The Common Randomness Capacity of a Network of Discrete Memoryless Channels

Sivarama Venkatesan, *Member, IEEE,* and Venkat Anantharam, *Fellow, IEEE*

Abstract—In this paper, we generalize our previous results on generating common randomness at two terminals to a situation where any *finite* number of agents, interconnected by an *arbitrary* network of independent, point-to-point, discrete memoryless channels, wish to generate common randomness by interactive communication over the network. Our main result is an exact characterization of the common randomness capacity of such a network, i.e., the maximum number of bits of randomness that all the agents can agree on per step of communication. As a by-product, we also obtain a purely combinatorial result, viz., a characterization of (the incidence vectors of) the spanning arborescences rooted at a specified vertex in a digraph, and having exactly one edge exiting the root, as precisely the extreme points of a certain unbounded convex polyhedron, described by a system of linear inequalities.

Index Terms—Common randomness, interactive communication, polyhedral characterization, spanning arborescences, strong converse.

I. INTRODUCTION

A. Previous Work

TWO or more communicating agents are said to have *common randomness* if there is a random experiment whose outcome is available to all of them. This notion of shared randomness turns out to be of significance in many problems of information theory. For example, common randomness available to a transmitter and receiver allows them to use random codes for data transmission, which can outperform deterministic codes in certain situations, e.g., with arbitrarily varying channels [1], [10], [8]. In the theory of identification over noisy channels [4]–[7], the maximum achievable identification rate is essentially determined by the amount of common randomness that the transmitter and receiver can set up. Common randomness can also significantly reduce the communication complexity of certain distributed computations [17], [20]. Finally, *secret* common randomness available to a transmitter and receiver allows them to communicate securely over a channel with eavesdroppers [2], [18], [19].

With these considerations in mind, Ahlswede and Csiszár initiated a systematic study of the role of common randomness in information theory and cryptography. To begin with, in [2], they addressed the problem of generating common randomness at two terminals without giving information about it to an eavesdropper (secret sharing) in various "source-type" and "channel-type" models, and proved bounds on the rates at which the two terminals could generate common randomness in these models.

Later, in [3], they studied the rates at which common randomness could be generated at two terminals without any secrecy requirements, but under various other resource constraints, such as access to side information and communication links. Interestingly, the common randomness results obtained in [3] also led to new insights into the data transmission capacity of an arbitrarily varying channel (AVC) with feedback.

In both [2] and [3], however, the possibility of exploiting *channel noise* to generate common randomness was not addressed, except in the simple case of two terminals connected by a noisy discrete memoryless channel with noiseless feedback. To see how channel noise actually could be useful in this context, consider a situation where there are two agents, Alice and Bob, connected to each other in both directions by a pair of channels. As an extreme case, assume that neither Alice nor Bob has access to any external sources of randomness, such as a random bit generator. Even then, they may be able to generate common randomness! The intuition is this: suppose Alice transmits some agreed upon input sequence to Bob. If the channel from Alice to Bob is noisy, then the resulting output sequence seen by Bob will be random. Since the input sequence is known, Bob could somehow "cancel" out its effect on the output sequence, and extract the randomness due to noise. Now, if the channel from Bob to Alice has positive Shannon capacity, then Bob could reliably convey to Alice the randomness thus obtained, by using suitable encoding techniques, thereby generating *common* randomness.

Of course, this procedure could be repeated, and Alice could simultaneously extract randomness from the output of the Bob-to-Alice channel (if it is noisy) and convey it reliably to Bob (if the Alice-to-Bob channel has positive Shannon capacity), thus generating more common randomness. It is clear that two conditions must be satisfied for Alice and Bob to be able to extract common randomness from channel noise as above: at least one of the channels must be noisy (otherwise, no randomness would be available), and at least one of them must have positive Shannon capacity (otherwise, the agents would not be able to agree on a common random output).

A natural and interesting question that arises in the above context is: what is the maximum *rate*, in bits per step of communica-

tion, at which Alice and Bob can generate common randomness from the noise on the two channels, i.e., what is the common randomness *capacity* of the pair of channels connecting them? This question was posed and answered in [23], under the assumption that the two channels are independently operating discrete memoryless channels (DMC's). The main result of [23] is that Alice and Bob can generate

$$\max_{X_A, X_B} \{\min[H(Y_B|X_B), I(X_A; Y_A)]$$
$$+ \min[H(Y_A|X_A), I(X_B; Y_B)]\} \quad (1)$$

bits of common randomness per step of communication over such a pair of channels. In (1), (X_A, Y_A) and (X_B, Y_B) are random input–output pairs for the Alice-to-Bob and Bob-to-Alice DMC's, respectively, and the maximum is over all possible distributions for the inputs X_A and X_B. The rate in (1) was shown to be optimal with a "strong" converse. Further, it was shown that the common randomness capacity in the presence of independent, discrete memoryless sources of external randomness at the two terminals could actually be derived from (1).

In the special case where both channels are binary-symmetric with crossover probabilities p and q, respectively, the common randomness capacity expression in (1) reduces to

$$\min\{h(p) + h(q), 2 - h(p) - h(q)\}$$

where

$$h(x) \stackrel{\text{def}}{=} -x \log x - (1 - x) \log(1 - x)$$

is the binary entropy function (with base 2 logarithms). This expression is 0 if and only if either $h(p) = h(q) = 0$ (no randomness from noise on either channel) or $h(p) = h(q) = 1$ (plenty of randomness available from channel noise but no ability to agree). Moreover, the common randomness capacity equals its maximum of 1 bit/step whenever

$$h(p) + h(q) = 1 - h(p) + 1 - h(q)$$

i.e., whenever the entropies and capacities of the two channels balance each other. For details and other examples, see [23].

B. Subject of this Paper

The results of [23] apply only to situations where common randomness is to be generated at *two* distant terminals. In this paper, we study a much more general problem where any *finite* number of agents, interconnected by an arbitrary network of point-to-point channels, wish to generate common randomness. As in the Alice–Bob case, we show that the agents can do so by communicating interactively over the network, and exploiting the noise on the channels. Our main result is an exact characterization of the common randomness capacity of any such network, i.e., the maximum number of bits of randomness that *all* the agents can agree on, per step of communication over the network.

We assume that the topology of the network is represented by a digraph $G = (V, E)$. The vertex set V of this digraph is just the set of agents, and the edge set $E \subseteq V \times V$ describes their interconnections—$(v, w) \in E$ means there is a channel

whose input is controlled by v, and whose output is seen by w. We assume that these channels are all discrete memoryless, that they operate independently, and that communication occurs simultaneously on all the channels, and in synchronism (i.e., there is a common clock).

The digraph G is allowed to have *self-loops*, i.e., edges of the form (v, v). Such edges can be used to incorporate into the network itself any external sources of randomness that the agents may have. For example, suppose agent v has a random bit generator providing R_v independent and unbiased bits in each step of communication. We could then assume that $(v, v) \in E$, and that the (v, v) DMC has only one input symbol, which produces 2^{R_v} equiprobable output symbols. In this way, v would get randomness at rate R_v from the output of the (v, v) DMC, instead of the external source. However, we do not restrict the self-loop DMC's to have this special form. In general, v could vary the distribution of the (v, v) DMC's output from step to step, just by varying the input symbol. Further, v could introduce memory in the sequence of outputs, by adapting the input in each step to all past outputs. With this in mind, we will assume that all available external sources have been incorporated into the network itself.

As in the two-agent case, there are basically two steps here in the process of generating common randomness. The first step is for the agents to bring channel noise into play by communicating over the channels, so that each agent can then extract randomness from the observed channel outputs. The second step is for each agent to convey reliably the randomness thus obtained to all the others, using suitable encoding techniques.

The mechanism for extracting randomness from noise is the same as in the two-agent case. However, several new features show up in the second step. In the two-agent case, there is only one path along which an agent can deliver randomness to the other, and this path consists of a single channel. Also, the flows of randomness originating from the two agents do not interact—each is confined to a different channel.

In contrast, in a general network of channels, there could be several paths from one agent to each of the others, many of these consisting of more than one channel, and all these paths could be used simultaneously to deliver randomness. Moreover, a given channel could be on several different paths, which means that the flows of randomness from different agents must interact.

For these reasons, the problem of optimally disseminating randomness from each agent to all the others is quite nontrivial in the general situation. In fact, the solution to this problem led to a purely combinatorial result of independent interest, viz., a *polyhedral characterization* of the spanning arborescences rooted at a specified vertex of a digraph, and containing at most d edges exiting the root; we proved that the incidence vectors of all such spanning arborescences are precisely the extreme points of a certain unbounded convex polyhedron, described by a system of linear inequalities.

Only the case $d = 1$ of the above result is relevant to the proof of the common randomness capacity theorem, and therefore we restrict ourselves in this paper to just that case. The simple proof given here for $d = 1$ does not seem to extend to higher values of d. Details of the statement and proof of the result for general d can be found in [22]. That result can be regarded as a generalization of Fulkerson's [14] well-known characterization of

all the spanning arborescences rooted at a specified vertex of a digraph (no degree constraints).

Polyhedral characterizations of the above kind are useful in the study of linear-objective combinatorial optimization (LOCO) problems (see, e.g., [16] or [15, Ch. 30]), since they allow such problems to be reduced to linear programs. Often, a polyhedral characterization, in conjunction with the LP duality theorem, leads naturally to an efficient algorithm for solving an associated LOCO problem.

The rest of the paper is organized as follows. The problem of generating common randomness over the network is formulated precisely in Section II. A formal definition of common randomness capacity is given in Section II-B. As with all capacity results in information theory, there are two parts to our characterization of the common randomness capacity—an "achievability" part which states that common randomness *can* be generated at a certain rate, and a "converse" part which states that no higher rate is achievable. These are stated in Sections II-D and II-E, respectively. From a technical point of view, the converse part, specifically the so-called "strong" converse, is the more interesting of the two. Its proof requires the development of a "typical sequence" machinery for interactive communication over a network of DMC's, which might be of independent interest. The polyhedral characterization referred to above is stated in Section II-G. Sections III—VI contain the proofs of the main theorems.

II. PROBLEM FORMULATION AND STATEMENT OF RESULTS

A. Notation and Preliminary Definitions

The following conventions will be in effect throughout: all logarithms and exponentials will be to the base two. If N is a positive integer, then $[N] \stackrel{\text{def}}{=} \{1, 2, \cdots, N\}$. $\lfloor z \rfloor$ will denote the largest integer not exceeding z. The standard sequence notation $z^k = (z_1, z_2, \cdots, z_k)$ will be employed. If S is a finite set, then $\boldsymbol{z} = (z_s : s \in S)$ will mean that $\boldsymbol{z}$ is a vector whose components are indexed by the elements of S. $\boldsymbol{R}_+$ will denote the set of all nonnegative real numbers, and $\boldsymbol{R}_+^S$ the set of all vectors $(z_s \in \boldsymbol{R}_+ : s \in S)$. If

$$\boldsymbol{W} = (W(y|x) : (x, y) \in \mathcal{X} \times \mathcal{Y})$$

is the matrix of transition probabilities of a discrete memoryless channel (DMC) with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$, P is a probability distribution on $\mathcal{X}$, and X and Y are random variables with the joint distribution

$$\Pr\{X = x, Y = y\} = P(x)W(y|x)$$

then we will write either $H(X)$ or $H(P)$ for the entropy of X, either $H(Y|X)$ or $H(\boldsymbol{W}|P)$ for the conditional entropy of Y given X, either $H(Y)$ or $H(P\boldsymbol{W})$ for the entropy of Y, and either $I(X;Y)$ or $I(P;\boldsymbol{W})$ for the mutual information between X and Y. (Note that $P\boldsymbol{W}$ is the distribution on $\mathcal{Y}$ given by $P\boldsymbol{W}(y) = \sum_x P(x)W(y|x)$.) As mentioned earlier, the network of DMC's connecting the agents will be assumed to be represented by the digraph $G = (V, E)$. Depending on the context, we will refer to elements of V as either vertices or agents, and to elements of E as either edges or channels. The DMC

corresponding to the edge $e \in E$ will be assumed to have finite input alphabet $\mathcal{X}_e$, finite output alphabet $\mathcal{Y}_e$, and transition probability matrix

$$\boldsymbol{W}_e = (W_e(y|x) : (x, y) \in \mathcal{X}_e \times \mathcal{Y}_e).$$

We will say that the edge (v, w) *exits* v and *enters* w. If $U \subseteq V$, then

$$\delta^-(U) \stackrel{\text{def}}{=} \{(v, w) \in E : v \notin U, w \in U\} \tag{2}$$

$$\delta^+(U) \stackrel{\text{def}}{=} \{(v, w) \in E : v \in U, w \notin U\} \tag{3}$$

$$\sigma(U) \stackrel{\text{def}}{=} \{(v, w) \in E : v \in U, w \in U\}. \tag{4}$$

Thus $\delta^-(U)$ (resp., $\delta^+(U)$) is the set of edges that exit (resp., enter) a vertex not in U and enter (resp., exit) one in U, while $\sigma(U)$ is the set of edges that exit *and* enter vertices in U. To simplify notation, we will write $\delta^-(v)$, $\delta^+(v)$, and $\sigma(v)$ for $\delta^-(\{v\})$, $\delta^+(\{v\})$, and $\sigma(\{v\})$, respectively. Note that $\sigma(v)$, if nonempty, contains exactly one edge, viz., a self-loop on v. It will also be convenient further to define

$$\delta_{\text{in}}(U) \stackrel{\text{def}}{=} \delta^-(U) \cup \sigma(U) \tag{5}$$

$$\delta_{\text{out}}(U) \stackrel{\text{def}}{=} \delta^+(U) \cup \sigma(U). \tag{6}$$

Thus $\delta_{\text{in}}(U)$ (resp., $\delta_{\text{out}}(U)$) is the set of edges that enter (resp., exit) a vertex in U. In particular, $\delta_{\text{in}}(v)$ (resp., $\delta_{\text{out}}(v)$) is just $\delta^-(v)$ (resp., $\delta^+(v)$) with the self-loop on v thrown in, if it exists.

B. Definition of Common Randomness Capacity

To generate common randomness, the agents communicate interactively over the network for a certain number of steps, say n, according to an agreed-upon set of rules. In each of these steps, each agent transmits an input symbol on each of his outgoing channels, based solely on the outputs of all his incoming channels in all previous steps. More elaborately, in step k $(1 \leq k \leq n)$, agent v does the following in sequence and in synchronism with all the other agents:

- He determines the input symbol to be transmitted in step k on each outgoing channel $e \in \delta_{\text{out}}(v)$, as a function

$$X_{e,k} = f_{e,k}\left(Y_{e'}^{k-1} : e' \in \delta_{in}(v)\right)$$

of the sequences of outputs $Y_{e'}^{k-1}$ received in the previous $k - 1$ steps on all incoming channels $e' \in \delta_{in}(v)$.

- He then transmits the symbols $X_{e,k}$ on their respective outgoing channels.

- Finally, he receives the outputs $Y_{e',k}$ corresponding to the symbols transmitted in step k on all his incoming channels.

After n steps, each agent either computes a random output taking values in a common finite set of size, say, K—without loss of generality, we will take this set to be $[K] \stackrel{\text{def}}{=} \{1, 2, \cdots, K\}$—or decides that the attempt to generate

common randomness failed. Each agent's decision is based solely on the output sequences available to him. Formally, agent v computes a *decision random variable* S_v that is a function of all the output sequences $Y_{e'}^n$, $e' \in \delta_{in}(v)$, and takes values in the set $\{*\} \cup [K]$. Here, $S_v = *$ is supposed to indicate that v declared failure to generate common randomness.

Let $f_e = (f_{e,1}, \cdots, f_{e,n})$, and $\boldsymbol{f} = (f_e\colon e \in E)$. Let $\boldsymbol{S} = (S_v\colon v \in V)$. Then the pair $(\boldsymbol{f}, \boldsymbol{S})$, which all the agents agree on before communication begins, sums up the set of rules according to which the agents communicate over the network and make their final decisions. We will refer to $(\boldsymbol{f}, \boldsymbol{S})$ as an (n, K) *protocol* for generating common randomness. Of course, the "amount" of randomness generated by this protocol, and the extent to which it is truly "common," are determined by the joint distribution of the decision random variables S_v, $v \in V$. Ideally, we would like to have

$$\Pr\{S_v = l \text{ for all } v \in V\} = \frac{1}{K}, \qquad \text{for each } l \in [K] \quad (7)$$

with K as large as possible. If (7) were true, then all the S_v's would be equal with probability 1, and uniformly distributed over $[K]$. (There would be no "failure" events of positive probability.) Such a protocol could reasonably be said to generate $\log K$ bits of common randomness in n steps of communication.

In general, however, it is not possible to satisfy (7) except in the trivial case $K = 1$. Therefore, we will have to settle for *approximate* equality and uniformity of the S_v's. To this end, we give the following definition:

Definition 2.1: $(\boldsymbol{f}, \boldsymbol{S})$ is an (n, K, λ) protocol if

$$\frac{1-\lambda}{K} \leq \Pr\{S_v = l \text{ for all } v \in V\} \leq \frac{1+\lambda}{K},$$
$$\text{for each } l \in [K]. \quad (8)$$

If $\lambda = 0$, then (8) reduces to (7), which, as we just mentioned, is too stringent a requirement. On the other hand, if $\lambda \geq 1$, then (8) is completely trivial; we can then satisfy it with arbitrarily large K, simply by letting each S_v equal $*$ with probability 1. For the above reasons, we will always assume from now on that $\lambda \in (0, 1)$. It is clear from (8) that λ is a measure of how far the joint distribution of the S_v's deviates from the ideal of (7). If λ is a small positive number, then the S_v's will be equal with high probability, and have marginal distributions close to uniform on $[K]$, conditional on the high-probability event $\{S_v \neq * \text{ for any } v\}$. The above considerations motivate the following definition of the common randomness capacity C of the given network of DMC's:

Definition 2.2: Let $K(n, \lambda)$ be the largest integer K such that an (n, K, λ) protocol exists for generating common randomness over the given network. The common randomness capacity of the network is

$$C \overset{\text{def}}{=} \lim_{\lambda \downarrow 0} \liminf_{n \to \infty} (1/n) \log K(n, \lambda). \quad (9)$$

Equivalently, C is the supremum of all rates R such that there exists a sequence of (n, K_n, λ_n) protocols for generating

common randomness over the network, with K_n and λ_n satisfying

$$\liminf_{n \to \infty} (1/n) \log K_n = R \quad \text{and} \quad \lim_{n \to \infty} \lambda_n = 0. \quad (10)$$

The equivalence of the two definitions of C given above is easily verified. We will say that rate R of generating common randomness is achievable if there exists a sequence of (n, K_n, λ_n) protocols satisfying (10). Thus C is the supremum of all achievable rates.

