Common Randomness and Secret Key Generation with a Helper

Imre Csiszár, *Fellow, IEEE,* and Prakash Narayan, *Senior Member, IEEE*

***Abstract*— We consider the generation of common randomness (CR), secret or not secret, by two user terminals with aid from a "helper" terminal. Each terminal observes a different component of a discrete memoryless multiple source. The helper aids the users by transmitting information to them over a noiseless public channel subject to a rate constraint. Furthermore, one of the users is allowed to transmit to the other user over a public channel under a similar rate constraint. We study the maximum rate of CR which can be thus generated, including under additional secrecy conditions when it must be concealed from a wiretapper. Lower bounds for the corresponding capacities are provided, and single-letter capacity formulas are obtained for several special cases of interest.**

***Index Terms*— Capacity, common randomness, correlated sources, multiuser information theory, private key, secret key, wiretapper.**

I. Introduction

We study the problem of determining the maximum amount of common randomness (CR) which can be generated by separate terminals under specified conditions and which may or may not involve secrecy requirements. The CR problem has been previously studied, after an early result due to Gács and Körner [7], by Maurer [8], [9], Ahlswede and Csiszár [2], [3], Bennett *et al*. [4], Ahlswede and Balakirsky [1], Csiszár [5], and Venkatesan and Anantharam [12], [13]. The generation of CR can be based on randomness represented by the outcomes of correlated sources available at the terminals (source-type models), or on randomness introduced by channel noise (channel-type models), or on both. In this paper, we shall focus on source-type models where the only means of information transmission are offered by noiseless public channels.

The main feature of this paper consists of a study of the generation of CR by user terminals with the aid of another party called the "helper," although new results for models not involving a helper are also obtained as special cases. The introduction of the helper is motivated by the role played by a "third party" (e.g., a centralized or trusted server in a key establishment protocol), which facilitates the generation of CR by user terminals by furnishing them additional correlated information. Also, the notion of a helper has potential significance for practical schemes for the generation of CR by three or more user terminals, wherein the different users can take turns serving as helper in successive rounds of CR generation. As another feature, we examine models for the generation of secret or nonsecret CR with rate constraints imposed on permissible transmissions, depicting bandwidth limitations associated with the use of shared public channels. This approach has been used in [3] for CR generation without secrecy. Our models tacitly assume that all public transmissions are impervious to any deliberate attempts at inserting corruption. In a cryptographic situation, this assumption implies, in effect, that the public transmissions are authenticated, or that an (adversarial) wiretapper is passive, i.e., unable to tamper with such transmissions. (For unauthenticated public transmissions, see [10], [11].) Much of the notation and terminology are from [2] and [3]. All logarithms and exponentiations are with respect to the base 2.

We consider first a discrete memoryless multiple source (DMMS) with three components, with alphabets $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ and corresponding generic random variables (rv's) (X, Y, Z). We adopt the practice in [2] of representing a terminal and its associated alphabet by the same symbol. Terminals $\mathcal{X}, \mathcal{Y}$, and $\mathcal{Z}$ observe the DMMS outputs $X^n = (X_1, \cdots, X_n)$, $Y^n = (Y_1, \cdots, Y_n)$, and $Z^n = (Z_1, \cdots, Z_n)$, respectively, of block lengths n. Terminals $\mathcal{Y}$ and $\mathcal{Z}$ represent the two users who wish to generate CR, secret or not secret, while terminal $\mathcal{X}$ plays the role of the helper. Terminal $\mathcal{X}$ can help the user terminals $\mathcal{Y}$ and $\mathcal{Z}$ in their tasks by sending them information over a noiseless public channel of capacity R_1, i.e., $\mathcal{X}$ can noiselessly transmit any function $f(X^n)$ of X^n to both $\mathcal{Y}$ and $\mathcal{Z}$ over a public channel subject to the rate constraint

$$\frac{1}{n} \log ||f|| \leq R_1 \tag{1.1}$$

where $||f||$ denotes the cardinality of the range of f. Furthermore, terminal $\mathcal{Y}$ is allowed to send information to terminal $\mathcal{Z}$ over a noiseless public channel of capacity R_2, i.e., $\mathcal{Y}$ can noiselessly transmit any function $g(Y^n, f)$ of Y^n and $f = f(X^n)$ to $\mathcal{Z}$ over a public channel subject to the rate constraint

$$\frac{1}{n} \log ||g|| \leq R_2. \tag{1.2}$$

No other resources are available to the three terminals. In particular, randomization is not permitted, i.e., f and g are deterministic mappings. (However, randomization can be incorporated into our model by the simple device of augmenting the source

Manuscript received August 5, 1998; revised September 14, 1999. The work of I. Csiszár was supported by the Hungarian National Foundation for Scientific Research under Grant T16386. The work of P. Narayan was supported by the Maryland Procurement Office under Grant MDA90497C3015 and the Center for Satellite and Hybrid Communication Networks, a NASA Commercial Space Center, under NASA Cooperative Agreement NCC3-528.

I. Csiszár is with the A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, POB 127, H-1364 Budapest, Hungary.

P. Narayan is with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA.

Communicated by V. Anantharam, Associate Editor for Communication Networks.

Publisher Item Identifier S 0018-9448(00)01348-1.

rv's X and Y by independent additional randomness.) We remark that this model generalizes the situation considered in [3] consisting of two terminals with two-way noiseless communication, the latter being recovered when $\mathcal{Z} = \mathcal{X}$.

A pair of rv's (K, L) represents ϵ-CR for the terminals $\mathcal{X}$ and $\mathcal{Y}$ if K and L are functions of the data available at $\mathcal{Y}$ and $\mathcal{Z}$, respectively, i.e.,

$$\begin{aligned} K &= K(f, Y^n) \\ L &= L(f, g, Z^n) \end{aligned} \tag{1.3}$$

K and L take values in the same finite set $\mathcal{K}$, and

$$\Pr\{K \neq L\} \leq \epsilon. \tag{1.4}$$

The rate of this CR is defined as $\frac{1}{n}H(K)$. For small ϵ, this differs only negligibly from $\frac{1}{n}H(L)$ by Fano's inequality, since (1.3) implies that the cardinality of the set $\mathcal{K}$ does not exceed $\exp(nc)$ for some $c > 0$ not depending on n. (Note that in [3] where the users were permitted to randomize, an analogous bound on the size of $\mathcal{K}$ was an assumption.)

Definition 1.1: A number H is called an achievable CR rate if for every $\epsilon > 0, \delta > 0$, and for all sufficiently large n, ϵ-CR can be generated with rate exceeding $H - \delta$, i.e., there exist functions f, g and rv's K, L satisfying (1.1)–(1.4) such that

$$\frac{1}{n}H(K) > H - \delta. \tag{1.5}$$

The largest achievable CR rate is called CR-capacity, and is denoted by $C_{\mathrm{CR}}(R_1, R_2)$.

We shall also consider a model involving a DMMS with four components with generic rv's (X, Y, Z, W). The users $\mathcal{Y}$ and $\mathcal{Z}$ wish to generate CR—with terminal $\mathcal{X}$ playing the role of helper—with the requirement that it be concealed from a wiretapper $\mathcal{W}$ who observes the public transmissions of f and g and, in addition, the $\mathcal{W}$-valued outputs $W^n = (W_1, \cdots, W_n)$ as side information. This model generalizes the "source-type model with wiretapper" of [2]. A pair of rv's (K, L) representing ϵ-CR for the users $\mathcal{Y}$ and $\mathcal{Z}$ constitutes an *ϵ-wiretap secret key* if it additionally satisfies the following secrecy condition:

$$\frac{1}{n} I(K \wedge f, g, W^n) \leq \epsilon. \tag{1.6}$$

Note that a wiretap secret key (WSK) is not necessarily concealed from the helper $\mathcal{X}$ which, thus, constitutes a "trusted third-party" in this situation.

Two special cases of wiretap secrecy as above will be examined separately. The first obtains when the users $\mathcal{Y}$ and $\mathcal{Z}$ wish to generate a WSK which is concealed from the helper $\mathcal{X}$ too. Accordingly, we define an *ϵ-private key* as an ϵ-wiretap secret key with W^n in (1.6) replaced by X^nW^n, whence the secrecy condition (1.6) now takes the more stringent form

$$\frac{1}{n} I(K \wedge g, X^n, W^n) \leq \epsilon \tag{1.7}$$

recalling that $f = f(X^n)$. This models the situation in which the users wish to maintain secrecy even from the centralized server.

A second special case of wiretap secrecy occurs when the users $\mathcal{Y}$ and $\mathcal{Z}$ wish to generate a secret key which is effectively concealed from an eavesdropper with access to the public transmissions of f and g, but not to any side information. Accordingly, we define an *ϵ-secret key* with (1.6) replaced by

$$\frac{1}{n} I(K \wedge f, g) \leq \epsilon. \tag{1.8}$$

Note that a secret key is not necessarily concealed from the helper $\mathcal{X}$ which now constitutes, once again, a "trusted third-party."

Definition 1.2: A number H is called an achievable WSK rate if for every $\epsilon > 0, \delta > 0$, and for all sufficiently large n, an ϵ-WSK can be generated with rate exceeding $H - \delta$, i.e., there exist functions f, g and rv's K, L satisfying (1.1)–(1.6). The largest achievable WSK rate is called the WSK–capacity and is denoted by $C_{\mathrm{WSK}}(R_1, R_2)$. As special cases, the private key (PK)-capacity and secret key (SK)-capacity, denoted by $C_{\mathrm{PK}}(R_1, R_2)$ and $C_{\mathrm{SK}}(R_1, R_2)$ are defined in an obvious manner with (1.6) being replaced by (1.7) and (1.8), respectively.

As in [2] and [3], we could require in all cases above that the CR be nearly uniformly distributed, in the sense that its distribution be close to the uniform distribution in variation distance. In fact, we shall also consider the notion of CR (secret or not secret) in a stronger sense, namely, by requiring

$$\sum_{k \epsilon \mathcal{K}} \left| \Pr\{K = k\} - \frac{1}{|\mathcal{K}|} \right| \leq \exp(-n\alpha) \tag{1.9}$$

and

$$\Pr\{K \neq L\} \leq \exp(-n\alpha) \tag{1.10}$$

for a suitable $\alpha > 0$, and replacing (1.6)–(1.8), respectively, by the conditions

$$I(K \wedge f, g, W^n) \leq \exp(-n\alpha) \tag{1.11}$$

$$I(K \wedge g, X^n, W^n) \leq \exp(-n\alpha) \tag{1.12}$$

and

$$I(K \wedge f, g) \leq \exp(-n\alpha). \tag{1.13}$$

We shall say that H is a strongly achievable CR, WSK, PK, or SK rate, if in Definitions 1.1 and 1.2, K is required to satisfy (1.9), and (1.4), (1.6)–(1.8) are replaced by the (stronger) conditions (1.10), and (1.11)–(1.13), respectively (with some $\alpha > 0$ depending on δ). The desirability of secrecy constraints in the strong sense has been pointed out by Maurer [9]. All our achievability results will be proved as strong achievability results. The fact that the stronger constraints do not reduce secrecy capacity has been demonstrated for certain models in [9], [4], and [5].

An early result on CR, due to Gács and Körner [7], states—in our terminology—that $C_{\mathrm{CR}}(0, 0) = H(V)$, where V is a maximal common function (c.f.) of Y and Z. A c.f. of Y and Z is any rv which equals both a function of Y and a function of Z; a c.f. V of Y and Z is maximal if every other c.f. of Y and Z is a function of V. We shall need the following sharpened version of this result.

Lemma 1.1:: Given a DMMS with generic rv's (Y, Z) whose maximal c.f. is V, for arbitrarily small $\xi > 0$ there exists $\epsilon > 0$ not depending on n, such that

$$\Pr\{K(Y^n) \neq L(Z^n)\} \leq \epsilon \tag{1.14}$$

can hold only when $K(Y^n)$ equals a function of V^n with probability exceeding $1 - \xi$.

Proof: The proof is an easy consequence of the results of Witsenhausen [14], and is given in Appendix F.

While this paper is devoted to the generation of CR at the two terminals $\mathcal{Y}$ and $\mathcal{Z}$ when a third terminal $\mathcal{X}$ serves as a helper, it is also of interest to consider the amount of CR (secret or not secret) which can be generated by all three terminals. Formally, the definitions above can be modified by replacing (1.4) by the condition

$$\Pr\{M = K = L\} \geq 1 - \epsilon \tag{1.15}$$

where $M = M(X^n)$ is a suitable function of X^n, and K, L are as in (1.3). The resulting "three-way" analogs of $C_{\mathrm{CR}}(R_1, R_2)$ and $C_{\mathrm{SK}}(R_1, R_2)$ will be denoted by $C^3_{\mathrm{CR}}(R_1, R_2)$ and $C^3_{\mathrm{SK}}(R_1, R_2)$, respectively. These will be determined in the special case $R_2 = 0$ (i.e., no communication from terminal $\mathcal{Y}$ to terminal $\mathcal{Z}$), and are relevant for our purpose, namely for the determination of $C_{\mathrm{CR}}(R_1, R_2)$ and $C_{\mathrm{SK}}(R_1, R_2)$ when R_1 or R_2 equals 0.

II. Summary of Results

Our main results are stated as Theorems 2.1–2.6 below. Although not specifically mentioned in the corresponding statements, all our achievability results hold, in fact, with strong achievability.

Theorem 2.1: The CR-capacity $C_{\mathrm{CR}}(R_1, R_2)$ with $R_1 > 0$, $R_2 > 0$, is bounded below according to

$$C_{\mathrm{CR}}(R_1, R_2) \geq \max_{U,V,t} \left[I(U \wedge X) + I(V \wedge Y|U) + t\right] \tag{2.1}$$

where the rv's U, V, and the number t satisfy the condition

$$0 \leq t \leq H(X|YU) \tag{2.2}$$

the rate conditions

$$I(U \wedge X) - \min\{I(U \wedge Y), I(U \wedge Z)\} + t \leq R_1 \tag{2.3}$$

$$I(V \wedge Y|U) - I(V \wedge Z|U) \leq R_2 \tag{2.4}$$

and the Markov conditions

$$U \multimap X \multimap YZ \tag{2.5}$$

$$V \multimap UY \multimap Z. \tag{2.6}$$

The following Theorem 2.2 is a special case of Theorem 2.5 below. However, we prefer to state it separately here because it is easier to understand, and because under certain conditions it represents a conclusive result, i.e., affords a converse.

Theorem 2.2: The SK-capacity $C_{\mathrm{SK}}(R_1, R_2)$ with $R_1 > 0$, $R_2 > 0$, is bounded below according to

$$C_{\mathrm{SK}}(R_1, R_2) \geq \max_{U,V} \left[\min\{I(U \wedge Y), I(U \wedge Z)\} + I(V \wedge Z|U)\right] \tag{2.7}$$

where the rv's U, V satisfy the rate conditions

$$I(U \wedge X) - \min\{I(U \wedge Y), I(U \wedge Z)\} \leq R_1 \tag{2.8}$$

and (2.4), and the Markov conditions (2.5), (2.6).

In order to interpret the bounds in (2.1) and (2.7), we note that $I(U \wedge X) + t$ (in (2.1)) and $\min\{I(U \wedge Y), I(U \wedge Z)\}$ (in (2.7)), respectively, represent the CR and SK rates which the users $\mathcal{Y}$ and $\mathcal{Z}$ can generate using the transmission $f = f(X^n)$ of helper $\mathcal{X}$ alone, while $I(V \wedge Y|U)$ (in (2.1)) and $I(V \wedge Z|U)$ (in (2.7)) represent the additional CR and SK rates which are enabled by the transmission $g = g(Y^n, f)$ of user $\mathcal{Y}$ (when using a specific scheme depending on the choice of U, V, and t, cf. the proof of Theorems 2.1 and 2.2 in Section III). A tradeoff between these two parts of CR and SK rates makes it conceivable that their sum attains its maximum when the first part does not (see Case 2 after Theorem 2.3 below).

It is not obvious *a priori* that the maxima in (2.1) and (2.7) are actually attained. This is true, however, by virtue of the fact proved in Appendix B that in (2.1) and (2.7), the rv's U, V can be assumed, without restricting generality, to take values in sets $\mathcal{U}, \mathcal{V}$ of cardinalities

$$|\mathcal{U}| \leq |\mathcal{X}| + 3 \qquad |\mathcal{V}| \leq |\mathcal{Y}|.$$

As in other results of similar form, these range constraints are also of conceptual importance, as they render the maxima in question computable, at least in principle (cf. [6, Sec. III-3]). The maxima in (2.1) and (2.7), as functions of R_1, R_2, will be discussed and related to the function

$$F_1(R_1, R_2) = \max_{U,V} \left[I(U \wedge X) + I(V \wedge Y|U)\right] \tag{2.9}$$

in Appendix E; the maximum in (2.9) is with respect to rv's U, V satisfying the rate conditions (2.8), (2.4), and the Markov conditions (2.5), (2.6).

Theorem 2.3: Suppose that the rv's X, Y, Z satisfy either of the Markov conditions $X \multimap Z \multimap Y$ or $X \multimap Y \multimap Z$. Then, the bound for $C_{\mathrm{CR}}(R_1, R_2)$ in (2.1) and the bound for $C_{\mathrm{SK}}(R_1, R_2)$ in (2.7) are tight for all $R_1 > 0$, $R_2 > 0$. Moreover, in the cases when $X \multimap Z \multimap Y$ or $X \multimap Y \multimap Z$ and $R_2 \leq H(Y|XZ)$, the maximum in (2.1) is attained with $t = 0$, so that $C_{\mathrm{CR}}(R_1, R_2) = F_1(R_1, R_2)$.

Although single-letter characterizations of $C_{\mathrm{CR}}(R_1, R_2)$ and $C_{\mathrm{SK}}(R_1, R_2)$ remain elusive in general, conclusive results are available in several special cases of independent interest which are discussed below. Note that $R_1 = \infty$ or $R_1 = 0$ (resp., $R_2 = \infty$ or $R_2 = 0$) means terminal $\mathcal{X}$ (resp., $\mathcal{Y}$) can transmit unfettered by any rate constraint or not at all.

Case 1. $C_{\mathrm{CR}}(\infty, R_2)$: In the absence of any rate constraint on the transmission from terminal $\mathcal{X}$, the CR-capacity is clearly the same as if terminals $\mathcal{Y}$ and $\mathcal{Z}$ observed X^nY^n and X^nZ^n,

and then terminal $\mathcal{Y}$ were allowed to transmit to terminal $\mathcal{Z}$ at rate not exceeding R_2. The latter CR-capacity has been determined in [3, Theorem 4.1] (cf. also Theorem 2.4 below), whence

$$C_{\mathrm{CR}}(\infty, R_2) = \max_V I(V \wedge XY) \le H(X) + \max_V I(V \wedge Y|X). \tag{2.10}$$

Here, the maximum is with respect to rv's V satisfying the rate condition $I(V \wedge XY) - I(V \wedge XZ) \le R_2$ which is equivalent to

$$I(V \wedge Y|X) - I(V \wedge Z|X) \le R_2 \tag{2.11}$$

and the Markov condition

$$V \multimap XY \multimap Z. \tag{2.12}$$

On the other hand, the lower bound in (2.1), with $U = X$, $t = 0$, gives that

$$C_{\mathrm{CR}}(R_1, R_2) \ge H(X) + \max_V I(V \wedge Y|X)$$

the maximum being subject to (2.11) and (2.12); thus the lower bound in (2.1) is tight in this case with $t = 0$, yielding

$$C_{\mathrm{CR}}(\infty, R_2) = F_1(\infty, R_2).$$

Remark: Note that $U = X$ is a permissible choice in (2.1), and hence

$$C_{\mathrm{CR}}(R_1, R_2) = C_{\mathrm{CR}}(\infty, R_2)$$

whenever $R_1 \ge \max\{H(X|Y), H(X|Z)\}$.

Case 2 .$C_{\mathrm{CR}}(R_1, \infty)$*:* As (1.1) and (1.3) imply that $\frac{1}{n}H(K)$ is bounded above by both $R_1 + H(Y)$ and $H(XY)$, the following trivial bound:

$$C_{\mathrm{CR}}(R_1, R_2) \le \min\{R_1, H(X|Y)\} + H(Y) \tag{2.13}$$

holds for every R_2, and, in particular, also for $R_2 = \infty$. On the other hand, the lower bound in (2.1) with $U =$ a constant, $V = Y$ gives that

$$C_{\mathrm{CR}}(R_1, \infty) \ge R_1 + H(Y), \qquad \text{if } R_1 \le H(X|Y). \tag{2.14}$$

It then follows from (2.13), (2.14), and the monotonicity of $C_{\mathrm{CR}}(R_1, R_2)$ as a function of R_1, that

$$C_{\mathrm{CR}}(R_1, \infty) = \min\{R_1, H(X|Y)\} + H(Y) \tag{2.15}$$

and the lower bound in (2.1) is tight in this case.

Remark: Note that $C_{\mathrm{CR}}(R_1, R_2) = C_{\mathrm{CR}}(R_1, \infty)$ whenever $R_2 \ge H(Y|Z)$.

Case 3. $C_{\mathrm{SK}}(\infty, R_2)$*:* If $R_1 = \infty$, or at least $R_1 \ge \max\{H(X|Y), H(X|Z)\}$, then $U = X$ is a permissible choice in (2.7). This gives the lower bound

$$C_{\mathrm{SK}}(\infty, R_2) \ge \min\{I(X \wedge Y), I(X \wedge Z)\} + \max_V I(V \wedge Z|X) \tag{2.16}$$

where the maximum is with respect to rv's V satisfying the conditions (2.11), (2.12).

