s_i(m)) \leq \Gamma \,, \tag{1}$$
where $\Lambda : \mathcal{S} \to \mathbb{R}^+$ is the input cost function. The sequence s^n is transmitted over n channel uses. The output sequences x^n, y^n and z^n are observed by Alice, Bob (legitimate receiver) and Eve (eavesdropper) respectively.
2) *(One-Way) Public Discussion*: After observing x^n, Alice generates a one-way public message $\phi = \phi(m, x^n) \in [1 : 2^{nR_\Phi}]$, and transmits it over a noiseless public channel.
3) *Key Generation*: Alice generates a key $k_{\mathrm{A}} = k_{\mathrm{A}}(m, x^n) \in \mathbb{N}$. After receiving his channel output y^n and the public message ϕ, Bob generates another key $k_{\mathrm{B}} = k_{\mathrm{B}}(y^n, \phi) \in \mathbb{N}$.

Definition 1 (Achievability)**.** *The secret key rate* $R_{\mathrm{SK}}(\Gamma)$ *is* achievable *if there exists a sequence of* $(2^{nR_M}, 2^{nR_\Phi}, n)$ *codes for the secret key generation protocol such that for every* $\epsilon > 0$*, the following constraints are satisfied:*
$$\mathsf{P}(K_{\mathrm{A}} \neq K_{\mathrm{B}}) < \epsilon \tag{2}$$
$$\frac{1}{n} I(K_{\mathrm{A}}; Z^n, \Phi) < \epsilon \tag{3}$$
$$\frac{1}{n} H(K_{\mathrm{A}}) > R_{\mathrm{SK}}(\Gamma) - \epsilon \tag{4}$$
for all $n \geq n_0(\epsilon, |\mathcal{S}|, |\mathcal{X}|, |\mathcal{Y}|, |\mathcal{Z}|)$.

Definition 2 (Secret Key Capacity)**.** *The* secret key capacity $C_{\mathrm{SK}}(\Gamma)$ *is the supremum of all achievable secret key rates.*

The first constraint on the code in (2) implies that we would like Alice's and Bob's keys to agree with high probability. The second constraint in (3), known as the *secrecy*

condition, requires that the eavesdropper should not be able to estimate the key $K_{\rm A}$ given sequence Z^n and public message Φ. This is manifested in fact that the normalized mutual information should be arbitrarily small so K_A and (Z^n, Φ) are statistically independent asymptotically. Finally, the rate condition in (4) implies that the entropy of $K_{\rm A}$ should be close to $R_{\rm SK}(\Gamma)$. In other words the pmf of $K_{\rm A}$ should be close to that of a uniform pmf supported on $[1 : 2^{nR_{\rm SK}(\Gamma)}]$.

Note the conditional distribution of $(X, Y, Z|S)$ can be factorized as $p(x|s)p(y, z|x, s)$. The first conditional distribution $p(x|s)$ can be thought of as Alice's influence on the channel state via the sounding signal s^n, while the second $p(y, z|x, s)$ can be thought of as a state-dependent channel. The variables S, X are available at Alice but she can only control the sounding signal S, which in turn triggers the channel. As mentioned in the Introduction, the model we study in this paper involves a probing mechanism. This is analogous to the model studied in [6, 7], in which the channel is influenced by a sequence of actions but there is no secrecy requirement. The main difference from [6, 7] is that in our model, we consider only one DMBC $p(x, y, z|s)$, thus the chosen channel input s^n does not depend on the observation x^n. However, the sender Alice uses *both* x^n and s^n to generate a key $k_{\rm A}$ in the subsequent public messaging step. A special class of channels in which Eve's observation is a degraded version of Bob's is defined as follows.

Definition 3 (Degradedness)**.** *We say that the DMBC $p(x, y, z|s)$ is* degraded *if $(X, S) - Y - Z$ form a Markov chain, i.e., $p(y, z|x, s) = p(y|x, s)p(z|y)$.*

Note that we do not differentiate between physical and stochastic degradedness [16, Ch. 5]. The capacity results will turn out to be identical for both cases.

III. Main Results

We present our main results in this section. We give a single-letter expression for the secret key capacity containing three auxiliary random variables. We also provide a looser upper bound in which there are no auxiliary random variables in the expression. The upper bound is tight when the system is such that the channel is degraded in favor of Bob, the legitimate receiver (as per Definition 3).

Theorem 1 (Secret Key Capacity)**.** *The secret key capacity of DMBC $(\mathcal{S}, p(x, y, z|s), \mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ is*

$$C_{\rm SK}(\Gamma) = \max\ [I(U, V; Y|W) - I(U, V; Z|W)] \ , \quad (5)$$

where the maximization is over joint distributions of the form

$$p(w, u, v, s, x, y, z) = p(u, w)p(s|u)p(v|w, u, x)p(x, y, z|s) \quad (6)$$

for some $p(u, w), p(s|u), p(v|w, u, x)$ such that $\mathsf{E}[\Lambda(S)] \leq \Gamma$.

Lemma 2 (Properties of $C_{\rm SK}(\Gamma)$)**.** *The function $C_{\rm SK} : (0, \infty) \to \mathbb{R}^+$ is non-decreasing, concave and continuous.*

The converse and the direct parts of Theorem 1 are detailed in Sections VI-A and VI-B respectively. We provide an operational proof of Lemma 2 in Section VI-C. Note that the joint distribution factorizes as in (6) if and only if the following Markov chains hold:

$$W - U - S - (X, Y, Z) \ , \quad (7)$$

$$V - (W, U, X) - (S, Y, Z) \ . \quad (8)$$

Furthermore, observe that the rate in (5) can be written as sum of two rates $R_{\rm SK} = R_{\rm ch} + R_{\rm src}$ where

$$R_{\rm ch} = I(U; Y|W) - I(U; Z|W) \ ,$$
$$R_{\rm src} = I(V; Y|W, U) - I(V; Z|W, U) \ .$$

The first rate $R_{\rm ch}$ can be interpreted as the confidential message rate of the wiretap channel $p(y, z|s)$ [4]. The second rate $R_{\rm src}$ is the secret key rate from excited correlated source (X, Y, Z) previously studied in [15] for a particular deterministic sounding signal s^n. Here the sounding signal S^n is randomly chosen by Alice based on her private source of randomness. As such, we can optimize over the distribution in (6) to find the largest "sum rate" $R_{\rm ch} + R_{\rm src}$. It turns out that there is a natural tradeoff between $R_{\rm ch}$ and $R_{\rm src}$. We see this using an example in Section IV-B.