Our main result is a "single-letter" characterization of C, the common randomness capacity, in terms of the topology of the network and the transition probability matrices of the channels constituting it. We prove this characterization in three steps. First, we prove that any rate not exceeding a certain C_* is achievable, so that $C \geq C_*$. Then we prove a "weak" converse result asserting that

$$\lim_{\lambda \downarrow 0} \limsup_{n \to \infty} (1/n) \log K(n, \lambda) \leq C^* \quad (11)$$

for a C^* that is apparently different from C_*. Finally, we prove that C_* and C^* are in fact equal. (It is the proof of $C_* = C^*$ that leads to the combinatorial result mentioned in Section I-B.) Since (11) obviously implies $C \leq C^*$, we then have $C_* = C = C^*$. We also prove that a "strong" converse result holds here, i.e.,

$$\limsup_{n \to \infty} (1/n) \log K(n, \lambda) \leq C^*, \qquad \text{for every } \lambda < 1. \quad (12)$$

Thus

$$\lim_{n \to \infty} (1/n) \log K(n, \lambda) = C$$

for *every* $\lambda \in (0, 1)$. Clearly, (12) says much more than (11), the "weak" converse. As is to be expected, its proof is also much harder, requiring an elaborate "typical sequence" machinery for dealing with interactive communication over a network of DMC's. However, such strong converses greatly strengthen the significance of coding theorems in information theory, as argued by Wolfowitz [25], and it is therefore desirable that they be proved wherever possible.

We provide separate proofs for (11) and (12), the weak and strong converses. Although (11) is clearly implied by (12), its proof is much simpler, and based on more conventional techniques, which is why we include it here. Also, the weak converse suffices to characterize the common randomness capacity C.

C. Definition of Spanning Arborescence

The graph-theoretic concept of a *spanning arborescence* will play a very important role in our characterization of the common randomness capacity C. Essentially, a spanning arborescence is the counterpart in a digraph of a spanning tree in an undirected graph. For a formal definition, we need to define path and circuit first. The following definitions are all with respect to the given digraph $G = (V, E)$.

A *path* from vertex u to vertex w is a set of $k \geq 1$ edges

$$\{(v_0, v_1), (v_1, v_2), (v_2, v_3), \cdots, (v_{k-1}, v_k)\}$$

with $v_0 = u$ and $v_k = w$. A *circuit* is a path from some vertex to itself. Note that a self-loop (v, v) constitutes a circuit by itself.

Definition 2.3: A spanning arborescence rooted at v is a set T of edges, with the following properties: i) no edge in T enters v; ii) for each $u \neq v$, exactly one edge in T enters u; iii) T does not contain any circuits.

It can be verified easily that in any spanning arborescence T rooted at v, there is a *unique* path from v to u, for each vertex $u \neq v$, and that the edges of T form a spanning tree in the undirected graph underlying G. We will denote the set of all spanning arborescences rooted at v by $\mathcal{T}(v)$, and let

$$\mathcal{T} \stackrel{\text{def}}{=} \bigcup_{v \in V} \mathcal{T}(v).$$

If $(v, w) \in T$, then we will say that v is the *parent of w in T*, and w is a *child of v in T*. Note that a vertex can have more than one child in T, but no more than one parent—in fact, every vertex other than the root has exactly one parent in T (the root has none). A vertex with no children in T will be called a *leaf of T* (every spanning arborescence has at least one leaf).

D. Achievability Result: $C \geq C_*$

For each $e \in E$, let P_e be a probability distribution on the input alphabet $\mathcal{X}_e$ of channel e, and let $\boldsymbol{P} = (P_e : e \in E)$. Let the vectors $\boldsymbol{H}(\boldsymbol{P}) \in \boldsymbol{R}_+^V$ and $\boldsymbol{I}(\boldsymbol{P}) \in \boldsymbol{R}_+^E$ be given by

$$H_v(\boldsymbol{P}) \stackrel{\text{def}}{=} \sum_{e \in \delta_{in}(v)} H(\boldsymbol{W}_e | P_e) \tag{13}$$

$$I_e(\boldsymbol{P}) \stackrel{\text{def}}{=} I(P_e; \boldsymbol{W}_e). \tag{14}$$

Let $\mathcal{R}(\boldsymbol{P})$ be the polyhedron of all vectors $\boldsymbol{r} \in \boldsymbol{R}_+^{\mathcal{T}}$ satisfying the following constraints:

$$\sum_{T \in \mathcal{T}(v)} r_T \leq H_v(\boldsymbol{P}), \qquad \text{for each } v \in V \tag{15}$$

$$\sum_{T : e \in T} r_T \leq I_e(\boldsymbol{P}), \qquad \text{for each } e \in E. \tag{16}$$

The "achievability" part of our characterization of the common randomness capacity essentially states that for any $\boldsymbol{P}$ and any $\boldsymbol{r} \in \mathcal{R}\boldsymbol{P})$ the rate $\sum_T r_T$ is achievable.

Theorem 2.1 (Achievability Result) : Let

$$C_*(\boldsymbol{P}) \stackrel{\text{def}}{=} \max_{\boldsymbol{r} \in (\mathcal{R}\boldsymbol{P})} \sum_{T \in \mathcal{T}} r_T \tag{17}$$

$$C_* \stackrel{\text{def}}{=} \max_{\boldsymbol{P}} C_*(\boldsymbol{P}). \tag{18}$$

Then

$$\liminf_{n \to \infty} (1/n) \log K(n, \lambda) \geq C_*, \qquad \text{for every } \lambda > 0. \tag{19}$$

In particular, the common randomness capacity $C \geq C_*$.

The intuition behind the achievability result is this: let $\boldsymbol{P} = (P_e : e \in E)$ be given. Suppose the agents communicate with each other over the network for a large number of steps in such a way that, on average, the input symbols of channel e are used

according to the distribution P_e. Then, on average, channel e corrupts each transmitted symbol by adding $H(\boldsymbol{W}_e | P_e)$ bits of noise. If the agents always use codewords from reliable block codes for communication, then they can recover this randomness due to channel noise from their respective incoming channels, by first decoding the transmitted codewords. This way, agent v can extract

$$\sum_{e \in \delta_{in}(v)} H(\boldsymbol{W}_e | P_e) \stackrel{\text{def}}{=} H_v(\boldsymbol{P})$$

bits of randomness from channel noise, in each step of communication.

To generate *common* randomness, each agent must then reliably convey as much of this randomness as possible, subject to capacity constraints, to all the other agents. This is where the concept of spanning arborescence comes in. A spanning arborescence rooted at v is just a minimal set of edges in which there is a path from v to every other vertex. Therefore, agent v could disseminate a part of the extracted randomness through each spanning arborescence T rooted at v, by sending a message to his children in T, each of whom relays it in turn to his own children in T, and so on, till the message reaches every agent.

Suppose v wants to send r_T bits of randomness per step this way through the spanning arborescence $T \in \mathcal{T}(v)$. Since v gets only $H_v(\bar{\boldsymbol{P}})$ bits of randomness per step from noise, we have the constraint

$$\sum_{T \in \mathcal{T}(v)} r_T \leq H_v(\boldsymbol{P})$$

for each $v \in V$. Also, the channel e can reliably convey at most

$$I(P_e; \boldsymbol{W}_e) \stackrel{\text{def}}{=} I_e(\boldsymbol{P})$$

bits per step if its input symbols are used according to the distribution P_e. Therefore, we also have the constraint

$$\sum_{T \ni e} r_T \leq I_e(\boldsymbol{P})$$

for each $e \in E$. These are just the conditions (15) and (16) defining the polyhedron $\mathcal{R}(\boldsymbol{P})$.

The above reasoning suggests that, as long as the vector of rates $\boldsymbol{r}$ is in $\mathcal{R}(\boldsymbol{P})$, the agents can generate common randomness at a total rate of $\sum_{v \in V} \sum_{T \in \mathcal{T}(v)} r_T$. The inner sum here is the rate at which common randomness originates at v (from the noise on the channels entering v), and is then conveyed reliably to all the other agents. In particular, the rate $R = C_*(\boldsymbol{P})$ must be achievable. This is a plausibility argument for the achievability result $C \geq C_*$. See Section III for its proof.

E. Converse Result: $C \leq C^*$

For each nonempty subset U of V, let

$$C_0(\boldsymbol{P}; U) \stackrel{\text{def}}{=} \sum_{v \in U} H_v(\boldsymbol{P}) + \sum_{e \in \delta^-(U)} I_e(\boldsymbol{P}) \tag{20}$$

and let

$$C_0(\boldsymbol{P}) \stackrel{\text{def}}{=} \min_U C_0(\boldsymbol{P}; U). \tag{21}$$

For each $m \geq 1$, and each collection $V_1, \cdots, V_{m+1}$ of nonempty and pairwise disjoint subsets of V, let

$$C_m(\boldsymbol{P}; V_1, \cdots, V_{m+1}) \overset{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I_e(\boldsymbol{P}) \qquad (22)$$

and let

$$C_m(\boldsymbol{P}) \overset{\text{def}}{=} \min_{V_1, \cdots, V_{m+1}} C_m(\boldsymbol{P}; V_1, \cdots, V_{m+1}). \qquad (23)$$

Theorem 2.2 (Converse Result): Let

$$C^*(\boldsymbol{P}) \overset{\text{def}}{=} \min_{m \geq 0} C_m(\boldsymbol{P}) \qquad (24)$$

$$C^* \overset{\text{def}}{=} \max_{\boldsymbol{P}} C^*(\boldsymbol{P}). \qquad (25)$$

Then

$$\limsup_{n \to \infty} (1/n) \log K(n, \lambda) \leq C^*, \qquad \text{for every } \lambda < 1. \quad (26)$$

In particular, the common randomness capacity $C \leq C^*$.

Here is the intuition behind the converse result: suppose that communication between the agents proceeds in such a way that the input symbols of channel e get used according to the distribution P_e, on average. Let $\boldsymbol{P} = (P_e : e \in E)$. Then, agent v gets $H_v(\boldsymbol{P})$ bits of randomness from channel noise, and channel e can reliably convey at most $I_e(\boldsymbol{P})$ bits, per step of communication.

Let U be any nonempty subset of V. Then, the total rate at which randomness from channel noise originates at vertices in U is $\sum_{v \in U} H_v(\boldsymbol{P})$. The total rate at which randomness could originate at vertices in $V - U$, and then be conveyed reliably to those in U, cannot exceed $\sum_{e \in \delta^-(U)} I_e(\boldsymbol{P})$, the sum of the capacities of all channels going from $V - U$ to U. Therefore, the total rate R at which the agents generate common randomness must satisfy

$$R \leq \sum_{v \in U} H_v(\boldsymbol{P}) + \sum_{e \in \delta^-(U)} I_e(\boldsymbol{P})$$
$$= C_0(\boldsymbol{P}; U). \qquad (27)$$

Next, let $m \geq 1$, and let $V_1, \cdots, V_{m+1}$ be nonempty and pairwise-disjoint subsets of V. If $1 \leq i, j \leq m+1$ and $i \neq j$, then each bit of common randomness that originates at a vertex in V_j must use up one bit of capacity on at least one edge in $\delta^-(V_i)$, in order to reach the vertices in V_i. Therefore, each bit of common randomness that originates at a vertex in $\bigcup_j V_j$ must use up at least m bits of capacity on edges in $\bigcup_i \delta^-(V_i)$. As for the common randomness that originates at a vertex not in $\bigcup_j V_j$ (if such a vertex exists), each such bit must actually use up $m+1$ bits of capacity on edges in $\bigcup_i \delta^-(V_i)$ (one bit on at least one edge in $\delta^-(V_i)$, for each i). This means that the rate R at which the agents generate common randomness cannot exceed $1/m$ times the sum of the capacities of edges in $\bigcup_i \delta^-(V_i)$, i.e.,

$$R \leq \frac{1}{m} \sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I_e(\boldsymbol{P})$$
$$= C_m(\boldsymbol{P}; V_1, \cdots, V_{m+1}). \qquad (28)$$

The bounds on R in (27) and (28) imply that $R \leq C^*(\boldsymbol{P})$. This is a plausibility argument for the converse result $C \leq C^*$. See Section IV for its proof.

F. $C_* = C^*$

To complete our characterization of the common randomness capacity, we need to prove that $C_* = C^*$. In fact, we will show that $C_*(\boldsymbol{P}) = C^*(\boldsymbol{P})$ for every $\boldsymbol{P}$.

Theorem 2.3: For every $\boldsymbol{P}$, $C_*(\boldsymbol{P}) = C^*(\boldsymbol{P})$. Consequently, C_* and C^* are both equal to the common randomness capacity C of the network.

We can gain some insight into the conclusion "$C_*(\boldsymbol{P}) = C^*(\boldsymbol{P})$ for every $\boldsymbol{P}$" of Theorem 2.3, by translating it to an equivalent statement about an extended digraph $\tilde{G}$, defined as follows:

$$\tilde{G} \overset{\text{def}}{=} (\tilde{V}, \tilde{E}),$$
$$\text{where } \tilde{V} \overset{\text{def}}{=} V \cup \{r\} \text{ and } \tilde{E} \overset{\text{def}}{=} E \cup \{(r, v) : v \in V\}. \quad (29)$$

In words, $\tilde{G}$ is obtained from G by adding a new vertex r, and adding new edges from r to all the vertices in G. Note that we may identify the edge (r, v) in $\tilde{G}$ with the vertex v in G. (It is mainly for the convenience of being able to treat vertices and edges of G on an equal footing that we have defined the extended digraph $\tilde{G}$.)

Some notation will be useful here. Let

$$\delta^+(r) \overset{\text{def}}{=} \{(r, v) : v \in V\} \qquad (30)$$

be the set of edges in $\tilde{G}$ that originate at r. For $U \subseteq V$, continue to define $\delta^-(U)$ as in (2). Further, let

$$\tilde{\delta}^-(U) \overset{\text{def}}{=} \delta^-(U) \cup \{(r, v) : v \in U\} \qquad (31)$$

be the set of edges in the extended digraph $\tilde{G}$ that exit vertices not in U and enter vertices in U.

We will denote the set of spanning arborescences in $\tilde{G}$ by $\tilde{\mathcal{T}}$. Note that *all* spanning arborescences in $\tilde{G}$ must be rooted at r, since no edge in $\tilde{G}$ enters r. Further, observe that there is a one-to-one correspondence between the spanning arborescences in $\tilde{G}$ that have exactly one edge exiting r, and the spanning arborescences in the original digraph G: if $T \in \tilde{\mathcal{T}}$, and (r, v) is the only edge in T that originates at r, then T corresponds to the spanning arborescence $T - \{(r, v)\}$ in G, which is rooted at v. Let

$$\tilde{\mathcal{T}}_1 \overset{\text{def}}{=} \{T \in \tilde{\mathcal{T}} : |T \cap \delta^+(r)| = 1\}. \qquad (32)$$

We will now derive the statement about $\tilde{G}$ that is equivalent to "$C_*(\boldsymbol{P}) = C^*(\boldsymbol{P})$ for every $\boldsymbol{P}$." To each $\boldsymbol{P} = (P_e : e \in E)$, there corresponds a vector $\boldsymbol{c} \in \boldsymbol{R}_+^{\tilde{E}}$ defined as follows:

$$c_e \overset{\text{def}}{=} \begin{cases} H_v(\boldsymbol{P}), & \text{if } e = (r, v) \text{ for some } v \in V \\ I_e(\boldsymbol{P}), & e \in E. \end{cases} \qquad (33)$$

Observe that $\boldsymbol{c}$ contains all the information in the vectors $\boldsymbol{H}(\boldsymbol{P})$ and $\boldsymbol{I}(\boldsymbol{P})$. Let

$$\mathcal{P}(\boldsymbol{c}) \overset{\text{def}}{=} \left\{ \boldsymbol{r} \in \boldsymbol{R}_+^{\tilde{\mathcal{T}}} : \sum_{T \ni e} r_T \leq c_e \text{ for each } e \in \tilde{E} \right\}. \qquad (34)$$

Then, it is easy to see that

$$C_*(\boldsymbol{P}) \overset{\text{def}}{=} \max_{\boldsymbol{r} \in \mathcal{R}(\boldsymbol{P})} \sum_{T \in \mathcal{T}} r_T$$
$$= \max_{\boldsymbol{r} \in \mathcal{P}(\boldsymbol{c})} \sum_{T \in \tilde{\mathcal{T}}_1} r_T. \qquad (35)$$

Next, for each nonempty subset U of V, let $\boldsymbol{z}(U) \in \boldsymbol{R}_+^{\tilde{E}}$ be given by

$$z_e(U) \overset{\text{def}}{=} \begin{cases} 1, & \text{if } e \in \tilde{\delta}^-(U) \\ 0, & \text{otherwise} \end{cases} \quad (36)$$

and, for each $m \geq 1$ and each collection $V_1, \cdots, V_{m+1}$ of nonempty and pairwise-disjoint subsets of V, let $\boldsymbol{z}(V_1, \cdots, V_{m+1}) \in \boldsymbol{R}_+^{\tilde{E}}$ be given by

$$z_e(V_1, \cdots, V_{m+1}) \overset{\text{def}}{=} \begin{cases} \dfrac{1}{m}, & \text{if } e \in \displaystyle\bigcup_{i=1}^{m+1} \delta^-(V_i) \\ 0, & \text{otherwise.} \end{cases} \quad (37)$$

Then, it is easy to see that

$$\begin{aligned} C_0(\boldsymbol{P}; U) &\overset{\text{def}}{=} \sum_{v \in U} H_v(\boldsymbol{P}) + \sum_{e \in \delta^-(U)} I_e(\boldsymbol{P}) \\ &= \sum_{e \in \tilde{E}} c_e z_e(U) \end{aligned} \quad (38)$$

and

$$\begin{aligned} C_m(\boldsymbol{P}; V_1, \cdots, V_{m+1}) &\overset{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I_e(\boldsymbol{P}) \\ &= \sum_{e \in \tilde{E}} c_e z_e(V_1, \cdots, V_{m+1}). \end{aligned} \quad (39)$$

If D_1 is the finite subset of $\boldsymbol{R}_+^{\tilde{E}}$ consisting of all the vectors $\boldsymbol{z}(U)$ and $\boldsymbol{z}(V_1, \cdots, V_{m+1})$ defined in (36) and (37), then by (38) and (39)

$$\begin{aligned} C^*(\boldsymbol{P}) &\overset{\text{def}}{=} \min_{m \geq 0} C_m(\boldsymbol{P}) \\ &= \min_{\boldsymbol{z} \in D_1} \sum_{e \in \tilde{E}} c_e z_e. \end{aligned} \quad (40)$$

From (35) and (40), it is clear that the statement about $\tilde{G}$ that is equivalent to "$C_*(\boldsymbol{P}) = C^*(\boldsymbol{P})$ for every $\boldsymbol{P}$" is

$$\max_{\boldsymbol{r} \in \mathcal{P}(\boldsymbol{c})} \sum_{T \in \tilde{T}_1} r_T = \min_{\boldsymbol{z} \in D_1} \sum_{e \in \tilde{E}} c_e z_e, \quad \text{for every } \boldsymbol{c} \in \boldsymbol{R}_+^{\tilde{E}}. \quad (41)$$

In the terminology of polyhedral combinatorics, (41) is a "max-min equality" (see, e.g., [15, Ch. 30]). It is clear that a proof of (41) would constitute a proof of Theorem 2.3 as well. Now (41) can, in turn, be regarded as a corollary of another combinatorial result, which we stated in Section II-G. Essentially this result is a characterization of (the incidence vectors of) the spanning arborescences in $\tilde{G}$ that use exactly one edge exiting r, as precisely the extreme points of a certain unbounded polyhedron defined by a system of linear inequalities; the system has one inequality for each vector in the finite set D_1. This is similar to a well-known result of Fulkerson [14] that gives an analogous characterization of *all* the spanning arborescences in $\tilde{G}$.