In order to obtain upper bounds, consider the modified situations in which terminal $\mathcal{Y}$ can observe X^n in addition to Y^n (resp., terminal $\mathcal{Z}$ can observe X^n in addition to Z^n); formally this entails replacing Y by XY (resp., Z by XZ). As these modifications lead to Markov settings and do not decrease SK-capacity, it follows from Theorem 2.3 that both (2.17) and (2.21) below constitute upper bounds for $C_{\mathrm{SK}}(\infty, R_2)$:

$$C_{\mathrm{SK}}(\infty, R_2) \le \max_{U,V} [I(U \wedge Z) + I(V \wedge Z|U)] \tag{2.17}$$

subject to the Markov conditions

$$U \multimap X \multimap YZ \tag{2.18}$$

$$V \multimap UXY \multimap Z \tag{2.19}$$

and the rate condition

$$I(V \wedge XY|U) - I(V \wedge Z|U) \le R_2; \tag{2.20}$$

and

$$C_{\mathrm{SK}}(\infty, R_2) \le \max_{U,V} [I(U \wedge Y) + I(V \wedge XZ|U)] \tag{2.21}$$

subject to the Markov conditions (2.18) and

$$V \multimap UY \multimap XZ \tag{2.22}$$

and the rate condition

$$I(V \wedge Y|U) - I(V \wedge XZ|U) \le R_2. \tag{2.23}$$

We prove in Appendix C that (2.17) and (2.21) give rise to the bounds

$$C_{\mathrm{SK}}(\infty, R_2) \le I(X \wedge Z) + \max_V I(V \wedge Z|X) \tag{2.24}$$

$$C_{\mathrm{SK}}(\infty, R_2) \le I(X \wedge Y) + \max_V I(V \wedge Z|X) \tag{2.25}$$

with V being subject to (2.11) and (2.12). This means that the lower bound for $C_{\mathrm{SK}}(\infty, R_2)$ in (2.16) is tight.

If additionally $R_2 = \infty$, or at least $R_2 \ge H(Y|XZ)$, the maximum in (2.16) equals $I(Y \wedge Z|X)$. Thus if both terminals $\mathcal{X}$ and $\mathcal{Y}$ can transmit without rate constraints, the corresponding SK-capacity is

$$C_{\mathrm{SK}}(\infty, \infty) = \min\{I(X \wedge Y), I(X \wedge Z)\} + I(Y \wedge Z|X) = \min\{I(XZ \wedge Y), I(XY \wedge Z)\}. \tag{2.26}$$

Recall from [8], [2] that in the absence of a "helper," the largest SK-rate achievable by a public transmission from terminal $\mathcal{Y}$ to terminal $\mathcal{Z}$ is equal to $I(Y \wedge Z)$. The result in (2.26) shows that if either of the Markov conditions $X \multimap Y \multimap Z$ or $X \multimap Z \multimap Y$ holds, the "helper" Terminal $\mathcal{X}$ cannot serve to achieve a larger SK-rate, but in all other cases it can.

Case 4. $C_{\mathrm{SK}}(R_1, \infty)$*:* This case remains open, in general.

The cases $R_1 = 0$ and $R_2 = 0$ are not covered by Theorems 2.1 and 2.2. Results for these simpler cases will follow as consequences of Theorem 2.4 below which provides single-letter characterizations of the "three-way" capacities $C^3_{\mathrm{CR}}(R_1, 0)$ and $C^3_{\mathrm{SK}}(R_1, 0)$ (cf. end of Section I).

Case 5. $C_{\rm CR}(R_1,0)$ and $C_{\rm SK}(R_1,0)$: Suppose that the rv's X, Y, Z have the property that the maximal c.f. of (X, Y) and (X, Z) is X. Then by Lemma 1.1, if (1.4) holds for K and L as in (1.3) with g = a constant, then necessarily $\Pr\{K = M\} > 1-\xi$ for a suitable function $M = M(X^n)$ of X^n, where $\xi > 0$ is as small as desired if $\epsilon > 0$ is sufficiently small. This implies that $C_{\rm CR}(R_1, 0) = C^3_{\rm CR}(R_1, 0)$ and $C_{\rm SK}(R_1, 0) = C^3_{\rm SK}(R_1, 0)$. Thus under the given condition on X, Y, Z, the capacities $C_{\rm CR}(R_1, 0)$ and $C_{\rm SK}(R_1, 0)$ are determined by Theorem 2.4 below.

Case 6. $C_{\rm CR}(0, R_2)$ and $C_{\rm SK}(0, R_2)$: These capacities are, respectively, the same as $C^3_{\rm CR}(R_2, 0)$ and $C^3_{\rm SK}(R_2, 0)$ in the special case $X = Y$, and are hence determined by Theorem 2.4 below. Note that $C_{\rm CR}(0, R_2)$ has already been determined in [3, Theorem 4.1].

Theorem 2.4: The "three-way" CR- and SK- capacities $C^3_{\rm CR}(R_1, 0)$ and $C^3_{\rm SK}(R_1, 0)$ are given by

$$C^3_{\rm CR}(R_1,0) = \max_U I\,(U \wedge X)$$
$$C^3_{\rm SK}(R_1, 0) = \max_U \min\{I(U \wedge Y), I(U \wedge Z)\}$$

where the maxima are with respect to rv's U satisfying the rate condition (2.8) and the Markov condition (2.5).

Remark: The result above for $C^3_{\rm CR}(R_1, 0)$ is related to [3, Theorem 4.2].

Finally, our results on WSK-capacity, $C_{\rm WSK}(R_1, R_2)$, and its special case of PK-capacity, $C_{\rm PK}(R_1, R_2)$, are stated below as Theorems 2.5 and 2.6 and their corollaries.

Theorem 2.5: The WSK-capacity $C_{\rm WSK}(R_1, R_2)$ with $R_1 > 0$ and $R_2 > 0$ is bounded below according to

$$\begin{aligned} C_{\rm WSK}(R_1, R_2) \ge \max_{U,V}[\,&|\min\{I(U \wedge Y), I(U \wedge Z)\} \\ &- I(U \wedge W)|^+ + |I(V \wedge Z|U) \\ &- I(V \wedge W|U)|^+] \end{aligned} \tag{2.27}$$

where the maximum is with respect to rv's U, V satisfying the rate conditions (2.8), (2.4), and the Markov conditions

$$U \multimap X \multimap YZW \qquad V \multimap UY \multimap ZW. \tag{2.28}$$

Corollary: The PK-capacity $C_{\rm PK}(R_1, R_2)$ with $R_1 > 0$ and $R_2 > 0$ is bounded below according to

$$C_{\rm PK}(R_1, R_2) \ge \max_{U,V} [I(V \wedge Z|U) - I(V \wedge XW|U)] \tag{2.29}$$

where the maximum is with respect to rv's U, V satisfying the rate conditions (2.8), (2.4), and the Markov conditions

$$U \multimap X \multimap YZW \qquad V \multimap UY \multimap XWZ. \tag{2.30}$$

Remark: The results above hold for achievability in the strong sense just as in Theorem 2.2.

The lower bounds in (2.27) and (2.29) are not tight in general; the next theorem and corollary show that (2.27) need not be tight for $R_1 = 0$, and (2.29) need not be tight for $R_1 = \infty$. However, Theorem 2.5 enables us to determine exactly $C_{\rm WSK}(0, R_2)$, though not by a direct substitution of $R_1 = 0$. This generalizes the corresponding result of [2] where no rate constraint was imposed on the public transmission from one user to the other. Also, this result enables us to determine exactly the PK-capacity when $R_1 = \infty$..

Theorem 2.6: The WSK-capacity without helper $C_{\rm WSK}(0, R_2)$ is

$$C_{\rm WSK}(0, R_2) = \max_{U,V} [I(V \wedge Z|U) - I(V \wedge W|U)] \tag{2.31}$$

where the maximum is with respect to rv's U, V satisfying the rate condition

$$I(V \wedge Y) - I(V \wedge Z) \le R_2 \tag{2.32}$$

and the Markov condition

$$U \multimap V \multimap Y \multimap ZW. \tag{2.33}$$

Corollary: The PK-capacity for $R_1 = \infty$ equals

$$C_{\rm PK}(\infty, R_2) = \max_{U,V} [I(V \wedge XZ|U) - I(V \wedge XW|U)] \tag{2.34}$$

where the maximum is with respect to rv's U, V satisfying the rate condition (2.11) and the Markov condition

$$U \multimap V \multimap XY \multimap ZW. \tag{2.35}$$

Special case: For PK-capacity when the eavesdropper has no side information (W = a constant), the Corollary yields

$$\begin{aligned} C_{\rm PK}(\infty, R_2) &= \max_{U,V} [I(V \wedge XZ|U) - I(V \wedge X|U)] \\ &= \max_{U,V} I(V \wedge Z|XU) \end{aligned} \tag{2.36}$$

subject to (2.11) and $U \multimap V \multimap XY \multimap Z$. The latter condition implies that

$$I(V \wedge Z|XU) \le I(V \wedge Z|X)$$

so that the maximum in (2.36) is achieved when U = a constant. Thus in this special case

$$C_{\rm PK}(\infty, R_2) = \max_V I(V \wedge Z|X) \tag{2.37}$$

subject to (2.11) and (2.12); in particular,

$$C_{\rm PK}(\infty, \infty) = I(Y \wedge Z|X) \tag{2.38}$$

which is implicit in [2].

III. Proofs

The proofs of Theorems 2.1–2.6 rely on techniques from multiuser information theory (cf. [6]), and are extensions of the proofs in [2] and [3]. All our achievability proofs are based on Lemmas A.2 and A.3 in Appendix A which elaborate the technique used in proving the forward parts of [3, Theorems 4.1 and 4.4].

We begin by introducing some notation, and refer the reader to [6] for standard terminology.

Types and joint types of n-length sequences will be represented, for convenience, by dummy rv's whose probability mass functions (pmf's) or joint pmf's coincide with the types or joint types under consideration. The type classes represented by the dummy rv's $\tilde{V}$ or $\tilde{V}, \tilde{X}$, etc., are denoted by $T^n_{\tilde{V}}$ or $T^n_{\tilde{V}\tilde{X}}$, etc. Given rv's V, X, etc. (taking values in the finite sets $\mathcal{V}$, $\mathcal{X}$, etc.), we denote by $T^n_{V,\xi}$ the set of sequences $\boldsymbol{v} \in \mathcal{V}^n$ which are V-typical with constant ξ, termed (V, ξ)-typical for brevity; similarly, we denote by $T^n_{VX,\xi}$, etc., the set of jointly (VX, ξ)-typical pairs $(\boldsymbol{v}, \boldsymbol{x}) \in \mathcal{V}^n \times \mathcal{X}^n$, etc. Thus for instance, $T^n_{VX,\xi}$ is the union of all type classes $T^n_{\tilde{V}\tilde{X}}$ such that

$$\max_{v,x} |P_{\tilde{V}\tilde{X}}(v,x) - P_{VX}(v,x)| \le \xi, P_{\tilde{V}\tilde{X}}(v,x) = 0, \quad \text{if } P_{VX}(v,x) = 0 \quad (3.1)$$

where $P_{\tilde{V}\tilde{X}}$ and P_{VX} are the joint pmf's of the rv's $\tilde{V}, \tilde{X}$, and V, X, respectively.

Whereas by $T^n_{V,\xi}, T^n_{VX,\xi}$, etc., we denote the same sets as denoted by $T^n_{[V]_\xi}, T^n_{[VX]_\xi}$, etc., in [6, p. 33], we shall use the notation $T^n_{\tilde{V}|\tilde{X}}(\boldsymbol{x})$ in a sense which differs somewhat from that in [6]. Specifically, we denote by $T^n_{\tilde{V}|\tilde{X}}(\boldsymbol{x})$ or $T^n_{V|X,\xi}(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathcal{X}^n$ (or, $T^n_{V|UX,\xi}(\boldsymbol{u}, \boldsymbol{x})$ with $\boldsymbol{u} \in \mathcal{U}^n$, etc.) the set of those $\boldsymbol{v} \in \mathcal{V}^n$ which have joint type $P_{\tilde{V}\tilde{X}}$ with $\boldsymbol{x}$ or are jointly (VX, ξ)-typical with $\boldsymbol{x}$ (or, are jointly (UVX, ξ)-typical with $(\boldsymbol{u}, \boldsymbol{x})$, etc.). Thus

$$\begin{aligned} T^n_{\tilde{V}|\tilde{X}}(\boldsymbol{x}) &= \{\boldsymbol{v} \in \mathcal{V}^n : (\boldsymbol{v}, \boldsymbol{x}) \in T^n_{\tilde{V}\tilde{X}}\} \\ T^n_{V|X,\xi}(\boldsymbol{x}) &= \{\boldsymbol{v} \in \mathcal{V}^n : (\boldsymbol{v}, \boldsymbol{x}) \in T^n_{VX,\xi}\} \\ T^n_{V|UX,\xi}(\boldsymbol{u}, \boldsymbol{x}) &= \{\boldsymbol{v} \in \mathcal{V}^n : (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x}) \in T^n_{UVX,\xi}\}. \end{aligned} \quad (3.2)$$

Note at this point that

$$(\boldsymbol{u}, \boldsymbol{x}) \in T^n_{UX,\xi} \Leftrightarrow \boldsymbol{u} \in T^n_{U|X,\xi}(\boldsymbol{x}) \Rightarrow \boldsymbol{u} \in T^n_{U,\xi|\mathcal{X}|} \quad (3.3)$$

and

$$\boldsymbol{x} \in T^n_{X,\xi} \Rightarrow T^n_{U|X,\xi}(\boldsymbol{x}) \neq \phi. \quad (3.4)$$

In addition to the standard $o(n)$ notation for any function of n whose absolute value does not exceed ϵn for any $\epsilon > 0$ if n is sufficiently large $(n \ge n_0(\epsilon))$, we shall denote by $o_\xi(n)$ any function of n and a parameter $\xi > 0$ whose absolute value does not exceed ϵn for any $\epsilon > 0$ if $n \ge n_0(\epsilon)$ and $\xi \le \xi_0(\epsilon)$. This notation will, on occasion, be used for different functions in the same equation or inequality, provided a common choice of the thresholds $n_o(\epsilon)$ and $\xi_o(\epsilon)$ is possible. With this convention, the standard bounds on the sizes of $T^n_{\tilde{X}}$ and $T^n_{\tilde{U}|\tilde{X}}(\boldsymbol{x})$, together with the continuity properties of the entropy function, permit us to write

$$\begin{aligned} |T^n_{\tilde{X}}| &= \exp\{nH(\tilde{X}) + o(n)\} \\ &= \exp\{nH(X) + o_\xi(n)\}, \quad \text{if } \phi \neq T^n_{\tilde{X}} \subset T^n_{X,\xi} \end{aligned} \quad (3.5)$$

and

$$\begin{aligned} |T^n_{\tilde{U}|\tilde{X}}(\boldsymbol{x})| &= \exp\{nH(\tilde{U}|\tilde{X}) + o(n)\} \\ &= \exp\{nH(U|X) + o_\xi(n)\}, \\ &\quad \text{if } \phi \neq T^n_{\tilde{U}|\tilde{X}}(\boldsymbol{x}) \subset T^n_{U|X,\xi}(\boldsymbol{x}). \end{aligned} \quad (3.6)$$

Also

$$P^n_X(\boldsymbol{x}) = \exp\{-nH(X) + o_\xi(n)\}, \quad \text{if } \boldsymbol{x} \in T^n_{X,\xi}. \quad (3.7)$$

If the typicality constant ξ in the hypotheses of (3.5)–(3.7) is replaced by $c\xi$ for some constant $c > 0$, the equations for the cardinalities and probabilities in (3.5)–(3.7) remain valid (since $o_{c\xi}(n)$ has the same significance as $o_\xi(n)$).

For all the proofs in this section, we observe that the claimed bounds depend continuously on the permissible transmission rates. This is proved in Lemma E.1 of Appendix E. In particular, in our bounds involving maxima under rate conditions and Markov conditions, the former can be imposed as strict inequalities if the maxima are replaced by suprema.

Proof of Theorems 2.1 and 2.2: Given the DMMS with generic rv's X, Y, Z, and auxiliary rv's U, V satisfying the Markov conditions (2.5), (2.6), and the rate conditions (2.8), (2.4) with strict inequalities, set

$$N_1 = \exp\{n[I(U \wedge X) - \min\{I(U \wedge Y), I(U \wedge Z)\} + \epsilon + \delta]\}. \quad (3.8)$$

$$N_2 = \exp\{n[\min\{I(U \wedge Y), I(U \wedge Z)\} - \epsilon]\} \quad (3.9)$$

$$N_3 = \exp\{n[I(V \wedge Y|U) - I(V \wedge Z|U) + \epsilon' + \delta']\} \quad (3.10)$$

and

$$N_4 = \exp\{n[I(V \wedge Z|U) - \epsilon']\} \quad (3.11)$$

(with the usual abuse of notation that our actual intent is to denote the smallest integers not less than the quantities on the right sides).

The idea behind the proof is as follows. First, U-typical sequences $\boldsymbol{u}_{ij}, 1 \le i \le N_1, 1 \le j \le N_2$, are selected at random. Then, with probability close to 1, to each typical $\boldsymbol{x}$ we can assign a $\boldsymbol{u}_{ij}$ jointly typical with $\boldsymbol{x}$ in such a way that if terminal $\mathcal{X}$ transmits the index i of $\boldsymbol{u}_{ij}$ assigned to $X^n = \boldsymbol{x}$, then terminals $\mathcal{Y}$ and $\mathcal{Z}$ can each reconstruct its index j (with small probability of error) by the joint typicality of $\boldsymbol{u}_{ij}$ with $Y^n = \boldsymbol{y}$ and $Z^n = \boldsymbol{z}$, respectively. Next, for each i, j, sequences $\boldsymbol{v}^{ij}_{k\ell}, 1 \le k \le N_3, 1 \le \ell \le N_4$, which are jointly typical with $\boldsymbol{u}_{ij}$ are selected at random. Then, for $\boldsymbol{y}$ and $\boldsymbol{u}_{ij}$ as above, there exists a $\boldsymbol{v}^{ij}_{k\ell}$ jointly typical with $(\boldsymbol{y}, \boldsymbol{u}_{ij})$ such that if terminal $\mathcal{Y}$ transmits the index k to terminal $\mathcal{Z}$, the latter can reconstruct ℓ by the joint typicality of $\boldsymbol{v}^{ij}_{k\ell}$ with $(\boldsymbol{z}, \boldsymbol{u}_{ij})$. We shall show that the entropy of the random quadruple (i, j, k, ℓ) (as a function of X^n, Y^n) is close to $\log N_1N_2N_3N_4$. Hence, this quadruple will represent CR of rate close to $\frac{1}{n} \log N_1N_2N_3N_4$, and the pair (j, ℓ) will represent secret CR of rate close to $\frac{1}{n} \log N_2N_4$, achievable with transmission rates $\frac{1}{n} \log N_1$ and $\frac{1}{n} \log N_3$. As the latter are less than R_1 and R_2 if ϵ, δ, ϵ', δ' are sufficiently small, this suffices to prove that $C_{\mathrm{CR}}(R_1, R_2) \ge F_1(R_1, R_2)$. (cf.(2.9)), and also Theorem 2.2. The full assertion of Theorem 2.1 will be

proved by showing that a suitable function of $\boldsymbol{x}$ can be added to (i, j, k, ℓ) to obtain a quintuple (i, j, k, ℓ, m) which achieves the CR rate claimed in (2.1).

We now provide the formal proof.