To find the secret key capacity for specific channels, three auxiliary random variables W, U and V solving (5) and satisfying (7) and (8) have to be identified. This may be a difficult task. In the next proposition, we provide an (albeit looser) upper bound which does not involve any auxiliary random variables. This result will turn out to be important in Section IV in which we present several channel models for which we identify the secret key capacities in closed-form.

Proposition 3 (Upper Bound in Secret Key Capacity)**.** *The secret key capacity is upper bounded by*

$$C_{\rm SK}(\Gamma) \leq \max\ I(X, S; Y|Z) \ , \quad (9)$$

where the maximization is over all input distributions $p(s)$ such that $\mathsf{E}[\Lambda(S)] \leq \Gamma$.

The proof of this proposition is given in Section VI-D. Roughly speaking, the expression in (9) can be interpreted as the secret key capacity when Alice and Bob have full knowledge (side information) of Eve's observation Z, hence the conditioning on Z. Indeed, in the case of degraded $p(x, y, z|s)$, the result in Proposition 3 is tight.

Corollary 4 (Secret Key Capacity of Degraded Channels)**.** *If the DMBC $p(x, y, z|s)$ is* degraded*, the secret key capacity is*

$$C_{\rm SK}(\Gamma) = \max\ [I(X, S; Y) - I(X, S; Z)] \ ,$$

where the maximization is over all input distributions $p(s)$ such that $\mathsf{E}[\Lambda(S)] \leq \Gamma$.

Proof: Let S have distribution $p(s)$ that achieves the upper bound in Proposition 3. The secret key capacity of the degraded DMBC can be upper bounded as

$$\begin{aligned} C_{\rm SK}(\Gamma) &\leq I(X, S; Y|Z) = I(X, S; Y, Z) - I(X, S; Z) \\ &= I(X, S; Y) - I(X, S; Z) \ . \end{aligned} \quad (10)$$

The last equality is due to the fact that $(X, S) - Y - Z$ form a Markov chain. On the other hand, the upper bound (10)

can be achieved by the specific choice of $W = \emptyset, U = S$, and $V = X$ in (5). □

IV. Examples

We consider two examples in this section. The first is an additive Gaussian interference channel where the interference of Alice and Bob is correlated. The second example is a binary on-off channel in which Eve receives a degraded version of Bob's (or Alice's) output.

A. Additive Gaussian Interference Channel

Consider the channel model

$$\begin{aligned} X &= S + I_1 + N_1 \\ Y &= S + I_2 + N_2 \\ Z &= S + I_3 + N_3 \end{aligned}$$

where $N_i, i = 1, 2, 3$, are independent Gaussian noises distributed as $\mathcal{N}\left(0, \sigma_i^2\right)$. The random variables $I_i, i = 1, 2, 3$ are distributed as $\mathcal{N}\left(0, \nu_i^2\right)$ and model interference at each receiver. The interferences I_i are independent of N_i. It is assumed that I_1, I_2, I_3 are jointly Gaussian with $\mathsf{E}[I_i I_j] = \rho_{ij}\nu_i\nu_j$ where $\rho_{ij} \in (-1, 1)$ is the correlation coefficient. It is further assumed that the channel is degraded in favor of Bob in the sense that $\nu_3^2 + \sigma_3^2 \geq \nu_2^2 + \sigma_2^2$. The input sequence S^n is subject to an average power constraint P, i.e., $\Lambda(s) = s^2$ and $\Gamma = P$.

By degradedness, we can define $Z' \triangleq Y + N_3'$ where N_3' is independent and distributed as $\mathcal{N}\left(0, \nu_3^2 + \sigma_3^2 - \nu_2^2 - \sigma_2^2\right)$. Note that $(X, S) - Y - Z'$ forms a Markov chain and Z' has the same marginal distribution as Z. Since (3) only depends on the marginal distribution $p(x, z|s)$, from Corollary 4 the secret key capacity is

$$C_{\mathrm{SK}} = \max_{p(s):\mathsf{E}[S^2]\leq P} I(X, S; Y) - I(X, S; Z) = R_{\mathrm{ch}} + R_{\mathrm{src}} \ .$$

Note that in this case, R_{src} is not a function of the input distribution $p(s)$. The optimal input distribution is $S \sim \mathcal{N}(0, P)$, which is same as that in the Gaussian wiretap channel [17]. Define $\mathrm{C}_0(x) \triangleq \frac{1}{2}\log(1+x)$ as the AWGN channel capacity with signal-to-noise ratio (SNR) x and $\mathrm{C}_1(\rho, \boldsymbol{\nu}_{ij}, \boldsymbol{\sigma}_{ij})$ as

$$\mathrm{C}_1(\rho, \boldsymbol{\nu}_{ij}, \boldsymbol{\sigma}_{ij}) \triangleq \mathrm{C}_0\left(\frac{\rho^2\nu_i^2\nu_j^2}{(\nu_i^2+\sigma_i^2)(\nu_j^2+\sigma_j^2) - \rho^2\nu_i^2\nu_j^2}\right) \ ,$$

where the parameters $\boldsymbol{\nu}_{ij}$ and $\boldsymbol{\sigma}_{ij}$ are defined as $\boldsymbol{\nu}_{ij} \triangleq (\nu_i, \nu_j)$ and $\boldsymbol{\sigma}_{ij} \triangleq (\sigma_i, \sigma_j)$. With these definitions, R_{ch} and R_{src} can be calculated as

$$\begin{aligned} R_{\mathrm{ch}} &= I(S; Y) - I(S; Z) \\ &= \mathrm{C}_0\left(\frac{P}{\nu_2^2+\sigma_2^2}\right) - \mathrm{C}_0\left(\frac{P}{\nu_3^2+\sigma_3^2}\right) \ , \\ R_{\mathrm{src}} &= I(X; Y|S) - I(X; Z|S) \\ &= \mathrm{C}_1(\rho_{12}, \boldsymbol{\nu}_{12}, \boldsymbol{\sigma}_{12}) - \mathrm{C}_1(\rho_{13}, \boldsymbol{\nu}_{13}, \boldsymbol{\sigma}_{13}) \ . \end{aligned}$$

Note that, in this case, R_{src} depends on the correlation between interference and is not a function of P. When we increase the input power P, only R_{ch} increases.