G. Combinatorial Result

For each $T \in \tilde{T}$, define an incidence vector $\boldsymbol{\xi}(T) \in \boldsymbol{R}_+^{\tilde{E}}$ as follows:

$$\xi_e(T) \overset{\text{def}}{=} \begin{cases} 1, & \text{if } e \in T \\ 0, & \text{otherwise.} \end{cases} \quad (42)$$

Let

$$A \overset{\text{def}}{=} \{\boldsymbol{\xi}(T) : T \in \tilde{T}\} \quad (43)$$

be the set of all such incidence vectors.

Consider the polyhedron $\text{conv}\,(A) + \boldsymbol{R}_+^{\tilde{E}}$, i.e., the set of vectors in $\boldsymbol{R}_+^{\tilde{E}}$ that lie on or above the convex hull of the incidence vectors of all spanning arborescences in $\tilde{G}$. An important and well-known result of Fulkerson [14] characterizes this polyhedron as the solution set of a certain system of linear inequalities:

Theorem 2.4 (Fulkerson): Let

$$\mathcal{A} \overset{\text{def}}{=} \{\boldsymbol{\xi} \in \boldsymbol{R}_+^{\tilde{E}} : \boldsymbol{z} \cdot \boldsymbol{\xi} \geq 1 \text{ for all } \boldsymbol{z} \in D\} \quad (44)$$

where D is the finite set consisting of all the vectors $\boldsymbol{z}(U)$ defined in (36). Then

$$\mathcal{A} = \text{conv}\,(A) + \boldsymbol{R}_+^{\tilde{E}}. \quad (45)$$

The extreme points of the polyhedron $\mathcal{A}$ are precisely the vectors in A.

We prove a result similar to Fulkerson's when there is a degree constraint on the root r: we characterize by linear inequalities the polyhedron $\text{conv}\,(A_1) + \boldsymbol{R}_+^{\tilde{E}}$, where

$$A_1 \overset{\text{def}}{=} \{\boldsymbol{\xi}(T) : T \in \tilde{T}_1\}. \quad (46)$$

Theorem 2.5 (Polyhedral Characterization): Let

$$\mathcal{A}_1 \overset{\text{def}}{=} \{\boldsymbol{\xi} \in \boldsymbol{R}_+^{\tilde{E}} : \boldsymbol{z} \cdot \boldsymbol{\xi} \geq 1 \text{ for all } \boldsymbol{z} \in D_1\} \quad (47)$$

where D_1 is the finite set consisting of all the vectors $\boldsymbol{z}(U)$ defined in (36), and all the vectors $\boldsymbol{z}(V_1, \cdots, V_{m+1})$ defined in (37). Then

$$\mathcal{A}_1 = \text{conv}\,(A_1) + \boldsymbol{R}_+^{\tilde{E}}. \quad (48)$$

The extreme points of the polyhedron $\mathcal{A}_1$ are precisely the vectors in A_1.

The following curious fact is worth noting: although $\mathcal{A}_1$ is obtained by adding more constraints to those defining $\mathcal{A}$, it actually has fewer extreme points than $\mathcal{A}$.

Theorem 2.5 is proved in Section VI. Here, we will show how to derive the max-min equality (41) from it. Using the definition of $\mathcal{A}_1$ in (47), we can rewrite the conclusion (48) of Theorem 2.5 as

$$\{\boldsymbol{\xi} \in \boldsymbol{R}_+^{\tilde{E}} : \boldsymbol{z} \cdot \boldsymbol{\xi} \geq 1 \text{ for each } \boldsymbol{z} \in D_1\} = \text{conv}\,(A_1) + \boldsymbol{R}_+^{\tilde{E}}. \quad (49)$$

We can paraphrase (49) as follows: an unbounded polyhedron whose *facets* are determined by D_1 equals an unbounded polyhedron whose *extreme points* are determined by A_1. This is essentially a relation between the finite sets D_1 and A_1. Rather interestingly, this relation is symmetric, i.e., from (49) we can conclude that

$$\{\boldsymbol{z} \in \boldsymbol{R}_+^{\tilde{E}} : \boldsymbol{\xi} \cdot \boldsymbol{z} \geq 1 \text{ for each } \boldsymbol{\xi} \in A_1\} = \text{conv}\,(D_1) + \boldsymbol{R}_+^{\tilde{E}}. \quad (50)$$

A paraphrase of (50) is obtained from that of (49) simply by interchanging *facets* and *extreme points*. The elegant "duality"

between (49) and (50) is a fundamental result in the theory of blocking polyhedra, developed by Fulkerson in [12] and [13]:

Lemma 2.1: Let A and D be finite subsets of $\boldsymbol{R}_+^n$, and let $\mathcal{D}$ and $\mathcal{A}$ be their respective "blockers":

$$\mathcal{D} \stackrel{\text{def}}{=} \{\boldsymbol{z} \in \boldsymbol{R}_+^n \colon \boldsymbol{\xi} \cdot \boldsymbol{z} \geq 1 \text{ for all } \boldsymbol{\xi} \in A\}$$

$$\mathcal{A} \stackrel{\text{def}}{=} \{\boldsymbol{\xi} \in \boldsymbol{R}_+^n \colon \boldsymbol{z} \cdot \boldsymbol{\xi} \geq 1 \text{ for all } \boldsymbol{z} \in D\}.$$

Then, $\mathcal{A} = \operatorname{conv}(A) + \boldsymbol{R}_+^n$ if and only if $\mathcal{D} = \operatorname{conv}(D) + \boldsymbol{R}_+^n$.

Proof: See, e.g., [15, Ch. 30, Theorem 6.3]. As with many other "duality" results, the proof is by an application of Farkas's lemma. □

Denoting the polyhedron

$$\{\boldsymbol{z} \in \boldsymbol{R}_+^{\tilde{E}} \colon \boldsymbol{\xi} \cdot \boldsymbol{z} \geq 1 \text{ for each } \boldsymbol{\xi} \in A_1\}$$

on the left-hand side (LHS) of (1) by $\mathcal{D}_1$, we have

$$\mathcal{D}_1 = \operatorname{conv}(D_1) + \boldsymbol{R}_+^{\tilde{E}}. \tag{51}$$

In the terminology of [13], $\mathcal{A}_1$ and $\mathcal{D}_1$ constitute a *blocking pair of polyhedra*, and the matrices whose rows are the vectors in D_1 and A_1, respectively, constitute a *blocking pair of matrices*. This blocking relation has many interesting consequences, one of which is the max–min equality we wish to prove (see [13] for others).

Consider the linear program (LP)

$$\max \sum_{T \in \tilde{\mathcal{T}}_1} r_T \quad \text{subject to} \quad \boldsymbol{r} \in \mathcal{P}(\boldsymbol{c}) \tag{52}$$

on the LHS of (41). We will now write down the *dual* to this LP. The dual LP has a variable $z_e \in \boldsymbol{R}_+$ for each $e \in \tilde{E}$. The dual feasible region is the set of all vectors $\boldsymbol{z} \in \boldsymbol{R}_+^{\tilde{E}}$ satisfying $\sum_{e \in T} z_e \geq 1$ for each $T \in \tilde{\mathcal{T}}_1$, which is just the polyhedron $\mathcal{D}_1$. The dual objective is to minimize $\sum_{e \in \tilde{E}} c_e z_e$.

By linear programming duality, the optimal values of the primal and dual programs are equal, i.e.,

$$\max_{\boldsymbol{r} \in \mathcal{P}(\boldsymbol{c})} \sum_{T \in \tilde{\mathcal{T}}_1} r_T = \min_{\boldsymbol{z} \in \mathcal{D}_1} \sum_{e \in \tilde{E}} c_e z_e. \tag{53}$$

Now, by (51) and the nonnegativity of $\boldsymbol{c}$, the minimum of $\sum_e c_e z_e$ over all $\boldsymbol{z} \in \mathcal{D}_1$ must equal the minimum of $\sum_e c_e z_e$ over all $\boldsymbol{z} \in \operatorname{conv}(D_1)$; by linearity, the latter must equal the minimum of $\sum_e c_e z_e$ over all $\boldsymbol{z} \in D_1$. Thus

$$\min_{\boldsymbol{z} \in \mathcal{D}_1} \sum_{e \in \tilde{E}} c_e z_e = \min_{\boldsymbol{z} \in D_1} \sum_{e \in \tilde{E}} c_e z_e. \tag{54}$$

From (53) and (54), we have (41).

III. PROOF OF THE ACHIEVABILITY RESULT

A. Equitype Codes and Equipartitions

The proof of the achievability result is essentially a formalization of the reasoning in Section II-D. It makes use of the following pair of pleasingly dual lemmas, which formed the basis for the achievability proof in [23] as well.

Lemma 3.1 (Equitype Code Lemma): Let

$$\boldsymbol{W} = (W(y|x) \colon (x,y) \in \mathcal{X} \times \mathcal{Y})$$

be a finite-alphabet DMC. For any $R' \geq 0$ and $L \leq \exp(nR')$, there exists a block code of block length n and size L for $\boldsymbol{W}$, all

of whose codewords have the same type P, and whose maximal error probability is at most $\exp\{-n\alpha(R', P) + o(n)\}$. Here

$$\alpha(R', P) \stackrel{\text{def}}{=} \min_{\boldsymbol{V}} [D(\boldsymbol{V}\|\boldsymbol{W}|P) + (I(P; \boldsymbol{V}) - R')^+]$$

is a continuous function of (R', P) that is positive if $R' < I(P; \boldsymbol{W})$ and zero otherwise. The $o(n)$ term does not depend on R' or P.

Proof: This is a classical result in channel coding—see [9, p. 165, Theorem 5.2] for a proof. □

Lemma 3.2 (Equipartition Lemma): Let

$$\boldsymbol{W} = (W(y|x) \colon (x,y) \in \mathcal{X} \times \mathcal{Y})$$

be a finite-alphabet DMC. Suppose $\boldsymbol{c} \in \mathcal{X}^n$ is of type P, and $\mathcal{C} \subseteq \mathcal{Y}^n$. Then, for any $R'' \geq 0$ and $M \leq \exp(nR'')$, there exists a partition of $\mathcal{C}$ into subsets $\mathcal{C}(*), \mathcal{C}(1), \cdots, \mathcal{C}(M)$, such that $W^n(\mathcal{C}(j)|\boldsymbol{c})$ has the same value for every $j \in [M]$, and

$$W^n(\mathcal{C}(*)|\boldsymbol{c}) \leq \exp\{-n\beta(R'', P) + o(n)\}.$$

Here

$$\beta(R'', P) \stackrel{\text{def}}{=} \min_{\boldsymbol{V}} [D(\boldsymbol{V}\|\boldsymbol{W}|P) + (H(\boldsymbol{V}|P) - R'')^+]$$

is a continuous function of (R'', P) that is positive if $R'' < H(\boldsymbol{W}|P)$ and zero otherwise. The $o(n)$ term does not depend on R'' or P.

Proof: See [23, Lemma 3.2]. □

Suppose $n \gg 1$ and P_e is an n-type on the input alphabet $\mathcal{X}_e$ of channel e. Fix an $\epsilon > 0$. Then, by Lemma 3.1, we can find an "equitype" block code of block length n for channel e, with $\exp\{\lfloor n(I(P_e; \boldsymbol{W}_e) - \epsilon)\rfloor\}$ codewords of type P_e, whose maximal error probability is bounded by $\exp(-\alpha n)$, for some $\alpha > 0$. Using this code, it is possible to convey one of $\exp\{\lfloor n(I(P_e; \boldsymbol{W}_e) - \epsilon)\rfloor\}$ messages reliably across the channel e.

By Lemma 3.2, each decoding set in the above code can further be partitioned into $\exp\{\lfloor n(H(\boldsymbol{W}_e|P_e) - \epsilon)\rfloor\}$ subsets that have *equal* probability when the corresponding codeword is transmitted, and a remaining "failure" set whose probability is bounded by $\exp(-\beta n)$, for some $\beta > 0$.

It is these "equipartitions" of the decoding sets that make it possible to generate randomness from noise. Whenever a message is transmitted across channel e using the block code, the agent at the receiving end can first decode the message reliably by observing which decoding set the channel output falls in. Then, by further observing which subset of the partition of that decoding set contains the channel output, he can obtain a "noise variable" B_e that is either a bit string of length $\lfloor n(H(\boldsymbol{W}_e|P_e) - \epsilon)\rfloor$, or the failure symbol $*$. The entropy of B_e is close to $n(H(\boldsymbol{W}_e|P_e) - \epsilon)$ because $\Pr\{B_e = *\} \leq \exp(-\beta n)$ and, conditional on $B_e \neq *$, all the bit strings are equiprobable.

Suppose the agents have agreed upon such equitype block codes, together with equipartitions of their decoding sets, for all the channels in the network. Let $\boldsymbol{P} = (P_e \colon e \in E)$ be the collection of chosen n-types, and let $\boldsymbol{r}$ be a vector in the interior of the polyhedron $\mathcal{R}(\boldsymbol{P})$. We will now describe a protocol for generating common randomness, using the agreed-upon codes

and partitions, that achieves the rate $\sum_{T \in \mathcal{T}} r_T$. This is sufficient to prove the achievability result.

B. Protocol for Generating Common Randomness

The protocol has n rounds, in each of which a codeword from the appropriate block code is transmitted on each channel in the network (these transmissions occur in synchronism). The total number of steps in the protocol is thus n^2. The codewords transmitted by an agent on his outgoing channels in any round are functions only of the channel outputs he received on all his incoming channels in the immediately preceding round—a "Markovian" property.

The following definitions will be useful here. The *depth of v in T*, denoted $d_T(v)$, is the number of edges in the unique path in T from the root to v. The root itself is defined to be at depth 0. The *depth of T* is $d_T \stackrel{\text{def}}{=} \max_v d_T(v)$. Finally, $d \stackrel{\text{def}}{=} \max_T d_T$.

Let the n rounds of the protocol be indexed by $0, 1, \cdots, n-1$. We will now describe the events in round k from the perspective of a particular agent, say v. The round has the following three stages:

1) First, for each spanning arborescence T of which v is not a leaf, agent v transmits a message $M_{T,v}(k)$ to all his children in T, using the appropriate block codes. The message $M_{T,v}(k)$ is either a bitstring of length $\lfloor n r_T \rfloor$, or the symbol $*$ (we will describe in Stage 3 how it is determined).

   Observe that the channel e must, therefore, carry one of $\exp(\lfloor n r_T \rfloor) + 1$ messages for each spanning arborescence T to which it belongs, so that the total number of possible messages is

   $$\prod_{T \ni e} (\exp(\lfloor n r_T \rfloor) + 1).$$

   If ϵ is small enough, this number is indeed smaller than $\exp\{\lfloor n(I(P_e; W_e) - \epsilon) \rfloor\}$—the number of codewords in the code for channel e—because r was assumed to be in the interior of $\mathcal{R}(P)$.

2) Next, for each spanning arborescence T of which v is not the root, agent v decodes the message sent in round k by his parent in T as, say, $\hat{M}_{T,v}(k)$. In the process of decoding these messages, he also obtains a noise variable $B_e(k)$ from each incoming channel $e \in \delta_{in}(v)$, as described earlier. Thus $B_e(k)$ is either a bitstring of length $\lfloor n(H(Q_e|P_e) - \epsilon) \rfloor$, or the failure symbol $*$.

3) Finally, for each spanning arborescence T of which v is not a leaf, agent v determines the messages $M_{T,v}(k+1)$ to be transmitted in the next round to his children in T. There are three cases to consider here:

   a) If T is not rooted at v, then

      $$M_{T,v}(k+1) \stackrel{\text{def}}{=} \hat{M}_{T,v}(k)$$

      the estimate that v made of the message sent in round k by his parent in T.
   b) If $k \geq n - d$, then

      $$M_{T,v}(k+1) \stackrel{\text{def}}{=} *$$

for each spanning arborescence T rooted at v. The reason is that a bit string transmitted by v in round $k+1$ will not reach the leaves of T by round $n-1$ if $d_T \geq n - k$, and will, therefore, not contribute to common randomness anyway.

   c) If $k < n - d$, then the messages $M_{T,v}(k+1)$ for the spanning arborescences T that are rooted at v are determined based on the noise variables $B_e(k)$, $e \in \delta_{in}(v)$, as follows:

      i. If $B_e(k) = *$ for some e, then $M_{T,v}(k+1) = *$ for every $T \in \mathcal{T}(v)$.
      ii. If $B_e(k) \neq *$ for all e, then the messages $M_{T,v}(k+1)$, $T \in \mathcal{T}(v)$, are taken to be disjoint substrings of the concatenation of all the bitstrings $B_e(k)$, with the length of $M_{T,v}(k+1)$ being $\lfloor n r_T \rfloor$.