Step 1: Apply Lemma A.2 with X and U in the roles of S and V, with N_1, N_2 as in (3.8) and (3.9); first with YZ, then with Y, and finally with Z in the role of W, with $\tilde{P} = P^n_{XYZ}$ or P^n_{XY} or P^n_{XZ}, respectively. It follows that upon randomly selecting $(U, \xi|\mathcal{X}|)$-typical sequences $\boldsymbol{u}_{ij}, 1 \le i \le N_1, 1 \le j \le N_2$, the following hold with probability tending to 1 exponentially as $n \to \infty$.

a) There exist mappings $f: \mathcal{X}^n \to \{1, \cdots, N_1\}$, $\varphi: \mathcal{X}^n \to \{1, \cdots, N_2\}$, such that /belowdisplayskip8pt

$$\begin{aligned} \left(\boldsymbol{u}_{f(\boldsymbol{x})\varphi(\boldsymbol{x})}, \boldsymbol{x}\right) &\in T^n_{UX,\xi}, \qquad \text{if } \boldsymbol{x} \in T^n_{X,\xi} \\ f(\boldsymbol{x}) = \varphi(\boldsymbol{x}) &= 1, \qquad \text{if } \boldsymbol{x} \notin T^n_{X,\xi}. \end{aligned} \tag{3.12}$$

b) For any f, φ as in (3.12)

$$P^n_{XYZ}\left\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\colon (\boldsymbol{u}_{f(\boldsymbol{x})\varphi(\boldsymbol{x})}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) \in T^n_{UXYZ,\eta}\right\} > 1 - \exp(-n\tau) \tag{3.13}$$

for some $\tau > 0$ depending on ξ and δ. This and c) below follow from Part b) of Lemma A.2, since the total P^n_{XYZ}-probability of the nontypical triples $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ is exponentially small.

c) There exist mappings $\tilde{\varphi}_i: \mathcal{Y}^n \to \{1, \cdots, N_2\}$, $\tilde{\tilde{\varphi}}_i: \mathcal{Z}^n \to \{1, \cdots, N_2\}, i = 1, \cdots, N_1$, such that for any f, φ as in (3.12)

$$\Pr\left\{\varphi(X^n) = \tilde{\varphi}_{f(X^n)}(Y^n) = \tilde{\tilde{\varphi}}_{f(X^n)}(Z^n)\right\} \ge 1 - \exp(-n\tau). \tag{3.14}$$

For the sake of brevity, we shall hereafter write $f(X^n) = f, \varphi(X^n) = \varphi$. Note that by (3.12), we have

$$(\boldsymbol{u}_{ij}, \boldsymbol{x}) \in \mathcal{T}^n_{UX,\xi}, \qquad \text{if } f = i, \varphi = j, (i,j) \ne (1,1).$$

Hence, using (3.6) and (3.7)

$$\begin{aligned} \Pr\{f = i, \varphi = j\} &\le \sum_{\boldsymbol{x} \in T^n_{X|U,\xi}(\boldsymbol{u}_{ij})} P^n_X(\boldsymbol{x}) \\ &\le |T^n_{X|U,\xi}(\boldsymbol{u}_{ij})| \exp\{-nH(X) + o_\xi(n)\} \\ &= \exp\{-nI(U \wedge X) + o_\xi(n)\}, \\ &\qquad \text{if } (i,j) \ne (1,1). \end{aligned} \tag{3.15}$$

This implies that

$$H(f, \varphi) \ge nI(U \wedge X) + o_\xi(n). \tag{3.16}$$

On account of (3.14) and (3.16), if terminal $\mathcal{X}$ transmits f (with rate $\frac{1}{n}\log N_1 \cong I(U \wedge X) - \min\{I(U \wedge Y), I(U \wedge Z)\}$) to terminals $\mathcal{Y}$ and $\mathcal{Z}$, then (f, φ) will represent CR for all three terminals, achieving CR rate $\cong I(U \wedge X)$. Furthermore, a comparison of (3.16) with (3.8), (3.9) shows that φ is nearly independent of f, i.e.,

$$I(f \wedge \varphi) \le n\delta + o_\xi(n) \tag{3.17}$$

and φ achieves SK rate $\cong \min\{I(U \wedge Y), I(U \wedge Z)\}$.

Step 2: Apply Lemma A.3 with $S = Y, W = Z$, with U, V as given, with $\boldsymbol{u}$ standing for any of $\boldsymbol{u}_{ij}$ of Step 1, and with $\tilde{P} = P_{ij}$ defined by

$$P_{ij}(\boldsymbol{y}, \boldsymbol{z}) = \Pr\{Y^n = \boldsymbol{y}, Z^n = \boldsymbol{z} | f = i, \varphi = j\}. \tag{3.18}$$

Let N_3 and N_4 in (3.10), (3.11) play the roles of N_1 and N_2, respectively, in (A.15). Let ϵ', η', δ', ξ' denote small positive numbers, related to each other as required in Lemma A.3, and related to $\epsilon, \eta, \delta, \xi$ of Step 1 as required in the proof below; in particular, we shall need

$$\eta \le \xi' |\mathcal{X}|^{-1} |\mathcal{Z}|^{-1}. \tag{3.19}$$

It follows by Lemma A.3 that for each $1 \le i \le N_1, 1 \le j \le N_2$, there exist sequences $\boldsymbol{v}^{ij}_{kl} \in T^n_{V|U,\xi'|\mathcal{Y}|}(\boldsymbol{u}_{ij})$, $k = 1, \cdots, N_3$, $\ell = 1, \cdots, N_4$, and functions $g(i, j, \boldsymbol{y})$ (with range $\{1, \cdots, N_3\}$) and $\gamma(i, j, \boldsymbol{y}), \tilde{\gamma}_k(i, j, \boldsymbol{z})$ (with range $\{1, \cdots, N_4\}$), such that $g(i, j, \boldsymbol{y}) = k$ and $\gamma(i, y, \boldsymbol{y}) = \ell$ satisfy

$$\begin{aligned} &(\boldsymbol{v}^{ij}_{kl}, \boldsymbol{y}) \in T^n_{VY|U,\xi'}(\boldsymbol{u}_{ij}), \\ &\qquad \text{if } \boldsymbol{y} \in T^n_{Y|U,\xi'}(\boldsymbol{u}_{ij}), \qquad k = \ell = 1, \text{otherwise} \end{aligned} \tag{3.20}$$

$$\begin{aligned} &\left(\boldsymbol{v}^{ij}_{kl}, \boldsymbol{y}, \boldsymbol{z}\right) \in T^n_{VYZ|U,\eta'}(\boldsymbol{u}_{ij}), \\ &\qquad \text{if } (\boldsymbol{y}, \boldsymbol{z}) \in T^n_{YZ|U,\xi'|\mathcal{Z}|^{-1}}(\boldsymbol{u}_{ij}) \backslash D_{ij} \end{aligned} \tag{3.21}$$

$$\begin{aligned} &\gamma(i, j, \boldsymbol{y}) = \tilde{\gamma}_k(i, j, \boldsymbol{z}), \\ &\qquad \text{if } (\boldsymbol{y}, \boldsymbol{z}) \in T^n_{YZ|U,\xi'|\mathcal{Z}|^{-1}}(\boldsymbol{u}_{ij}) \backslash D_{ij} \end{aligned} \tag{3.22}$$

where

$$P_{ij}(D_{ij}) < 2\exp(-n\delta'). \tag{3.23}$$

Then, (3.18) together with (3.22), (3.23) yield

$$\begin{aligned} \Pr\{\gamma(f, \varphi, Y^n) &= \tilde{\gamma}_{g(f,\varphi,Y^n)}(f, \varphi, Z^n)\} \\ &= \sum_{i,j} \Pr\{f = i, \varphi = j\} \\ &\quad \cdot P_{ij}\{(\boldsymbol{y}, \boldsymbol{z})\colon \gamma(i, j, \boldsymbol{y}) = \tilde{\gamma}_{g(i,j,\boldsymbol{y})}(i, j, \boldsymbol{z})\} \\ &\ge P^n_{XYZ}\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})\colon (\boldsymbol{y}, \boldsymbol{z}) \\ &\qquad \in T^n_{U|YZ,\xi'|\mathcal{Z}|^{-1}}(\boldsymbol{u}_{f(\boldsymbol{x})\varphi(\boldsymbol{x})})\} \\ &\quad - 2\exp(-n\delta'). \end{aligned} \tag{3.24}$$

On account of (3.13) and (3.19), the lower bound in (3.24) is exponentially close to 1. Further, by (3.14), the probability

that replacing $\varphi = \varphi(X^n)$ by either $\tilde{\varphi}_f = \tilde{\varphi}_{f(X^n)}(Y^n)$ or $\tilde{\tilde{\varphi}}_f = \tilde{\tilde{\varphi}}_{f(X^n)}(Z^n)$, leads to an error, is exponentially small. Hence, (3.24) implies

$$\gamma(f, \tilde{\varphi}_f, Y^n) = \tilde{\gamma}_{g(f,\tilde{\varphi}_f,Y^n)}(f, \tilde{\tilde{\varphi}}_f, Z^n) \tag{3.25}$$

with probability exponentially close to 1.

Let terminal $\mathcal{Y}$ transmit $g = g(f, \tilde{\varphi}_f, Y^n)$, with rate

$$(1/n)\ \log N_3 \cong I(V \wedge Y|U) - I(V \wedge Z|U)$$

after it and terminal $\mathcal{Z}$ have received f. Then, by (3.25), $\gamma = \gamma(f, \tilde{\varphi}_f, Y^n)$ can be recovered by terminal $\mathcal{Z}$ with exponentially small probability of error. Thus (f, φ, g, γ) represents CR for terminals $\mathcal{Y}$ and $\mathcal{Z}$. We claim that $\frac{1}{n}H(f, \varphi, g, \gamma)$ is close to /belowdisplayskip8.6pt

$$\frac{1}{n}\ \log N_1 N_2 N_3 N_4 \cong I(U \wedge X) + I(V \wedge Y|U).$$

This will establish that (f, φ, g, γ) achieves CR rate $\cong I(U \wedge X) + I(V \wedge Y|U)$, and, in view of (3.8)–(3.11), also that (φ, γ) is nearly independent of $(f,\ g)$ and achieves SK rate $\cong \min\{I(U \wedge Y), I(U \wedge Z)\} + I(V \wedge Z|U)$ (just as (3.16) led to (3.17)).

Now, set

$$E = \{(\boldsymbol{x}, \boldsymbol{y})\colon (\boldsymbol{u}_{f(\boldsymbol{x})\varphi(\boldsymbol{x})}, \boldsymbol{x}, \boldsymbol{y}) \in T^n_{UXY,\eta|\mathcal{Z}|}\}. \tag{3.26}$$

Then, by (3.13)

$$P^n_{XY}(E)\ >\ 1 - \exp(-n\tau). \tag{3.27}$$

On account of (3.19), $(\boldsymbol{x}, \boldsymbol{y}) \in E$ implies that $\boldsymbol{y} \in T^n_{U|Y,\xi'}(\boldsymbol{u}_{ij})$ for $i = f(\boldsymbol{x}), j = \varphi(\boldsymbol{x})$, and, hence, in turn that

$$(\boldsymbol{u}_{ij}, \boldsymbol{v}^{ij}_{kl}, \boldsymbol{y}) \in T^n_{UVY,\xi'}, \qquad \text{for } k = g(i, j, \boldsymbol{y}),\ \ell = \gamma(i, j, \boldsymbol{y})$$

by (3.20). It follows that

$$\begin{aligned}
&\Pr\{f = i, \varphi = j, g = k, \gamma = \ell, (X^n, Y^n) \in E\} \\
&\quad \le \sum_{\boldsymbol{y} \in T^n_{Y|UV,\xi'}(\boldsymbol{u}_{ij}, \boldsymbol{v}^{ij}_{kl})}\ \sum_{\boldsymbol{x} \in T^n_{X|UY,\eta|\mathcal{Z}|}(\boldsymbol{u}_{ij}, \boldsymbol{y})} P^n_{XY}(\boldsymbol{x}, \boldsymbol{y}) \\
&\quad \le \exp\{nH(Y|UV) + o_{\xi'}(n)\} \\
&\qquad \cdot \exp\{nH(X|UY) + o_\eta(n)\} \\
&\qquad \cdot \exp\{-nH(X, Y) + o_\eta(n)\} \\
&\quad = \exp\{-n(I(U \wedge X) + I(V \wedge Y|U)) + o_{\xi'}(n)\}.
\end{aligned} \tag{3.28}$$

Here, the second inequality is a consequence of (3.6), (3.7), while the last equality follows by simple algebra using the fact that η is less than ξ' multiplied by a constant. Finally, (3.27) and (3.28) yield

$$\begin{aligned}
H(f, \varphi, g, \gamma) &\ge P^n_{XY}(E) H(f, \varphi, g, \gamma|(X^n, Y^n) \in E) \\
&\ge n[I(U \wedge X) + I(V \wedge Y|U)] + o_{\xi'}(n)
\end{aligned} \tag{3.29}$$

establishing our claim.

This completes the proof of $C_{\mathrm{CR}}(R_1, R_2) \ge F_1(R_2, R_2)$, and also of Theorem 2.2 (albeit not with the strong achievability of the SK-rate, to which we shall soon return). In order to complete the proof of Theorem 2.1, viz. (2.1), it suffices to show that if $t \ge 0$ satisfies (2.3) and (2.2) with strict inequality, there exists a function $\psi\colon \mathcal{X}^n \to \{1, \cdots, N_5\}$ with $N_5 = \exp(nt)$, such that $\psi(X^n)$ is almost uniformly distributed and is almost independent of $(f(X^n), Y^n)$. Indeed, then $\psi = \psi(X^n)$ is almost independent of (f, φ, g, γ). Hence, if terminal $\mathcal{X}$ transmits ψ—in addition to f—with total rate $\frac{1}{n}\ \log N_1 N_5 \le R_1$, then $(f, \varphi, g, \gamma, \psi)$ will represent CR of rate close to /belowdisplayskip9pt

$$\frac{1}{n}H(f, \varphi, g, \gamma) + t \cong I(U \wedge X) + I(V \wedge Y|Z) + t$$

as required.

The existence of ψ with the properties above follows by Lemma A.4 in Appendix A. In order to apply that lemma, we now show that the probability that (X^n, Y^n) belongs to

$$G = \{(\boldsymbol{x}, \boldsymbol{y})\colon \Pr\{X^n = \boldsymbol{x}|f(X^n) = f(\boldsymbol{x}), Y^n = \boldsymbol{y}\}\ >\ L^{-1}\} \tag{3.30}$$

is arbitrarily small (for n sufficiently large) if

$$L \le \exp\{n(H(X|YU) - \epsilon)\}. \tag{3.31}$$

On account of (3.27), it suffices to bound $P^n_{XY}(G \cap E)$. Since the definition of E in (3.26) implies that pairs $(\boldsymbol{x}, \boldsymbol{y}) \in E$ are jointly typical (with constant $\eta|\mathcal{U}||\mathcal{Z}|$), the probability of any such pair is $\exp\{-nH(XY) + o_\eta(n)\}$. It follows that for

$$\boldsymbol{x} \in E(i, \boldsymbol{y}) = \{\boldsymbol{x}\colon f(\boldsymbol{x}) = i, (\boldsymbol{x}, \boldsymbol{y}) \in E\}$$

we have

$$\begin{aligned}
&\Pr\{X^n = \boldsymbol{x}|f(X^n) = f(\boldsymbol{x}), Y^n = \boldsymbol{y}\} \\
&\qquad \ge \frac{P^n_{XY}(\boldsymbol{x}, \boldsymbol{y})}{\sum\limits_{\boldsymbol{x}' \in E(i,\boldsymbol{y})} P^n_{XY}(\boldsymbol{x}', \boldsymbol{y})} \\
&\qquad = |E(i, \boldsymbol{y})|^{-1} \exp\{o_\eta(n)\}.
\end{aligned} \tag{3.32}$$

Thus if $\boldsymbol{x} \in E(i, \boldsymbol{y})$, a necessary condition for $(\boldsymbol{x}, \boldsymbol{y}) \in G$ is $|E(i, \boldsymbol{y})| < L \exp\{o_\eta(n)\}$. It follows that

$$\begin{aligned}
P^n_{XY}(G \cap E) &= \sum_{(i,\boldsymbol{y})}\ \sum_{\substack{\boldsymbol{x} \in E(i,\boldsymbol{y}), \\ (\boldsymbol{x},\boldsymbol{y}) \in G}} P^n_{XY}(\boldsymbol{x}, \boldsymbol{y}) \\
&\le |\{(i, \boldsymbol{y})\colon E(i, \boldsymbol{y}) \ne \phi\}| L \\
&\qquad \cdot \exp\{-nH(XY) + o_\eta(n)\}.
\end{aligned} \tag{3.33}$$

Since $\boldsymbol{x} \in E(i, \boldsymbol{y})$ means by definition that $(\boldsymbol{u}_{ij}, \boldsymbol{x}, \boldsymbol{y})$, with $i = f(\boldsymbol{x}),\ j = \varphi(\boldsymbol{x})$, is jointly typical (with constant $\eta|\mathcal{Z}|$), the set $E(i, \boldsymbol{y})$ can be nonempty only for $\exp\{nH(Y) + o_\eta(n)\}$ (typical) sequences $\boldsymbol{y}$. Moreover, if for a given $\boldsymbol{y}$ there exists any i with $E(i, \boldsymbol{y}) \ne \phi$, the number of such i's is bounded

above by the number of those sequences $\boldsymbol{u}_{ij}$ from the $N_1N_2 = \exp\{n(I(X \wedge Y) + \delta)\}$ randomly selected sequences which are jointly typical with $\boldsymbol{y}$. By Lemma A.1 in Appendix A, that number is /belowdisplayskip7pt

$$\exp\{n(I(X \wedge Y) + \delta - I(U \wedge Y)) + o_{\xi+\eta}(n)\}$$

(with probability exponentially close to 1). Hence, (3.33) gives, for L as in (3.31), that

$$\begin{aligned} P_{XY}^n(G \cap E) &\le \exp\{nH(Y) + o_\eta(n)\} \\ &\quad \cdot \exp\{n(I(U \wedge X) + \delta - I(U \wedge Y)) + o_{\xi+\eta}(n))\}. \\ &\quad \cdot \exp\{n(H(X|YU) - \epsilon)\} \exp\{-nH(XY) + o_\eta(n)\} \\ &= \exp\{-n(\epsilon - \delta) + o_{\xi+\eta}(n)\}. \end{aligned} \tag{3.34}$$

Recalling that ϵ, η, δ, and ξ can be arbitrarily chosen subject to the conditions in Lemma A.2, for an appropriate choice of these constants the bound in (3.34) establishes our claim regarding the smallness of $\Pr\{(X^n, Y^n) \in G\}$. Setting $L = \exp(nt)$, the last result implies, by Lemma A.4, the existence of $\psi : \mathcal{X}^n \to \{1, \cdots, N_5\}$ as claimed. This completes the proof of Theorem 2.1.

In order to prove that Theorem 2.2 also holds with strong achievability, we now show that we can use a function $F = F(\varphi, \gamma)$ of effectively the same rate as φ, γ, for which $I(f, g \wedge F)$ is exponentially small. To this end, we again turn to Lemma A.4 in Appendix A and choose

$$\begin{aligned} A &= \{(j,\ell) \colon 1 \le j \le N_2,\ 1 \le \ell \le N_4\} \\ B &= \{(i,k) \colon 1 \le i \le N_1, 1 \le k \le N_3\} \\ \tilde{P}(i,j,k,\ell) &= \Pr\{f = i, \varphi = j, g = k, \gamma = \ell | (X^n, Y^n) \in E\}. \end{aligned}$$

Also, write

$$\begin{aligned} C = \{(i,k) \colon \Pr\{f = i, g = k | (X^n, Y^n) \in E\} \\ \ge N_1^{-1} N_2^{-1} \exp(-n\delta')\}. \end{aligned} \tag{3.35}$$

Then, by (3.28)

$$\begin{aligned} \Pr\{\varphi = j, \gamma = \ell | f = i, g = k, (X^n, Y^n) \in E\} \\ \le \exp\{-n[I(U \wedge X) + I(V \wedge Y|U)] \\ + o_\xi(n)\} N_1 N_2 \exp(n\delta') \end{aligned} \tag{3.36}$$

if $(i, k) \in C$. Since there exist N_1N_3 possible pairs (i, k), (3.35) implies that the $\tilde{P}$-probability of $(i, k) \notin C$ is less than $\exp(-n\delta')$. This, together with (3.36), means that the hypothesis (A.16) of Lemma A.4 is met with

$$L = \exp\{n[I(U \wedge X) + I(V \wedge Y|U) - \delta'] + o_\xi(n)\} N_1^{-1} N_3^{-1}$$

and $\epsilon = \exp(-n\delta'/2)$. It follows that there exists $F : A \to \mathcal{K}$ satisfying (A.17), with rate $\frac{1}{n} \log|\mathcal{K}|$ which differs only by an arbitrarily small amount from $\frac{1}{n} \log L$ and, hence (by (3.8) and (3.10)), also from

$$\min\{I(U \wedge Y), I(U \wedge Z)\} + I(V \wedge Z|U).$$

Although this has been established for $\tilde{P}$ (the conditional joint pmf of (f, φ, g, γ), conditioned on the event $(X^n, Y^n) \in E$), the same result also holds for the unconditional joint pmf on account of (3.27) and the corollary of Lemma A.4. Of course, the fact that the variation distance $d_{\text{av}}(F)$ of the joint pmf of $F = F(\varphi, \gamma)$ and (f, g) from the product of the marginal pmf's, is exponentially small implies that $I(f, g \wedge F)$ is exponentially small as claimed (cf. e.g., [5, Lemma 1]).

Proof of Theorem 2.3:

Part a): We first claim that the assertions concerning $C_{\text{CR}}(R_1, R_2)$ will follow if we show that $C_{\text{CR}}(R_1, R_2)$ is bounded above by

$$F_3(R_1, R_2) = \max_{U,V,t} [I(U \wedge X) + I(V \wedge Y|U) + t] \tag{3.37}$$

where the rv's U, V satisfy the rate conditions (2.3), (2.4), and the Markov conditions (2.5), (2.6), and $t \ge 0$; note that the condition (2.2) on t is not imposed here. Indeed, we can assume that $R_1 \le \max\{H(X|Y), H(X|Z)\}$, since the complementary case has already been addressed (cf. Case 1 in Section II). Then, as shown in Appendix E, Lemma E.3, $F_3(R_1, R_2)$ is equal to $F_1(R_1, R_2)$ defined by (2.9), if either $X \multimap Z \multimap Y$ or $X \multimap Y \multimap Z$ and $R_2 \le H(Y|XZ)$. In the remaining case, when $X \multimap Y \multimap Z$ and $R_2 > H(Y|XZ)$, we can assume that $R_2 \le H(Y|Z)$ since the complementary case has already been covered (cf. Case 2 in Section II). Then, by Lemma E.4 in Appendix E, the right-hand side of (2.1), denoted by $F_2(R_1, R_2)$, is either equal to $F_3(R_1, R_2)$ or else to $H(XY)$. Since $H(XY)$ is an obvious upper bound for $C_{\text{CR}}(R_1, R_2)$, our first claim is established.