The secret key capacity for the specific choice of parameters $\rho_{12} = 0.8, \rho_{13} = 0.3$, $\nu_1 = \nu_2 = 1, \nu_3 = 2$, and $\sigma_i = 1$ for all $i \in \{1, 2, 3\}$ is plotted against the input power P in Fig. 2. As Lemma 2 suggests, $C_{\mathrm{SK}}(P)$ is non-decreasing and concave[1] in P. When the allowed input power P is small, extracting common secret information from correlated interference is important, evidenced by the fact that $R_{\mathrm{src}} \geq R_{\mathrm{ch}}$ at low power levels P. On the other hand, when the input power is large, we can simply use the wiretap channel $p(y, z|s)$ without any significant loss of rate.

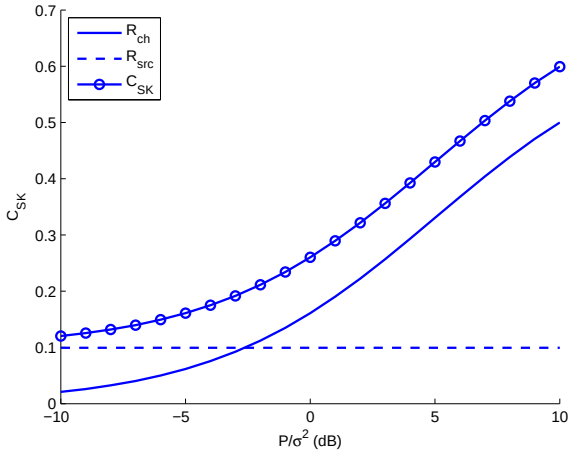

Fig. 2. Secret key capacity of the Gaussian additive interference channel. $\rho_{12} = 0.8, \rho_{13} = 0.3$. $\nu_1 = \nu_2 = 1, \nu_3 = 2$. $\sigma_i = 1$ for all $i \in \{1, 2, 3\}$.

B. Binary On-off Channel

In our second example, we consider the binary on-off model

$$\begin{aligned} X &= H \cdot S \oplus N_1 \\ Y &= H \cdot S \oplus N_2 \\ Z &= (\tilde{H} \cdot H) \cdot S \oplus N_3 \ , \end{aligned}$$

where all the variables are binary and where the operations are performed in the field of size 2. Hence, the addition above is is binary modulo-2 addition. The "channel gain" H is $\mathrm{Bern}\,(q)$ and $\tilde{H}$ is $\mathrm{Bern}\,(\tilde{q})$. Noise N_i is $\mathrm{Bern}\,(\delta_i)$ and the N_i are mutually independent. The channel describes a model in which, in the absence of noise, Eve's observation is strictly worse than that of Alice's and Bob's since $\tilde{H}$ is present.

If $\delta_1 = \delta_2 = \delta$ and $\tilde{q}\delta < \delta_3$, then Eve's channel output is a degraded version of Bob's. In this case, there exists a $Z' \triangleq \tilde{H}' \cdot Y \oplus N_3'$ for some $\tilde{H}'$, with the same distribution as $\tilde{H}$, and independent $N_3' \sim \mathrm{Bern}\,(\delta_3')$ such that $(X, S) - Y - Z'$, where

$$\delta_3' = \frac{\delta_3 - \tilde{q}\delta}{1 - 2\tilde{q}\delta} \ .$$

Let $S \sim \mathrm{Bern}\,(\alpha)$. The first term of R_{ch} is

$$\begin{aligned} &I(S; Y) = H(Y) - H(Y|S) \\ &= H_{\mathrm{b}}(\alpha q * \delta) - [\alpha H(Y|S=1) + (1-\alpha)H(Y|S=0)] \\ &= H_{\mathrm{b}}(\alpha q * \delta) - \alpha H_{\mathrm{b}}(q * \delta) - (1-\alpha)H_{\mathrm{b}}(\delta) \ , \end{aligned}$$

[1]The concavity would be more apparent if the horizontal axis of Fig. 2 is linear but we find that it is more convenient to plot it in dB.

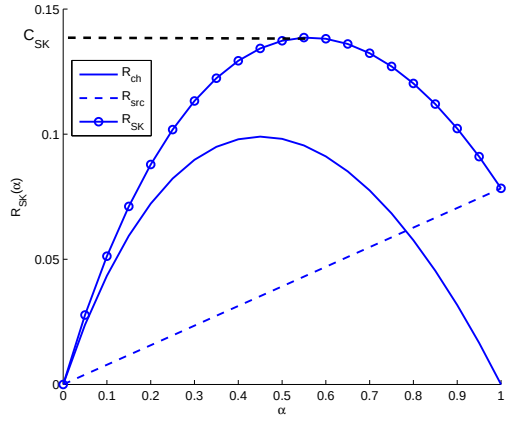

Fig. 3. Secret key rate of the binary on-off channel as a function of α. The input $S \sim \text{Bern}(\alpha)$. The parameters are $q = 0.5, \tilde{q} = 0.8$, $\delta = 0.1, \delta_3 = 0.2$. Note that $C_{\text{SK}} = \max_{\alpha \in [0,1]} R_{\text{SK}}(\alpha)$ and the maximizing $\alpha^* \approx 0.59$.

where $H_{\text{b}}(\cdot)$ is the binary entropy function and the operation $a * b \triangleq a(1-b) + (1-a)b$. Similarly, the second term of R_{ch} can be expressed as

$$I(S;Z) = H_{\text{b}}(\alpha\tilde{q}q * \delta_3) - \alpha H_{\text{b}}(\tilde{q}q * \delta_3) - (1-\alpha)H_{\text{b}}(\delta_3) \ .$$

The secret key rate due to source X can be calculated as

$$\begin{aligned} R_{\text{src}} &= I(X;Y|S) - I(X;Z|S) \\ &= \alpha[I(X;Y|S=1) - I(X;Z|S=1)] \\ &= \alpha[H_{\text{b}}(q * \delta) - H_{\text{b}}(\delta * \delta) - H_{\text{b}}(\tilde{q}q * \delta_3) \\ &\qquad + (1 - q * \delta)H_{\text{b}}(\delta_3') + (q * \delta)H_{\text{b}}(\tilde{q} * \delta_3')] \ . \end{aligned}$$

The second equality is because if $S = 0$, the source is not observed and so there is no mutual information between X and Y (and also between X and Z).