   In Case ii above, observe that we need a total of $\sum_{T \in \mathcal{T}(v)} \lfloor n r_T \rfloor$ bits to define the messages $M_{T,v}(k+1)$. If ϵ is small enough, this number is indeed smaller than

   $$\sum_{e \in \delta_{in}(v)} \lfloor n(H(Q_e|P_e) - \epsilon) \rfloor$$

   the length of the concatenation of all the $B_e(k)$; because r was assumed to be in the interior of $\mathcal{R}(P)$.

It remains to specify how the messages $M_{T,v}(0)$ are chosen. Actually, the purpose of round 0 is only to initiate the process of extracting randomness from noise, and there are really no messages to be transmitted. We may therefore define $M_{T,v}(0)$ arbitrarily, say $M_{T,v}(0) = *$, for all T and v.

To complete the description of the protocol, we must also describe how the agents compute their decision random variables at the end of round $n - 1$. Observe that in round $k+1$, $0 \leq k < n - d$, the root of the spanning arborescence T transmits a message—derived in round k from the noise variables on his incoming channels—to all his children in T; each of these children forms an estimate of that message in round $k+1$, and in turn conveys that estimate in round $k+2$ to his own children in T, and so on.

In general, agent v forms an estimate $\hat{M}_{T,v}(k+d_T(v))$, in round $k + d_T(v)$, of the message transmitted by the root of T in round $k+1$. If no decoding errors occur, then the estimates formed by all the agents coincide with the original message. Moreover, if there are no failures in extracting randomness from noise (i.e., $B_e(k) \neq *$ for any edge e and any $k \geq 0$), then all such messages are bit strings contributing to common randomness. Observe also that if there are no decoding errors then $\hat{M}_{T,v}(k) = *$ for $k < d_T(v)$ and $k \geq n - d + d_T(v)$.

For spanning arborescences T rooted at v, it will be convenient to define $\hat{M}_{T,v}(k)$ as $M_{T,v}(k+1)$. With this convention, let

$$\hat{M}_v \stackrel{\text{def}}{=} (\hat{M}_{T,v}(k + d_T(v)) : 0 \leq k < n - d, T \in \mathcal{T}), \qquad v \in V. \tag{55}$$

Then, by the remarks made in the preceding paragraph, $\hat{M}_v$ is the vector of v's estimates of the messages sent by the roots of all spanning arborescences in rounds $1, 2, \cdots, n - d$ (for spanning arborescences rooted at v itself, these estimates are the same as

the actual messages, by our convention). Clearly, if no decoding errors occur in any round on any of the channels, then the $\hat{\boldsymbol{M}}_v$'s will all be equal.

Here is how agent v determines his decision random variable S_v. If some component of $\hat{\boldsymbol{M}}_v$ equals $*$, say

$$\hat{M}_{T,v}(k + d_T(v)) = *$$

(for some $T \in \mathcal{T}$ and some $0 \le k < n - d$), then agent v assumes that the root of T failed to generate randomness from noise in round k, and therefore sets $S_v = *$. Agent v also sets $S_v = *$ if $\hat{M}_{T,v}(k) \ne *$ for some $T \in \mathcal{T}$ and some $k < d_T(v)$ or $k \ge n - d + d_T(v)$, as this obviously indicates a decoding error on some channel in some round.

Otherwise, agent v takes S_v to be, say, $1+$ the integer whose binary representation is the concatenation—in some agreed-upon order—of all the bit strings $\hat{M}_{T,v}(k + d_T(v))$, $0 \le k < n - d$, $T \in \mathcal{T}$. Thus, the S_v's all take values in $\{*\} \cup [K]$, where

$$K \stackrel{\text{def}}{=} \exp\left\{ (n-d) \sum_{T \in \mathcal{T}} \lfloor nr_T \rfloor \right\}$$
$$= \exp\left\{ n^2 \left(\sum_{T \in \mathcal{T}} r_T \right) - o(n^2) \right\}. \qquad (56)$$

It remains to prove that the protocol just described does achieve the rate $\sum_{T \in \mathcal{T}} r_T$. We will do this in Section III-C by showing that the protocol satisfies (8) with

$$\lambda \stackrel{\text{def}}{=} n|E|[\exp(-\alpha n) + \exp(-\beta n)] \qquad (57)$$
$$= o(1). \qquad (58)$$

The RHS of (57) can be understood as a union bound on the probability that some "bad" event occurs during the protocol—here, a "bad" event is either a decoding error, or a failure to extract randomness from noise (i.e., $B_e(k) = *$), on some channel in some round.

C. Analysis of the Protocol

For $0 \le k < n$, define the event $\mathcal{D}_k$ as follows:

$$\mathcal{D}_k \stackrel{\text{def}}{=} \{\text{No block-code decoding errors occur}$$
$$\text{in round } k\}.$$

For each $0 \le k < n - d$ and each $T \in \mathcal{T}$, let $\hat{m}_T(k)$ be some fixed bit string of length $\lfloor nr_T \rfloor$. Define the event $\mathcal{M}_k$ as follows:

$$\mathcal{M}_k \stackrel{\text{def}}{=} \{\text{For each } T, \text{ the root of } T \text{ sends the message}$$
$$\hat{m}_T(k) \text{ in round } k+1\}.$$

The following notation will be convenient:

$$\mathcal{D}^k \stackrel{\text{def}}{=} \bigcap_{i \le k} \mathcal{D}_k$$

and

$$\mathcal{M}^k \stackrel{\text{def}}{=} \bigcap_{i \le k} \mathcal{M}_k.$$

Observe that $\mathcal{D}^{n-1} \cap \mathcal{M}^{n-d-1}$ is the event that, for each $0 \le k < n - d$ and each $T \in \mathcal{T}$, the root of T transmits the bit

string $\hat{m}_T(k)$ in round $k+1$, and, for every $v \in V$, the estimate $\hat{M}_{T,v}(k + d_T(v))$ that agent v forms in round $k + d_T(v)$ of that message is correct.

Therefore, to prove that the protocol described in the last section achieves the λ defined in (57), it suffices for us to prove that

$$\frac{1-\lambda}{K} \le \Pr[\mathcal{D}^{n-1} \cap \mathcal{M}^{n-d-1}] \le \frac{1+\lambda}{K}. \qquad (59)$$

The main idea in the proof of (59) is to use the "Markovian" nature of the protocol to decompose the probability of interest. To begin with

$$\Pr[\mathcal{D}^{n-1} \cap \mathcal{M}^{n-d-1}] = \prod_{k=0}^{n-d-1} \Pr[\mathcal{D}_k \cap \mathcal{M}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{k-1}]$$
$$\cdot \prod_{k=n-d}^{n-1} \Pr[\mathcal{D}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{n-d-1}]. \qquad (60)$$

We will bound each of the terms in (60) separately. For $0 \le k < n - d$, let

$$\mathcal{B}_k \stackrel{\text{def}}{=} \{B_e(k) \ne * \text{ for all } e\}.$$

Thus $\mathcal{B}_k$ is the event that there is no failure to extract randomness from noise on any channel in round k.

Then observe that, for $0 \le k < n - d$,

$$\Pr[\mathcal{D}_k \cap \mathcal{M}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{k-1}]$$
$$= \Pr[\mathcal{B}_k \cap \mathcal{D}_k \cap \mathcal{M}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{k-1}] \qquad (61)$$
$$= \Pr[\mathcal{B}_k \cap \mathcal{D}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{k-1}]$$
$$\cdot \Pr[\mathcal{M}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{k-1} \cap \mathcal{B}_k \cap \mathcal{D}_k]. \qquad (62)$$

The equality in (61) is because $\mathcal{M}_k \subseteq \mathcal{B}_k$. Now, the event $\mathcal{D}^{k-1} \cap \mathcal{M}^{k-1}$ determines the codewords transmitted on all the channels in round k. Conditional on these codewords being transmitted, the probability that either a decoding error occurs on channel e, or $B_e(k) = *$, is at most $\exp(-n\alpha) + \exp(-n\beta)$. Since the events on different channels are independent once their inputs are specified, we have

$$\prod_{e \in E} (1 - 2^{-n\alpha} - 2^{-n\beta}) \le \Pr[\mathcal{B}_k \cap \mathcal{D}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{k-1}] \le 1. \qquad (63)$$

Further, by the equipartition property

$$\Pr[\mathcal{M}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{k-1} \cap \mathcal{B}_k \cap \mathcal{D}_k] = \prod_{T \in \mathcal{T}} \frac{1}{2^{\lfloor nr_T \rfloor}}. \qquad (64)$$

Combining (62)–(64), we have

$$\frac{\prod_{e \in E} (1 - 2^{-n\alpha} - 2^{-n\beta})}{\prod_{T \in \mathcal{T}} 2^{\lfloor nr_T \rfloor}}$$
$$\le \Pr[\mathcal{D}_k \cap \mathcal{M}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{k-1}]$$
$$\le \frac{1}{\prod_{T \in \mathcal{T}} 2^{\lfloor nr_T \rfloor}} \qquad (65)$$

for each $0 \le k < n - d$. Next, suppose $n - d \le k < n$. Then the event $\mathcal{D}^{k-1} \cap \mathcal{M}^{n-d-1}$ determines the codewords transmitted on all the channels in round k (recall that the roots of all

spanning arborescences transmit only the symbol $*$ in rounds $n - d + 1, \cdots, n - 1$). Therefore,

$$\prod_{e \in E} (1 - 2^{-n\alpha}) \leq \Pr[\mathcal{D}_k | \mathcal{D}^{k-1} \cap \mathcal{M}^{n-d-1}] \leq 1 \quad (66)$$

for each $n - d \leq k < n$. From (60), (65), and (66), we can conclude that

$$\Pr[\mathcal{D}^{n-1} \cap \mathcal{M}^{n-d-1}] \leq \left(\frac{1}{\prod_{T \in \mathcal{T}} 2^{\lfloor n r_T \rfloor}} \right)^{n-d}$$

$$= \frac{1}{K} \quad (67)$$

and

$$\Pr[\mathcal{D}^{n-1} \cap \mathcal{M}^{n-d-1}]$$

$$\geq \left(\frac{\prod_{e \in E} (1 - 2^{-n\alpha} - 2^{-n\beta})}{\prod_{T \in \mathcal{T}} 2^{\lfloor n r_T \rfloor}} \right)^{n-d} \left(\prod_{e \in E} (1 - 2^{-n\alpha}) \right)^d$$

$$\geq \frac{\prod_{e \in E} (1 - 2^{-n\alpha} - 2^{-n\beta})^n}{K}$$

$$\geq \frac{1 - n|E|(2^{-n\alpha} + 2^{-n\beta})}{K}. \quad (68)$$

The desired result (59) follows from (67) and (68).

IV. PROOF OF THE WEAK CONVERSE

A. Preliminaries

Let $(\boldsymbol{f}, \boldsymbol{S})$ be any (n, K, λ) protocol for generating common randomness, with $K = \exp\{nR + o(n)\}$; this means that $(\boldsymbol{f}, \boldsymbol{S})$ is the nth term in a *sequence* of (n, K_n, λ) protocols, with

$$\limsup_{n \to \infty} (1/n) \log K_n = R.$$

We will then prove that $R \leq C^* + o_\lambda(1)$. Here, $o_\lambda(1)$ denotes a quantity that converges to zero as $\lambda \downarrow 0$.

Let $X_{e,k}$ and $Y_{e,k}$ be the random variables representing the input and output, respectively, on channel e in step k, when the agents communicate according to the n-step "strategy" $\boldsymbol{f}$. For each $e \in E$, define P_e to be the average of the distributions of $X_{e,1}, \cdots, X_{e,n}$

$$P_e(x) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{k=1}^{n} \Pr\{X_{e,k} = x\}.$$

It will be convenient to define random variables X_e and Y_e, for each $e \in E$, with the following joint distribution:

$$\Pr\{X_e = x, Y_e = y\} = P_e(x) W_e(y|x), \quad (x, y) \in \mathcal{X}_e \times \mathcal{Y}_e. \quad (69)$$

Let $\boldsymbol{P} = (P_e : e \in E)$. We will actually prove the weak converse by showing that $R \leq C^*(\boldsymbol{P}) + o_\lambda(1)$. In the proof, the following notation will be useful: if $E' \subseteq V \times V$ then

$$\boldsymbol{X}_{E',k} \stackrel{\text{def}}{=} (X_{e,k} : e \in E' \cap E) \text{ and } \boldsymbol{X}_{E'}^k \stackrel{\text{def}}{=} (\boldsymbol{X}_{E',1}, \cdots, \boldsymbol{X}_{E',k})$$

$$\boldsymbol{Y}_{E',k} \stackrel{\text{def}}{=} (Y_{e,k} : e \in E' \cap E) \text{ and } \boldsymbol{Y}_{E'}^k \stackrel{\text{def}}{=} (\boldsymbol{Y}_{E',1}, \cdots, \boldsymbol{Y}_{E',k}).$$

Further, if $U \subseteq V$, then $\boldsymbol{S}_U \stackrel{\text{def}}{=} (S_v : v \in U)$.

By the definition of $C^*(\boldsymbol{P})$ in (24), to establish $R \leq C^*(\boldsymbol{P}) + o_\lambda(1)$, we must show that $R \leq C_0(\boldsymbol{P}; U) + o_\lambda(1)$ for every nonempty subset U of V, and $R \leq C_m(\boldsymbol{P}; V_1, \cdots, V_{m+1}) + o_\lambda(1)$ for every $m \geq 1$ and every collection $V_1, \cdots, V_{m+1}$ of nonempty and pairwise-disjoint subsets of V. Now

$$C_0(\boldsymbol{P}; U) = \sum_{v \in U} H_v(\boldsymbol{P}) + \sum_{e \in \delta^-(U)} I_e(\boldsymbol{P}) \quad (70)$$

$$= \sum_{v \in U} \sum_{e \in \delta_{in}(v)} H(Y_e|X_e) + \sum_{e \in \delta^-(U)} I(X_e; Y_e) \quad (71)$$

$$= \sum_{e \in \delta_{in}(U)} H(Y_e|X_e) + \sum_{e \in \delta^-(U)} I(X_e; Y_e) \quad (72)$$

and

$$C_m(\boldsymbol{P}; V_1, \cdots, V_{m+1}) = \frac{1}{m} \sum_{j=1}^{m+1} \sum_{e \in \delta^-(V_j)} I_e(\boldsymbol{P}) \quad (73)$$

$$= \frac{1}{m} \sum_{j=1}^{m+1} \sum_{e \in \delta^-(V_j)} I(X_e; Y_e). \quad (74)$$

Here, (70) is by (20); (71) is by (13), (14), and (69); (72) is because $\delta_{in}(U) = \bigcup_{v \in U} \delta_{in}(U)$; (73) is by (22); and (74) is again by (14) and (69).

Because of (72) and (74), it suffices for us to show that

$$R \leq \sum_{e \in \delta_{in}(U)} H(Y_e|X_e) + \sum_{e \in \delta^-(U)} I(X_e; Y_e) + o_\lambda(1) \quad (75)$$

and

$$R \leq \frac{1}{m} \sum_{j=1}^{m+1} \sum_{e \in \delta^-(V_j)} I(X_e; Y_e) + o_\lambda(1). \quad (76)$$

We will prove (75) and (76) in the next two subsections. There, we will need a few simple inequalities. First, by (8)

$$\Pr\{\text{There exists } l \in [K] \text{ such that } S_v = l \text{ for all } v \in V\} \geq 1 - \lambda. \quad (77)$$

Next, note that

$$H(\boldsymbol{S}_V) \stackrel{\text{def}}{=} - \sum_{\boldsymbol{s}} \Pr\{\boldsymbol{S}_V = \boldsymbol{s}\} \log \Pr\{\boldsymbol{S}_V = \boldsymbol{s}\}$$

$$\geq - \sum_{l=1}^{K} \Pr\{S_v = l \text{ for all } v \in V\}$$

$$\cdot \log \Pr\{S_v = l \text{ for all } v \in V\}$$

$$\geq \sum_{l=1}^{K} \left(\frac{1 - \lambda}{K} \right) \log \left(\frac{K}{1 + \lambda} \right)$$

$$= (1 - \lambda) \log K - (1 - \lambda) \log(1 + \lambda)$$

$$\geq (1 - \lambda) \log K - 1. \quad (78)$$

Here, the first inequality holds because $-z \log z \geq 0$ for $z \in [0, 1]$, and the second inequality is by (8). In (78), we have used $\lambda \in (0, 1)$.

Now (77) implies $\Pr\{S_u \neq S_{u'}\} \leq \lambda$ for all u and u'. Hence $H(S_{u'}|S_u) \leq 1 + \lambda \log K$, by Fano's inequality. In fact, for all nonempty sets U and U' of vertices, we have

$$H(\boldsymbol{S}_{U'}|\boldsymbol{S}_U) \leq \sum_{u' \in U'} H(S_{u'}|\boldsymbol{S}_U)$$

$$\leq |U'|(1 + \lambda \log K). \quad (79)$$

The inequalities above are by the chain rule for entropy and the fact that conditioning cannot increase entropy.