Now, let $\delta > 0$ be given. Consider any pair of functions f, g satisfying the rate constraints (1.1), (1.2), and any pair of rv's (K, L) representing ϵ-CR, i.e., satisfying (1.3) and (1.4). We show that if either of the Markov hypotheses $X \multimap Z \multimap Y$ or $X \multimap Y \multimap Z$ is satisfied, there exists rv's U, V which satisfy the Markov conditions (2.5), (2.6), and the rate conditions

$$I(U \wedge X) - \min\{I(U \wedge Y), I(U \wedge Z)\} + t \le R_1 \tag{3.38}$$

$$I(V \wedge Y|U) - I(V \wedge Z|U) \le R_2 + \delta \tag{3.39}$$

for some $t \ge 0$ such that

$$\frac{1}{n} H(K) \le I(U \wedge X) + I(V \wedge Y|U) + t \tag{3.40}$$

if ϵ in (1.4) is sufficiently small and n is suitably large. This will be established by showing that the rv's U, V in (3.43) below satisfy the conditions (2.5), (2.6), and (3.38)–(3.40) with the rv's X, Y, Z replaced by the rv's X_J, Y_J, Z_J, whose joint distribution equals that of X, Y, Z. We start by representing $H(K|fZ^n)$ as follows:

$$\begin{aligned} H(K|fZ^n) &= H(K|fZ^n) - H(K|fY^n), \\ &\qquad \text{since } K = K(f, Y^n) \\ &= I(K \wedge Y^n|f) - I(K \wedge Z^n|f) \\ &= \sum_{i=1}^n [I(K \wedge Y_i|fY^{i-1}Z_{i+1}^n) \\ &\qquad - I(K \wedge Z_i|fY^{i-1}Z_{i+1}^n)] \end{aligned} \tag{3.41}$$

with $Z_{i+1}^n = (Z_{i+1}, \cdots, Z_n)$, where the previous identity is a standard tool in multiuser information theory (cf. e.g., [2, Lemma 4.1]). Equation (3.41) can be written as /belowdisplayskip6pt

$$H(K|fZ^n) = n[I(V \wedge Y_J|U) - I(V \wedge Z_J|U)] \tag{3.42}$$

where J is an rv distributed uniformly on $\{1, \cdots, n\}$ and independent of (X^n, Y^n, Z^n), and

$$U = fY^{J-1}Z_{J+1}^n J \qquad V = K. \tag{3.43}$$

The rv's U, V in (3.43) satisfy the rate condition (3.39) on account of (1.2), (3.42), and

$$\begin{aligned} \log \|g\| &\geq H(g) \geq H(g|fZ^n) \\ &= H(gK|fZ^n) - H(K|fgZ^n) \\ &\geq H(K|fZ^n) - H(K|L) \\ &\geq H(K|fZ^n) - n\delta \end{aligned} \tag{3.44}$$

where the last inequality follows by Fano's inequality, provided that $\epsilon > 0$ is sufficiently small in (1.4). In order to show that U, V satisfy the Markov conditions, viz.

$$U \multimap X_J \multimap Y_J Z_J \tag{3.45}$$

$$V \multimap UY_J \multimap Z_J \tag{3.46}$$

note that (3.45) follows from (3.43) on account of the Markov relation

$$X^n Y^{i-1} Z_{i+1}^n \multimap X_i \multimap Y_i Z_i \tag{3.47}$$

which is obvious since $X^{i-1}X_{i+1}^n Y^{i-1}Z_{i+1}^n$ is independent of $X_iY_iZ_i$. In order to establish (3.46), recalling (3.43) and $K = K(f, Y^n)$, it suffices to show that

$$Y_{i+1}^n \multimap fY^i Z_{i+1}^n \multimap Z_i. \tag{3.48}$$

This can be verified by direct calculation, using the hypothesis $X \multimap Z \multimap Y$ or $X \multimap Y \multimap Z$ (cf. Appendix D). We remark that this Markov hypothesis on X, Y, Z is needed at this point only, and will not be used in the remainder of the proof. In particular, we shall not replace $\min\{I(U \wedge Y), I(U \wedge Z)\}$ by $I(U \wedge Y)$ or $I(U \wedge Z)$ according to the prevailing Markov assumption, since doing so would not lead to any significant simplification.

In order to establish that U, V satisfy the rate condition (3.38), observe first that

$$nR_1 \geq \log \|f\| \geq H(f) = I(f \wedge X^n). \tag{3.49}$$

This is bounded below by

$$\begin{aligned} &I(f \wedge X^n) - I(f \wedge Y^n) \\ &\quad = \sum_{i=1}^n [I(f \wedge X_i|X_{i+1}^n Y^{i-1}) - I(f \wedge Y_i|X_{i+1}^n Y^{i-1})] \\ &\quad = \sum_{i=1}^n [I(fX_{i+1}^n Y^{i-1} \wedge X_i) - I(fX_{i+1}^n Y^{i-1} \wedge Y_i)] \\ &\quad = \sum_{i=1}^n [I(fX_{i+1}^n Y^{i-1} Z_{i+1}^n \wedge X_i) \\ &\qquad\quad - I(fX_{i+1}^n Y^{i-1} Z_{i+1}^n \wedge Y_i)] \\ &\quad = \sum_{i=1}^n [I(fY^{i-1} Z_{i+1}^n \wedge X_i) - I(fY^{i-1} Z_{i+1}^n \wedge Y_i)] \\ &\qquad + \sum_{i=1}^n [I(X_{i+1}^n \wedge X_i|fY^{i-1}Z_{i+1}^n) \\ &\qquad\quad - I(X_{i+1}^n \wedge Y_i|fY^{i-1}Z_{i+1}^n)] \\ &\quad \geq \sum_{i=1}^n [I(fY^{i-1} Z_{i+1}^n \wedge X_i) - I(fY^{i-1} Z_{i+1}^n \wedge Y_i)] \\ &\quad = n[I(U \wedge X_J|J) - I(U \wedge Y_J|J)] \\ &\quad = n[I(U \wedge X_J) - I(U \wedge Y_J)]. \end{aligned} \tag{3.50}$$

Here, the first equality above holds by the identity used also in (3.41), the second by the independence of (X_i, Y_i) from (X_{i+1}^n, Y^{i-1}), and the third by the Markov relation

$$Z_{i+1}^n \multimap fX_{i+1}^n Y^{i-1} \multimap X_i Y_i;$$

the latter is implied by

$$Z_{i+1}^n \multimap X_{i+1}^n \multimap fX_i Y^i$$

which follows from the consequence

$$Z_{i+1}^n \multimap X_{i+1}^n \multimap X^n Y^i$$

of the independence of X^iY^i and $X_{i+1}^n Z_{i+1}^n$. The inequality in (3.50) holds since

$$I(X_{i+1}^n \wedge X_i|fY^{i-1}Z_{i+1}^n) = I(X_{i+1}^n \wedge X_iY_i|fY^{i-1}Z_{i+1}^n)$$

owing to the Markov relation

$$X_{i+1}^n \multimap fY^{i-1}Z_{i+1}^n X_i \multimap Y_i;$$

the latter is implied by

$$X_{i+1}^n fY^{i-1}Z_{i+1}^n \multimap X_i \multimap Y_i$$

which is a consequence of (3.47). Finally, the last equalities in (3.50) follow by (3.43), and the independence of X_J and Y_J from J. Similarly, the right-hand side of (3.49) is also bounded below by

$$\begin{aligned} &I(f \wedge X^n) - I(f \wedge Z^n) \\ &\qquad = \sum_{i=1}^n [I(f \wedge X_i|X^{i-1}Z_{i+1}^n) - I(f \wedge Z_i|X^{i-1}Z_{i+1}^n)] \end{aligned}$$

which, in comparison with (3.50), entails going "backward." Proceeding along the lines which led to (3.50), we get

$$I(f \wedge X^n) - I(f \wedge Z^n) \geq n[I(U \wedge X_J) - I(U \wedge Z_J)]. \tag{3.51}$$

A comparison of (3.49)–(3.51) shows that

$$\frac{1}{n}H(f) = I(U \wedge X_J) - \min\{I(U \wedge Y_J), I(U \wedge Z_J)\} + t \tag{3.52}$$

where $t \geq 0$ satisfies (3.38). From (3.52) and (3.49), it follows that U, V are in conformity with (3.38).

It remains to bound $\frac{1}{n}H(K)$ as in (3.40). To this end, we proceed as follows:

$$\begin{aligned} H(K) &= I(K \wedge fY^n), \quad \text{since } K = K(f, Y^n) \\ &= I(K \wedge f) + I(K \wedge Y^n|f) \\ &\leq H(f) + \sum_{i=1}^n I(K \wedge Y_i|fY^{i-1}) \end{aligned} \tag{3.53}$$

$$\begin{aligned}&\le H(f)+\sum_{i=1}^{n} I(KZ_{i+1}^n \wedge Y_i|fY^{i-1})\\ &\le H(f)+\sum_{i=1}^{n} [I(fY^{i-1}Z_{i+1}^n \wedge Y_i)\\ &\qquad + I(K\wedge Y_i|fY^{i-1}Z_{i+1}^n)]\\ &= H(f)+n[I(U\wedge Y_J)+I(V\wedge Y_J|U)]\end{aligned}$$

so that

$$\frac{1}{n}H(K) \le \frac{1}{n}H(f)+I(U\wedge Y_J)+I(V\wedge Y_J|U). \quad (3.54)$$

Note further that

$$\begin{aligned}H(K) &= I(K\wedge fZ^n)+H(K|fZ^n)\\ &= I(K\wedge f)+I(K\wedge Z^n|f)+H(K|fZ^n). \quad (3.55)\end{aligned}$$

Similarly as the right-hand side of (3.53) was bounded above by (3.54), we have

$$\begin{aligned}&I(K\wedge f)+I(K\wedge Z^n|f)\\ &\qquad \le n[H(f)+I(U\wedge Z_J)+I(V\wedge Z_J|U)]. \quad (3.56)\end{aligned}$$

Hence, and by (3.42), it follows from (3.55) that

$$\frac{1}{n}H(K) \le \frac{1}{n}H(f)+I(U\wedge Z_J)+I(V\wedge Y_J|U). \quad (3.57)$$

Combining (3.54) and (3.57), we get

$$\begin{aligned}\frac{1}{n}H(K) \le \frac{1}{n}\,H(f)&+\min\{I(U\wedge Y_J), I(U\wedge Z_J)\}\\ &+I(V\wedge Y_J|U). \quad (3.58)\end{aligned}$$

The desired upper bound for $\frac{1}{n}H(K)$ in (3.40) is then obtained by substituting in (3.58) the expression for $\frac{1}{n}H(f)$ from (3.52)

$$\frac{1}{n}H(K) \le I(U\wedge X_J)+I(V\wedge Y_J|U)+t \quad (3.59)$$

where $t \ge 0$ satisfies (3.38).

Part b):: Consider any pair of functions of f, g satisfying the rate constraints (1.1), (1.2), and any pair of rv's (K, L) which satisfy (1.3), (1.4), and the secrecy condition (1.8).

Since the rv's $K' = (K, f, g)$ and $L' = (L, f, g)$ are, in effect, functions of f, Y^n and f, g, Z^n, respectively, and (K', L') satisfies (1.4) as (K, L) does so, clearly (K', L') represents ϵ-CR. Hence, (3.59) applies with K' in the role of K, yielding

$$\frac{1}{n}\,H(K') \le I(U\wedge X_J)+I(V\wedge Y_J|U)+t \quad (3.60)$$

where the rv's U, V are given by (3.43) with K' replacing K, and $t \ge 0$ satisfies (3.38).

Next, we bound below $\frac{1}{n}H(K')$ as follows:

$$\begin{aligned}\frac{1}{n}\,H(K') &= \frac{1}{n}\,H(K,f,g) && (3.61)\\ &= \frac{1}{n}\,H(K)+\frac{1}{n}\,H(fg|K) && (3.62)\\ &\ge \frac{1}{n}\,H(K)+\frac{1}{n}\,H(fg)-\epsilon, \quad \text{by (1.8)} && (3.63)\\ &= \frac{1}{n}\,H(K)+\frac{1}{n}\,H(f)+\frac{1}{n}\,H(g|f)-\epsilon. && (3.64)\end{aligned}$$

The lower and upper bounds for $\frac{1}{n}H(K')$ in (3.64) and (3.60), respectively, yield

$$\begin{aligned}\frac{1}{n}\,H(K)+\frac{1}{n}\,H(f)+\frac{1}{n}\,H(g|f)-\epsilon&\\ \le I(U\wedge X_J)+I(V\wedge Y_J|U)+t& \quad (3.65)\end{aligned}$$

whence

$$\begin{aligned}\frac{1}{n}\,H(K) \le I(U\wedge X_J)+I(V\wedge Y_J|U)+t&\\ -\frac{1}{n}H(f)-\frac{1}{n}H(g|f)+\epsilon.& \quad (3.66)\end{aligned}$$

To complete the proof, we replace $\frac{1}{n}\,H(f)$ and $\frac{1}{n}\,H(g|f)$ in (3.66) by appropriate terms. Specifically, we obtain from (3.52) that

$$\frac{1}{n}H(f) = I(U\wedge X_J)-\min\{I(U\wedge Y_J), I(U\wedge Z_J)\}+t, \quad (3.67)$$

and from (3.44) and (3.42) that

$$\begin{aligned}\frac{1}{n}H(g|f) &\ge \frac{1}{n}H(g|fZ^n) && (3.68)\\ &\ge I(V\wedge Y_J|U)-I(V\wedge Z_J|U)-n\delta. && (3.69)\end{aligned}$$

Upon substituting (3.67) and (3.69) in (3.66), we finally obtain

$$\begin{aligned}\frac{1}{n}H(K) \le \min\{I(U\wedge Y_J), I(U\wedge Z_J)\}&\\ +I(V\wedge Z_J|U)+\epsilon+\delta.& \quad (3.70)\end{aligned}$$

The asserted tightness of the lower bound for $C_{\mathrm{SK}}(R_1, R_2)$ in (2.7) is now evident from (3.70).

Proof of Theorem 2.4::

Forward Part: The forward part follows simply by Step 1 of the proof of Theorems 2.1 and 2.2 above. Specifically, (f, φ) and φ from said step represent, respectively, the "three-way" CR and "three-way" SK with the desired rates.

Converse Part: We first prove the converse part pertaining to the "three-way" CR-capacity $C_{\mathrm{CR}}^3(R_1, 0)$. Consider any function $f = f(X^n)$ which satisfies the rate constraint (1.1), and take $g = g(Y^n, f) =$ a constant (since $R_2 = 0$). Consider a triple of rv's (M, K, L) as in (1.15). We shall show the existence of an rv U satisfying the rate condition (2.8) with R_1 replaced by $R_1+\delta$, and the Markov condition (2.5), such that

$$\frac{1}{n}\,H(M) = I(U\wedge X) \quad (3.71)$$

where $\delta > 0$ is arbitrarily small if ϵ in (1.15) is sufficiently small. As in the proof of Theorem 2.3, the conditions above will be established with X_J, Y_J, Z_J in the roles of X, Y, Z.

We can bound $H(f)$ below according to

$$\begin{aligned}H(f) \ge H(f|Y^n) &= H(fM|Y^n)-H(M|fY^n)\\ &\ge H(fM|Y^n)-H(M|K),\\ &\qquad \text{since } K = K(f, Y^n)\\ &\ge H(M|Y^n)-H(M|K)\\ &\ge H(M|Y^n)-n\delta \quad (3.72)\end{aligned}$$

where the last inequality follows by (1.15) and Fano's inequality, provided that $\epsilon > 0$ is sufficiently small. Now

$$H(M|Y^n) = H(Y^n|M) + H(M) - H(Y^n) \tag{3.73}$$

and

$$\begin{aligned} H(Y^n|M) &= \sum_{i=1}^{n} H(Y_i|MY^{i-1}) \\ &\geq \sum_{i=1}^{n} H(Y_i|MY^{i-1}X^{i-1}) \\ &= \sum_{i=1}^{n} H(Y_i|MX^{i-1}) \\ &= nH(Y_J|U) \end{aligned} \tag{3.74}$$

where J is uniformly distributed on $\{1, \cdots, n\}$ and is independent of (X^n, Y^n), and $U = MX^{J-1}J$. Also

$$\begin{aligned} H(M) &= I(M \wedge X^n), \quad \text{since } M = M(X^n) \\ &= \sum_{i=1}^{n} I(M \wedge X_i|X^{i-1}) \\ &= nI(M \wedge X_J|X^{J-1}J) \\ &= nI(MX^{J-1}J \wedge X_J) \\ &= nI(U \wedge X_J). \end{aligned} \tag{3.75}$$

Substituting (3.73)–(3.75) in (3.72), we get

$$\begin{aligned} \frac{1}{n}\, H(f) &\geq H(Y_J|U) + I(U \wedge X_J) - H(Y_J) - \delta \\ &= I(U \wedge X_J) - I(U \wedge Y_J) - \delta. \end{aligned} \tag{3.76}$$

A bound similar to (3.76) with Z_J replacing Y_J can be obtained by commencing as in (3.72) with Z^n in lieu of Y^n. Combining these two bounds, we obtain

$$\frac{1}{n}H(f) \geq I(U \wedge X_J) - \min\{I(U \wedge Y_J), I(U \wedge Z_J)\} - \delta. \tag{3.77}$$

Since $\frac{1}{n}H(f) \leq (1/n)\, \log \|f\| \leq R_1$ (cf. (1.1)), we see from (3.77) and (3.75) that the rv $U = MX^{J-1}J$ meets the claims set forth in the first passage above. Note that U satisfies the Markov condition (3.45) on account of (3.47). This completes the desired converse for $C^3_{\mathrm{CR}}(R_1, 0)$.

We turn next to the converse part for the "three-way" SK-capacity $C^3_{\mathrm{SK}}(R_1, 0)$. Consider the functions f, g as above, and let the triple of rv's (M, K, L) in (1.15) additionally satisfy the secrecy condition (1.8). Then the triple of rv's $M' = (M, f)$, $K' = (K, f)$ and $L' = (L, f)$ represent "three-way" ϵ-CR in the sense of (1.15). Consequently, by (3.75)

$$\frac{1}{n}\, H(M') = I(U \wedge X_J) \tag{3.78}$$

where the rv U is now given by $U = M'X^{J-1}J$. Also

$$\begin{aligned} \frac{1}{n}\, H(M') &= \frac{1}{n}\, H(M, f) = \frac{1}{n}H(M) + \frac{1}{n}H(f|M) \\ &\geq \frac{1}{n}\, H(M) + \frac{1}{n}H(f) - \epsilon - \delta \end{aligned} \tag{3.79}$$

where the inequality follows from (1.8), (1.15), and Fano's inequality provided that $\epsilon > 0$ is sufficiently small. Combining (3.77)–(3.79) we obtain

$$\frac{1}{n}H(M) \leq \min\{I(U \wedge Y_J), I(U \wedge Z_J)\} + \epsilon + 2\delta. \tag{3.80}$$

The desired converse is then provided by (3.80) and (3.77), noting that the rv $U = M'X^{J-1}J$ satisfies the Markov condition (3.45).

Proof of Theorem 2.5: Given X, Y, Z, W, and auxiliary rv's U, V satisfying the Markov conditions (2.28), we proceed to construct functions $f(x), \varphi(x), g(i, j, \boldsymbol{y}), \gamma(i, j, \boldsymbol{y})$ as in the proof of Theorems 2.1 and 2.2. In short, the approach entails that terminal $\mathcal{X}$ will transmit $f = f(X^n)$ enabling both terminals $\mathcal{Y}$ and $\mathcal{Z}$ to reconstruct $\varphi = \varphi(X^n)$ with exponentially small probability of error; then, terminal $\mathcal{Y}$ will transmit $g = g(f, \varphi, Y^n)$ enabling terminal $\mathcal{Z}$ to reconstruct $\gamma = \gamma(f, \varphi, Y^n)$. Next, a function $F = F(\varphi, \gamma)$ is sought which remains secret in the strong sense from an eavesdropper who observes f, g, and additionally, W^n, in that $I(f, g, W^n \wedge F)$ is exponentially small, and this F is almost uniformly distributed on a set $\mathcal{K}$ with $\frac{1}{n}\, \log|\mathcal{K}|$ close to

$$\begin{aligned} &|\min\{I(U \wedge Y), I(U \wedge Z)\} - I(U \wedge W)|^+ \\ &\qquad + |I(V \wedge Z|U) - I(V \wedge W|U)|^+. \end{aligned}$$

This function $F = F(\varphi, \gamma)$ will be obtained using Lemma A.4, as was done in a simpler situation toward the end of the proof of Theorem 2.2. As might be expected, however, in the case $I(V \wedge Z|U) \leq (V \wedge W|U)$ a simpler strategy is required, with terminal $\mathcal{Y}$ not transmitting at all, and then a suitable $F = F(\varphi)$ suffices with $I(f, W^n \wedge F)$ exponentially small.

In preparation for the proof, the following additions are needed to the basic construction used in the proof of Theorems 2.1 and 2.2.