The secret key rate when the input is a $\text{Bern}(\alpha)$ source is $R_{\text{SK}}(\alpha) = R_{\text{ch}}(\alpha) + R_{\text{src}}(\alpha)$ and is shown in Figure 3 as a function of α for the set of parameters $q = 0.5, \tilde{q} = 0.8$, $\delta = 0.1, \delta_3 = 0.2$. Note that R_{ch} is a concave function of α and R_{src} is a linear function of α. Also, if $\alpha = 0$, then $R_{\text{src}} = 0$ in this example. In contrast, R_{src} is positive at all powers for the additive Gaussian interference channel. When $\alpha = 1$ (S^n is the all ones sequence), the input excites all common randomness due to the *common* on-off coefficient H. However, when $\alpha = 1$, the secrecy rate of the wiretap channel $R_{\text{ch}} = 0$. Thus, we observe that there is a inherent tradeoff in between the amount of common randomness and the wiretap secrecy rate.

V. Extensions

In this paper, we derived capacity theorems for the secret key agreement problem when the sender excites the channel model. Several questions arise naturally from this work:

- What happens if we place a constraint on the rate of the public channel as was done in [5]? In fact, a lower bound for the capacity of this problem can be derived based on the coding scheme described in Section VI-B. The upper bound is still an open problem.
- What changes if we allow Alice and Bob to communicate over multiple rounds, i.e, multi-way public discussion is permitted?
- Can we derive reliability and secrecy exponents along the lines of [15]?
- Can we derive upper and/or lower bounds for the secret key capacity when the transition probability is state-dependent, i.e., when it is a function of an underlying state s_0 so the transition probability is $p(x, y, z|s, s_0)$? The state s_0 assumed to be known at the encoder.

VI. Proofs of Theorems

A. Proof of Converse of Theorem 1

We start with a lemma [1, Lemma 4.1], which is a consequence of the Csiszár sum identity [16, Ch. 2].

Lemma 5. *The following equality holds for arbitrary random variables K, Φ, Y^n, Z^n:*

$$\begin{aligned} &I(K;Y^n|\Phi) - I(K;Z^n|\Phi) \\ &\quad = \sum_{i=1}^{n} I(K;Y_i|Y^{i-1}, Z_{i+1}^n, \Phi) - I(K;Z_i|Y^{i-1}, Z_{i+1}^n, \Phi) \ . \end{aligned}$$

The converse follows from the following steps:

$$\begin{aligned} nR_{\text{SK}} &\overset{(a)}{\le} I(K_{\text{A}};Y^n,\Phi) + n\epsilon_n \\ &\overset{(b)}{\le} I(K_{\text{A}};Y^n,\Phi) - I(K_{\text{A}};Z^n,\Phi) + 2n\epsilon_n \\ &= I(K_{\text{A}};Y^n|\Phi) - I(K_{\text{A}};Z^n|\Phi) + 2n\epsilon_n \\ &\overset{(c)}{=} \sum_{i=1}^{n} I(K_{\text{A}};Y_i|Y^{i-1}, Z_{i+1}^n, \Phi) \\ &\qquad - I(K_{\text{A}};Z_i|Y^{i-1}, Z_{i+1}^n, \Phi) + 2n\epsilon_n \\ &\overset{(d)}{=} \sum_{i=1}^{n} I(K_{\text{A}};Y_i|W_i) - I(K_{\text{A}};Z_i|W_i) + 2n\epsilon_n \\ &\overset{(e)}{=} \sum_{i=1}^{n} I(U_i,V_i;Y_i|W_i) - I(U_i,V_i;Z_i|W_i) + 2n\epsilon_n \ , \end{aligned}$$

where (a) is due to Fano's inequality ($\epsilon_n \to 0$ as $n \to \infty$), (b) is due to the secrecy condition (3), (c) by applying Lemma 5, (d) follows from defining the auxiliary random variable $W_i \triangleq (Y^{i-1}, Z_{i+1}^n, \Phi)$, (e) follows by defining the auxiliary random variables $U_i \triangleq (K_{\text{A}}, W_i)$ and $V_i \triangleq K_{\text{A}}$. The chosen variables W_i, U_i, V_i satisfy the Markov conditions $W_i - U_i - (Y_i, Z_i)$ and $V_i - (W_i, U_i, X_i) - (Y_i, Z_i)$ as required in (7) and (8). Finally, by using the definition and concavity of $C_{\text{SK}}(\Gamma)$ (see Lemma 2), we have

$$\begin{aligned} nR_{\text{SK}} &\le \sum_{i=1}^{n} C_{\text{SK}}(\mathsf{E}[\Lambda(S_i)]) + 2n\epsilon_n \\ &\le n \sum_{i=1}^{n} C_{\text{SK}}\left(\frac{1}{n}\mathsf{E}[\Lambda(S_i)]\right) + 2n\epsilon_n \le n[C_{\text{SK}}(\Gamma) + 2\epsilon_n]. \end{aligned}$$

This completes the proof of the converse. □

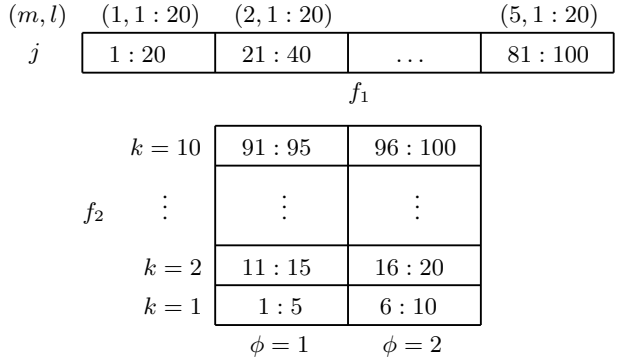

Fig. 4. Illustration of the functions f_1 and f_2 for $|\mathcal{M}| = 5$, $|\mathcal{L}| = 20$, $|\mathcal{J}| = 100$, $|\Phi| = 2$ and $|\mathcal{K}| = 10$. In the top figure, the function f_1 unwraps the $|\mathcal{M}|\times|\mathcal{L}| = |\mathcal{J}| = 100$ indices into a single row. In the bottom figure, the function f_2 partitions the 100 indices into a two-dimensional grid of bins where each bin contains $T = |\mathcal{J}|/(|\Phi||\mathcal{K}|) = 5$ indices.

B. Proof of Achievability of Theorem 1

We use the notion of typicality in [16, Ch. 2]. The typical set is denoted as $\mathcal{T}_\varepsilon^{(n)}$. Fix $\varepsilon > \varepsilon' > \varepsilon'' > 0$ and also fix the distributions $p(u,w), p(s|u), p(v|w,u,x)$ that achieve $C_{\mathrm{SK}}(\frac{\Gamma}{1+\varepsilon})$ in (5). By marginalization, this choice of distributions induces $p(u,w), p(s|u)$ and $p(v|u,w)$. We first prove achievability for $W = \emptyset$. At the end, we generalize the result.