From (78) and (79), it follows that for any nonempty $U \subset V$

$$H(\boldsymbol{S}_U) = H(\boldsymbol{S}_V) - H(\boldsymbol{S}_{V-U}|\boldsymbol{S}_U)$$
$$\geq (1 - |V|\lambda)\log K - |V|. \tag{80}$$

Finally, note also that by (79) and (80), with U replaced by U' in (80)

$$I(\boldsymbol{S}_U; \boldsymbol{S}_{U'}) = H(\boldsymbol{S}_{U'}) - H(\boldsymbol{S}_{U'}|\boldsymbol{S}_U)$$
$$\geq (1 - 2|V|\lambda)\log K - 2|V|. \tag{81}$$

B. Proof of $R \leq C_0(\boldsymbol{P}; U) + o_\lambda(1)$

Let U be a nonempty set of vertices. Then, by (80)

$$H(\boldsymbol{S}_U) \geq (1 - |V|\lambda)\log K - |V|$$
$$= nR(1 - o_\lambda(1)) - o(n).$$

So (75) will be proved if we show that $H(\boldsymbol{S}_U)$ is bounded above by n times the right-hand side (RHS) of (75), which we proceed to do now

$$H(\boldsymbol{S}_U) \leq H(\boldsymbol{Y}^n_{\delta_{in}(U)}) \tag{82}$$

$$= \sum_{k=1}^{n} H(\boldsymbol{Y}_{\delta_{in}(U),k}|\boldsymbol{Y}^{k-1}_{\delta_{in}(U)}) \tag{83}$$

$$= \sum_{k=1}^{n} H(\boldsymbol{Y}_{\sigma(U),k}, \boldsymbol{Y}_{\delta^-(U),k}|\boldsymbol{Y}^{k-1}_{\delta_{in}(U)}) \tag{84}$$

$$= \sum_{k=1}^{n} H(\boldsymbol{Y}_{\sigma(U),k}, \boldsymbol{Y}_{\delta^-(U),k}|\boldsymbol{Y}^{k-1}_{\delta_{in}(U)}, \boldsymbol{X}_{\sigma(U),k}) \tag{85}$$

$$\leq \sum_{k=1}^{n} [H(\boldsymbol{Y}_{\sigma(U),k}|\boldsymbol{X}_{\sigma(U),k}) + H(\boldsymbol{Y}_{\delta^-(U),k})] \tag{86}$$

$$\leq \sum_{k=1}^{n} \left[\sum_{e \in \sigma(U)} H(Y_{e,k}|X_{e,k}) + \sum_{e \in \delta^-(U)} H(Y_{e,k}) \right] \tag{87}$$

$$= n \left[\sum_{e \in \sigma(U)} \left(\frac{1}{n}\sum_{k=1}^{n} H(Y_{e,k}|X_{e,k}) \right) + \sum_{e \in \delta^-(U)} \left(\frac{1}{n}\sum_{k=1}^{n} H(Y_{e,k}) \right) \right]$$

$$\leq n \left[\sum_{e \in \sigma(U)} H(Y_e|X_e) + \sum_{e \in \delta^-(U)} H(Y_e) \right] \tag{88}$$

$$= n \left[\sum_{e \in \sigma(U)} H(Y_e|X_e) + \sum_{e \in \delta^-(U)} H(Y_e|X_e) + \sum_{e \in \delta^-(U)} I(X_e; Y_e) \right]$$

$$= n \left[\sum_{e \in \delta_{in}(U)} H(Y_e|X_e) + \sum_{e \in \delta^-(U)} I(X_e; Y_e) \right]. \tag{89}$$

Here, (82) is because $\boldsymbol{S}_U$ is a function of $\boldsymbol{Y}^n_{\delta_{in}(U)}$; (83) is by the chain rule for entropy; (84) is by (5); (85) is because $\boldsymbol{X}_{\sigma(U),k}$ is a function of $\boldsymbol{Y}^{k-1}_{\delta_{in}(U)}$; (86) and (87) are by the chain rule and the fact that conditioning cannot increase entropy; (88) is by the concavity of the entropy and conditional entropy functions; and (89) is because $\sigma(U) \cup \delta^-(U) = \delta_{in}(U)$.

C. Proof of $R \leq C_m(\boldsymbol{P}; V_1, \cdots, V_{m+1}) + o_\lambda(1)$

Let $V_1, \cdots, V_{m+1}$ be nonempty and pairwise disjoint sets of vertices. Let

$$V_0 \stackrel{\text{def}}{=} V - \bigcup_{i=}^{m+1} V_i \tag{90}$$

and, for $1 \leq j \leq m+1$, let

$$U_j \stackrel{\text{def}}{=} \bigcup_{i=1}^{j} V_i. \tag{91}$$

Then, by (81), for $j = 1, 2, \cdots, m$

$$I(\boldsymbol{S}_{U_j}; \boldsymbol{S}_{V_{j+1}}) \geq (1 - 2|V|\lambda)\log K - 2|V|$$
$$= nR(1 - o_\lambda(1)) - o(n).$$

To prove (76), it will therefore suffice to prove that $(1/m)\sum_{j=1}^{m} I(\boldsymbol{S}_{U_j}; \boldsymbol{S}_{V_{j+1}})$ is bounded above by n times the RHS of (76). To this end, note that

$$\sum_{j=1}^{m} I(\boldsymbol{S}_{U_j}; \boldsymbol{S}_{V_{j+1}}) \leq \sum_{j=1}^{m} I(\boldsymbol{Y}^n_{\delta_{in}(U_j)}; \boldsymbol{Y}^n_{\delta_{in}(V_{j+1})}) \tag{92}$$

$$\leq \sum_{j=1}^{m} I(\boldsymbol{Y}^n_{\delta_{in}(U_j)}, \boldsymbol{Y}^n_{U_j \times V_0}; \boldsymbol{Y}^n_{\delta_{in}(V_{j+1})}, \boldsymbol{Y}^n_{V_{j+1} \times V_0}) \tag{93}$$

$$= \sum_{j=1}^{m} [H(\boldsymbol{Y}^n_{\delta_{in}(U_j)}, \boldsymbol{Y}^n_{U_j \times V_0}) + H(\boldsymbol{Y}^n_{\delta_{in}(V_{j+1})}, \boldsymbol{Y}^n_{V_{j+1} \times V_0}) - H(\boldsymbol{Y}^n_{\delta_{in}(U_{j+1})}, \boldsymbol{Y}^n_{U_{j+1} \times V_0})] \tag{94}$$

$$= H(\boldsymbol{Y}^n_{\delta_{in}(U_1)}, \boldsymbol{Y}^n_{U_1 \times V_0}) + \sum_{j=1}^{m} H(\boldsymbol{Y}^n_{\delta_{in}(V_{j+1})}, \boldsymbol{Y}^n_{V_{j+1} \times V_0}) - H(\boldsymbol{Y}^n_{\delta_{in}(U_{m+1})}, \boldsymbol{Y}^n_{U_{m+1} \times V_0}) \tag{95}$$

$$= \sum_{j=1}^{m+1} H(\boldsymbol{Y}^n_{\delta_{in}(V_j)}, \boldsymbol{Y}^n_{V_j \times V_0}) - H(\boldsymbol{Y}^n_{\delta_{in}(U_{m+1})}, \boldsymbol{Y}^n_{U_{m+1} \times V_0}). \tag{96}$$

Here, (92) is because $\boldsymbol{S}_{U_j}$ is a function of $\boldsymbol{Y}^n_{\delta_{in}(U_j)}$, and $\boldsymbol{S}_{V_{j+1}}$ is a function of $\boldsymbol{Y}^n_{\delta_{in}(V_{j+1})}$; (93) is because

$$I(Z_1; Z_2) \leq I(Z_1, Z_3; Z_2, Z_4);$$

(94) is obtained using the formula

$$I(Z_1; Z_2) = H(Z_1) + H(Z_2) - H(Z_1, Z_2)$$

and the definition (91); (95) is obtained by "telescoping" the sum in (94); and in (96) we have used the fact that $U_1 = V_1$, by

the definition in (91). We will bound each of the terms in (96) separately. First

$$H(\boldsymbol{Y}^n_{\delta_{in}(V_j)}, \boldsymbol{Y}^n_{V_j \times V_0})$$

$$= \sum_{k=1}^{n} H(\boldsymbol{Y}_{\delta_{in}(V_j), k}, \boldsymbol{Y}_{V_j \times V_0, k} | \boldsymbol{Y}^{k-1}_{\delta_{in}(V_j)}, \boldsymbol{Y}^{k-1}_{V_j \times V_0}) \quad (97)$$

$$= \sum_{k=1}^{n} H(\boldsymbol{Y}_{\delta^-(V_j), k}, \boldsymbol{Y}_{\sigma(V_j), k},$$
$$\boldsymbol{Y}_{V_j \times V_0, k} | \boldsymbol{Y}^{k-1}_{\delta_{in}(V_j)}, \boldsymbol{Y}^{k-1}_{V_j \times V_0}) \quad (98)$$

$$\leq \sum_{k=1}^{n} H(\boldsymbol{Y}_{\delta^-(V_j), k}, \boldsymbol{Y}_{\sigma(V_j), k}, \boldsymbol{Y}_{V_j \times V_0, k} | \boldsymbol{Y}^{k-1}_{\delta_{in}(V_j)},$$
$$\boldsymbol{X}_{\sigma(V_j), k}, \boldsymbol{X}_{V_j \times V_0, k}) \quad (99)$$

$$\leq \sum_{k=1}^{n} [H(\boldsymbol{Y}_{\delta^-(V_j), k}) + H(\boldsymbol{Y}_{\sigma(V_j), k} | \boldsymbol{X}_{\sigma(V_j), k})$$
$$+ H(\boldsymbol{Y}_{V_j \times V_0, k} | \boldsymbol{X}_{V_j \times V_0, k})] \quad (100)$$

$$\leq \sum_{k=1}^{n} \left[\sum_{e \in \delta^-(V_j)} H(Y_{e, k}) + \sum_{e \in \sigma(V_j)} H(Y_{e, k} | X_{e, k}) \right.$$
$$\left. + \sum_{e \in (V_j \times V_0) \cap E} H(Y_{e, k} | X_{e, k}) \right]. \quad (101)$$

Here, (97) is by the chain rule; (98) is by (5); (99) is obtained from (98) by first conditioning on $\boldsymbol{X}_{\sigma(V_j), k}$ and $\boldsymbol{X}_{V_j \times V_0, k}$ also (this does not change (98) since these are both functions of $\boldsymbol{Y}^{k-1}_{\delta_{in}(V_j)}$), and then dropping the conditioning on $\boldsymbol{Y}^{k-1}_{V_j \times V_0}$ (conditioning cannot increase entropy); and (100) and (101) are by the chain rule and the fact that conditioning cannot increase entropy.

Next, we will bound $H(\boldsymbol{Y}^n_{\delta_{in}(U_{m+1})}, \boldsymbol{Y}^n_{U_{m+1} \times V_0})$ from below. For convenience, let

$$E' \overset{\text{def}}{=} \delta_{in}(U_{m+1}) \cup (U_{m+1} \times V_0).$$

Then

$$H(\boldsymbol{Y}^n_{\delta_{in}(U_{m+1})}, \boldsymbol{Y}^n_{U_{m+1} \times V_0}) = H(\boldsymbol{Y}^n_{E'}) \quad (102)$$

$$= \sum_{k=1}^{n} H(\boldsymbol{Y}_{E', k} | \boldsymbol{Y}^{k-1}_{E'}) \quad (103)$$

$$\geq \sum_{k=1}^{n} H(\boldsymbol{Y}_{E', k} | \boldsymbol{Y}^{k-1}_{E}) \quad (104)$$

$$= \sum_{k=1}^{n} H(\boldsymbol{Y}_{E', k} | \boldsymbol{Y}^{k-1}_{E}, \boldsymbol{X}_{E', k}) \quad (105)$$

$$= \sum_{k=1}^{n} \sum_{e \in E'} H(Y_{e, k} | X_{e, k}) \quad (106)$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{m+1} \left[\sum_{e \in \delta_{in}(V_j)} H(Y_{e, k} | X_{e, k}) \right.$$
$$\left. + \sum_{e \in (V_j \times V_0) \cap E} H(Y_{e, k} | X_{e, k}) \right] \quad (107)$$

and

$$H(\boldsymbol{Y}^n_{\delta_{in}(U_{m+1})}, \boldsymbol{Y}^n_{U_{m+1} \times V_0})$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{m+1} \left[\sum_{e \in \delta^-(V_j)} H(Y_{e, k} | X_{e, k}) \right.$$
$$+ \sum_{e \in \sigma(V_j)} H(Y_{e, k} | X_{e, k})$$
$$\left. + \sum_{e \in (V_j \times V_0) \cap E} H(Y_{e, k} | X_{e, k}) \right]. \quad (108)$$

Here, (103) is again by the chain rule; (104) is because conditioning cannot increase entropy; (105) is because $\boldsymbol{X}_{E', k}$ is a function of $\boldsymbol{Y}^{k-1}_E$; (106) is by the chain rule and the fact that, given $X_{e, k}$, $Y_{e, k}$ is conditionally independent of $\boldsymbol{Y}^{k-1}_E$, $\boldsymbol{X}_{E-\{e\}, k}$, and $\boldsymbol{Y}_{E-\{e\}, k}$; (107) is by the definition

$$E' = \delta_{in}(U_{m+1}) \cup (U_{m+1} \times V_0)$$

and (91); and (108) is by (5).

From (96), (101), and (108), we have, after cancelling common terms

$$\frac{1}{m} \sum_{j=1}^{m} I(\boldsymbol{S}_{U_j}; \boldsymbol{S}_{V_{j+1}})$$

$$\leq \frac{1}{m} \sum_{k=1}^{n} \sum_{j=1}^{m+1} \sum_{e \in \delta^-(V_j)} [H(Y_{e, k}) - H(Y_{e, k} | X_{e, k})]$$

$$= \frac{1}{m} \sum_{j=1}^{m+1} \sum_{e \in \delta^-(V_j)} \left[\sum_{k=1}^{n} I(X_{e, k}; Y_{e, k}) \right]$$

$$\leq n \left[\frac{1}{m} \sum_{j=1}^{m+1} \sum_{e \in \delta^-(V_j)} I(X_e; Y_e) \right]. \quad (109)$$

The inequality in (109) is by the concavity of the mutual information function. This completes the proof of the weak converse.

V. PROOF OF THE STRONG CONVERSE

A. Preliminaries

Let $(\boldsymbol{f}, \boldsymbol{S})$ be any (n, K, λ) protocol for generating common randomness, with $K = \exp\{nR + o(n)\}$; as before, this means that $(\boldsymbol{f}, \boldsymbol{S})$ is the nth term in a sequence of (n, K_n, λ) protocols, with $\limsup_{n \to \infty}(1/n) \log K_n = R$. Suppose $\lambda < 1$. We will then identify an appropriate $\boldsymbol{P}$ such that $R \leq C^*(\boldsymbol{P})$. This will prove the strong converse. By the same reasoning that led to (75) and (76), $R \leq C^*(\boldsymbol{P})$ will be proved if we show that

$$R \leq \sum_{e \in \delta_{in}(U)} H(\boldsymbol{W}_e | P_e) + \sum_{e \in \delta^-(U)} I(P_e; \boldsymbol{W}_e) \quad (110)$$

for every nonempty subset U of V, and

$$R \leq \frac{1}{m} \sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I(P_e; \boldsymbol{W}_e) \quad (111)$$

for every $m \geq 1$ and every collection $V_1, \cdots, V_{m+1}$ of nonempty and pairwise disjoint subsets of V.

Much of the effort required to prove the strong converse will go into the development of a "typical sequence" machinery for interactive communication over a network of DMC's.

B. Typical Sequences for Interactive Communication

Let $Y_{e,k}$ be the random variable representing the output of channel e in step k when the agents communicate according to the n-step "strategy" $\boldsymbol{f}$. Let

$$\boldsymbol{Y}^k \overset{\text{def}}{=} (Y_e^k : e \in E).$$

Then $\boldsymbol{Y}^n$ completely determines the corresponding vector $(X_e^n : e \in E)$ of channel input sequences: if $e \in \delta_{\text{out}}(v)$ then

$$X_{e,k} = f_{e,k}(Y_{e'}^{k-1} : e' \in \delta_{in}(v)), \qquad 1 \le k \le n.$$

We will abuse notation and write, for each $e \in E$

$$X_{e,k} = f_{e,k}(\boldsymbol{Y}^{k-1}) \quad \text{and} \quad X_e^n = f_e(\boldsymbol{Y}^n).$$

For each nonempty subset U of V, let

$$\boldsymbol{Y}_U^k \overset{\text{def}}{=} (Y_e^k : e \in \delta_{in}(U)).$$

Then observe that $(X_e^n : e \in \delta_{out}(U))$ is actually determined completely by $\boldsymbol{Y}_U^n$ itself. We will, therefore, abuse notation further and write, for each $e \in \delta_{\text{out}}(U)$

$$X_{e,k} = f_{e,k}(\boldsymbol{Y}_U^{k-1}) \quad \text{and} \quad X_e^n = f_e(\boldsymbol{Y}_U^n).$$

We will denote the probability that $\boldsymbol{Y}^n = \boldsymbol{y}^n$ by $\boldsymbol{W}(\boldsymbol{y}^n)$, and, more generally, the probability that $\boldsymbol{Y}^n \in \mathcal{E}$ by $\boldsymbol{W}(\mathcal{E})$. Observe that

$$\boldsymbol{W}(\boldsymbol{y}^n) = \prod_{e \in E} W_e^n(y_e^n | f_e(\boldsymbol{y}^n))$$
$$= \prod_{k=1}^n \prod_{e \in E} W_e(y_{e,k} | f_{e,k}(\boldsymbol{y}^{k-1})).$$

Definition 5.1: If π_e is an n-type on $\mathcal{X}_e \times \boldsymbol{Y}_e$ (i.e., $n\pi_e(a_e, b_e)$ is an integer for all $(a_e, b_e) \in \mathcal{X}_e \times \boldsymbol{Y}_e$), and $\boldsymbol{\pi} = (\pi_e : e \in E)$, then

$$\mathcal{T}_{\boldsymbol{\pi}} \overset{\text{def}}{=} \{\boldsymbol{y}^n : \text{the joint type of } f_e(\boldsymbol{y}^n) \text{ and } y_e^n \text{ is } \pi_e, \text{ for each } e \in E\}. \tag{112}$$

We will simply say that a vector $\boldsymbol{y}^n \in \mathcal{T}_{\boldsymbol{\pi}}$ has the type $\boldsymbol{\pi}$. Obviously, the set of all vectors $\boldsymbol{y}^n = (y_e^n : e \in E)$ is the disjoint union of the "type classes" $\mathcal{T}_{\boldsymbol{\pi}}$. Moreover, the number of these type classes can be upper-bounded by $(n+1)^c$, where $c \overset{\text{def}}{=} \sum_e |\mathcal{X}_e| |\boldsymbol{Y}_e|$. Let P_{π_e} and Q_{π_e} be the input and output marginals of π_e:

$$P_{\pi_e}(a_e) \overset{\text{def}}{=} \sum_{b_e} \pi_e(a_e, b_e) \tag{113}$$

$$Q_{\pi_e}(b_e) \overset{\text{def}}{=} \sum_{a_e} \pi_e(a_e, b_e). \tag{114}$$

If $\boldsymbol{y}^n \in \mathcal{T}_{\boldsymbol{\pi}}$, then y_e^n has type Q_{π_e} and $f_e(\boldsymbol{y}^n)$ has type P_{π_e}, for each $e \in E$.

Definition 5.2: $\boldsymbol{\pi}$ is θ-typical if

$$|\pi_e(a_e, b_e) - P_{\pi_e}(a_e) W_e(b_e | a_e)| \le \theta \sqrt{\frac{W_e(b_e | a_e)}{n}} \tag{115}$$

for all $e \in E$ and $(a_e, b_e) \in \mathcal{X}_e \times \boldsymbol{Y}_e$.