In Step 1, we apply Lemma A.2 also with YW playing the role of W, whereby we obtain in a manner similar to (3.13) that

$$\begin{aligned} P^n_{XYW}\{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})\colon (\boldsymbol{u}_{f(\boldsymbol{x})\varphi(\boldsymbol{x})}, f(\boldsymbol{x}), \boldsymbol{y}, \boldsymbol{w}) \in T^n_{UXYW,\eta}\} \\ > 1 - \exp(-n\tau). \end{aligned} \tag{3.81}$$

In Step 2, we apply Lemma A.3 also with (Y in the role of S and) the present W in the role of W, and with the pmf P'_{ij} on $\mathcal{Y}^n \times \mathcal{W}^n$ defined by

$$P'_{ij}(\boldsymbol{y}, \boldsymbol{w}) = \Pr\{Y^n = \boldsymbol{y}, W^n = \boldsymbol{w}|f = i, \varphi = j\} \tag{3.82}$$

playing the role of $\tilde{P}$. Then, similarly as in (3.21) and (3.23), we obtain that $g(i, j, \boldsymbol{y}) = k$ and $\gamma(i, j, \boldsymbol{y}) = \ell$ satisfy

$$\begin{aligned} \left(\boldsymbol{v}^{ij}_{k\ell}, \boldsymbol{y}, \boldsymbol{w}\right) &\in T^n_{VYW|U,\eta'}(\boldsymbol{u}_{ij}), \qquad \text{if } (\boldsymbol{y}, \boldsymbol{w}) \\ &\in T^n_{YW|U,\xi'|\mathcal{W}|^{-1}}(\boldsymbol{u}_{ij}) \backslash D'_{ij} \end{aligned} \tag{3.83}$$

where

$$P'_{ij}(D'_{ij}) < 2\exp(-n\delta'). \tag{3.84}$$

Instead of (3.26), we now define

$$\begin{aligned} E = \{(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w})\colon (\boldsymbol{u}_{f(\boldsymbol{x})\varphi(\boldsymbol{x})}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}) \\ \in T^n_{UXYW,\eta}, (\boldsymbol{y}, \boldsymbol{w}) \notin D'_{f(\boldsymbol{x})\varphi(\boldsymbol{x})}\}. \end{aligned} \tag{3.85}$$

In addition to (3.19), we now also assume that $\eta \le \xi'|\mathcal{X}|^{-1}|\mathcal{W}|^{-1}$, whereby the first condition in the definition of E above implies $(\boldsymbol{y},\boldsymbol{w}) \in T^n_{YW|U,\xi'|\mathcal{W}|^{-1}}(\boldsymbol{u}_{f(\boldsymbol{x})\varphi(\boldsymbol{x})})$. With this, (3.83) gives

$$(\boldsymbol{u}_{ij},\boldsymbol{v}^{ij}_{k\ell},\boldsymbol{y},\boldsymbol{w}) \in T^n_{UVYW,\eta'}, \qquad \text{if } (\boldsymbol{x},\boldsymbol{y},\boldsymbol{w}) \in E. \quad (3.86)$$

Also, (3.81) and (3.84) show that $P^n_{XYW}(E)$ is exponentially close to 1.

For the application of Lemma A.4, it will be convenient to use the pmf on quintuples $(i, j, k, \ell, \boldsymbol{w})$ defined by

$$\tilde{P}(i,j,k,\ell,\boldsymbol{w}) = \Pr\{f=i,\varphi=j,g=k,\gamma=\ell, W^n=\boldsymbol{w}|(X^n,Y^n,W^n)\in E\}. \quad (3.87)$$

Then, in the final result, we get rid of the conditioning by the corollary of Lemma A.4, using the fact that $P^n_{XYW}(E)$ is exponentially close to 1, just as in the proof of Theorem 2.2.

A necessary condition for $\tilde{P}(i, j, k, \ell, \boldsymbol{w}) > 0$ is

$$(\boldsymbol{u}_{ij},\boldsymbol{v}^{ij}_{k\ell},\boldsymbol{w}) \in T^n_{UVW,\eta'|\mathcal{Y}|} \quad (3.88)$$

(cf. (3.86)). The quintuples $(i, j, k, \ell, \boldsymbol{w})$ satisfying (3.88) will be called possible quintuples. Using (3.6) and (3.8)–(3.11), their number is

$$\begin{aligned} M &= N_1N_2N_3N_4\exp\{nH(W|UV)+o_{\eta'}(n)\} \\ &= \exp\{n[I(U\wedge X)+I(V\wedge Y|U)+H(W|UV) \\ &\qquad +\delta+\delta']+o_{\eta'}(n)\}. \end{aligned} \quad (3.89)$$

Observe next by (3.85), (3.86), and (3.6), (3.7), that

$$\begin{aligned} &\tilde{P}(i,j,k,\ell,\boldsymbol{w}) \\ &\quad\le \sum_{\boldsymbol{y}\in T^n_{Y|UVW,\eta'}(\boldsymbol{u}_{ij},\boldsymbol{v}^{ij}_{k\ell},\boldsymbol{w})} \ \sum_{\boldsymbol{x}\in T^n_{X|UYW,\eta}(\boldsymbol{u}_{ij},\boldsymbol{y},\boldsymbol{w})} \\ &\qquad \cdot P^n_{XYW}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{w})/P^n_{XYW}(E) \\ &\quad\le \exp\{nH(Y|UVW)+o_{\eta'}(n)\} \\ &\qquad \cdot \exp\{nH(X|UYW)+o_{\eta}(n)\} \\ &\qquad \cdot \exp\{-nH(XYW)+o_{\eta}(n)\}. \end{aligned}$$

Here

$$\begin{aligned} &H(Y|UVW)+H(X|UYW)-H(XYW) \\ &\quad = H(YUVW)-H(UVW)+H(XUYW) \\ &\qquad -H(UYW)-H(XYW) \\ &\quad = H(V|UYW)+H(U|XYW)-H(UVW) \\ &\quad = H(V|UY)+H(U|X)-H(U) \\ &\qquad -H(V|U)-H(W|UV) \\ &\quad = -I(V\wedge Y|U)-I(U\wedge X)-H(W|UV) \end{aligned}$$

where the third equality follows by the Markov conditions (2.28). Comparing this with (3.89), the last bound in $\tilde{P}(i, j, k, \ell, \boldsymbol{w})$ yields

$$\tilde{P}(i,j,k,\ell,\boldsymbol{w}) \le M^{-1}\exp\{n\mu\} \quad (3.90)$$

where $\mu > 0$ can be arbitrarily small if δ, δ', η, η' are sufficiently small (and $n \ge n_0(\mu)$).

To proceed further, we distinguish between three cases:

i) $\min\{I(U\wedge Y), I(U\wedge Z)\} > I(U\wedge W), I(V\wedge Z|U) > I(V\wedge W|U)$;

ii) $\min\{I(U\wedge Y), (U\wedge Z)\} \le I(U\wedge W), I(V\wedge Z|U) > I(V\wedge W|U)$;

iii) $\min\{I(U\wedge Y), I(U\wedge Z)\} > I(U\wedge W), I(V\wedge Z|U) \le I(V\wedge W|U)$.

The logically possible fourth case is clearly irrelevant.

In case i), we shall apply Lemma A.4 with the choices

$$\begin{aligned} A &= \{(j,\ell)\colon 1\le j\le N_2, 1\le \ell\le N_4\} \\ B &= \{(i,k,\boldsymbol{w})\colon 1\le i\le N_1, 1\le k\le N_3, \\ &\qquad \boldsymbol{w}\in T_{W,\eta'\mathcal{U}\|\mathcal{V}\|\mathcal{Y}|}\}. \end{aligned}$$

Let $M(i,k,\boldsymbol{w})$ denote the number of those pairs (j,ℓ) for which $(i,j,k,\ell,\boldsymbol{w})$ is a possible quintuple, and let

$$\begin{aligned} C = \{(i,k,\boldsymbol{w})\colon &\tilde{P}(i,j,k,\ell,\boldsymbol{w}) \ge M^{-1}\exp(-n\mu) \\ &\text{for at least } M(i,k,\boldsymbol{w})\exp(-2n\mu) \text{ pairs } (j,\ell)\}. \end{aligned} \quad (3.91)$$

Then, the probability of $B\backslash C$ under (the marginal on B of) $\tilde{P}$ is bounded above by the total $\tilde{P}$-probability for all quintuples with $\tilde{P}(i,j,k,\ell,\boldsymbol{w}) < M^{-1}\exp(-n\mu)$ plus $M^{-1}\exp(-n\mu)$ times the maximum value of $\tilde{P}(i,j,k,\ell,\boldsymbol{w})$, and, hence, by $2\exp(-n\mu)$ (cf. (3.90)).

For $(i,j,\boldsymbol{w}) \in C$, the conditional probabilities

$$P(j,\ell|i,k,\boldsymbol{w}) = \frac{\tilde{P}(i,j,k,\ell,\boldsymbol{w})}{\sum_{j',\ell'}\tilde{P}(i,j',k,\ell',\boldsymbol{w})}$$

can be bounded—using (3.90)—according to

$$\begin{aligned} &P(j,\ell|i,k,\boldsymbol{w}) \\ &\quad\le \frac{M^{-1}\exp(n\mu)}{M(i,k,\boldsymbol{w})\exp(-2n\mu)M^{-1}\exp(-n\mu)} \\ &\quad = \frac{\exp(4n\mu)}{M(i,k,\boldsymbol{w})}. \end{aligned} \quad (3.92)$$

In order to bound below the term $M(i,k,\boldsymbol{w})$ on the right-hand side of (3.92), we invoke Lemma A.1, recalling that the sequences $\boldsymbol{u}_{ij}$ and $\boldsymbol{v}^{ij}_{k\ell}$ were obtained by random selection (from $T_{U,\xi|\mathcal{X}|}$) and $T^n_{V|U,\xi'|\mathcal{Y}|}(\boldsymbol{u}_{ij})$, respectively). Fix any joint type class $T^n_{\tilde{U}\tilde{V}\tilde{W}} \subset T^n_{UVW,\eta'|\mathcal{Y}|}$ which contains $(\boldsymbol{u}_{ij},\boldsymbol{v}^{ij}_{k\ell},\boldsymbol{w})$ for some possible quintuple with the given $\boldsymbol{w}$. Then, by Lemma A.1 a), among the $N_2 = \exp(n[\min\{I(U\wedge Y), I(U\wedge Z)\}-\epsilon])$ randomly selected $\boldsymbol{u}_{ij}$'s with a fixed i, there are

$$\exp(n[\min\{I(U\wedge Y), I(U\wedge Z)\}-\epsilon-I(U\wedge W)]+o_{\xi+\eta'}(n))$$

many such that $(\boldsymbol{u}_{ij},\boldsymbol{w}) \in T^n_{\tilde{U}\tilde{W}}$, except for doubly exponentially small probability. (Note that in case i), the present N_2 satisfies the hypothesis of the lemma, provided that ϵ is sufficiently small.) Further, by fixing any such $\boldsymbol{u}_{ij}$ and applying Lemma A.1 b), we obtain that among the

$$N_4 = \exp(n[I(V\wedge Z|U)-\epsilon'])$$

randomly selected $\boldsymbol{v}^{ij}_{k\ell}$'s with a given k, there are

$$\exp(n[I(V\wedge Z|U)-I(V\wedge W|U)-\epsilon']+o_{\xi'+\eta'}(n))$$

many such that $(\boldsymbol{u}_{ij},\boldsymbol{v}^{ij}_{k\ell},\boldsymbol{w}) \in T^n_{\tilde{U}\tilde{V}\tilde{W}}$, except for doubly exponentially small probability. (Again, the present N_4 satisfies the

hypothesis of the lemma, provided that ϵ' is sufficiently small.) Thus we arrive at the lower bound

$$M(i,k,\boldsymbol{w}) \geq \exp(n[\min\{I(U\wedge Y), I(U\wedge Z)\} - I(U\wedge W) + I(V\wedge Z|U) - I(V\wedge W|U) - \epsilon - \epsilon'] + o_{\xi+\xi'+\eta'}(n)). \tag{3.93}$$

Substitution of this lower bound into (3.92) shows that (except for $(i,k,\boldsymbol{w}) \notin C$, an event whose probability is exponentially small), the conditional probabilities $P(j,\ell|i,k,\boldsymbol{w})$ are bounded above by L^{-1} with $\frac{1}{n}\log L$ arbitrarily close to the desired expression. Now, applying Lemma A.4 and its corollary, the proof is completed.

In case ii), Lemma A.4 is applied differently by choosing

$$A = \{\ell\colon 1 \leq \ell \leq N_4\},$$
$$B = \{(i,j,k,\boldsymbol{w})\colon 1 \leq i \leq N_1, 1 \leq j \leq N_2, 1 \leq k \leq N_3, \boldsymbol{w} \in T^n_{W,\eta'|\mathcal{U}||\mathcal{V}||\mathcal{Y}|}\}.$$

Accordingly, we consider the number $M(i,j,k,\boldsymbol{w})$ of those ℓ's for which $(i,j,k,\ell,\boldsymbol{w})$ is a possible quintuple. We obtain the analog

$$P(\ell|i,j,k,\boldsymbol{w}) \leq \frac{\exp(4n\mu)}{M(i,j,k,\boldsymbol{w})} \tag{3.94}$$

of (3.92), for $(i,j,k,\boldsymbol{w})$ belonging to the analog of the set C which is now defined by replacing, in (3.91), "$(i,k,\boldsymbol{w})$" by "$(i,j,k,\boldsymbol{w})$" and "pairs (j,ℓ)" by "symbols ℓ." The term $M(i,j,k,\boldsymbol{w})$ is bounded below by the number of those among the N_4 randomly selected $\boldsymbol{v}^{ij}_{k\ell}$'s (for given i,j,k) which satisfy $(\boldsymbol{u}_{ij}, \boldsymbol{v}^{ij}_{k\ell}, \boldsymbol{w}) \in T^n_{\tilde{U}\tilde{V}\tilde{W}}$ as above, yielding

$$M(i,j,k,\boldsymbol{w}) \geq \exp(n[I(V\wedge Z|U) - I(V\wedge W|U) - \epsilon'] + o_{\xi'+\eta'}(n)). \tag{3.95}$$

The proof is then completed as above. Observe that it yields more than the stated assertion, in that now $F = F(\gamma)$, and even $I(f,\varphi,g,W^n\wedge F)$ is exponentially small.

The remaining case iii) is the simplest. Now, the need is obviated for Step 2 of the basic construction and terminal $\mathcal{Y}$ will not transmit at all. Lemma A.2 is applied to

$$A = \{j\colon 1 \leq j \leq N_2\}$$
$$B = \{(i,\boldsymbol{w})\colon 1 \leq i \leq N_1, \boldsymbol{w} \in T_{W,\eta|U||\mathcal{X}||\mathcal{Y}|}\}$$

bounding $P(j|i,\boldsymbol{w})$ by counting the number of $\boldsymbol{u}_{ij}$'s with fixed i which have given joint type with $\boldsymbol{w}$, as was done in case i) above.

Proof of Theorem 2.6:

Forward part: Observe that for any $R'_1 > 0, R'_2 > 0$, with $R'_1 + R'_2 = R_2$, a lower bound for $C_{\mathrm{WSK}}(0, R_2)$ is provided by $C_{\mathrm{WSK}}(R'_1, R'_2)$ for the case $X = Y$. By Theorem 2.5, the latter is bounded below by $I(V\wedge Z|U) - I(V\wedge W|U)$ for any rv's U, V satisfying the Markov conditions

$$U \multimap Y \multimap ZW \qquad V \multimap UY \multimap ZW \tag{3.96}$$

and the rate conditions

$$I(U\wedge Y) - I(U\wedge Z) \leq R'_1$$
$$I(V\wedge Y|U) - I(V\wedge Z|U) \leq R'_2. \tag{3.97}$$

Now, given U, V satisfying the rate condition (2.32) with strict inequality, and the Markov condition (2.33), we see that (3.96) holds as an obvious consequence of (2.33), and (3.97) also holds for suitable $R'_1 > 0, R'_2 > 0$, with $R'_1 + R'_2 = R_2$, because

$$I(V\wedge Y) = I(UV\wedge Y) = I(U\wedge Y) + I(V\wedge Y|U)$$
$$I(V\wedge Z) = I(UV\wedge Z) = I(U\wedge Z) + I(V\wedge Z|U).$$

As observed above, this ensures that $C_{\mathrm{WSK}}(0, R_2)$ is bounded below by $I(V\wedge Z|U) - I(V\wedge W|U)$, establishing the forward part.

Converse part: Consider $f = f(X^n) =$ a constant (since $R_1 = 0$) and any function $g = g(Y^n)$ which satisfies the rate constraint (1.2), and any pair of rv's (K, L) which represent ϵ-WSK, i.e., satisfy (1.3), (1.4), and the secrecy condition (1.6). We shall show the existence of rv's U, V satisfying the rate condition (2.32) with R_2 replaced by $R_2 + \delta$, and the Markov condition (2.33), such that

$$\frac{1}{n}H(K) \leq I(V\wedge Z|U) - I(V\wedge W|U) + \delta \tag{3.98}$$

where $\delta > 0$ is arbitrarily small if $\epsilon > 0$ in (1.4) and (1.6) is sufficiently small.

To this end, observe that $I(K\wedge gW^n) = o_\epsilon(n)$ by (1.6) and $H(K|gZ^n) \leq H(K|L) = o_\epsilon(n)$ by (1.3), (1.4). Thus

$$\begin{aligned} H(K) &= H(K|gW^n) + I(K\wedge gW^n) \\ &= H(K|gW^n) - H(K|gZ^n) + o_\epsilon(n), \\ &= I(K\wedge Z^n|g) - I(K\wedge W^n|g) + o_\epsilon(n) \\ &= \sum_{i=1}^n [I(K\wedge Z_i|gZ^{i-1}W^n_{i+1}) \\ &\qquad - I(K\wedge W_i|gZ^{i-1}W^n_{i+1})] + o_\epsilon(n) \\ &\qquad\qquad \text{(cf. e.g., in (3.41))} \\ &= n[I(K\wedge Z_J|U) - I(K\wedge W_J|U) + o_\epsilon(n) \end{aligned} \tag{3.99}$$

where J is uniformly distributed on $\{1,\cdots,n\}$ and is independent of (Y^n, Z^n, W^n), and $U = gZ^{J-1}W^n_{J+1}J$. Observe that the last expression in (3.99) does not change if K is replaced by $V = UK$. This establishes the desired bound in (3.98) with rv's U, V that satisfy the asserted Markov condition

$$U \multimap V \multimap Y_J \multimap Z_JW_J. \tag{3.100}$$

Indeed, since

$$V = UK = g(Y^n)Z^{J-1}W^N_{J+1}JK(Y^N)$$

the nontrivial part

$$V \multimap Y_J \multimap Z_JW_J$$

of (3.100) follows from

$$Y^{i-1}Y^n_{i+1}Z^{i-1}W^n_{i+1} \multimap Y_i \multimap Z_iW_i$$

which is true since (Y_i, Z_i, W_i) is independent of $(Y^{i-1}, Y_{i+1}^n, Z^{i-1}, W_{i+1}^n)$.

It remains to show that U, V satisfy the desired rate condition. For this purpose, we proceed as follows (using again that $H(K|gZ^n) = o_\epsilon(n)$):

$$\begin{aligned} nR_2 &\geq H(g) \geq H(g|Z^n) \\ &= H(gK|Z^n) + o_\epsilon(n) \\ &= H(gK|Z^n) - H(gK|Y^n) + o_\epsilon(n) \\ &= I(gK \wedge Y^n) - I(gK \wedge Z^n) + o_\epsilon(n). \end{aligned} \tag{3.101}$$

The expression $I(gK \wedge Y^n) - I(gK \wedge Z^n)$ is the exact analog of $I(f \wedge X^n) - I(f \wedge Y^n)$ which was bounded in (3.50), the roles of f, X, Y, Z now being played by gK, Y, Z, W, respectively. Consequently, recalling the definition $V = KU$, we obtain the analog of (3.50) as

$$I(gK \wedge Y^n) - I(gK \wedge Z^n) \geq n[I(V \wedge Y_J) - I(V \wedge Z_J)].$$

This, together with (3.101), completes the proof of Theorem 2.7.

Turning to the Corollary, observe that since the rate of f is not constrained, we can pick $f = f(X^n) = X^n$ as an optimal choice to achieve PK-capacity; this choice of f is permitted by the privacy condition (1.7). Then the PK-capacity with helper $C_{\rm PK}(\infty, R_2)$ is the same as the WSK-capacity without helper $C_{\rm WSK}(0, R_2)$ with XY and XZ in the roles of Y and Z, respectively, and XW in the role of W. Applying Theorem 2.6, we get

$$C_{\rm PK}(\infty, R_2) = \max_{U,V} [I(V \wedge XZ|U) - I(V \wedge XW|U)] \tag{3.102}$$

where the maximum is with respect to rv's U, V which satisfy the rate condition

$$I(V \wedge XY) - I(V \wedge XZ) \leq R_2 \tag{3.103}$$

and the Markov condition

$$U \multimap V \multimap XY \multimap ZW. \tag{3.104}$$

The proof is completed by noting that the rate constraint in (3.103) is equivalent to (2.11).

IV. Discussion

We have derived (strongly) achievable common randomness (CR) and wiretap secret key (WSK) rates, and in particular, private key (PK) and secret key (SK) rates, for source-type models involving two user terminals abetted by a helper, subject to transmission rate constraints, as a continuation of previous work [2], [3], [5]. Our results on achievable CR and SK rates are tight when the generic rv's of the given discrete memoryless multiple source satisfy the Markov conditions $X \multimap Z \multimap Y$ or $X \multimap Y \multimap Z$, thus determining the CR- and SK-capacities in these situations. In particular, our CR- and SK-capacity results are tight if $X = Z$, when our model reduces to that without a helper but with two-way communication permitted between the users $\mathcal{Y}$ and $\mathcal{Z}$. For the latter, the CR-capacity has been determined in [3], while the SK-capacity result is new. Else, the issue of tightness, in general, remains unresolved. Nevertheless, single-letter characterizations of CR-, SK-, WSK-, and PK-capacities have been obtained in many special cases of interest when one of the bounds on the transmission rates is 0 (no communication) or ∞ (no rate constraint). These special cases also yield new results for previously studied models.