1) Codebook generation: Define the five index sets:

$$\begin{aligned}
\mathcal{M} &\triangleq [1 : 2^{n(I(U;Y)-2\delta)}] \\
\mathcal{L} &\triangleq [1 : 2^{n(I(V;X|U)+\delta)}] \\
\Phi &\triangleq [1 : 2^{n(I(V;X|U)-I(V;Y|U)+2\delta)}] \\
\mathcal{K} &\triangleq [1 : 2^{n(I(U,V;Y)-I(U,V;Z))}] \\
\mathcal{J} &\triangleq [1 : 2^{n(I(V;X|U)+I(U;Y)-\delta)}] \ .
\end{aligned} \tag{11}$$

Note that $|\mathcal{J}| = |\mathcal{M}||\mathcal{L}|$. The set $\mathcal{K}$ represents the alphabet of Alice's and Bob's key. The secret key rate is $\frac{1}{n}\log|\mathcal{K}|$ [compare to (5) with $W = \emptyset$].

Randomly and independently generate $2^{n(I(U;Y)-2\delta)}$ sequence pairs $(u^n(m), s^n(m)), m \in \mathcal{M}$ drawn according to $\prod_{i=1}^n p(u_i, s_i)$. For each $m \in \mathcal{M}$, randomly and conditionally independently generate $2^{n(I(V;X|U)+\delta)}$ sequences $v^n(m,l), l \in \mathcal{L}$ according to $\prod_{i=1}^n p(v_i|u_i(m))$.

Let $f : \mathcal{M}\times\mathcal{L} \to \Phi\times\mathcal{K}$ be a *deterministic* binning function defined as follows. First, let $f_1 : \mathcal{M}\times\mathcal{L} \to \mathcal{J}$ be the function[2] defined as $f_1(m,l) \triangleq (m-1)|\mathcal{L}| + l$. Second, let $f_2 : \mathcal{J} \to \Phi\times\mathcal{K}$ be a function that induces a partition of the indices in $\mathcal{J}$ such that each *sub-bin*, doubly-indexed by (ϕ, k), contains an equal number of (m,l) pairs, namely[3] $T = |\mathcal{J}|/(|\Phi||\mathcal{K}|) \doteq 2^{n(I(U,V;Z)-3\delta)}$. More precisely, we may set $\phi \in \Phi$ and $k \in \mathcal{K}$ as follows: $f_2(j) = (\phi, k)$ if $k = \lceil j/|\mathcal{K}|\rceil$ and $\phi = \lceil((j-1) \bmod |\mathcal{K}|)/T\rceil$. Then, we define the composite function $f(m,l) \triangleq f_2(f_1(m,l))$. See Fig. 4 for an illustration of the function f. In addition, if $f(m,l) = (\phi, k)$, we also define the two (projection) maps $\phi(m,l) \triangleq \phi$ and $k(m,l) \triangleq k$. These are called the first bin (public message) and second bin (key) respectively. Define

$$\begin{aligned}
\mathcal{B}(\phi) &\triangleq \{(u^n(m), v^n(m,l)) : \phi(m,l) = \phi\} \\
\mathcal{C}(k) &\triangleq \{(u^n(m), v^n(m,l)) : k(m,l) = k\} \ ,
\end{aligned}$$

to be the set of pairs of sequences with first bin index equal to ϕ and the set of pairs of sequences with second bin index equal to k respectively. Note that the number of pairs of sequences in $\mathcal{B}(\phi)$ is $|\mathcal{B}(\phi)| = |\mathcal{J}|/|\Phi| \doteq 2^{n(I(U,V;Y)-3\delta)}$. It can similarly verified that $|\mathcal{C}(k)| \doteq 2^{n(I(V;X|U)-I(V;Y|U)+I(U,V;Z)-\delta)}$.

2) Encoding: Alice selects $m \in \mathcal{M}$ uniformly at random, sets $u^n = u^n(m)$ and selects inputs $s^n = s^n(m)$ to channel. Let the event that the codewords are atypical be

$$\mathcal{E}_0 \triangleq \{(u^n, s^n) \notin \mathcal{T}_{\varepsilon''}^{(n)}\} \cup \{\Lambda(s^n) > \Gamma\} \ .$$

After observing the channel output x^n, Alice finds an $l \in \mathcal{L}$ such that $(v^n(m,l), u^n, x^n) \in \mathcal{T}_{\varepsilon''}^{(n)}$. If no such index is found, declare $l = 1$. Define

$$\mathcal{E}_1 \triangleq \{(v^n(m,l), u^n, x^n) \notin \mathcal{T}_{\varepsilon''}^{(n)}, \forall\, l \in \mathcal{L}\}$$

to be the encoding error event. Alice generates the public message as $\phi = \phi(m,l)$ and sets her key $k_{\mathrm{A}} = k(m,l)$, where the functions ϕ and k are defined in the codebook generation step.

3) Decoding: Bob receives y^n from the channel output and finds a unique $\hat{m} \in \mathcal{M}$ such that $(u^n(\hat{m}), y^n) \in \mathcal{T}_{\varepsilon'}^{(n)}$. If more than one such $\hat{m} \in \mathcal{M}$ is found, declare $\hat{m}$ to be the smallest such index. The error events are

$$\mathcal{E}_2 \triangleq \{(u^n(m), y^n) \notin \mathcal{T}_{\varepsilon'}^{(n)}\} \ ,$$

where m corresponds to the m used in encoding and

$$\mathcal{E}_3 \triangleq \{\exists\, \tilde{m} \in \mathcal{M} : \tilde{m} \neq m, (u^n(\tilde{m}), y^n) \in \mathcal{T}_{\varepsilon'}^{(n)}\}$$

is the first decoding error. Define $\hat{u}^n \triangleq u^n(\hat{m})$.

After receiving ϕ from the public channel, Bob also finds an index $\hat{l} \in \mathcal{L}$ such that $v^n(\hat{m}, \hat{l}) \in \mathcal{B}(\phi)$ and $(v^n(\hat{m}, \hat{l}), \hat{u}^n, y^n) \in \mathcal{T}_\varepsilon^{(n)}$. If more than one such $\hat{l}$ is found, choose the one with the smallest index. The error events are

$$\mathcal{E}_4 \triangleq \{(v^n(m,l), u^n(m), y^n) \notin \mathcal{T}_\varepsilon^{(n)}\} \ ,$$

where l corresponds to the l used in encoding and

$$\mathcal{E}_5 \triangleq \{\exists\, \tilde{l} \neq l : v^n(\hat{m}, \tilde{l}) \in \mathcal{B}(\phi) \text{ and } (v^n(\tilde{l}), \hat{u}^n, y^n) \in \mathcal{T}_\varepsilon^{(n)}\}$$

is the second decoding error. Bob generates his key as $k_{\mathrm{B}} = k(\hat{m}, \hat{l})$, where the function k is defined in the codebook generation step.