Remark: Later, when proving the strong converse, we will choose the typicality parameter θ to be an appropriately large constant. For now, we will simply assume that $\theta \ge 1$. This is just

so that we can upper-bound $\sqrt{\theta}$ by θ, and thus simplify several expressions in the lemmas of this section.

Suppose $\boldsymbol{\pi}$ is θ-typical, and $\boldsymbol{y}^n \in \mathcal{T}_{\boldsymbol{\pi}}$. Then, by (115), the conditional type of y_e^n with respect to (w.r.t.) $f_e(\boldsymbol{y}^n)$ must be "close" to W_e, for each $e \in E$. Intuitively, we expect the set of all such $\boldsymbol{y}^n$ to carry most of the probability under the distribution $\boldsymbol{W}$. This is confirmed in the following lemma, which is analogous to [9, Lemma 2.12].

Lemma 5.1: Let $\mathcal{E}_\theta$ be the set of vectors $\boldsymbol{y}^n$ whose type $\boldsymbol{\pi}$ is θ-typical

$$\mathcal{E}_\theta \overset{\text{def}}{=} \bigcup_{\theta\text{-typical } \boldsymbol{\pi}} \mathcal{T}_{\boldsymbol{\pi}}. \tag{116}$$

Then

$$\boldsymbol{W}(\mathcal{E}_\theta) \ge 1 - \frac{c}{\theta^2}. \tag{117}$$

Here, $c \overset{\text{def}}{=} \sum_e |\mathcal{X}_e| |\boldsymbol{Y}_e|$.

Proof: Recall that X_e^n and Y_e^n are the random input and output sequences on channel e when the agents communicate according to the strategy $\boldsymbol{f}$. Let $\boldsymbol{\pi}$ be the (random) type of $\boldsymbol{Y}^n = (Y_e^n : e \in E)$. We will prove that, for any given $e \in E$ and $(a_e, b_e) \in \mathcal{X}_e \times \boldsymbol{Y}_e$, the probability that this random $\boldsymbol{\pi}$ violates the condition (115) is at most $1/\theta^2$. A union bound over all e and (a_e, b_e) will then give (117).

For each $1 \le k \le n$, define the random variable Z_k to be 1 if $X_{e,k} = a_e$ and $Y_{e,k} = b_e$, and 0 otherwise

$$Z_k \overset{\text{def}}{=} \mathbf{1}\{X_{e,k} = a_e, Y_{e,k} = b_e\}. \tag{118}$$

Then

$$\frac{1}{n} \sum_{k=1}^n Z_k = \pi_e(a_e, b_e). \tag{119}$$

Also

$$E[Z_k | \boldsymbol{Y}^{k-1}] = \mathbf{1}\{X_{e,k} = a_e\} W_e(b_e | a_e) \tag{120}$$

so that

$$\frac{1}{n} \sum_{k=1}^n E[Z_k | \boldsymbol{Y}^{k-1}] = P_{\pi_e}(a_e) W_e(b_e | a_e). \tag{121}$$

Let

$$\tilde{Z}_k \overset{\text{def}}{=} Z_k - E[Z_k | \boldsymbol{Y}^{k-1}], \qquad 1 \le k \le n.$$

Then, by (119) and (121), the probability that the condition (115) is violated equals

$$\Pr\left\{\left|\sum_k \tilde{Z}_k\right| > \theta \sqrt{n W_e(b_e | a_e)}\right\}.$$

But

$$\Pr\left\{\left|\sum_{k=1}^n \tilde{Z}_k\right| \theta \sqrt{n W_e(b_e | a_e)}\right\} \le \frac{\text{Var}\left(\sum_k \tilde{Z}_k\right)}{n\theta^2 W_e(b_e | a_e)} \tag{122}$$

$$= \frac{\sum_k \text{Var}(\tilde{Z}_k)}{n\theta^2 W_e(b_e | a_e)}. \tag{123}$$

Here, (122) is by Chebyshev's inequality, and (123) is by the easily verified fact that the $\tilde{Z}_k$'s are pairwise uncorrelated.

To finish the proof, we only have to show that

$$E[\tilde{Z}_k^2] \leq W_e(b_e|a_e) \tag{124}$$

for each k. Now

$$\begin{aligned}
\tilde{Z}_k &= Z_k - E[Z_k|\boldsymbol{Y}^{k-1}] \\
&= \mathbf{1}\{X_{e,k} = a_e\}[\mathbf{1}\{Y_{e,k} = b_e\} - W_e(b_e|a_e)] \tag{125}
\end{aligned}$$

and so

$$\begin{aligned}
\tilde{Z}_k^2 = \mathbf{1}\{X_{e,k} = a_e\}[\mathbf{1}\{Y_{e,k} = b_e\} &- 2W_e(b_e|a_e) \\
\cdot \mathbf{1}\{Y_{e,k} = b_e\} &+ W_e^2(b_e|a_e)]. \tag{126}
\end{aligned}$$

Hence

$$\begin{aligned}
E[\tilde{Z}_k^2|\boldsymbol{Y}^{k-1}] &= \mathbf{1}\{X_{e,k} = a_e\}[W_e(b_e|a_e) - W_e^2(b_e|a_e)] \\
&\leq W_e(b_e|a_e) \tag{127}
\end{aligned}$$

which obviously implies (124). $\square$

It is worth noting that the set of θ-typical types $\boldsymbol{\pi}$ does not depend on the strategy $\boldsymbol{f}$ being used, but the set $\mathcal{E}_\theta$ of vectors $\boldsymbol{y}^n$ whose type is θ-typical does.

In the next few lemmas, we will derive some consequences of θ-typicality that will be needed in the proof of the strong converse. But first we will state a simple and useful result on the continuity of the entropy function. This is a weaker but more convenient form of the standard bound in [9, p. 33, Lemma 2.7].

Lemma 5.2 (Continuity of the entropy function) : If P_1 and P_2 are probability distributions on a finite set $\mathcal{Z}$, and $|P_1(z) - P_2(z)| \leq \beta$ for all $z \in \mathcal{Z}$, then $|H(P_1) - H(P_2)| \leq 2|\mathcal{Z}|\sqrt{\beta}$.
    Proof: See [24]. $\square$

Lemma 5.3: If $\boldsymbol{\pi}$ is θ-typical, then

$$|H(Q_{\pi_e}) - H(P_{\pi_e}\boldsymbol{W}_e)| \leq (2|\mathcal{X}_e\|\boldsymbol{Y}_e|) \frac{\theta}{n^{1/4}} \tag{128}$$

and

$$\left|\sum_{(a_e,b_e)} \pi_e(a_e,b_e) \log W_e(b_e|a_e) + H(\boldsymbol{W}_e|P_{\pi_e})\right|$$
$$\leq (2|\mathcal{X}_e\|\boldsymbol{Y}_e|) \frac{\theta}{n^{1/4}} \tag{129}$$

for all $e \in E$.
    Proof: By summing both sides of (115) over all $a_e \in \mathcal{X}_e$, we have

$$|Q_{\pi_e}(b_e) - P_{\pi_e}\boldsymbol{W}_e(b_e)| \leq \frac{\theta|\mathcal{X}_e|}{\sqrt{n}}$$

for any $b_e \in \boldsymbol{Y}_e$. An application of Lemma 5.2 now gives

$$\begin{aligned}
|H(Q_{\pi_e}) - H(P_{\pi_e}\boldsymbol{W}_e)| &\leq 2|\boldsymbol{Y}_e|\sqrt{\frac{\theta|\mathcal{X}_e|}{\sqrt{n}}} \\
&\leq (2|\mathcal{X}_e\|\boldsymbol{Y}_e|) \frac{\theta}{n^{1/4}}
\end{aligned}$$

because $|\mathcal{X}_e| \geq 1$ and $\theta \geq 1$. This proves (128). Next, to prove (129), observe that

$$\left|\sum_{(a_e,b_e)} \pi_e(a_e,b_e) \log W_e(b_e|a_e) + H(\boldsymbol{W}_e|P_{\pi_e})\right|$$

$$= \left|\sum_{(a_e,b_e)} [\pi_e(a_e,b_e) - P_{\pi_e}(a_e)W_e(b_e|a_e)] \log W_e(b_e|a_e)\right|$$

$$\leq \sum_{(a_e,b_e)} |\pi_e(a_e,b_e) - P_{\pi_e}(a_e)W_e(b_e|a_e)| \, |\log W_e(b_e|a_e)|$$

$$\leq \sum_{(a_e,b_e)} \theta\sqrt{\frac{W_e(b_e|a_e)}{n}} \, |\log W_e(b_e|a_e)|$$

$$\leq (2|\mathcal{X}_e\|\boldsymbol{Y}_e|) \frac{\theta}{n^{1/4}}.$$

In the last step, we have used $|\sqrt{x} \log x| \leq (2 \log e)/e \leq 2$, for $0 \leq x \leq 1$, and $n^{1/4} \leq \sqrt{n}$. $\square$

Given any type $\boldsymbol{\pi}$ and any nonempty $U \subseteq V$, define the probability distribution $\boldsymbol{W}_{\boldsymbol{\pi},U}$ on the set of all vectors

$$\boldsymbol{y}_U^n = (y_e^n : e \in \delta_{in}(U))$$

in the following way:

$$\boldsymbol{W}_{\boldsymbol{\pi},U}(\boldsymbol{y}_U^n) \stackrel{\text{def}}{=} \prod_{e \in \delta^-(U)} Q_{\pi_e}^n(y_e^n) \prod_{e \in \sigma(U)} W_e^n(y_e^n|f_e(\boldsymbol{y}_U^n)). \tag{130}$$

That $\boldsymbol{W}_{\boldsymbol{\pi},U}$ is indeed a probability distribution is clear when we expand the RHS of (130) to get

$$\begin{aligned}
\boldsymbol{W}_{\boldsymbol{\pi},U}(\boldsymbol{y}_U^n) = \prod_{k=1}^{n} &\left[\prod_{e \in \delta^-(U)} Q_{\pi_e}(y_{e,k})\right] \\
&\cdot \left[\prod_{e \in \sigma(U)} W_e(y_{e,k}|f_{e,k}(\boldsymbol{y}_U^{k-1}))\right].
\end{aligned}$$

Observe also that when $U = V$, $\boldsymbol{W}_{\boldsymbol{\pi},U}$ reduces to $\boldsymbol{W}$ for all $\boldsymbol{\pi}$.

Lemma 5.4: If $\boldsymbol{\pi}$ is θ-typical, and $\boldsymbol{y}^n \in \mathcal{T}_{\boldsymbol{\pi}}$, then

$$\left|\log \boldsymbol{W}_{\boldsymbol{\pi},U}(\boldsymbol{y}_U^n) + n\left[\sum_{e \in \delta_{in}(U)} H(\boldsymbol{W}_e|P_{\pi_e})\right.\right.$$
$$\left.\left. + \sum_{e \in \delta^-(U)} I(P_{\pi_e}; \boldsymbol{W}_e)\right]\right| \leq 2c\theta n^{3/4} \tag{131}$$

for every nonempty $U \subseteq V$. In particular

$$\left|\log \boldsymbol{W}(\boldsymbol{y}^n) + n\left[\sum_{e \in E} H(\boldsymbol{W}_e|P_{\pi_e})\right]\right| \leq 2c\theta n^{3/4}. \tag{132}$$

Here

$$c \stackrel{\text{def}}{=} \sum_e |\mathcal{X}_e\|\boldsymbol{Y}_e|.$$

Proof: By the definition in (130), and the fact that $\boldsymbol{y}^n \in \mathcal{T}_{\boldsymbol{\pi}}$, we have

$$
\begin{aligned}
\log &\boldsymbol{W}_{\boldsymbol{\pi}, U}(\boldsymbol{y}_U^n) \\
&= \sum_{e \in \delta^-(U)} \log Q_{\pi_e}^n(y_e^n) + \sum_{e \in \sigma(U)} \log W_e^n(y_e^n | f_e(\boldsymbol{y}_U^n)) \\
&= \sum_{e \in \delta^-(U)} [-n H(Q_{\pi_e})] \\
&\quad + \sum_{e \in \sigma(U)} \left[\sum_{(a_e, b_e)} n \pi_e(a_e, b_e) \log W_e(b_e | a_e) \right]. \quad (133)
\end{aligned}
$$

If $\boldsymbol{\pi}$ is θ-typical, then we can bound each of the terms in (133) using (128) and (129), to get

$$
\begin{aligned}
&\left| \log \boldsymbol{W}_{\boldsymbol{\pi}, U}(\boldsymbol{y}_U^n) + n \left[\sum_{e \in \delta^-(U)} H(P_{\pi_e} \boldsymbol{W}_e) \right. \right. \\
&\qquad \left. \left. + \sum_{e \in \sigma(U)} H(\boldsymbol{W}_e | P_{\pi_e}) \right] \right| \\
&\leq \sum_{e \in \delta^-(U)} 2 |\mathcal{X}_e| |\boldsymbol{Y}_e| \theta n^{3/4} + \sum_{e \in \sigma(U)} 2 |\mathcal{X}_e| |\boldsymbol{Y}_e| \theta n^{3/4} \\
&\leq 2 c \theta n^{3/4}.
\end{aligned}
$$

From this, (131) follows because

$$
\begin{aligned}
\sum_{e \in \delta^-(U)} H(P_{\pi_e} \boldsymbol{W}_e) &+ \sum_{e \in \sigma(U)} H(\boldsymbol{W}_e | P_{\pi_e}) \\
&= \sum_{e \in \delta^-(U)} I(P_{\pi_e}; \boldsymbol{W}_e) + \sum_{e \in \delta_{in}(U)} H(\boldsymbol{W}_e | P_{\pi_e}).
\end{aligned}
$$

By taking $U = V$ in (131), we get (132). $\qquad \square$

For each nonempty $U \subseteq V$, let $\mathcal{T}_{\boldsymbol{\pi}, U}$ be the projection of the type class $\mathcal{T}_{\boldsymbol{\pi}}$ on the coordinates in $\delta_{in}(U)$; i.e., let

$$
\begin{aligned}
\mathcal{T}_{\boldsymbol{\pi}, U} \stackrel{\text{def}}{=} \{ \boldsymbol{y}_U^n : &\text{ there exists a } \tilde{\boldsymbol{y}}^n \in \mathcal{T}_{\boldsymbol{\pi}} \text{ s.t.} \\
&\tilde{y}_e^n = y_e^n \text{ for all } e \in \delta_{in}(U) \}. \quad (134)
\end{aligned}
$$

Lemma 5.5: If $\boldsymbol{\pi}$ is θ-typical, then

$$
\begin{aligned}
|\mathcal{T}_{\boldsymbol{\pi}, U}| \leq \exp \left\{ n \left[\sum_{e \in \delta_{in}(U)} H(\boldsymbol{W}_e | P_{\pi_e}) \right. \right. \\
\left. \left. + \sum_{e \in \delta^-(U)} I(P_{\pi_e}; \boldsymbol{W}_e) \right] + 2 c \theta n^{3/4} \right\} \quad (135)
\end{aligned}
$$

for every nonempty $U \subseteq V$.

Proof: If $\boldsymbol{\pi}$ is θ-typical, and $\boldsymbol{y}_U^n \in \mathcal{T}_{\boldsymbol{\pi}, U}$, then

$$
\begin{aligned}
\boldsymbol{W}_{\boldsymbol{\pi}, U}(\boldsymbol{y}_U^n) \geq \exp \left\{ -n \left[\sum_{e \in \delta_{in}(U)} H(\boldsymbol{W}_e | P_{\pi_e}) \right. \right. \\
\left. \left. + \sum_{e \in \delta^-(U)} I(P_{\pi_e}; \boldsymbol{W}_e) \right] - 2 c \theta n^{3/4} \right\}
\end{aligned}
$$

by Lemma 5.4. This, together with $\boldsymbol{W}_{\boldsymbol{\pi}, U}(\mathcal{T}_{\boldsymbol{\pi}, U}) \leq 1$, gives (135). $\qquad \square$

Lemma 5.6: Suppose that $m \geq 1$, $V_1, \cdots, V_{m+1}$ are nonempty and pairwise-disjoint subsets of V, and

$$
V_0 \stackrel{\text{def}}{=} V - \bigcup_{i=1}^{m+1} V_i.
$$

If $\boldsymbol{\pi}$ is θ-typical, and $\boldsymbol{y}^n \in \mathcal{T}_{\boldsymbol{\pi}}$, then

$$
\begin{aligned}
\prod_{i=1}^{m+1} \boldsymbol{W}_{\boldsymbol{\pi}, V_i}&(\boldsymbol{y}_{V_i}^n) \prod_{e \in \delta_{in}(V_0)} W_e^n(y_e^n | f_e(\boldsymbol{y}^n)) \\
&\geq \boldsymbol{W}(\boldsymbol{y}^n) \exp \left\{ -n \left[\sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I(P_{\pi_e}; \boldsymbol{W}_e) \right] - 8 m c \theta n^{3/4} \right\}.
\end{aligned}
$$

Proof: By Lemma 5.4, applied to each V_i, $1 \leq i \leq m+1$

$$
\begin{aligned}
\prod_{i=1}^{m+1} &\boldsymbol{W}_{\boldsymbol{\pi}, V_i}(\boldsymbol{y}_{V_i}^n) \\
&\geq \prod_{i=1}^{m+1} \exp \left\{ -n \left[\sum_{e \in \delta_{in}(V_i)} H(\boldsymbol{W}_e | P_{\pi_e}) \right. \right. \\
&\qquad\qquad\qquad \left. \left. + \sum_{e \in \delta^-(V_i)} I(P_{\pi_e}; \boldsymbol{W}_e) \right] - 2 c \theta n^{3/4} \right\} \\
&= \exp \left\{ -n \sum_{i=1}^{m+1} \left[\sum_{e \in \delta_{in}(V_i)} H(\boldsymbol{W}_e | P_{\pi_e}) \right. \right. \\
&\qquad\qquad \left. \left. + \sum_{e \in \delta^-(V_i)} I(P_{\pi_e}; \boldsymbol{W}_e) \right] - 2(m+1) c \theta n^{3/4} \right\}.
\end{aligned}
$$