This paper offers new evidence of the strong achievability of secrecy capacity (i.e., WSK-, PK-, and SK-capacities), observed previously in [9], [4], and [5]. Another—and perhaps surprising—observation is that randomization does not enhance secrecy capacity, at least in those instances in which the latter could be determined. Indeed, in all these cases, achievability has been established without recourse to randomization (unlike in previous works), and allowing randomization does not alter our converse results. Note that the CR-capacity, $C_{\rm CR}(R_1, R_2)$, can be enhanced by randomization, although only in a trival fashion (see [3]): if either rate constraint is inoperative, i.e., R_1 or R_2 could be decreased without reducing $C_{\rm CR}(R_1, R_2)$, the available excess rate can be utilized to generate additional CR through the transmission of random bits.

Turning next to the issue of the extent of allowed public communication between the user terminals, it is worth noting that our SK-capacity without rate constraints, viz. $C_{\rm SK}(\infty, \infty) = \min\{I(XZ \wedge Y), I(XY \wedge Z)\}$ (cf. (2.26)), does not increase even if communication in all directions in any number of rounds is permitted; this follows from a similar result for SK-capacity without helper (see [8], [2]). Another secrecy capacity with helper which is not improved by unlimited communication is the PK-capacity without rate constraints in the special case when the eavesdropper lacks side information, viz. $C_{\rm PK}(\infty, \infty) = I(Y \wedge Z|X)$ (cf. (2.37)); this is implicit in [8]. In general, however, greater freedom allowed in communication between the users might lead to enhanced capacities, an issue which merits further examination.

Finally, we discuss certain decomposition properties of the various capacities which are suggested by heuristic considerations and are confirmed by our results. First, it is plausible that the maximum achievable rate of CR with $\mathcal{X}$ and $\mathcal{Y}$ transmitting at rates R_1 and R_2, respectively, obtains when these transmissions result in the maximum achievable rate of secret CR, so that the total achievable CR rate equals the achievable secret CR rate augmented by $R_1 + R_2$ representing the CR extracted from these public transmissions. Hence, it is to be expected that

$$C_{CR}(R_1, R_2) = R_1 + R_2 + C_{SK}(R_1, R_2) \tag{4.1}$$

provided that transmissions at rates R_1 and R_2 are "indeed needed" to achieve $C_{\rm SK}(R_1, R_2)$ (i.e., when $C_{\rm SK}(R_1, R_2)$ is reduced upon decreasing R_1 or R_2).

In the Markov cases $X \multimap Z \multimap Y$ and $X \multimap Y \multimap Z$, when the CR- and SK-capacities could be determined, (4.1) follows as expected from Theorem 2.3, using Lemma E.3 in Appendix E, whenever

$$\begin{aligned} R_1 &\leq R_1^* = \max\{H(X|Y), H(X|Z)\} \\ R_2 &\leq R_2^* = H(Y|XZ). \end{aligned}$$

Note that if R_1 and R_2 are "indeed needed" to achieve $C_{\rm SK}(R_1, R_2)$, then surely $R_1 \leq R_1^*$, while $R_2 \leq R_2^*$ has to

hold if $X \multimap Z \multimap Y$ but not necessarily if $X \multimap Y \multimap Z$. Additional effort would show that Theorem 2.3 implies (4.1) also when $X \multimap Y \multimap Z$ and $R_2 \geq R_2^*$, provided, of course, that R_1 and R_2 are "indeed needed." Also, for arbitrary X, Y, Z, when $R_1 = R_1^*, R_2 \leq R_2^*$, the formulas for

$$C_{\mathrm{CR}}(\infty, R_2) = C_{\mathrm{CR}}(R_1^*, R_2)$$

and

$$C_{\mathrm{SK}}(\infty, R_2) = C_{\mathrm{SK}}(R_1^*, R_2)$$

obtained in Cases 1 and 3 of Section II, imply (4.1).

The SK-capacity formula

$$C_{\mathrm{SK}}(\infty, R_2) = \min\{I(X \wedge Y), I(X \wedge Z)\} + \max_V I(V \wedge Z|X) \tag{4.2}$$

where the maximization is subject to (2.11), (2.12) (Case 3 of Section II), also lends itself to a heuristic interpretation: with no rate constraint on the helper $\mathcal{X}$, the optimum SK-rate for the users $\mathcal{Y}$ and $\mathcal{Z}$ can be achieved by first generating a three-way SK for $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ using the helper's transmission alone (the first term on the right-hand side of (4.2); see Theorem 2.4), followed by generating a PK for $\mathcal{Y}$ and $\mathcal{Z}$ using also the transmission of $\mathcal{Y}$ (the second term on the right-hand side of (4.2); see the special case—eavesdropper without side information—of the corollary of Theorem 2.6 (cf. (2.36))). Upon combining this with (4.1), and recalling that the absence of a rate constraint on the helper is effectively equivalent to letting $R_1 = R_1^* = \max\{H(X|Y), H(X|Z)\}$, we also get the following decomposition of the CR-capacity, $C_{\mathrm{CR}}(\infty, R_2)$,whenever $R_2 \leq R_2^* = H(Y|XZ)$:

$$C_{\mathrm{CR}}(\infty, R_2) = R_1^* + R_2 + C_{\mathrm{SK}}^3(\infty, 0) + C_{\mathrm{PK}}(\infty, R_2) \tag{4.3}$$

with an obvious heuristic interpretation. In the special case $R_2 = R_2^*$, effectively equivalent to $R_2 = \infty$ when $R_1 = \infty$, (4.3) reduces to a decomposition of $C_{\mathrm{CR}}(\infty, \infty) = H(XY)$ as

$$\begin{aligned} H(XY) = {} & \max\{H(X|Y), H(X|Z)\} + H(Y|XZ) \\ & + \min\{I(X \wedge Y), I(X \wedge Z)\} + I(Y \wedge Z|X) \end{aligned}$$

and provides an interesting interpretation of this simple identity.

APPENDIX A

Lemma A.1: Let S, U, and V be rv's with finite ranges $\mathcal{S}, \mathcal{U}$, and $\mathcal{V}$, respectively.

a) Select at random $N = \lceil \exp(nR) \rceil$ sequences $\boldsymbol{v}_i$ from $T_{V,\tau}^n, 1 \leq i \leq N$, where $R > I(S \wedge V)$. Then, for each joint type class $T_{\tilde{S}\tilde{V}}^n \subset T_{SV,\xi}^n$ with $T_{\tilde{V}}^n \subset T_{V,\tau}^n$, we have for all $\boldsymbol{s} \in T_{\tilde{S}}^n$

$$|\{i\colon (\boldsymbol{s}, \boldsymbol{v}_i) \in T_{\tilde{S}\tilde{V}}^n\}| = \exp\{n(R - I(S \wedge V)) + o_{\xi+\tau}(n)\} \tag{A.1}$$

except with probability going to 0 doubly exponentially as $n \to \infty$.

b) Pick $\boldsymbol{u} \in \mathcal{U}^n$ with $T_{V|U,\tau}^n(\boldsymbol{u}) \neq \phi$, and select at random $N = \lceil \exp(nR) \rceil$ sequences $\boldsymbol{v}_i, i = 1, \cdots, N$, from $T_{V|U,\tau}^n(\boldsymbol{u})$, where $R > I(S \wedge V|U)$. Then, for each joint type class $T_{\tilde{S}\tilde{U}\tilde{V}}^n \subset T_{SUV,\xi}^n$ with $\boldsymbol{u} \in T_{\tilde{U}}^n, T_{\tilde{U}\tilde{V}}^n \subset T_{UV,\tau}^n$, we have for all $\boldsymbol{s} \in T_{\tilde{S}|\tilde{U}}^n(\boldsymbol{u})$ that

$$\begin{aligned} & |\{i\colon (\boldsymbol{s}, \boldsymbol{u}, \boldsymbol{v}_i) \in T_{\tilde{S}\tilde{U}\tilde{V}}^n\}| \\ & \qquad = \exp\{n(R - I(S \wedge V|U)) + o_{\xi+\tau}(n)\}. \end{aligned} \tag{A.2}$$

Proof:

a) Given a joint type class $T_{\tilde{S}\tilde{V}}^n$ with $T_{\tilde{V}}^n \subset T_{V,\tau}^n$, and any $\boldsymbol{s} \in T_{\tilde{S}}^n$, a randomly selected $\boldsymbol{v} \in T_{V,\tau}^n$ will satisfy $(\boldsymbol{s}, \boldsymbol{v}) \in T_{\tilde{S}\tilde{V}}^n$ with probability

$$p_1 = \frac{|T_{\tilde{V}|\tilde{S}}^n(\boldsymbol{s})|}{|T_{V,\tau}^n|}.$$

Additionally, if $T_{\tilde{S}\tilde{V}}^n \subset T_{SV,\xi}^n$, then by (3.5), (3.6), and the fact that the number of possible type classes is $\exp(o(n))$, we obtain that

$$\begin{aligned} p_1 &= \frac{\exp\{nH(V|S) + o_\xi(n)\}}{\exp\{nH(V) + o_\tau(n)\}} \\ &= \exp\{-nI(S \wedge V) + o_{\xi+\tau}(n)\}. \end{aligned} \tag{A.3}$$

The standard large deviation bounds for the binomial distribution then yield

$$\begin{aligned} & \Pr\{|\{i : (\boldsymbol{s}, \boldsymbol{v}_i) \in T_{SV,\xi}^n\}| < (1-\epsilon)Np_1\} \\ & \qquad \leq \exp\{-ND((1-\epsilon)p_1 \| p_1)\} \end{aligned}$$

and

$$\begin{aligned} & \Pr\{|\{i : (\boldsymbol{s}, \boldsymbol{v}_i) \in T_{SV,\xi}^n\}| > (1+\epsilon)Np_1\} \\ & \qquad \leq \exp\{-ND((1+\epsilon)p_1 \| p_1)\} \end{aligned}$$

where (the divergence)

$$D(p\|q) = p \log \frac{p}{q} + (1-p) \log \frac{(1-p)}{(1-q)}.$$

As in [3, Proof of Lemma 1.1]), both divergences in the bounds above are $\geq \epsilon^2 p_1 / \ln 2$ if $\epsilon \leq 1/2$. Hence, with $\epsilon = 1/2$, say, we obtain that (A.1) fails with probability at most $\exp\{-Np_1/4\ln 2\}$, which decays to 0 doubly exponentially owing to (A.3) and the assumption that $R > I(S \wedge V)$.

b) Only obvious modifications are needed in the proof above.

Lemma A.2: Given the rv's V, S, W, where $V \multimap S \multimap W$, with finite ranges $\mathcal{V}, \mathcal{S}, \mathcal{W}$, respectively, and positive numbers $\epsilon, \eta < \eta_0, \delta < \delta_0, \xi < \xi_0$, where

$$\eta_0 = \eta_0(\epsilon), \quad \delta_0 = \delta_0(\epsilon, \eta), \quad \xi_0 = \xi_0(\epsilon, \eta, \delta)$$

are suitable thresholds, select at random $N_1 N_2$ sequences $\boldsymbol{v}_{ij} \in T_{V,\xi|\mathcal{S}|}^n, 1 \leq i \leq N_1, 1 \leq j \leq N_2$, with

$$\begin{aligned} N_1 N_2 &= \exp\{n(I(S \wedge V) + \delta)\} \\ N_2 &\leq \exp\{n(I(V \wedge W) - \epsilon)\}. \end{aligned} \tag{A.4}$$

Then, given any pmf $\tilde{P}$ on $\mathcal{S}^n \times \mathcal{W}^n$, the following hold except with probability less than $\exp(-n\delta/3)$ if n is sufficiently large, $n \geq n_0(\epsilon, \delta)$..

a) There exist mappings

$$f\colon \mathcal{S}^n \to \{1, \cdots, N_1\}, \varphi\colon \mathcal{S}^n \to \{1, \cdots, N_2\}$$

such that

$$(\boldsymbol{s}, \boldsymbol{v}_{f(\boldsymbol{s})\varphi(\boldsymbol{s})}) \in T^n_{SV,\xi} \quad \text{if } \boldsymbol{s} \in T^n_{S,\xi};$$
$$f(\boldsymbol{s}) = \varphi(\boldsymbol{s}) = 1 \quad \text{if } \boldsymbol{s} \notin T_{S,\xi}. \quad \text{(A.5)}$$

b) There exist mappings

$$\tilde{\phi}_i \colon \mathcal{W}^n \to \{1, \cdots, N_1\}, \qquad i = 1, \cdots, N_2$$

and a set $D \subset \mathcal{S}^n \times \mathcal{W}^n$ with $\tilde{P}(D) < 2\exp(-n\delta)$ such that for any (f, φ) as in (A.5), each $(\boldsymbol{s}, \boldsymbol{w}) \in T^n_{SW,\xi|\mathcal{W}|^{-1}} \backslash D$ satisfies

$$(\boldsymbol{s}, \boldsymbol{v}_{f(\boldsymbol{s})\varphi(\boldsymbol{s})}, \boldsymbol{w}) \in T^n_{SVW,\eta} \quad \text{(A.6)}$$

and

$$\varphi(\boldsymbol{s}) = \tilde{\phi}_{f(\boldsymbol{s})}(\boldsymbol{w}). \quad \text{(A.7)}$$

Proof:

a) Observe that $\boldsymbol{s} \in T^n_{S,\xi}$ implies, by (3.4), the existence of a joint type class $T^n_{\tilde{S}\tilde{V}} \subset T^n_{SV,\xi}$ with $\boldsymbol{s} \in T^n_{\tilde{S}}$, which then automatically satisfies $T_{\tilde{V}} \subset T_{V,\xi|\mathcal{S}|}$. Thus by (A.1), $(\boldsymbol{s}, \boldsymbol{v}_{ij}) \in T^n_{\tilde{S}\tilde{V}} \subset T^n_{SV,\xi}$ for $\exp(n\delta + o_\xi(n))$ sequences $\boldsymbol{v}_{ij}$, except with doubly exponentially small probability. This is an even stronger result than that claimed.

b) Denote by D_1 the set of those $(SW, \xi|\mathcal{W}|^{-1})$-typical pairs $(\boldsymbol{s}, \boldsymbol{w})$ for which there exist i and j such that

$$(\boldsymbol{s}, \boldsymbol{v}_{ij}) \in T^n_{SV,\xi} \qquad (\boldsymbol{s}, \boldsymbol{v}_{ij}, \boldsymbol{w}) \notin T^n_{SVW,\eta}. \quad \text{(A.8)}$$

For $(SW, \xi|\mathcal{W}|^{-1})$-typical pairs $(\boldsymbol{s}, \boldsymbol{w})$, note that $\boldsymbol{s}$ is (S, ξ)-typical by (3.3) so that $(\boldsymbol{s}, \boldsymbol{v}_{f(\boldsymbol{s})\varphi(\boldsymbol{s})}) \in T^n_{SV,\xi}$ by a). Thus (A.6) holds if $(\boldsymbol{s}, \boldsymbol{w}) \in D_1$, which leads us to the task of bounding $\tilde{P}(D_1)$.

We first bound the probability that, for fixed i, j, the randomly selected $\boldsymbol{v}_{ij}$ satisfies (A.8) for a fixed $(SW, \xi|\mathcal{W}|^{-1})$-typical pair $(\boldsymbol{s}, \boldsymbol{w})$. For such $(\boldsymbol{s}, \boldsymbol{w})$ (A.8), is tantamount (cf. (3.1)) to the joint type of $(\boldsymbol{s}, \boldsymbol{v}_{ij}, \boldsymbol{w})$ being represented by dummy rv's $\tilde{S}, \tilde{V}, \tilde{W}$, such that

$$\max_{s,v} |P_{\tilde{S}\tilde{V}}(s,v) - P_{SV}(s,v)| \le \xi$$
$$\max_{s,w} |P_{\tilde{S}\tilde{W}}(s,w) - P_{SW}(s,w)| \le \xi|\mathcal{W}|^{-1}$$
$$\max_{s,v,w} |P_{\tilde{S}\tilde{V}\tilde{W}}(s,v,w) - P_{SVW}(s,v,w)| > \eta. \quad \text{(A.9)}$$

Note that since $S \multimap V \multimap W$, the joint typicality of $(\boldsymbol{v}_{ij}, \boldsymbol{s})$ and of $(\boldsymbol{s}, \boldsymbol{w})$ exclude the possibility that $P_{\tilde{S}\tilde{V}\tilde{W}}(s,v,w) > 0$ when $P_{SVW}(s,v,w) = 0$; thus $(\boldsymbol{s}, \boldsymbol{v}_{ij}, \boldsymbol{w})$ cannot fail in this manner to be jointly typical.

For given $\tilde{S}, \tilde{V}, \tilde{W}$ representing a joint type as above, we have by (3.5) and (3.6) that

$$\begin{aligned}
&\Pr\{(\boldsymbol{s}, \boldsymbol{v}_{ij}, \boldsymbol{w}) \in T^n_{\tilde{S}\tilde{V}\tilde{W}}\} \\
&\quad = \frac{|T^n_{\tilde{V}|\tilde{S}\tilde{W}}(\boldsymbol{s}, \boldsymbol{w})|}{|T^n_{V,\xi|\mathcal{S}|}|} \\
&\quad = \frac{\exp\{nH(\tilde{V}|\tilde{S}\tilde{W}) + o(n)\}}{\exp\{nH(V) + o_\xi(n)\}} \\
&\quad = \frac{\exp\{n[H(\tilde{V}|\tilde{S}) + I(\tilde{V} \wedge \tilde{W}|\tilde{S})] + o(n)\}}{\exp\{nH(V) + o_\xi(n)\}} \\
&\quad \le \exp\{-n[I(S \wedge V) + \psi(\xi, \eta)] + o_\xi(n)\} \quad \text{(A.10)}
\end{aligned}$$

where $\psi(\xi, \eta)$ denotes the minimum of $I(\tilde{V} \wedge \tilde{W}|\tilde{S})$ for $\tilde{V}, \tilde{S}, \tilde{W}$ not necessarily representing a joint type, and satisfying (A.9) with the last "$>$" replaced by "$\ge$" (to ensure that the minimum is attained). Clearly, $\psi(0, \eta) > \psi(0,0) = 0$ for every $\eta > 0$. Hence, by continuity

$$\psi(\xi, \eta) > \tfrac{1}{2}\psi(0, \eta) > 0 \text{ for } \xi \text{ sufficiently small.} \quad \text{(A.11)}$$

On account of (A.4), (A.10), and (A.11), the probability that a fixed $(SW, \xi|\mathcal{W}|^{-1})$-typical pair $(\boldsymbol{s}, \boldsymbol{w})$ satisfies (A.8) for some i, j, i.e., that $(\boldsymbol{s}, \boldsymbol{w})$ belongs to the random set D_1, is bounded according to

$$\begin{aligned}
\Pr\{(\boldsymbol{s}, \boldsymbol{w}) \in D_1\} &\le N_1 N_2 \exp\{-n[I(S \wedge V) \\
&\qquad + \psi(\xi, \eta)] + o_\xi(n)\} \\
&< \exp\left\{n\delta - n\frac{\psi(0,\eta)}{2} + o_\xi(n)\right\} \\
&< \exp\left(-\frac{3}{2}\, n\delta\right)
\end{aligned}$$

if $\delta < \frac{1}{6}\psi(0, \eta), \xi$ is sufficiently small, and $n > n_0(\delta)$. Furthermore

$$\begin{aligned}
E(\tilde{P}(D_1)) &= \sum_{(\boldsymbol{s},\boldsymbol{w}) \in T^n_{SW,\xi|\mathcal{W}|^{-1}}} \tilde{P}(\boldsymbol{s}, \boldsymbol{w}) \Pr\{\boldsymbol{s}, \boldsymbol{w}) \in D_1\} \\
&\le \exp\left(-\frac{3}{2}\, n\delta\right)
\end{aligned}$$

and, hence

$$\Pr\{\tilde{P}(D_1) > \exp(-n\delta)\} < \exp\left(-\frac{n\delta}{2}\right). \quad \text{(A.12)}$$

Finally, let D_2 denote the set of those pairs $(\boldsymbol{s}, \boldsymbol{w}) \in \mathcal{S}^n \times \mathcal{W}^n$ for which

$$(\boldsymbol{s}, \boldsymbol{v}_{ij}) \in T^n_{SV,\xi} \qquad (\boldsymbol{v}_{ik}, \boldsymbol{w}) \in T^n_{VW,\eta|\mathcal{S}|} \quad \text{(A.13)}$$

for some i and $j \ne k$. For fixed $(\boldsymbol{s}, \boldsymbol{w}), i$ and $j \ne k$, the randomly selected $\boldsymbol{v}_{ij}$ and $\boldsymbol{v}_{ik}$ satisfy (A.13) with probability

$$\exp\{-nI(S \wedge V) + o_\xi(n)\} \exp\{-nI(V \wedge W) + o_\eta(n)\}$$

if (A.13) is at all possible for the given $(\boldsymbol{s}, \boldsymbol{w})$. This, together with (A.4), yield that for any fixed $(\boldsymbol{s}, \boldsymbol{w})$

$$\begin{aligned}
&\Pr\{(\boldsymbol{s}, \boldsymbol{w}) \in D_2\} \\
&\quad \le N_1 N_2 \exp\{-nI(S \wedge V) + o_\xi(n)\} \\
&\qquad \cdot \exp\{-nI(V \wedge W) + o_\eta(n)\} \\
&\quad \le \exp\{n\delta - n\epsilon + o_{\xi+\eta}(n)\} \le \exp\left\{n\delta - \frac{n\epsilon}{2}\right\} \\
&\quad < \exp\left(-\frac{3}{2}\, n\delta\right)
\end{aligned}$$

if η, ξ are sufficiently small, $n > n_0(\epsilon)$, and $\delta < \epsilon/5$. This implies, as above, that

$$\Pr\{\tilde{P}(D_2) > \exp(-n\delta)\} < \exp\left(-\frac{n\delta}{2}\right). \quad \text{(A.14)}$$

Define the mappings

$$\tilde{\phi}_i : \mathcal{W}^n \to \{1, \cdots, N_2\}, \qquad i = 1, \cdots, N_1$$

such that

$$\tilde{\varphi}_i(\boldsymbol{w}) \in \{j \colon (\boldsymbol{v}_{ij}, \boldsymbol{w}) \in T^n_{VW,\eta|\mathcal{S}|}\}$$

whenever the last set is nonempty. Clearly, if (A.6) holds (implying $(\boldsymbol{v}_{f(\boldsymbol{s})\varphi(\boldsymbol{s})}, \boldsymbol{w}) \in T_{VW,\eta|\mathcal{S}|}$), then (A.7) must hold whenever $(\boldsymbol{s}, \boldsymbol{w}) \notin D_2$. On account of (A.12) and (A.14), this completes the proof, with $D = D_1 \cup D_2$.