4) Error probability analysis: The error probability can be decomposed as

$$\mathsf{P}\left[\bigcup_{i=0}^5 \mathcal{E}_i\right] = \sum_{i=0}^5 \mathsf{P}\left[\mathcal{E}_i \cap (\bigcup_{j=0}^{i-1} \mathcal{E}_j)^c\right] \ .$$

Firstly, $\mathsf{P}[\mathcal{E}_0] \to 0$ by the law of large numbers and the Typical Average Lemma [16]. Secondly, for each m,

[2]Similar to Matlab's (:) notation, the raster scan f_1 "unwraps" the pair of indices (m,l). Equivalently, f_1 may be defined as $f_1(m,l) \triangleq (l-1)|\mathcal{M}|+m$.

[3]We say that $a_n \doteq b_n$ if $\lim_{n\to\infty} n^{-1}\log(a_n/b_n) = 0$.

the number of v^n sequences is $\doteq 2^{n(I(V;X|U)+\delta)}$. By the Covering Lemma [16], for sufficiently small ε (the typicality tolerance) relative to δ, $\mathsf{P}[\mathcal{E}_1 \cap \mathcal{E}_0^c] \to 0$. Thirdly, by the Conditional Typicality Lemma [16], $\mathsf{P}[\mathcal{E}_2 \cap \mathcal{E}_0^c] \to 0$ and $\mathsf{P}[\mathcal{E}_4 \cap \mathcal{E}_0^c \cap \mathcal{E}_1^c] \to 0$. To bound the error event $\mathcal{E}_3$, since $|\mathcal{M}| \doteq 2^{n(I(U;Y)-2\delta)}$, by the Packing Lemma [16], $\mathsf{P}[\mathcal{E}_3 \cap (\bigcup_{j=0}^{2} \mathcal{E}_j)^c] \to 0$. Finally, since $\frac{|\mathcal{B}(\phi)|}{|\mathcal{M}|} \doteq 2^{n(I(V;Y|U)-\delta)}$, by the Packing Lemma $\mathsf{P}[\mathcal{E}_5 \cap (\bigcup_{j=0}^{4} \mathcal{E}_j)^c] \to 0$. For this final step, note that we have conditioned on the event that Bob decoded m correctly so the "cloud center" u^n is known.

5) Equivocation rate analysis: The information leakage rate in (3) can be rewritten as

$$\begin{aligned}\frac{1}{n} I(K_{\mathrm{A}}; Z^n, \Phi) &= \frac{1}{n} H(K_{\mathrm{A}}) - \frac{1}{n} H(K_{\mathrm{A}}|Z^n, \Phi) \\ &\leq I(U,V;Y) - I(U,V;Z) - \frac{1}{n} H(K_{\mathrm{A}}|Z^n, \Phi) \ . \end{aligned} \tag{12}$$

The inequality is due to the code construction, namely that $|\mathcal{K}| \doteq 2^{n(I(U,V;Y)-I(U,V;Z))}$. It remains to show that

$$H(K_{\mathrm{A}}|Z^n, \Phi) \geq n(I(U,V;Y) - I(U,V;Z)) - n\epsilon_n \ . \tag{13}$$

where $\epsilon_n \to 0$ as $n \to \infty$. This is because substituting (13) into (12) yields $I(K_{\mathrm{A}}; Z^n, \Phi) \leq n\epsilon_n$, satisfying the secrecy condition in (3). To show (13), firstly write the equivocation in (12) as a difference of two terms:

$$\begin{aligned} H(K_{\mathrm{A}}|Z^n, \Phi) &= H(K_{\mathrm{A}}, U^n, V^n|Z^n, \Phi) \\ &\quad - H(U^n, V^n|K_{\mathrm{A}}, Z^n, \Phi) \ . \end{aligned} \tag{14}$$

We bound each term on the right hand side separately. For the first term, consider

$$\begin{aligned} H(K_{\mathrm{A}}, U^n, V^n|Z^n, \Phi) &= H(K_{\mathrm{A}}, U^n, V^n, \Phi|Z^n) - H(\Phi|Z^n) \\ &\geq H(U^n, V^n|Z^n) - H(\Phi|Z^n) \ , \end{aligned} \tag{15}$$

Now, we lower bound the multi-letter entropy (first term) in (15) as follows. Let $\mathfrak{C}$ be a particular realization of the *random* codebook. Then consider,

$$\begin{aligned} &H(U^n, V^n|Z^n, \mathfrak{C}) = H(U^n, V^n, Z^n|\mathfrak{C}) - H(Z^n|\mathfrak{C}) \\ &= H(Z^n|U^n, V^n, \mathfrak{C}) + H(U^n, V^n|\mathfrak{C}) - H(Z^n|\mathfrak{C}) \\ &= \sum_{i=1}^{n} H(Z_i|Z^{i-1}, U^n, V^n, \mathfrak{C}) + H(U_i, V_i|U^{i-1}, V^{i-1}, \mathfrak{C}) \\ &\qquad - H(Z_i|Z^{i-1}, \mathfrak{C}) \\ &\geq \sum_{i=1}^{n} H(Z_i|U_i, V_i, \mathfrak{C}) + H(U_i, V_i|U^{i-1}, V^{i-1}, \mathfrak{C}) - H(Z_i|\mathfrak{C}) \ , \end{aligned}$$

where the inequality comes from conditioning reduces entropy and the fact that $(U^{n\setminus i}, V^{n\setminus i}, Z^{i-1}) - (U_i, V_i) - Z_i$ form a Markov chain conditioned on a specific realization of the codebook $\mathfrak{C}$. If we average over all codebooks, we have

$$\sum_{\mathfrak{C}} p(\mathfrak{C}) H(U_i, V_i|U^{i-1}, V^{i-1}, \mathfrak{C}) = H(U_i, V_i) \ ,$$

so the multi-letter conditional entropy can be lower bounded as $H(U^n, V^n|Z^n) \geq nH(U,V|Z)$. Continuing from (15),

$$\begin{aligned} &H(K_{\mathrm{A}}, U^n, V^n|Z^n, \Phi) \\ &\geq nH(U,V|Z) - H(\Phi|Z^n) \\ &\geq n[H(U|Z) + H(V|U,Z)] - H(\Phi) \\ &\overset{(a)}{\geq} n[H(U|Z) + H(V|U,Z) - I(V;X|U) + I(V;Y|U) - 2\delta] \\ &\overset{(b)}{\geq} n[H(U|Z) - H(U|Y) + H(V|U,Z) - H(V|U,X) \\ &\qquad - I(V;X|U) + I(V;Y|U) - 2\delta] \\ &= n[I(U;Y) - I(U;Z) + I(V;X|U) - I(V;Z|U) \\ &\qquad - I(V;X|U) + I(V;Y|U) - 2\delta] \\ &= n[I(U,V;Y) - I(U,V;Z) - 2\delta] \ , \end{aligned}$$

where (a) comes from the fact that $H(\Phi) \leq \log|\Phi| = I(V;X|U) - I(V;Y|U) + 2\delta$ by the code construction and (b) because $H(U|Y) \geq 0$ and $H(V|U,X) \geq 0$.