And, by (129) applied to each $e \in \delta_{in}(V_0)$,

$$
\begin{aligned}
\prod_{e \in \delta_{in}(V_0)} &W_e^n(y_e^n | f_e(\boldsymbol{y}^n)) \\
&= \prod_{e \in \delta_{in}(V_0)} \prod_{(a_e, b_e)} W_e(b_e | a_e)^{n \pi_e(a_e, b_e)} \\
&\geq \prod_{e \in \delta_{in}(V_0)} \exp \{ -n H(\boldsymbol{W}_e | P_{\pi_e}) - 2 |\mathcal{X}_e| |\boldsymbol{Y}_e| \theta n^{3/4} \} \\
&\geq \exp \left\{ -n \sum_{e \in \delta_{in}(V_0)} H(\boldsymbol{W}_e | P_{\pi_e}) - 2 c \theta n^{3/4} \right\}.
\end{aligned}
$$

Finally, by (132),

$$
\boldsymbol{W}(\boldsymbol{y}^n) \leq \exp \left\{ -n \sum_{e \in E} H(\boldsymbol{W}_e | P_{\pi_e}) + 2 c \theta n^{3/4} \right\}. \quad (136)
$$

Therefore,

$$
\frac{1}{n} \left[\log \boldsymbol{W}(\boldsymbol{y}^n) - \log \left(\prod_{i=1}^{m+1} \boldsymbol{W}_{\boldsymbol{\pi}, V_i}(\boldsymbol{y}_{V_i}^n) \right) \right. \\
\left. - \log \left(\prod_{e \in \delta_{in}(V_0)} W_e^n(y_e^n | f_e(\boldsymbol{y}^n)) \right) \right]
$$

$$\leq \sum_{i=1}^{m+1} \left[\sum_{e \in \delta_{in}(V_i)} H(\boldsymbol{W}_e | P_{\pi_e}) + \sum_{e \in \delta^-(V_i)} I(P_{\pi_e}; \boldsymbol{W}_e) \right]$$

$$+ \sum_{e \in \delta_{in}(V_0)} H(\boldsymbol{W}_e | P_{\pi_e}) - \sum_{e \in E} H(\boldsymbol{W}_e | P_{\pi_e})$$

$$+ 2(m+3)c\theta n^{-1/4}$$

$$= \sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I(P_{\pi_e}; \boldsymbol{W}_e) + 2(m+3)c\theta n^{-1/4}$$

$$\leq \sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I(P_{\pi_e}; \boldsymbol{W}_e) + 8mc\theta n^{-1/4}. \qquad (137)$$

In (137), we have used $m \geq 1$. $\qquad \square$

Lemma 5.7: Suppose that $m \geq 1$, $V_1, \cdots, V_{m+1}$ are nonempty and pairwise-disjoint subsets of $V_{,,}$ and

$$V_0 \stackrel{\text{def}}{=} V - \bigcup_{i=1}^{m+1} V_i.$$

For each $0 \leq i \leq m+1$, let $\mathcal{D}_i$ be an arbitrary set of vectors $\boldsymbol{y}_{V_i}^n = (y_e^n : e \in \delta_{in}(V_i))$. Let $\mathcal{D}$ be the Cartesian product of $\mathcal{D}_0$, $\mathcal{D}_1, \cdots, \mathcal{D}_{m+1}$:

$$\mathcal{D} \stackrel{\text{def}}{=} \{\boldsymbol{y}^n : \boldsymbol{y}_{V_i}^n \in \mathcal{D}_i, i = 0, 1, \cdots, m+1\}.$$

Then, for any θ-typical $\boldsymbol{\pi}$

$$\prod_{i=1}^{m+1} \boldsymbol{W}_{\boldsymbol{\pi}, V_i}(\mathcal{D}_i)$$

$$\geq \boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}} \cap \mathcal{D}) \exp \left\{ -n \left[\sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I(P_{\pi_e}; \boldsymbol{W}_e) \right] \right.$$

$$\left. - 8mc\theta n^{3/4} \right\}. \qquad (138)$$

Proof: For convenience, let

$$\alpha_n \stackrel{\text{def}}{=} n \left[\sum_{i=1}^{m+1} \sum_{e \in \delta^-(V_i)} I(P_{\pi_e}; \boldsymbol{W}_e) \right] + 8mc\theta n^{3/4}. \quad (139)$$

We have

$$2^{-\alpha_n} \boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}} \cap \mathcal{D}) = \sum_{\boldsymbol{y}^n \in \mathcal{T}_{\boldsymbol{\pi}} \cap \mathcal{D}} 2^{-\alpha_n} \boldsymbol{W}(\boldsymbol{y}^n)$$

$$\leq \sum_{\boldsymbol{y}^n \in \mathcal{D}} \left[\prod_{i=1}^{m+1} \boldsymbol{W}_{\boldsymbol{\pi}, V_i}(\boldsymbol{y}_{V_i}^n) \right]$$

$$\cdot \left[\prod_{e \in \delta_{in}(V_0)} W_e^n(y_e^n | f_e(\boldsymbol{y}^n)) \right]$$

$$= \sum_{\substack{(\boldsymbol{y}_{V_1}^n, \cdots, \boldsymbol{y}_{V_{m+1}}^n) \\ \boldsymbol{y}_{V_i}^n \in \mathcal{D}_i}} \left[\prod_{i=1}^{m+1} \boldsymbol{W}_{\boldsymbol{\pi}, V_i}(\boldsymbol{y}_{V_i}^n) \right]$$

$$\cdot \sum_{\boldsymbol{y}_{V_0}^n \in \mathcal{D}_0} \left[\prod_{e \in \delta_{in}(V_0)} W_e^n(y_e^n | f_e(\boldsymbol{y}^n)) \right]$$

$$\qquad (140)$$

and

$$2^{-\alpha_n} \boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}} \cap \mathcal{D}) \leq \sum_{\substack{(\boldsymbol{y}_{V_1}^n, \cdots, \boldsymbol{y}_{V_{m+1}}^n) \\ \boldsymbol{y}_{V_i}^n \in \mathcal{D}_i}} \left[\prod_{i=1}^{m+1} \boldsymbol{W}_{\boldsymbol{\pi}, V_i}(\boldsymbol{y}_{V_i}^n) \right]$$

$$= \prod_{i=1}^{m+1} \left[\sum_{\boldsymbol{y}_{V_i}^n \in \mathcal{D}_i} \boldsymbol{W}_{\boldsymbol{\pi}, V_i}(\boldsymbol{y}_{V_i}^n) \right]$$

$$= \prod_{i=1}^{m+1} \boldsymbol{W}_{\boldsymbol{\pi}, V_i}(\mathcal{D}_i). \qquad (141)$$

Here, (140) is by Lemma 5.6. $\qquad \square$

C. Rest of the Proof

For each $l \in [K]$, let

$$\mathcal{D}_l \stackrel{\text{def}}{=} \{\boldsymbol{y}^n : S_v(\boldsymbol{y}^n) = l \text{ for each } v \in V\}. \qquad (142)$$

In words, $\mathcal{D}_l$ is the set of vectors of output sequences $\boldsymbol{y}^n$ that result in all the agents agreeing on the random output l. Since $(\boldsymbol{f}, \boldsymbol{S})$ is an (n, K, λ) protocol, we must have

$$\frac{1-\lambda}{K} \leq \boldsymbol{W}(\mathcal{D}_l) \leq \frac{1+\lambda}{K}, \qquad \text{for each } l \in [K], \quad (143)$$

and hence

$$\boldsymbol{W}\left(\bigcup_{l \in [K]} \mathcal{D}_l\right) \geq 1 - \lambda. \qquad (144)$$

For the rest of the proof, fix $\theta \geq 1$ at a large enough value that

$$\boldsymbol{W}(\mathcal{E}_\theta) \geq \frac{1+\lambda}{2}. \qquad (145)$$

This is possible by (117) and the hypothesis $\lambda < 1$, e.g., with $\theta = \sqrt{2c/(1-\lambda)}$.

Lemma 5.8: There exists a θ-typical type $\boldsymbol{\pi}^*$ satisfying

$$\boldsymbol{W}\left(\mathcal{T}_{\boldsymbol{\pi}^*} \cap \bigcup_{l \in [K]} \mathcal{D}_l\right) \geq \frac{1-\lambda}{2(n+1)^c}. \qquad (146)$$

Proof: By (144) and (145)

$$\boldsymbol{W}\left(\mathcal{E}_\theta \cap \bigcup_{l \in [K]} \mathcal{D}_l\right) \geq \left(\frac{1+\lambda}{2}\right) - \lambda$$

$$= \left(\frac{1-\lambda}{2}\right). \qquad (147)$$

Since $\mathcal{E}_\theta$ is the disjoint union of the type classes $\mathcal{T}_{\boldsymbol{\pi}}$ corresponding to θ-typical $\boldsymbol{\pi}$, and the number of such types $\boldsymbol{\pi}$ is at most $(n+1)^c$, there must exist a θ-typical $\boldsymbol{\pi}^*$ satisfying (146). $\qquad \square$

From now on, we will focus attention only on those $\boldsymbol{y}^n$ that have the type $\boldsymbol{\pi}^*$ given by Lemma 5.8. In fact, we will prove (110) and (111), and hence $R \leq C^*(\boldsymbol{P})$, with $\boldsymbol{P} = (P_{\pi_e^*} : e \in E)$ (the vector of input marginals associated with $\boldsymbol{\pi}^*$).

The first step is to prove that many of the "decision regions" $\mathcal{D}_l$ must intersect significantly with $\mathcal{T}_{\boldsymbol{\pi}^*}$.

Lemma 5.9: Let $\mathcal{L}$ be the set of $l \in [K]$ for which

$$\boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}^*} \cap \mathcal{D}_l) \geq \frac{1-\lambda}{4K(n+1)^c}.$$

Then

$$|\mathcal{L}| \geq \frac{K(1-\lambda)}{8(n+1)^c}. \tag{148}$$

Proof: We have

$$\frac{(1-\lambda)}{2(n+1)^c} \leq \boldsymbol{W}\left(\mathcal{T}_{\boldsymbol{\pi}^*} \cap \bigcup_{l \in [K]} \mathcal{D}_l\right) \tag{149}$$

$$= \sum_{l \in \mathcal{L}} \boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}^*} \cap \mathcal{D}_l) + \sum_{l \notin \mathcal{L}} \boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}^*} \cap \mathcal{D}_l)$$

$$\leq \sum_{l \in \mathcal{L}} \boldsymbol{W}(\mathcal{D}_l) + (K - |\mathcal{L}|)\left(\frac{1-\lambda}{4K(n+1)^c}\right) \tag{150}$$

$$\leq |\mathcal{L}|\left(\frac{1+\lambda}{K}\right) + \frac{1-\lambda}{4(n+1)^c} \tag{151}$$

$$\leq |\mathcal{L}|\left(\frac{2}{K}\right) + \frac{1-\lambda}{4(n+1)^c}. \tag{152}$$

Here, (149) is by (146); (150) is by the definition of the set $\mathcal{L}$; and (151) is by (143). From (152), (148) follows immediately.$\qquad\square$

Theorem 5.1:

$$K \leq \frac{8(n+1)^c}{1-\lambda} \exp\left\{n\left[\sum_{e \in \delta_{in}(U)} H(\boldsymbol{W}_e | P_{\pi_e^*})\right.\right.$$

$$\left.\left. + \sum_{e \in \delta^-(U)} I(P_{\pi_e^*}; \boldsymbol{W}_e)\right] + 2c\theta n^{3/4}\right\} \tag{153}$$

for every nonempty $U \subseteq V$.

Proof: If $l \in \mathcal{L}$, then $\boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}^*} \cap \mathcal{D}_l) > 0$, and so $\mathcal{T}_{\boldsymbol{\pi}^*} \cap \mathcal{D}_l$ is nonempty. This means that $\mathcal{T}_{\boldsymbol{\pi}^*, U} \cap \mathcal{D}_l$ is nonempty for every $l \in \mathcal{L}$ and every nonempty $U \subseteq V$. But the sets $\mathcal{T}_{\boldsymbol{\pi}^*, U} \cap \mathcal{D}_l$, $l \in [K]$, are pairwise-disjoint. Therefore, we must have $|\mathcal{L}| \leq |\mathcal{T}_{\boldsymbol{\pi}^*, U}|$ for every nonempty $U \subseteq V$. The conclusion (153) now follows from the bounds on $|\mathcal{L}|$ and $|\mathcal{T}_{\boldsymbol{\pi}^*, U}|$ proved in Lemmas 5.9 and 5.5, respectively.$\qquad\square$

In the proof of the next theorem, we will need the following useful inequality. Observe that it reduces to the usual Cauchy–Schwarz inequality when $p = l = 2$.

Lemma 5.10 (Generalized Cauchy–Schwarz Inequality): For any integer $p \geq 1$,

$$\sum_i \left(\prod_j a_{ij}\right) \leq \prod_j \left(\sum_i a_{ij}^p\right)^{1/p} \tag{154}$$

if the a_{ij}, $1 \leq i \leq k, 1 \leq j \leq l$, are nonnegative numbers.

Proof: The proof is by induction on p. It is easy to verify (154) when $p = 1$. Assume, therefore, that $p \geq 2$ and that (154) is true with p replaced by $p - 1$. Then

$$\sum_i \left(\prod_j a_{ij}\right) = \sum_i a_{i1}\left(\prod_{j>1} a_{ij}\right)$$

$$\leq \left[\sum_i a_{i1}^p\right]^{1/p}\left[\sum_i \left(\prod_{j>1} a_{ij}\right)^{p/p-1}\right]^{p-1/p} \tag{155}$$

$$= \left[\sum_i a_{i1}^p\right]^{1/p}\left[\sum_i \left(\prod_{j>1} a_{ij}^{p/p-1}\right)\right]^{p-1/p}$$

$$\leq \left[\sum_i a_{i1}^p\right]^{1/p}\left[\prod_{j>1} \left(\sum_i a_{ij}^p\right)^{1/p-1}\right]^{p-1/p}$$

$$= \prod_j \left(\sum_i a_{ij}^p\right)^{1/p}. \tag{156}$$

Here, (155) is by Hölder's inequality, and (156) is by the induction hypothesis.$\qquad\square$

Theorem 5.2:

$$K \leq \left(\frac{8(n+1)^c}{1-\lambda}\right)^3$$

$$\cdot \exp\left\{n\left[\frac{1}{m}\sum_{i=1}^{m+1}\sum_{e \in \delta^-(V_i)} I(P_{\pi_e^*}; \boldsymbol{W}_e)\right] + 8c\theta n^{3/4}\right\} \tag{157}$$

for every $m \geq 1$, and every collection $V_1, \cdots, V_{m+1}$ of nonempty and pairwise-disjoint subsets of V.

Proof: Fix a collection $V_1, \cdots, V_{m+1}$ as above, and let

$$V_0 = V - \bigcup_{i=1}^{m+1} V_i.$$

For $0 \leq i \leq m + 1$, let $\mathcal{D}_{l, V_i}$ be the projection of the decision region $\mathcal{D}_l$ on the coordinates in $\delta_{in}(V_i)$; i.e., let

$$\mathcal{D}_{l, V_i} \stackrel{\text{def}}{=} \{\boldsymbol{y}_{V_i}^n : \text{there exists a } \tilde{\boldsymbol{y}}^n \in \mathcal{D}_l \text{ s.t. } \tilde{y}_e^n = y_e^n$$
$$\text{for all } e \in \delta_{in}(V_i)\}. \tag{158}$$

$\mathcal{D}_{l, V_i}$ is the set of vectors $\boldsymbol{y}_{V_i}^n$ that result in all the agents in V_i deciding on the random output l. Obviously, the sets $\mathcal{D}_{l, V_i}$, $l \in [K]$, are pairwise-disjoint for each i. It is also important to observe that $\mathcal{D}_l$ is the Cartesian product of the sets $\mathcal{D}_{l, V_i}$, $0 \leq i \leq m + 1$, i.e.,

$$\mathcal{D}_l = \{\boldsymbol{y}^n : \boldsymbol{y}_{V_i}^n \in \mathcal{D}_{l, V_i}, i = 0, 1, \cdots, m+1\}, \qquad l \in [K].$$

This reflects the fact that E is the disjoint union of the sets $\delta_{in}(V_i)$, $0 \leq i \leq m+1$. The desired result can be obtained now from the following chain of inequalities:

$$1 \geq \prod_{i=1}^{m+1} \left[\sum_{l=1}^{K} \boldsymbol{W}_{\boldsymbol{\pi}^*, V_i}(\mathcal{D}_{l, V_i}) \right] \tag{159}$$

$$\geq \left(\sum_{l=1}^{K} \left[\prod_{i=1}^{m+1} \boldsymbol{W}_{\boldsymbol{\pi}^*, V_i}(\mathcal{D}_{l, V_i}) \right]^{1/m+1} \right)^{m+1} \tag{160}$$

$$\geq \left(\sum_{l=1}^{K} [\boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}^*} \cap \mathcal{D}_l) 2^{-\alpha_n}]^{1/m+1} \right)^{m+1} \tag{161}$$

$$\geq 2^{-\alpha_n} \left(\sum_{l \in \mathcal{L}} [\boldsymbol{W}(\mathcal{T}_{\boldsymbol{\pi}^*} \cap \mathcal{D}_l)]^{1/m+1} \right)^{m+1}$$

$$\geq 2^{-\alpha_n} \left[\sum_{l \in \mathcal{L}} \left(\frac{1-\lambda}{4K(n+1)^c} \right)^{1/m+1} \right]^{m+1} \tag{162}$$

$$\geq 2^{-\alpha_n} \left(\frac{1-\lambda}{4K(n+1)^c} \right) \left(\frac{K(1-\lambda)}{8(n+1)^c} \right)^{m+1} \tag{163}$$

$$\geq 2^{-\alpha_n} \left(\frac{1-\lambda}{8(n+1)^c} \right)^{3m} K^m. \tag{164}$$

Here, (159) is by the disjointness of the sets $\mathcal{D}_{l, V_i}$, $l \in [K]$, for each i; (160) is by Lemma 5.10; (161) is by Lemma 5.7, with α_n as defined in (139); (162) is by the definition of the set $\mathcal{L}$; (163) is by (148); and (164) is because $m \geq 1$. $\qquad\square$

From (153) and (157), we have (110) and (111), respectively. This completes the proof of the strong converse.

VI. PROOF OF THE POLYHEDRAL CHARACTERIZATION

A. Preliminaries

This section is devoted to the proof of Theorem 2.5, the polyhedral characterization of the spanning arborescences in $\tilde{G}$ with root-degree equal to one. Observe that $\mathcal{A} \supseteq A_1$ and $\mathcal{A} \supseteq \mathcal{A}_1$, so that $\mathcal{A}_1 = \text{conv}(A_1) + \boldsymbol{R}_+^{\tilde{E}}$ does not follow immediately by a "sandwich" argument from Fulkerson's result (45).