Lemma A.3: Given the rv's S, U, V, W, with $V \multimap SU \multimap W$ and arbitrary $\boldsymbol{u} \in \mathcal{U}^n$ such that $T_{V|U,\xi|\mathcal{S}|} \neq \phi$, all the assertions of Lemma A.2 remain valid for randomly selected sequences $\boldsymbol{v}_{ij} \in T^n_{UV,\xi|\mathcal{S}|}(\boldsymbol{u}), i \leq i \leq N_1, 1 \leq j \leq N_2$, for arbitrary $\boldsymbol{u} \in \mathcal{U}^n$ such that $T^n_{V|U,\xi|\mathcal{S}|}(\boldsymbol{u}) \neq \phi$, with the following changes: (A.4) is replaced by

$$N_1 N_2 = \exp\{n[I(S \wedge V|U) + \delta]\}$$
$$N_2 \leq \exp\{n[I(V \wedge W|U) - \epsilon]\} \tag{A.15}$$

and the sets of form

$$T^n_{S,\tau}, T^n_{SV,\tau}, T^n_{SW,\tau}, T^n_{VW,\tau}, T^n_{SVW,\tau}$$

(where τ stands for the appropriate typicality constant) are replaced by

$$T^n_{S|U,\tau}(\boldsymbol{u}), T^n_{SV|U,\tau}(\boldsymbol{u}), T^n_{SW|U,\tau}(\boldsymbol{u}), T^n_{VW|U,\tau}(\boldsymbol{u}), T^n_{SVW|U,\tau}(\boldsymbol{u}).$$

Proof: The proof is identical to that of Lemma A.2, with the obvious changes.

The following "almost independence" lemma was proved in [5] as a consequence of the "coloring" lemma of [3].

Lemma A.4 [5]: Given finite sets A, B, and a pmf $\tilde{P}$ on $A \times B$, denote by P and $P(\cdot|b), b \in B$, the corresponding marginal pmf and conditional pmf on A. Suppose that for some $0 < \epsilon < 1/9$ and $L > 0$

$$\sum_{(a,b):\, P(a|b) > L^{-1}} \tilde{P}(a,b) \leq \epsilon^2. \tag{A.16}$$

Then, there exists a mapping $F: A \to \mathcal{K}$ with $|\mathcal{K}| = \lfloor \epsilon^2 L / 3\log(2|B|) \rfloor$ such that $F(\cdot)$ is almost uniformly distributed, and is almost independent of b in the sense that

$$\sum_{k \in \mathcal{K}} |P(F^{-1}(k)) - |\mathcal{K}|^{-1}| < 5\epsilon \qquad d_{\text{av}}(F) < 10\epsilon. \tag{A.17}$$

Here, $d_{\text{av}}(F)$ denotes the variation distance of the joint pmf $\tilde{P}_F$ from the product of its marginals, where $\tilde{P}_F$ is defined on $\mathcal{K} \times B$ by

$$\tilde{P}_F(k,b) = \sum_{a:\, F(a)=k} \tilde{P}(a,b).$$

Corollary: If $\tilde{P}$ satisfies (A.16) and $\tilde{P}'$ is another pmf on $A \times B$ whose variation distance from $\tilde{P}$ is less than δ, the assertion of Lemma A.4 holds also for $\tilde{P}$ upon replacing 5ϵ and 10ϵ in (A.17) by $5\epsilon + \delta$ and $10\epsilon + 3\delta$, respectively.

Proof: Let $\tilde{P}'_F$ denote the pmf on $\mathcal{K} \times B$ defined as $\tilde{P}_F$ above upon replacing $\tilde{P}$ by $\tilde{P}'$. Clearly, the variation distance of $\tilde{P}'_F$ from $\tilde{P}_F$ is also less than δ, and the same is true of their marginals on $\mathcal{K}$ and B. Thus the Corollary follows using the triangle inequality and the fact that the variation distance between product distributions does not exceed the sum of those of their marginals.

APPENDIX B

In order to establish the claim in the paragraph following Theorem 2.2, it suffices to prove

Lemma B: Given arbitrary rv's U, V satisfying the Markov conditions (2.5), (2.6), there exist rv's $\tilde{U}, \tilde{V}$ satisfying the same Markov conditions and taking values in sets $\tilde{\mathcal{U}}, \tilde{\mathcal{V}}$ of sizes $|\tilde{\mathcal{U}}| \leq |\mathcal{X}| + 3, |\tilde{\mathcal{V}}| \leq |\mathcal{Y}|$, such that

$$I(\tilde{U} \wedge X) - I(\tilde{U} \wedge Y) = I(U \wedge X) - I(U \wedge Y) \tag{B.1}$$

$$(\tilde{U} \wedge X) - I(\tilde{U} \wedge Z) = I(U \wedge X) - I(U \wedge Z) \tag{B.2}$$

$$I(\tilde{V} \wedge Y|\tilde{U}) - I(\tilde{V} \wedge Z|\tilde{U}) \leq I(V \wedge Y|U) - I(V \wedge Z|U) \tag{B.3}$$

and

$$I(\tilde{U} \wedge X) + I(\tilde{V} \wedge Y|\tilde{U}) \geq I(U \wedge X) + I(V \wedge Y|U). \tag{B.4}$$

Proof: The proof is similar to that of the analogous result in [3]. First, the given U can be assumed to satisfy

$$P_{X|U=u_1} \neq P_{X|U=u_2} \text{ whenever } u_1 \neq u_2. \tag{B.5}$$

Indeed, if (B.5) were not to hold, we can replace U, V by $U' = f(U), V' = UV$ without changing the relevant mutual information quantities, and consistent with the Markov conditions (2.5), (2.6), where $f(U)$ denotes the equivalence class of u under the equivalence relation $u_1 \sim u_2$ iff $P_{X|U=u_1} = P_{X|U=u_2}$. Next, (B.5) permits us to define a $|\mathcal{Y}| \times |\mathcal{V}|$ matrix-valued continuous function $F(P)$ on the set $\mathcal{P}(\mathcal{X})$ of probability distributions on $\mathcal{X}$, such that the (y, v)-entry $F(P)$ equals $\Pr\{V = v|Y = y, U = u\}$ if $P = P_{X|U=u}$. Then, denoting by W_1 and W_2 the conditional probability matrices $P_{Y|X}$ and $P_{Z|X}$,, and writing $P_u = P_{X|U=u}$, we have

$$I(U \wedge X) = H(X) - \sum_u \Pr\{U = u\} H(P_u)$$
$$I(U \wedge Y) = H(Y) - \sum_u \Pr\{U = u\} H(P_u W_1)$$
$$I(U \wedge Z) = H(Z) - \sum_u \Pr\{U = u\} H(P_u W_2)$$
$$I(V \wedge Y|U) = \sum_u \Pr\{U = u\} I(P_u W_1, F(P_u))$$

and

$$I(V \wedge Z|U) = \sum_u \Pr\{U = u\} I(P_u W_2, G(P_u) F(P_u))$$

where $G(P)$ denotes the $|\mathcal{Z}| \times |\mathcal{Y}|$ matrix whose (z, y)-entry equals

$$\sum_x P(x) \Pr\{Y = y, Z = z | X = x\} / \sum_x P(x) W_2(z|x).$$

Applying the Support Lemma [6, p. 310] to the following $|\mathcal{X}|+3$ continuous functions of $P \in \mathcal{P}(\mathcal{X})$

$$f_1(P) = H(PW_1) - H(P)$$
$$f_2(P) = H(PW_2) - H(P)$$
$$f_3(P) = I(PW_1, F(P)) - I(PW_2, G(P)F(P))$$
$$f_4(P) = I(PW_1, F(P)) - H(P)$$
$$f_j(P) = P(x_{j-4}), \quad 5 \leq j \leq |\mathcal{X}| + 3, \quad \text{where } \mathcal{X} = \{x_1, \cdots, x_{|\mathcal{X}|}\}$$

it follows as in [3] that an rv $\tilde{U}$ taking values in $\{1, \cdots, |\mathcal{X}|+3\}$ and satisfying $\tilde{U} \multimap X \multimap YZ$ exists such that, with the natural definition of a corresponding $\tilde{V}$ satisfying $\tilde{V} \multimap \tilde{U}Y \multimap Z$,, the mutual information differences on the right-hand sides of (B.1)–(B.4) remain unchanged when U, V is replaced by $\tilde{U}, \tilde{V}$. Finally, it remains to show that $\tilde{V}$ can be replaced by $\tilde{V}'$ taking no more than $|\mathcal{Y}|$ values such that (B.3), (B.4) are satisfied; this can be done exactly as in [3].

APPENDIX C

In this appendix, we prove the bounds in (2.24) and (2.25).

Suppose first that U, V satisfy the conditions (2.18) –(2.20). Since

$$\begin{aligned} I(U \wedge Z) + I(V \wedge Z|U) &= I(UV \wedge Z) \le I(UVX \wedge Z) \\ &= I(X \wedge Z) + I(UV \wedge Z|X) \end{aligned} \tag{C.1}$$

in order to show that (2.17) implies (2.24), it suffices to check that $V' = UV$ satisfies the conditions (2.11), (2.12). Now, the Markov condition (2.12) is an obvious consequence of (2.19). Further, (2.11) is checked as follows:

$$\begin{aligned} &I(UV \wedge Y|X) - I(UV \wedge Z|X) \\ &\quad = I(UV \wedge XY) - I(UV \wedge XZ) \\ &\quad = I(V \wedge XY|U) - I(V \wedge XZ|U) \\ &\quad \le R_2. \end{aligned} \tag{C.2}$$

The second equality above holds because the Markov condition (2.18) implies that $I(U \wedge XY) = I(U \wedge XZ) (= I(U \wedge X))$; the inequality above follows from (2.20).

Before proceeding further, we provide the following lemma which is also needed elsewhere.

Lemma C.1: Given rv's U, Y, Z, consider the maximum of $I(V \wedge Y|U)$ and $I(V \wedge Z|U)$ with respect to rv's V which satisfy the Markov condition

$$V \multimap UY \multimap Z \tag{C.3}$$

and the rate condition

$$I(V \wedge Y|U) - I(V \wedge Z|U) \le R \tag{C.4}$$

for a given $R \ge 0$. Both maxima are attained for $V = Y$ if $R \ge H(Y|UZ)$, while in the case $R < H(Y|UZ)$ there exists V attaining both maxima and satisfying (C.4) with equality. In particular, we have for any $R \ge 0$ that

$$\begin{aligned} &\max_V I(V \wedge Y|U) \\ &\qquad = \min\{R, H(Y|UZ)\} + \max_V I(V \wedge Z|U) \end{aligned} \tag{C.5}$$

where the maxima are subject to (C.3) and (C.4).

Proof: Note that $I(V \wedge Y|U)$ is maximized by $V = Y$ if that choice is permitted by condition (C.4), i.e., if $R \ge H(Y|UZ)$. Also, $I(V \wedge Z|U)$ is maximized by $V = Y$, when permissible, since

$$I(V \wedge Z|U) \le I(VY \wedge Z|U) = I(Y \wedge Z|U) \tag{C.6}$$

for all V satisfying the Markov condition (C.3). In particular, (C.5) holds trivially when $R \ge H(Y|UZ)$.

To prove the lemma for $R < H(Y|UZ)$, it suffices to show that $\max I(V \wedge Z|U)$ is attained for some V which satisfies (C.4) with equality. Indeed, this V then also attains $\max I(V \wedge Y|U)$, since by (C.4), that maximum certainly does not exceed $R + \max I(V \wedge Z|U)$. The proof will be completed by showing that if V_1 maximizes $I(V \wedge Z|U)$ subject to (C.3), (C.4) and

$$I(V_1 \wedge Y|U) - I(V_1 \wedge Z|U) < R \tag{C.7}$$

then there exists V also satisfying (C.3), (C.4), the latter with equality, such that

$$I(V \wedge Z|U) \ge I(V_1 \wedge Z|U). \tag{C.8}$$

To this end, let J be an rv with values in $\{1, 2\}$, with $\Pr\{J = 1\} = \alpha = 1 - \Pr\{J = 2\}$, and independent of U, Y, Z, V_1. Set $V = (V_J, J)$, where $V_2 = Y$. Then

$$\begin{aligned} I(V \wedge Z|U) &= I(V_J \wedge Z|UJ) \\ &= \alpha I(V_1 \wedge Z|U) + (1-\alpha) I(V_2 \wedge Z|U) \\ &\ge I(V_1 \wedge Z|U) \end{aligned}$$

on account of (C.6). Clearly, $V = (V_J, J)$ satisfies (C.3); and it will satisfy (C.4) with equality for a suitable choice of $0 < \alpha < 1$ owing to (C.7) and

$$I(V_2 \wedge Y|U) - I(V_2 \wedge Z|U) = H(Y|UZ) > R.$$

This completes the proof of Lemma C.1.

Returning to the proof of the bound in (2.25), apply Lemma C.1 with XZ in the role of Z, which is permissible owing to (2.22). Since the Markov condition (2.18) implies

$$H(Y|UXZ) = H(Y|XZ) \tag{C.9}$$

it follows that

$$\begin{aligned} &\max_V I(V \wedge XZ|U) \\ &\qquad = \max_V I(V \wedge Y|U) - \min\{R_2, H(Y|XZ)\} \end{aligned} \tag{C.10}$$

where the maximizations are subject to (2.22) and (2.23).

It can be seen similarly as for (C.1) that $V' = UV$ satisfies the conditions (2.11), (2.12) also under the present assumptions (2.18), (2.22), and (2.23) on U, V. On this occasion, the inequality (C.2) follows from (2.23) since the Markov condition (2.22) implies $I(V \wedge X|UY) = 0$. Hence, using (C.10), we obtain that

$$\begin{aligned} &I(U \wedge Y) + \max_V I(V \wedge XZ|U) \\ &\quad = \max_V I(UV \wedge Y) - \min\{R_2, H(Y|XZ)\} \\ &\quad \le \max_{V'} I(V' \wedge Y) - \min\{R_2, H(Y|XZ)\} \end{aligned} \tag{C.11}$$

where the last maximum is with respect to rv's satisfying the conditions (2.11), (2.12). Here

$$\begin{aligned} \max_V I(V' \wedge Y) &\le \max_{V'} I(V'X \wedge Y) \\ &= I(X \wedge Y) + \max_{V'} I(V' \wedge Y|X) \\ &= I(X \wedge Y) + \max_{V'} I(V' \wedge Z|X) \\ &\quad + \min\{R_2, H(Y|XZ)\} \end{aligned} \tag{C.12}$$

where the last equality holds by Lemma C.1 when applied with X in the role of U. A comparison of (C.11) and (C.12) shows that (2.21) implies (2.25), as claimed.

APPENDIX D

Proof of (3.48)*:* We first claim that the Markov hypothesis $X -\!\circ\!- Z -\!\circ\!- Y$ implies

$$Y_{i+1}^n -\!\circ\!- Z_{i+1}^n -\!\circ\!- X^n Y^i Z_i \tag{D.1}$$

whence—recalling that $f = f(X^n)$—we get

$$Y_{i+1}^n -\!\circ\!- Z_{i+1}^n -\!\circ\!- f Y^i Z_i \tag{D.2}$$

which is a stronger property than (3.48). The claim in (D.1) is a straightforward consequence of

$$Y_{i+1}^n -\!\circ\!- Z_{i+1}^n -\!\circ\!- X_{i+1}^n$$

(by the hypothesis $X -\!\circ\!- Z -\!\circ\!- Y$), and the fact that $(X_{i+1}^n, Y_{i+1}^n, Z_{i+1}^n)$ is independent of (X^i, Y^i, Z^i).

We next claim that the Markov hypothesis $X -\!\circ\!- Y -\!\circ\!- Z$ implies

$$X^n Y^n Z_{i+1}^n -\!\circ\!- Y_i -\!\circ\!- Z_i \tag{D.3}$$

whence

$$f Y^n Z_{i+1}^n -\!\circ\!- Y_i -\!\circ\!- Z_i \tag{D.4}$$

which is again a stronger property than (3.48). The claim in (D.3) is seen to be a straightforward consequence of

$$X_i -\!\circ\!- Y_i -\!\circ\!- Z_i$$

and the fact that $(X^{i-1}, X_{i+1}^n, Y^{i-1}, Y_{i+1}^n, Z_{i+1}^n)$ is independent of (X_i, Y_i, Z_i).

APPENDIX E

Recall the functions $F_1(R_1, R_2)$ (cf. (2.9)), $F_2(R_1, R_2)$ (as the right-hand side of (2.1)), and $F_3(R_1, R_2)$ (cf. (3.37)). Also, denote by $F_4(R_1, R_2)$ the function of $R_1 \geq 0,\ R_2 \geq 0$ defined by the maximum in (2.7).

Lemma E.1: $F_i(R_1, R_2), i = 1, 2, 3, 4$, are concave functions of (R_1, R_2). In particular, these functions are continuous for $R_1 > 0, R_2 > 0$.

Proof: It suffices to show that for any $0 < \alpha < 1$ and $(U_1, V_1), (U_2, V_2)$ satisfying the Markov conditions (2.5), (2.6), we can find $(\overline{U}, \overline{V})$ satisfying the same Markov conditions such that

$$I(\overline{U} \wedge X) = \alpha I(U_1 \wedge X) + (1-\alpha) I(U_2 \wedge X) \tag{E.1}$$

$$I(\overline{U} \wedge Y) = \alpha I(U_1 \wedge Y) + (1-\alpha) I(U_2 \wedge Y) \tag{E.2}$$

$$I(\overline{U} \wedge Z) = \alpha I(U_1 \wedge Z) + (1-\alpha) I(U_2 \wedge Z) \tag{E.3}$$

$$I(\overline{V} \wedge Y|\overline{U}) = \alpha I(V_1 \wedge Y|U_1) + (1-\alpha) I(V_2 \wedge Y|U_2) \tag{E.4}$$

$$I(\overline{V} \wedge Z|\overline{U}) = \alpha I(V_1 \wedge Z|U_1) + (1-\alpha) I(V_2 \wedge Z|U_2). \tag{E.5}$$

Indeed, (E.2) and (E.3) imply that

$$\begin{aligned} &\min\{I(\overline{U} \wedge Y), I(\overline{U} \wedge Z)\} \\ &\quad \geq \alpha \min\{I(U_1 \wedge Y), I(U_1 \wedge Z)\} \\ &\qquad + (1-\alpha) \min\{I(U_2 \wedge Y), I(U_2 \wedge Z)\} \end{aligned} \tag{E.6}$$

and the concavity inequality

$$\begin{aligned} &F_i(\alpha R_1^1 + (1-\alpha) R_1^2, \alpha R_2^1 + (1-\alpha) R_2^2) \\ &\qquad \geq \alpha F_i(R_1^1, R_2^1) + (1-\alpha) F_i(R_1^2, R_2^2) \end{aligned} \tag{E.7}$$

$i = 1, 2, 3, 4$, follows from (E.1), (E.6), (E.4), (E.5) (where, for the case $i = 2$, we also observe that (2.5) implies $H(X|YU) = H(X|Y) - I(U \wedge X|Y) = H(X|Y) - I(U \wedge X) + I(U \wedge Y)$). The proof of Lemma E.1 is completed noting that $(\overline{U}, \overline{V})$ with the required properties is obtained by setting $\overline{U} = (U_J, J), \overline{V} = V_J$, where J is an rv independent of $X, Y, Z, U_1, V_1, U_2, V_2$, and taking values in $\{1, 2\}$ with $\Pr\{J = 1\} = \alpha = 1 - \Pr\{J = 2\}$.