Now we bound the second term in (14). We claim that

$$H(U^n, V^n|K_{\mathrm{A}}, Z^n, \Phi) \leq n\epsilon_n \tag{16}$$

for some sequence $\epsilon_n \to 0$. For this purpose, we show that there exists a decoding function $(\tilde{M}, \tilde{L}) \triangleq g(K_{\mathrm{A}}, Z^n, \Phi)$ such that $\mathsf{P}[(U^n, V^n) \neq (u^n(\tilde{M}), v^n(\tilde{M}, \tilde{L}))] \to 0$ as $n \to \infty$. Then, by applying Fano's inequality, we get (16).

Let $g : \mathcal{Z}^n \times \mathcal{K} \times \Phi \to \mathcal{M} \times \mathcal{L}$ be a joint typicality decoder. More precisely, declare $g(z^n, k_{\mathrm{A}}, \phi) = (\tilde{m}, \tilde{l})$ if there is a unique pair of sequences $(u^n(\tilde{m}), v^n(\tilde{m}, \tilde{l}))$ such that $(u^n(\tilde{m}), v^n(\tilde{m}, \tilde{l}), z^n) \in \mathcal{T}_{\varepsilon}^{(n)}$ and $f(\tilde{m}, \tilde{l}) = (k_{\mathrm{A}}, \phi)$, where f is defined in the code construction. Otherwise, set $g(k_{\mathrm{A}}, z^n, \phi)$ to be $(1,1)$. Since there are $|\mathcal{J}|/(|\Phi||\mathcal{K}|) \doteq 2^{n(I(U,V;Z)-3\delta)}$ sequences in each sub-bin, by the Packing Lemma, $\mathsf{P}[(U^n, V^n) \neq (u^n(\tilde{M}), v^n(\tilde{M}, \tilde{L}))] \to 0$ as $n \to \infty$.

To complete the proof, let $p(w)$ be the optimizing W distribution in (5). Let $p_n(w)$ be a sequence of n-types such that $p_n(w) \to p(w)$ for every $w \in \mathcal{W}$. Fix w^n as some sequence in the type class of $p_n(w)$, i.e., the type of w^n is equal to $p_n(w)$. The sequence w^n is appended to the codebook and thus known to all parties. Then we follow the previous proof by replacing the marginal distributions with the conditional distributions in the codebook generation step (as in [1, Lemma A]). More precisely, we use $\prod_{i=1}^{n} p(u_i, s_i|w_i)$ in place of $\prod_{i=1}^{n} p(u_i, s_i)$ and $\prod_{i=1}^{n} p(v_i|u_i, w_i)$ in place of $\prod_{i=1}^{n} p(v_i|u_i)$. This achieves the rate $I(U,V;Y|W) - I(U,V;Z|W)$ as desired.

Hence, the rate $C_{\mathrm{SK}}(\frac{\Gamma}{1+\varepsilon})$ is achievable. The proof of the achievability of (5) completed by taking $\varepsilon \to 0$ and appealing to the continuity of C_{SK} (Lemma 2).

C. Proof of Lemma 2

The fact that $C_{\mathrm{SK}}(\Gamma)$ is non-decreasing is evident from its definition. We now show that $C_{\mathrm{SK}}(\Gamma)$ is concave. Fix two length-n codes $\mathfrak{C}_1$ and $\mathfrak{C}_2$ that achieve $C_{\mathrm{SK}}(\Gamma_1)$ and $C_{\mathrm{SK}}(\Gamma_2)$ respectively. Consider the length-$2n$ code $\mathfrak{C}$ that is the concatenation of $\mathfrak{C}_1$ and $\mathfrak{C}_2$. That is, for the first n

channel uses, we use $\mathfrak{C}_1$ and for the next n, we use $\mathfrak{C}_2$. Then the total expected cost of $\mathfrak{C}$ is

$$\mathsf{E}\left[\sum_{i=1}^{2n}\Lambda(S_i)\right] = \mathsf{E}\left[\sum_{i=1}^{n}\Lambda(S_i) + \sum_{i=n+1}^{2n}\Lambda(S_i)\right] \leq n(\Gamma_1+\Gamma_2)\,,$$

since the first and second codes have expected costs smaller than Γ_1 and Γ_2 respectively. Hence, $\mathfrak{C}$ satisfies $\frac{1}{2n}\mathsf{E}[\sum_{i=1}^{2n}\Lambda(S_i)] \leq \frac{1}{2}(\Gamma_1+\Gamma_2)$. We have constructed a codebook with rate $\frac{1}{2}(C_{\mathrm{SK}}(\Gamma_1)+C_{\mathrm{SK}}(\Gamma_2))$ and with expected cost $\leq \frac{1}{2}(\Gamma_1+\Gamma_2)$. Thus, $C_{\mathrm{SK}}(\frac{1}{2}(\Gamma_1+\Gamma_2)) \geq \frac{1}{2}(C_{\mathrm{SK}}(\Gamma_1)+C_{\mathrm{SK}}(\Gamma_2))$, i.e., it is mid-point concave. Since $C_{\mathrm{SK}}(\Gamma)$ is non-decreasing, its level sets are intervals and so it is Lebesgue measurable (Sierpinski's theorem [18, pp. 12]). Combining this with the fact that it is mid-point concave, we conclude that $C_{\mathrm{SK}}(\Gamma)$ is concave. Since a concave function on an open set is also continuous, $C_{\mathrm{SK}}(\Gamma)$ is continuous on $(0,\infty)$.