The polyhedron $\mathcal{A}_1$ can be decomposed as the vector sum of the convex hull of its extreme points, and the cone generated by its extreme directions. (Every polyhedron of nonnegative vectors can be so decomposed; see, e.g., [21].) Now, the cone generated by the extreme directions of $\mathcal{A}_1$ equals all of $\boldsymbol{R}_+^{\tilde{E}}$ because $\mathcal{A}_1$ is unbounded along every coordinate direction: if $\boldsymbol{\xi} \in \mathcal{A}_1$ and $\boldsymbol{\xi}' \geq \boldsymbol{\xi}$ then $\boldsymbol{\xi}' \in \mathcal{A}_1$. Therefore, to establish Theorem 2.5, it suffices to prove that A_1 is precisely the set of extreme points of $\mathcal{A}_1$.

We will do this in two steps. First, in Section VI-B, we will prove that $\mathcal{A}_1 \supset A_1$, and every vector in A_1 is an extreme point of $\mathcal{A}_1$. This will imply $\mathcal{A}_1 \supseteq \text{conv}(A_1) + \boldsymbol{R}_+^{\tilde{E}}$. Then, in Section VI-C, we will use Theorem 2.4 and the classical max-flow min-cut theorem [11], to prove that $\mathcal{A}_1$ has no extreme points other than the vectors in A_1. This will imply $\mathcal{A}_1 \subseteq \text{conv}(A_1) + \boldsymbol{R}_+^{\tilde{E}}$, and prove Theorem 2.5.

B. Proof of $\mathcal{A}_1 \supseteq \text{conv}(A_1) + \boldsymbol{R}_+^{\tilde{E}}$

Parts a) and b) of the following lemma show that $\mathcal{A}_1 \supseteq A_1$, i.e., $\boldsymbol{z} \cdot \boldsymbol{\xi} \geq 1$ for all $\boldsymbol{z} \in D_1$ and $\boldsymbol{\xi} \in A_1$. Part c) states that much more is true: each vector in A_1 is actually an extreme point of $\mathcal{A}_1$.

Lemma 6.1: Let $T \in \tilde{\mathcal{T}}$. Then
  a) for any nonempty subset U of V, $\boldsymbol{z}(U) \cdot \boldsymbol{\xi}(T) \geq 1$, i.e., T contains at least one edge from $\tilde{\delta}^-(U)$;
  b) if $T \in \tilde{\mathcal{T}}_1$ then, for any $m \geq 1$ and any collection of nonempty and pairwise-disjoint subsets $V_1, \cdots, V_{m+1}$ of V, $\boldsymbol{z}(V_1, \cdots, V_{m+1}) \cdot \boldsymbol{\xi}(T) \geq 1$, i.e., T contains at least m edges from $\bigcup_{i=1}^{m+1} \delta^-(V_i)$;
  c) if $T \in \tilde{\mathcal{T}}_1$, then $\boldsymbol{\xi}(T)$ is an extreme point of $\mathcal{A}_1$.

  Proof:
  a) Pick any $v \in U$. Then there is a path in T from r to v. Since $r \notin U$, there must be an edge in this path that exits a vertex not in U and enters one in U, and this edge belongs to $\tilde{\delta}^-(U)$.
  b) By the result of Part a), T has an edge from $\tilde{\delta}^-(V_i)$, for each $1 \leq i \leq m+1$. These $m+1$ edges must be distinct because the V_i's are pairwise-disjoint. Since $|T \cap \delta^+(r)| = 1$, at most one of these edges can belong to $\delta^+(r)$, which means that at least m of them must belong to $\bigcup_{i=1}^{m+1} \delta^-(V_i)$.
  c) By Parts a) and b), if $T \in \tilde{\mathcal{T}}_1$, then $\boldsymbol{\xi}(T) \in \mathcal{A}_1$. To prove that $\boldsymbol{\xi}(T)$ is actually an extreme point of $\mathcal{A}_1$, we must find a subset of the inequalities defining $\mathcal{A}_1$, whose *unique* solution, when converted to equalities, is $\boldsymbol{\xi}(T)$. Consider the following set of $|\tilde{E}|$ inequalities:

$$\xi_e \geq 0, \qquad e \in \tilde{E} - T \tag{165}$$

$$\boldsymbol{z}(U_v) \cdot \boldsymbol{\xi} \geq 1, \qquad v \in V. \tag{166}$$

Here, for each $v \in V$, $U_v \subseteq V$ is defined to be the set of vertices in the subarborescence of T that is rooted at v (including v); in other words, U_v is the set of all vertices u such that the path from r to u in T passes through v. Observe that T has exactly one edge from $\tilde{\delta}^-(U_v)$, viz., the edge that enters v. From this, it is obvious that $\boldsymbol{\xi}(T)$ is the unique solution of the equations obtained from (165) and (166). $\qquad\square$

C. Proof of $\mathcal{A}_1 \subseteq \text{conv}(A_1) + \boldsymbol{R}_+^{\tilde{E}}$

We will prove that

$$\min\{\boldsymbol{c} \cdot \boldsymbol{\xi} : \boldsymbol{\xi} \in A_1\} \leq \min\{\boldsymbol{c} \cdot \boldsymbol{\xi} : \boldsymbol{\xi} \in \mathcal{A}_1\},$$
$$\text{for } any \text{ vector } \boldsymbol{c} \in \boldsymbol{R}_+^{\tilde{E}}. \tag{167}$$

Since the minimum of any nonnegative linear functional over $\mathcal{A}_1$ must occur at one of its extreme points, and since $A_1 \subset \mathcal{A}_1$ is already known, it will follow from (167) that A_1 contains every extreme point of $\mathcal{A}_1$.

The main idea in the proof of (167), besides Theorem 2.4, is the following lemma, whose proof is based on a corollary of the max-flow min-cut theorem.

Lemma 6.2 (Extreme Point Lemma): Every extreme point $\boldsymbol{\xi}$ of the polyhedron $\mathcal{A}_1$ satisfies $\sum_{e \in \delta^+(r)} \xi_e = 1$.

Proof: See Section VI-D. $\qquad\qquad\qquad\square$

We will now proceed to prove (167), using Theorem 2.4 and Lemma 6.2. Let any $\boldsymbol{c} \in \boldsymbol{R}_+^{\tilde{E}}$ be given, and define a vector $\boldsymbol{c}' \in \boldsymbol{R}_+^{\tilde{E}}$ as follows:

$$c_e' \stackrel{\text{def}}{=} \begin{cases} c_e + M, & \text{if } e \in \delta^+(r) \\ c_e, & \text{otherwise.} \end{cases} \qquad (168)$$

Later, we will take M to be a large positive number. Now, since $\sum_{e \in \delta^+(r)} \xi_e = 1$ for any $\boldsymbol{\xi} \in A_1$, we have $\boldsymbol{c} \cdot \boldsymbol{\xi} = \boldsymbol{c}' \cdot \boldsymbol{\xi} - M$ for any $\boldsymbol{\xi} \in A_1$, and hence

$$\min_{\boldsymbol{\xi} \in A_1} \boldsymbol{c} \cdot \boldsymbol{\xi} = \min_{\boldsymbol{\xi} \in A_1} \boldsymbol{c}' \cdot \boldsymbol{\xi} - M. \qquad (169)$$

By Lemma 6.2, $\boldsymbol{c} \cdot \boldsymbol{\xi} = \boldsymbol{c}' \cdot \boldsymbol{\xi} - M$ for any extreme point $\boldsymbol{\xi}$ of $\mathcal{A}_1$, so that

$$\min_{\boldsymbol{\xi} \in \mathcal{A}_1} \boldsymbol{c} \cdot \boldsymbol{\xi} = \min_{\boldsymbol{\xi} \in \mathcal{A}_1} \boldsymbol{c}' \cdot \boldsymbol{\xi} - M. \qquad (170)$$

Because of (169) and (170), (167) will be proved if we show that

$$\min_{\boldsymbol{\xi} \in A_1} \boldsymbol{c}' \cdot \boldsymbol{\xi} \leq \min_{\boldsymbol{\xi} \in \mathcal{A}_1} \boldsymbol{c}' \cdot \boldsymbol{\xi}. \qquad (171)$$

To prove (171), we must bring Theorem 2.4 into play. The key to doing this is the following observation: if M is very large, then

$$\min_{\boldsymbol{\xi} \in A_1} \boldsymbol{c}' \cdot \boldsymbol{\xi} = \min_{\boldsymbol{\xi} \in A} \boldsymbol{c}' \cdot \boldsymbol{\xi} \qquad (172)$$

where A is as defined in (43). The reason is this: the RHS of (172) is the minimum of $\sum_{e \in T} c_e'$ over all spanning arborescences T in $\tilde{G}$. If M is very large—say, $M > \sum_{e \in \tilde{E}} c_e$—then any spanning arborescence T with more than one edge exiting r would have a much larger value of $\sum_{e \in T} c_e'$ than one with exactly one edge exiting r (note that each T must use at least one edge exiting r). Therefore, the RHS of (172) must actually equal the the minimum of $\sum_{e \in T} c_e'$ restricted to spanning arborescences T with exactly one edge exiting r, which is just the LHS of (172).

By the conclusion of Theorem 2.4, viz., (45), and the fact that $\mathcal{A} \supseteq \mathcal{A}_1$, we have

$$\min_{\boldsymbol{\xi} \in A} \boldsymbol{c}' \cdot \boldsymbol{\xi} = \min_{\boldsymbol{\xi} \in \mathcal{A}} \boldsymbol{c}' \cdot \boldsymbol{\xi}$$
$$\leq \min_{\boldsymbol{\xi} \in \mathcal{A}_1} \boldsymbol{c}' \cdot \boldsymbol{\xi}. \qquad (173)$$

From (172) and (173), we obtain the desired conclusion (171). This completes the proof.

D. Proof of the Extreme Point Lemma

We will actually prove the following stronger result: every *minimal* vector $\boldsymbol{\xi} \in \mathcal{A}_1$ satisfies $\sum_{e \in \delta^+(r)} \xi_e = 1$. Here, $\boldsymbol{\xi}$ is defined to be minimal if $\boldsymbol{\xi}' \in \mathcal{A}_1$ and $\boldsymbol{\xi}' \leq \boldsymbol{\xi}$ imply that $\boldsymbol{\xi}' = \boldsymbol{\xi}$. Clearly, every extreme point of $\mathcal{A}_1$ is minimal.

Let $\boldsymbol{\xi} \in \mathcal{A}_1$ be a given minimal vector. Now $\mathcal{A}_1$ is defined by a set of constraints, one for each vector in D_1. From the constraint $\boldsymbol{z}(V) \cdot \boldsymbol{\xi} \geq 1$, we have $\sum_{e \in \delta^+(r)} \xi_e \geq 1$. Let $\xi_{r, v_1}, \cdots, \xi_{r, v_k}$ be the positive terms in this sum, so that

$$\sum_{e \in \delta^+(r)} \xi_e = \sum_{i=1}^{k} \xi_{r, v_i}.$$

By the minimality of $\boldsymbol{\xi}$, decreasing $\boldsymbol{\xi}$ in the component corresponding to (r, v_i) results in a vector not in $\mathcal{A}_1$. This means that, for each $1 \leq i \leq k$, there is a constraint involving ξ_{r, v_i} that is "tight." In other words, there exist subsets $U_1, \cdots, U_k$ of V, such that $v_i \in U_i$ and

$$\boldsymbol{z}(U_i) \cdot \boldsymbol{\xi} = 1. \qquad (174)$$

We claim that

$$\boldsymbol{z}(U_j \cup U_{j'}) \cdot \boldsymbol{\xi} = 1, \qquad \text{if } U_j \cap U_{j'} \text{ is nonempty.} \qquad (175)$$

The proof of this claim is by a neat trick adapted from [14], which uses the max-flow min-cut theorem. Let $u \in U_j \cap U_{j'}$. Suppose we think of $\tilde{G}$ as a flow network with source r and sink u, in which the capacity of edge e is ξ_e. The constraints $\boldsymbol{z}(U) \cdot \boldsymbol{\xi} \geq 1$ then imply that every cut $(\tilde{V} - U, U), u \in U \subseteq V$, separating r and u has capacity at least 1. So, by (174) applied to U_j and $U_{j'}$, both $(\tilde{V} - U_j, U_j)$ and $(\tilde{V} - U_{j'}, U_{j'})$ are min-cuts. But by [11, Corollary I.5.4], this means that

$$(\tilde{V} - (U_j \cup U_{j'}), U_j \cup U_{j'})$$

is also a min-cut, which is the same as saying $\boldsymbol{z}(U_j \cup U_{j'}) \cdot \boldsymbol{\xi} = 1$.

Now, by repeatedly combining pairs of intersecting sets, we can express $U_1 \cup \cdots \cup U_k$ as the union of pairwise-disjoint sets $V_1, \cdots, V_{m+1}$, with $0 \leq m < k$. Here, each V_j is the union of certain of the U_i's. By applying (175) repeatedly to each pair of sets that are combined, we can conclude that $(\tilde{V} - V_j, V_j)$ is also a min-cut, for each j; i.e.,

$$\boldsymbol{z}(V_j) \cdot \boldsymbol{\xi} = 1, \qquad 1 \leq j \leq m+1. \qquad (176)$$

We can now prove $\sum_{e \in \delta^+(r)} \xi_e \leq 1$, and hence $\sum_{e \in \delta^+(r)} \xi_e = 1$, as follows. First of all

$$\sum_{e \in \delta^+(r)} \xi_e = \sum_{i=1}^{k} \xi_{r, v_i}$$
$$\leq \sum_{j=1}^{m+1} \sum_{v \in V_j} \xi_{r, v}. \qquad (177)$$

The inequality above holds because $v_1, \cdots, v_k$ are all in $V_1 \cup \cdots \cup V_{m+1}$. If $m = 0$, then

$$\sum_{j=1}^{m+1} \sum_{v \in V_j} \xi_{r, v} = \sum_{v \in V_1} \xi_{r, v}$$
$$\leq \sum_{e \in \tilde{\delta}^-(V_1)} \xi_e$$
$$= \boldsymbol{z}(V_1) \cdot \boldsymbol{\xi}$$
$$= 1. \qquad (178)$$

The last equality above is by (176). On the other hand, if $m > 0$, then

$$
\begin{aligned}
\sum_{j=1}^{m+1} \sum_{v \in V_j} \xi_{r,v} &= \sum_{j=1}^{m+1} \left[\sum_{e \in \tilde{\delta}^-(V_j)} \xi_e - \sum_{e \in \delta^-(V_j)} \xi_e \right] \\
&= \sum_{j=1}^{m+1} \mathbf{z}(V_j) \cdot \boldsymbol{\xi} - m[\mathbf{z}(V_1, \cdots, V_{m+1}) \cdot \boldsymbol{\xi}] \\
&\leq (m+1) - m \\
&= 1.
\end{aligned}
\tag{179}
$$

Here, the second equality is by the definitions (36) and (37). The inequality is by (176), and the constraint $\mathbf{z}(V_1, \cdots, V_{m+1})$ $\cdot \boldsymbol{\xi} \geq 1$, which holds because $\boldsymbol{\xi} \in \mathcal{A}_1$ and the V_j's are nonempty and pairwise-disjoint.

The desired result, $\sum_{e \in \delta^+(r)} \xi_e \leq 1$, now follows from (177)–(179).

REFERENCES

[1] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Z. Wahrsch. Verw. Gebiete*, vol. 33, pp. 159–175, 1978.

[2] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography, Part 1: Secret sharing," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1121–1132, July 1993.

[3] ——, "Common randomness in information theory and cryptography, Part 2: CR capacity," *IEEE Trans. Inform. Theory*, vol. 44, pp. 225–240, Jan. 1998.

[4] R. Ahlswede and G. Dueck, "Identification in the presence of feedback—A discovery of new capacity formulas," *IEEE Trans. Inform. Theory*, vol. 35, Jan. 1989.

[5] ——, "Identification via channels," *IEEE Trans. Inform. Theory*, vol. 35, Jan. 1989.

[6] R. Ahlswede and B. Verboven, "On identification via multiway channels with feedback," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1519–1526, Nov.. 1991.

[7] R. Ahlswede and Z. Zhang, "New directions in the theory of identification via channels," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1040–1050, July 1995.

[8] D. Blackwell, L. Breiman, and A. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, pp. 558–567, 1960.

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

[10] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inform. Theory*, vol. 34, pp. 181–193, Mar. 1988.

[11] L. R. Ford and D. R. Fulkerson, *Flows in Networks*.   Princeton, NJ: Princeton Univ. Press, 1962.

[12] D. R. Fulkerson, "Blocking polyhedra," in *Graph Theory and its Applications*, B. Harris, Ed.   New York: Academic, 1970.

[13] ——, "Blocking and anti-blocking pairs of polyhedra," *Math. Programming*, vol. 1, pp. 168–194, 1971.

[14] ——, "Packing rooted directed cuts in a weighted directed graph," *Math. Programming*, vol. 6, pp. 1–13, 1974.

[15] R. L. Graham, M. Grötschel, and L. Lovász, *Handbook of Combinatorics*.   Cambridge, MA: MIT Press, 1995.

[16] A. Hoffman, "Polyhedral aspects of discrete optimization," *Ann. Discr. Math.*, pp. 183–190, 1979.

[17] L. Lovász, "Communication complexity: A survey," in *Paths, Flows and VLSI Layout*, B. H. Korte *et al.*, Eds.   New York: Spriner-Verlag, 1990.

[18] U. M. Maurer, "Perfect cryptographic security from partially independent channels," in *Proc. 23rd Annu. ACM Symp. Theory of Computing*, 1991.

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

[20] A. Orlitsky and A. El Gamal, "Communication complexity," in *Complexity in Information Theory*, Y. Abu-Mostafa, Ed.   New York: Springer-Verlag, 1988.

[21] A. Schrijver, *Theory of Linear and Integer Programming*.   New York: Wiley, 1986.

[22] S. Venkatesan, "Generating common randomness from channel noise: Capacity formulas and combinatorial results," Ph.D. dissertation, Cornell University, Ithaca, NY, Jan. 1998.

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

[24] ——, "Identification plus transmission over channels with perfect feedback," *IEEE Trans. Inform. Theory*, vol. 44, pp. 284–290, Jan. 1998.

[25] J. Wolfowitz, *Coding Theorems of Information Theory*.   New York: Springer-Verlag, 1978.