Lemma E.2: For fixed $R_2 \geq 0$, write

$$g(U) = \min\{I(U \wedge Y), I(U \wedge Z)\} + \max_V I(V \wedge Y|U) \tag{E.8}$$

where the maximum is taken with respect to rv's V satisfying the Markov condition (2.6) and the rate condition (2.4). Let $\overline{U}$ maximize $g(U)$ subject to the Markov condition (2.8), and set

$$\overline{R}_1 = I(\overline{U} \wedge X) - \min\{I(\overline{U} \wedge Y), I(\overline{U} \wedge Z)\}. \tag{E.9}$$

Then

$$F_3(R, R_2) = \begin{cases} F_1(R_1, R_2), & \text{if } R_1 \leq \overline{R}_1 \\ R_1 + g(\overline{U}), & \text{if } R_1 \geq \overline{R}_1 \end{cases} \tag{E.10}$$

and

$$F_2(R_1, R_2) = F_3(R_1, R_2) \quad \text{if } R_1 \leq \overline{R}_1 + H(X|\overline{U}Y). \tag{E.11}$$

Proof: By definition, $F_1(R_1, R_2)$ equals the maximum of $I(U \wedge X) + I(V \wedge Y|Z)$ subject to the Markov conditions (2.5), (2.6), and the rate conditions (2.8), (2.4). Hence

$$\begin{aligned} F_1(R_1, R_2) &= \max_U \big[I(U \wedge X) \\ &\qquad - \min\{I(U \wedge Y), I(U \wedge Z)\} + g(U)\big] \\ &\leq R_1 + \max_U g(U) \leq R_1 + g(\overline{U}) \end{aligned} \tag{E.12}$$

where the maxima are with respect to rv's U satisfying (2.8), (2.5). With $\overline{U}$ and $\overline{R}_1$ as in the lemma, both inequalities in (E.12) become equalities for $R_1 = \overline{R}_1$, i.e.,

$$\begin{aligned} F_1(\overline{R}_1, R_2) &= I(\overline{U} \wedge X) + \max_V I(V \wedge Y|\overline{U}) \\ &= \overline{R}_1 + g(\overline{U}). \end{aligned} \tag{E.13}$$

By (E.12) and (E.13), the graph of $F_1(R_1, R_2)$ as a function of R_1—concave by Lemma E.1—meets its supporting line of slope 1 at $R_1 = \overline{R}_1$. This implies (E.10), since on account of the definition of $F_3(R_1, R_2)$, the graph of $F_3(R_1, R_2)$ as a function of R_1 is the upper envelope of all straight lines of slope 1 starting (upwards) from points of the graph of $F_1(R_1, R_2)$. The equality (E.11) for $R_1 \leq \overline{R}_1$ immediately follows from (E.10), and

$$F_1(R_1, R_2) \leq F_2(R_1, R_2) \leq F_3(R_1, R_2). \tag{E.14}$$

Further, by (E.13) and the definition of $F_2(R_1, R_2)$,

$$F_2(\overline{R}_1 + t, R_2) \geq F_1(\overline{R}_1, R_2) + t = \overline{R}_1 + t + g(\overline{U}) \tag{E.15}$$

if $0 \le t \le H(X|\overline{U}Y)$. As $\overline{R}_1 + t + g(\overline{U}) = F_3(\overline{R}_1 + t, R_2)$ by (E.10), this completes the proof of (E.11).

Lemma E.3: For all $R_1 \le \max\{H(X|Y), H(X|Z)\}$, we have that

$$F_1(R_1, R_2) = F_3(R_1, R_2) \tag{E.16}$$

if either $H(X|Y) \ge H(X|Z)$ or $R_2 \le H(Y|XZ)$. Moreover

$$F_3(R_1, R_2) \le R_1 + R_2 + F_4(R_1, R_2), \tag{E.17}$$

with equality holding iff the maximum in (2.7) defining $F_4(R_1, R_2)$ is achieved by some U, V which satisfy the rate condition (2.4) with equality. In particular, equality always holds in (E.17) if $R_2 \le H(Y|XZ)$, or if

$$R_2 \le \min\{\tilde{R}_2 \colon F_4(R_1, \tilde{R}_2) = F_4(R_1, \infty)\}.$$

Further, equality in (E.17) for some (R_1, R_2) implies the same for (R_1, R_2') whenever $R_2' < R_2$.

Proof:

i) By Lemma E.2, the equality (E.16) holds for all

$$R_1 \le \max\{H(X|Y), H(X|Z)\}$$

if $\overline{U}$ in that Lemma can be chosen as $\overline{U} = X$. Note that in (E.8)

$$I(U \wedge Y) + \max_V I(V \wedge Y|U) = \max_V I(UV \wedge Y) \tag{E.18}$$

and, using Lemma C.1 in Appendix C

$$\begin{aligned} &I(U \wedge Z) + \max_V I(V \wedge Y|U) \\ &\quad = \max_V I(UV \wedge Z) + \min\{R_2, H(Y|UZ)\}. \end{aligned} \tag{E.19}$$

We claim that both $\max_V I(UV \wedge Y)$ and $\max_V (UV \wedge Z)$ in (E.18) and (E.19) are maximized with respect to U (subject to (2.5), i.e., $U \multimap X \multimap YZ$) by $\overline{U} = X$. Establishing this will prove that $\overline{U} = X$ attains the maximum of $g(U)$ in (E.8) if either (E.18) does not exceed (E.19) when $U = X$, or if $\min\{R_2, H(Y|UZ)\}$ does not depend on U. Since $H(X|Y) \ge H(X|Z)$ implies the former, and $R_2 \le H(Y|XZ)$ implies the latter, as then

$$\min\{R_2, H(Y|UZ)\} = R_2 \tag{E.20}$$

(on account of $U \multimap X \multimap YZ$), this will complete the proof of (E.16) in both cases.

Since $I(UV \wedge Y) \le I(UVX \wedge Y), I(UV \wedge Z) \le I(UVX \wedge Z)$, our claim will be established if we show that the maximum of $I(UV \wedge Y)$ (resp., $I(UV \wedge Z)$) with respect to V satisfying (2.4) and (2.6) is achieved by $\overline{V}$ such that $U' = X$ and $V' = U\overline{V}$ satisfy (2.4), (2.6).

Now, any V satisfying (2.6) can be replaced by $\overline{V}$ satisfying the stronger Markov condition

$$\overline{V} \multimap UY \multimap XZ \tag{E.21}$$

without changing the relevant mutual information quantities, simply by letting the conditional distribution of $\overline{V}$ given $UXYZ$ equal that of V given UYZ (or, equivalently, UY). Hence, the maximizer $\overline{V}$ of $I(UV \wedge Y)$ (resp., $I(UV \wedge Z)$) as above can be chosen to satisfy (E.21). Then, the Markov conditions $U \multimap X \multimap YZ$ and (E.21) imply that $U\overline{V} \multimap XY \multimap Z$, i.e., $U' = X$ and $V' = U\overline{V}$ satisfy (2.6). Further, as $U \multimap X \multimap YZ$ implies $I(U \wedge Y|X) = I(U \wedge Z|X) = 0$, and (E.21) implies $I(\overline{V} \wedge X|UY) = 0$, we have

$$\begin{aligned} &I(U\overline{V} \wedge Y|X) - I(U\overline{V} \wedge Z|X) \\ &\quad = I(\overline{V} \wedge Y|UX) - I(\overline{V} \wedge Z|UX) \\ &\quad = I(\overline{V} \wedge XY|U) - I(\overline{V} \wedge XZ|U) \\ &\quad = I(\overline{V} \wedge Y|U) - I(\overline{V} \wedge XZ|U) \end{aligned} \tag{E.22}$$

This shows that $U' = X$ and $V' = U\overline{V}$ satisfy (2.4) if $U, \overline{V}$ do so, completing the proof of the first assertion of Lemma E.3.

ii) Let U, V, t achieve the maximum in (3.37) defining $F_3(R_1, R_2)$. Then

$$\begin{aligned} F_3(R_1, R_2) &= I(U \wedge X) + t + I(V \wedge Y|U) \\ &\le \min\{I(U \wedge Y), I(U \wedge Z)\} \\ &\quad + R_1 + I(V \wedge Z|U) + R_2 \\ &\le F_4(R_1, R_2) + R_1 + R_2. \end{aligned}$$

where the first inequality holds by the rate conditions (2.3), (2.4), and the second by the definition of $F_4(R_1, R_2)$ as the maximum of (2.7) (noting that the rate condition (2.3) in the definition of $F_3(R_1, R_2)$ implies the rate condition (2.8) in the definition of $F_4(R_1, R_2)$). This proves (E.17), and also shows that equality in (E.17) holds only when U, V achieve the maximum defining $F_4(R_1, R_2)$, and they satisfy the rate condition (2.4) with equality, i.e.,

$$I(V \wedge Y|U) - I(V \wedge Z|U) = R_2. \tag{E.23}$$

On the other hand, if the maximum defining $F_4(R_1, R_2)$ is achieved by some U, V which satisfy (E.23), and we set

$$t = R_1 - I(U \wedge X) + \min\{I(U \wedge Y), I(U \wedge Z)\}$$

(nonnegative by (2.8) and obviously satisfying the rate condition (2.3)), we have

$$\begin{aligned} &F_4(R_1, R_2) + R_1 + R_2 \\ &\quad = \min\{I(U \wedge Y), I(U \wedge Z)\} + I(V \wedge Z|U) + R_1 + R_2 \\ &\quad = I(U \wedge X) + t + I(V \wedge Y|U) \\ &\quad \le F_3(R_1, R_2). \end{aligned}$$

We thus see in this case that equality holds in (E.17).

Note next that if U, V achieve the maximum defining $F_4(R_1, R_2)$, then V maximizes $I(V \wedge Z|U)$ subject to the rate condition (2.4) and the Markov condition (2.6). Hence, by Lemma C.1, V satisfies (E.23) if $R_2 \le H(Y|ZU)$. Since the Markov condition (2.5) implies that $H(Y|ZU) \ge H(Y|XZ)$, the condition $R_2 \le H(Y|ZU)$ is certainly fulfilled if $R_2 \le H(Y|XZ)$.

Denoting by R_2^* the smallest R_2-rate with $F_4(R_1, R_2) = F_4(R_1, \infty)$, the concavity of the nondecreasing function $F_4(R_1, R_2)$ (with R_1 fixed) implies that $F_4(R_1, R_2') < F_4(R_1, R_2)$ whenever $R_2' < R_2 \le R_2^*$. This means that for $R_2 \le R_2^*$, the maximum defining $F_4(R_1, R_2)$ cannot be achieved with rv's U, V satisfying the rate condition (2.4) with strict inequality, i.e., U, V must satisfy (E.23).

It remains to prove the last assertion in the case $R_2^* < R_2' < R_2$. Now, the assumed equality in (E.17) for R_2 implies the existence of rv's U_1, V_1 satisfying the constraints in the definition of $F_4(R_1, R_2)$ that achieve the maximum in (2.7) (now equal

to $F_4(R_1,\infty)$), and satisfy (E.23). By the previous paragraph, there exist rv's U_0, V_0 with the same properties when R_2 is replaced by R_2^*. Representing R_2' as $R_2' = \alpha R_2^* + (1-\alpha)R_2$, take $U' = (U_J, J), V' = V_J$, where J is a $\{0,1\}$-valued rv independent of all the others, with

$$\Pr\{J=0\} = \alpha = 1 - \Pr\{J=1\}.$$

Then, U', V' satisfy the constraints in the definition of $F_4(R_1, R_2')$, achieve the maximum defining $F_4(R_1, R_2')$ (since it equals $F_4(R_1,\infty)$), and they satisfy (E.23).

This completes the proof of Lemma E.3.

Lemma E.4: Suppose that $X \multimap Y \multimap Z$ and $H(Y|XZ) \le R_2 \le H(Y|Z)$. Then

$$F_2(R_1,R_2) = \begin{cases} F_3(R_1,R_2), & \text{if } R_1 \le H(XY|Z) - R_2 \\ H(XY), & \text{if } R_1 \ge H(XY|Z) - R_2. \end{cases} \tag{E.24}$$

Proof: Our Markov assumption causes the condition (2.5) to become

$$U \multimap X \multimap Y \multimap Z \tag{E.25}$$

and $g(U)$ in (E.8) to reduce to (E.19). It follows that

$$g(U) \le I(Y \wedge Z) + \min\{R_2, H(Y|UZ)\}, \tag{E.26}$$

since the Markov conditions (2.6) and (E.25) imply that

$$I(UV \wedge Z) \le I(UY \wedge Z) = I(Y \wedge Z). \tag{E.27}$$

It is also seen from (E.19) that (E.26) holds with equality if $H(Y|UZ) = R_2$. Thus the maximum of $g(U)$ subject to (E.25) equals

$$g(\overline{U}) = I(Y \wedge Z) + R_2 \tag{E.28}$$

and is achieved for any $\overline{U}$ which satisfies $H(Y|\overline{U}Z) = R_2$ in addition to the Markov condition (E.25). A suitable choice is

$$\overline{U} = (U_J, J) \text{ with } U_1 = \text{constant},\ U_2 = X \tag{E.29}$$

where J is an rv independent of X, Y, Z, with $\Pr\{J=1\} = \alpha = 1 - \Pr\{J=2\}$, such that

$$\alpha H(Y|Z) + (1-\alpha)H(Y|XZ) = R_2. \tag{E.30}$$

Since (E.9), with (E.29), gives that

$$\begin{aligned} \overline{R}_1 &= I(\overline{U} \wedge X) - I(\overline{U} \wedge Z) \\ &= I(U_J \wedge X|J) - I(U_J \wedge Z|J) \\ &= (1-\alpha)H(X|Z) \end{aligned} \tag{E.31}$$

and (E.30) is equivalent to $\alpha I(X \wedge Y|Z) = R_2 - H(Y|XZ)$, it follows that

$$\begin{aligned} \overline{R}_1 + H(X|\overline{U}Y) &= (1-\alpha)H(X|Z) + \alpha H(X|Y) \\ &= H(X|Z) - \alpha I(X \wedge Y|Z) \\ &= H(X|Z) - R_2 + H(Y|XZ) \\ &= H(XY|Z) - R_2. \end{aligned} \tag{E.32}$$

Finally, (E.24) follows from (E.11) and (E.32), since

$$F_2(R_1,R_2) \le H(XY)$$

for all R_1 as easily seen from the definition of $F_2(R_1,R_2)$, and (E.10), (E.32), and (E.28) give that

$$\begin{aligned} &F_3(\overline{R}_1 + H(X|\overline{U}Y), R_2) \\ &\quad = \overline{R}_1 + H(X|\overline{U}Y) + g(\overline{U}) \\ &\quad = H(XY|Z) - R_2 + I(Y \wedge Z) + R_2 \\ &\quad = H(XY) \end{aligned}$$

where the last equality follows from the Markov assumption $X \multimap Y \multimap Z$.

APPENDIX F

The maximum correlation $S(Y,Z)$ of arbitrary rv's Y, Z is defined as the supremum of $E[f(Y)g(Z)]$ for real-valued functions f, g of Y, Z,, respectively, such that

$$E[f(Y)] = E[g(Z)] = 0, E[f^2(Y)] = E[g^2(Z)] = 1.$$

We shall only consider rv's taking values in finite sets; then, as shown in [14], $S(Y,Z) < 1$ iff Y and Z have no nonconstant c.f.'s. Further [14, Theorem 2], for any mappings $\varphi : \mathcal{Y} \to \{0,1\}$, $\psi : \mathcal{Z} \to \{0,1\}$, the probability that $\varphi(Y) \ne \psi(Z)$ is bounded below as

$$\begin{aligned} &\Pr\{\varphi(Y) \ne \psi(Z)\} \\ &\quad \ge 2(1 - S(Y,Z))[\Pr\{\varphi(Y)=0\}\Pr\{\phi(Y)=1\} \\ &\qquad \Pr\{\psi(Z)=0\}\Pr\{\psi(Z)=1\}]^{1/2}. \end{aligned} \tag{F.1}$$

A key property of maximum correlation [14] is that for $Y^n = (Y_1, \cdots, Y_n), Z^n = (Z_1, \cdots, Z_n)$, such that the n pairs (Y_i, Z_i) are mutually independent but not necessarily independent and identically distributed (i.i.d.)

$$S(Y^n, Z^n) = \max_{1 \le i \le n} S(Y_i, Z_i). \tag{F.2}$$

Proof of Lemma 1.1: Consider first the case when Y and Z have no c.f.'s except the constants, i.e., $S(Y,Z) < 1$. Then we have to show that (1.14) with $\epsilon > 0$ sufficiently small implies that $K(Y^n)$ is constant with probability $> 1 - \xi$. Suppose instead that $K(Y^n)$ does not equal a constant with probability $> 1 - \xi$. Then, there exists a set A of possible values of $K(Y^n)$ such that

$$\xi < \Pr\{K(Y^n) \in A\} < 1 - \xi. \tag{F.3}$$

Define $\{0,1\}$-valued functions of Y^n and Z^n by

$$\varphi(Y^n) = \begin{cases} 1, & \text{if } K(Y^n) \in A \\ 0, & \text{otherwise} \end{cases}$$

$$\psi(Z^n) = \begin{cases} 1, & \text{if } L(Z^n) \in A \\ 0, & \text{otherwise.} \end{cases}$$

Then (1.14) implies that

$$\Pr\{\varphi(Y^n) = \psi(Z^n)\} < \epsilon \tag{F.4}$$

and by (F.3), (F.4), we have

$$\begin{aligned} \xi < \Pr\{\varphi(Y^n)=1\} &< 1 - \xi, \\ \xi - \epsilon < \Pr\{\psi(Z^n)=1\} &< 1 - \xi + \epsilon. \end{aligned} \tag{F.5}$$

Since $S(Y^n, Z^n) = S(Y, Z) < 1$ by (F.2), we get from (F.1) and (F.5) a contradiction to (F.4) if $\epsilon > 0$ is small.

Turning to the case when the maximal c.f. V of Y and Z is nonconstant, for an arbitrary fixed value $\boldsymbol{v} = (v_1, \cdots, v_n)$ of V^n, assign to $(Y_1, Z_1), \cdots, (Y_n, Z_n)$ the conditional distribution under the condition $V^n = \boldsymbol{v}$. As $(Y_i, Z_i), i = 1, \cdots, n$, are independent, though not i.i.d. under this assignment, (F.2) gives for the corresponding maximum correlation that

$$S(Y^n, Z^n | V^n = \boldsymbol{v}) = \max_{1 \le i \le n} S(Y_i, Z_i | V_i = v_i) \le \max_v S(Y, Z | V = v) < 1. \quad \text{(F.6)}$$

Here, the last inequality follows since V is a maximal c.f. of Y, Z. Indeed, the latter implies that if the conditional distribution under the condition $V = v$ is assigned to Y, Z for arbitrary v, then under this assignment each c.f. of Y, Z must be constant with probability 1.

Using (F.6), it follows as in the first part of the proof that for any $\xi > 0$, if

$$\Pr\{K(Y^n) = L(Z^n) | V^n = \boldsymbol{v}\} > 1 - \epsilon \quad \text{(F.7)}$$

holds with $\epsilon > 0$ sufficiently small (not depending on n or $\boldsymbol{v}$), then $K(Y^n)$ equals a constant, say $M(\boldsymbol{v})$, with $\Pr\{\cdot | V^n = \boldsymbol{v}\}$-probability $> 1 - \xi$.

Finally, (1.14) implies that the set D of $\boldsymbol{v}$'s for which (F.7) holds with ϵ replaced by $\epsilon\xi^{-1}$, has probability $\Pr\{V^n \in D\} > 1 - \xi$. Clearly, the conclusion in the previous paragraph remains unaffected if ϵ in (F.7) is replaced by $\epsilon\xi^{-1}$. Thus we have proved that (1.14) implies

$$\Pr\{K(Y^n) = M(V^n)\} \ge \Pr\{V^n \in D\} \Pr\{K(Y^n) = M(V^n) | V^n \in D\} > (1 - \xi)^2$$

where $\xi > 0$ can be arbitrarily small if $\epsilon > 0$ in (1.14) is sufficiently small. This completes the proof of Lemma 1.1.

REFERENCES

[1] R. Ahlswede and V. B. Balakirsky, "Identification under random processes," *Probl. Pered. Inform. (Special issue devoted to M. S. Pinsker)*, vol. 32, no. 1, pp. 144–160, 1996.

[2] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography, Part I: Secret sharing," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, July 1993.

[3] ——, "Common randomness in information theory and cryptography, Part II: CR capacity," *IEEE Trans. Inform. Theory*, vol. 44, pp. 225–240, Jan. 1998.

[4] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer, "Generalized privacy amplification," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1915–1923, Nov. 1995.

[5] I. Csiszár, "Almost independence and secrecy capacity," *Probl. Pered. Inform. (Special issue devoted to M. S. Pinsker)*, vol. 32, no. 1, pp. 48–57, 1996.

[6] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York, NY: Academic, 1981.

[7] P. Gács and J. Körner, "Common information is far less than mutual information," *Probl. Contr. Inform. Theory*, vol. 21, pp. 149–162, 1973.

[8] U. M. Maurer, "Secret key agreement by public discussion from common information," *IEEE Trans. Inform. Theory*, vol. 39, pp. 733–742, May 1993.

[9] ——, "The strong secret key rate of discrete random triples," in *Communications and Cryptography: Two Sides of One Tapestry*, R. E. Blahut *et al.*, Eds. Norwell, MA: Kluwer, 1994, ch. 26, pp. 271–285.

[10] ——, "Information-theoretically secure secret-key agreement by NOT authenticated public discusion," in *Advances in Cryptology-EUROCRYPT '97 Lecture Notes in Computer Science*, W. Fumy, Ed. New York, NY: Springer, 1997, pp. 209–225.

[11] U. Maurer and S. Wolf, "Privacy amplification secure against active adversaries," in *Advances in Cryptology-CRYPTO '97 Lecture Notes in Computer Science*, B. Kaliski, Ed. New York, NY: Springer, 1997, pp. 307–321.

[12] S. Venkatesan and V. Anantharam, "The common radomness capacity of a pair of independent discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 44, pp. 215–224, Jan. 1998.

[13] ——, "The common randomness capacity of a network of discrete memoryless channels," *IEEE Trans. Inform. Theory*, to appear.

[14] H. S. Witsenhausen, "On sequences of pairs of dependent random variables," *SIAM J. Appl. Math.*, vol. 38, pp. 100–113, 1975.