Note that the above proof is an *operational* one and does not depend on the functional form of $C_{\mathrm{SK}}(\Gamma)$ in (5). □

D. Proof of Proposition 3

We prove the upper bound in (9). Consider the inequalities:

$$\begin{aligned}
nR_{\mathrm{SK}} &\overset{(a)}{\leq} I(K_{\mathrm{A}};Y^n,\Phi)+n\epsilon_n \leq I(K_{\mathrm{A}};Y^n,\Phi,Z^n)+n\epsilon_n\\
&\overset{(b)}{\leq} I(K_{\mathrm{A}};Y^n|\Phi,Z^n)+2n\epsilon_n\\
&\leq I(K_{\mathrm{A}},\Phi;Y^n|Z^n)+2n\epsilon_n\\
&\overset{(c)}{\leq} I(X^n,M;Y^n|Z^n)+2n\epsilon_n\\
&= I(X^n;Y^n|Z^n)+I(M;Y^n|X^n,Z^n)+2n\epsilon_n\\
&\overset{(d)}{\leq} I(X^n;Y^n|Z^n)+I(S^n;Y^n|X^n,Z^n)+2n\epsilon_n\\
&= I(S^n;Y^n|Z^n)+I(X^n;Y^n|S^n,Z^n)+2n\epsilon_n,
\end{aligned} \tag{17}$$

where (a) follows Fano's inequality, (b) is due to the secrecy condition (3), (c) follows because (K_{A},Φ) is a function of (X^n,M) and (d) follows because the channel only depends on S^n so $M-S^n-(X^n,Y^n,Z^n)$.[4] Now the first term (17) can be upper bounded as follows

$$\begin{aligned}
&I(S^n;Y^n|Z^n) = H(Y^n|Z^n)-H(Y^n|S^n,Z^n)\\
&=\sum_{i=1}^{n} H(Y_i|Y^{i-1},Z^n)-H(Y_i|Y^{i-1},S^n,Z^n)\\
&\leq \sum_{i=1}^{n} H(Y_i|Z_i)-H(Y_i|S_i,Z_i) = \sum_{i=1}^{n} I(S_i;Y_i|Z_i)\,,
\end{aligned} \tag{18}$$

where the inequality follows by conditioning reduces entropy and the Markov chain $(Y^{i-1},Z^{n\setminus i},S^{n\setminus i})-(S_i,Z_i)-Y_i$. The second term in (17) can be written as a sum:

$$I(X^n;Y^n|S^n,Z^n) = \sum_{i=1}^{n} I(X_i;Y_i|S_i,Z_i) \tag{19}$$

[4]In fact, (d) holds with equality because $S^n = S^n(M)$ in addition to the stated Markov relationship.

because the channel $p(x,y,z|s)$ is memoryless. Substituting (18) and (19) into (17) yields

$$\begin{aligned}
nR_{\mathrm{SK}} &\leq \sum_{i=1}^{n} I(S_i;Y_i|Z_i)+I(X_i;Y_i|S_i,Z_i)+2n\epsilon_n\\
&= \sum_{i=1}^{n} I(X_i,S_i;Y_i|Z_i)+2n\epsilon_n\\
&\overset{(e)}{\leq} \sum_{i=1}^{n} \max_{p(s)}\ I(X,S;Y|Z)+2n\epsilon_n\,,
\end{aligned}$$

where (e) holds due to the fact that mutual information is concave function of input distribution $p(s)$. This completes the proof of (9). □

Acknowledgements

VYFT would like to thank Mukul Agarwal (MIT) for discussions that led to the operational proof of Lemma 2.

REFERENCES

[1] R. Ahlswede and I. Csiszár. Common randomness in information theory and cryptography part I: Secret sharing. *IEEE Trans. on Inf. Th.*, 39(4):1121–1132, 1993.

[2] U. M. Maurer. Secret key agreement by public discussion from common information. *IEEE Trans. on Inf. Th.*, 39(3):733–742, 1993.

[3] A. D. Wyner. The wire-tap channel. *The Bell Systems Technical Journal*, 54:1355 – 1387, 1975.

[4] I. Csiszár and J. Körner. Broadcast channels with confidential messages. *IEEE Trans. on Inf. Th.*, 24(3):339–348, 1978.

[5] I. Csiszár and P. Narayan. Common randomness and secret key generation with a helper. *IEEE Trans. on Inf. Th.*, 46(2):344366, 2000.

[6] T. Weissman. Capacity of channels with action-dependent states. *IEEE Trans. on Inf. Th.*, 56(11):5396–5411, 2010.

[7] H. Asnani, H. Permuter, and T. Weissman. Probing Capacity. *arXiv:1010.1309*.

[8] K. Kittichokechai, T. J. Oechtering, M. Skoglund, and R. Thobaben. Source and channel coding with action-dependent partially known two-sided state information. In *Proc. Int. Symp. Inform. Theory*, pages 629–633, June 2010.

[9] A. Khisti, S. Diggavi, and G. Wornell. Secret-key generation with correlated sources and noisy channels. In *Proc. Int. Symp. Inform. Theory*, pages 1005–1009, July 2008.

[10] V. Prabhakaran, K. Eswaran, and K. Ramchandran. Secrecy via sources and channels – a secret key-secret message rate tradeoff region. In *Proc. Int. Symp. Inform. Theory*, pages 1010–1014, July 2008.

[11] Y. Chen and A. J. Han Vinck. Wiretap channel with side information. *IEEE Trans. on Inf. Th.*, 54(1):395 –402, Jan. 2008.

[12] W. Liu and B. Chen. Wiretap channel with two-sided channel state information. In *Proc. Asilomar Conf. Signals, Systems and Computers, 2007*, pages 893 –897, Nov. 2007.

[13] A. Khisti, S. Diggavi, and G. Wornell. Secret key agreement using asymmetry in channel state knowledge. In *Proc. Int. Symp. Inform. Theory*, pages 2286–2290, 2009.

[14] Y. K. Chia and A. El Gamal. Wiretap channel with causal state information. *arXiv:1001.2327*.

[15] T.-H. Chou, S. Draper, and A. Sayeed. Key generation using an external source of excitation: Capacity, reliability, and secrecy exponent. *submitted to the IEEE Trans. on Inf. Th.*, 2010.

[16] A. El Gamal and Y.-H. Kim. Lecture Notes on Network Information Theory. *arXiv:1001.3404*.

[17] S. Leung-Yan-Cheong and M. Hellman. The Gaussian wire-tap channel. *IEEE Trans. on Inf. Th.*, 24(4):451–456, Jul. 1978.

[18] *Distributions and Fourier transforms*. Academic Press, 1969.