GENERATING COMMON RANDOMNESS IN AN ARBITRARY NETWORK OF CHANNELS: CAPACITY FORMULAS AND SOME COMBINATORIAL RESULTS

by

Sivarama Venkatesan and Venkat Anantharam

Memorandum No. UCB/ERL M97/9

24 January 1997

GENERATING COMMON RANDOMNESS IN AN ARBITRARY NETWORK OF CHANNELS: CAPACITY FORMULAS AND SOME COMBINATORIAL RESULTS

by

Sivarama Venkatesan and Venkat Anantharam

Memorandum No. UCB/ERL M97/9

24 January 1997

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Generating Common Randomness in an Arbitrary Network of Channels: Capacity Formulas and some Combinatorial Results *†

Sivarama Venkatesan
School of Electrical Engg.
Cornell University
Ithaca, NY 14853

Venkat Anantharam
Dept. of EECS
Univ. of California
Berkeley, CA 94720

E-mail: {svenkat,ananth}@vyasa.eecs.berkeley.edu

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 description by linear inequalities of the blocking-type polyhedron whose extreme points are precisely the incidence vectors of all arborescences in a digraph, with a prescribed root of out-degree 1.

Keywords: Common randomness, interactive communication, blocking polyhedra, arborescences in a digraph.

*Research supported by NSF IRI 9005849, IRI 9310670, NCR 9422513, and the AT&T Foundation.

†Address all correspondence to the first author (currently at U.C. Berkeley as an Exchange Scholar): 211-134, Cory Hall, Dept. of EECS, Univ. of California, Berkeley, CA 94720.

1 Introduction

Two or more communicating agents are said to have *common randomness* if there is a random variable, e.g., a random bitstring, whose value is known 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], [8]. In the theory of identification over noisy channels [4], [5], [6], 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 [12], [15]. Finally, *secret* common randomness available to a transmitter and receiver allows them to communicate securely over a channel with eavesdroppers [2], [13], [14].

For all these reasons, Ahlswede and Csiszar initiated a systematic study of the role of common randomness in information theory and cryptography. In [2], they addressed the problem of generating common randomness at two terminals without giving information about it to an eavesdropper, in both a "source-type" model and a "channel-type" model. Then, 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.

However, in both [2] and [3], the possibility of exploiting *channel noise* to generate common randomness was not explored, except in the simple case of two terminals connected by a discrete memoryless channel with noiseless feedback. To see how channel noise could actually be useful in this context, consider a situation where there are two agents A and B connected to each other in both directions by a pair of channels. As an extreme case, assume that neither agent has access to any external sources of randomness (such as a random bit generator). Even then, A and B may be able to generate common randomness! The intuition is this: suppose A transmits some known input sequence to B. If the channel from A to B is noisy, then the resulting output sequence seen by B will be random. Since the input sequence is known, B could somehow "cancel" out its effect on the output sequence, and extract the randomness due to noise. Now, if the channel from B to A has positive Shannon capacity, then B could reliably convey the randomness thus obtained to A, using suitable encoding techniques; A and B would then have *common* randomness.

Of course, this procedure could be repeated, and A could simultaneously extract randomness from the output of the B-to-A channel (if it is noisy) and convey it reliably to B (if the A-to-B channel has positive Shannon capacity), thus generating more common randomness. For this scheme to work, it is clear that 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 question that arises here is: what is the maximum *rate*, in bits per step of communication, at which the two agents can generate common randomness this way from the noise on the two channels, i.e., what is the common randomness

capacity of the given pair of channels? This question was posed and answered in [17], under the assumption that the two channels are independently operating discrete memoryless channels (DMC's). The main result of [17] is that A and B can generate

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

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 A-to-B and B-to-A 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, which was proved by developing a novel typical sequence machinery for interactive communication. 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 expression in (1) reduces to $\min\{h(p) + h(q), 2 - h(p) - h(q)\}$, where $h(\cdot)$ is the binary entropy function. Observe that this is 0 if and only if either $h(p) = h(q) = 0$ (no randomness on either channel) or $h(p) = h(q) = 1$ (plenty of randomness but no ability to agree); moreover the capacity equals its maximum of 1 bit per step when $h(p) = h(q) = 1/2$. For details and other examples, see [17].

1.1 Subject of this paper

The results of [17] 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 by communicating interactively. 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 — $(u, v) \in E$ means there is a channel whose input is controlled by u, and whose output is seen by v. (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.) As in [17], we assume that these channels are all DMC's, and that they operate independently. Communication occurs simultaneously on all the channels, and in synchronism (i.e., there is a common clock).

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 each of the other agents, using suitable encoding techniques.

The mechanism for extracting randomness from noise extends in a straightforward manner from the two-agent case to the general situation considered here. How-

ever, 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 an arbitrary 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 non-trivial in the general situation. In fact, the solution to this problem leads to some purely combinatorial results about blocking polyhedra, which are of independent interest. Some of these results are described in Section 4 of this paper, including a characterization by linear inequalities of the blocking-type polyhedron whose extreme points are precisely the incidence vectors of all arborescences in a digraph, with a given root of out-degree 1 — a result akin to Fulkerson's in [10]. Other related results, including an algorithm for finding a minimum-weight arborescence with a given root of prescribed out-degree (and a corollary "packing" result similar to Fulkerson's optimum arborescence theorem [10]), will appear elsewhere.

Note that 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 $\exp(R_v)$ equiprobable output symbols. 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. In fact, it is not even necessary to assume that the external sources available to different agents are independent. By introducing self-loops *and* modifying the channels in the network appropriately, we could cover the situation where agent v gets an external random input $Z_v^{(k)}$ in step k, and the random vectors $(Z_v^{(k)} : v \in V)$, $k = 1, 2, \ldots$, are i.i.d. with an *arbitrary* common distribution. With these considerations in mind, we will assume from now on that all available external sources have been incorporated into the network itself.

The rest of the paper is organized as follows. In Section 2, we formulate the problem more precisely, and state our main results in Theorems 2.1, 2.2, and 2.3. Theorem 2.1 is an "achievability" result, stating that common randomness can be generated over the network at a certain rate C_*. Theorem 2.3 is a "converse" result, stating that no rate higher than a certain C^* is achievable. Theorem 2.2, whose proof yields the combinatorial results mentioned above, is then used to show that $C_* = C^*$ (this result has no non-trivial counterpart in [17]). These theorems are

proved in Sections 3, 4, and 5.

2 Statement of problem and results

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

If $\mathbf{Q} = (Q(y|x) : (x, y) \in \mathcal{X} \times \mathcal{Y})$ 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 $Pr\{X = x, Y = y\} = P(x)Q(y|x)$, then we will also write $H(\mathbf{Q}|P)$ and $I(P; \mathbf{Q})$ for $H(Y|X)$ and $I(X; Y)$, respectively.

2.1 Definition of common randomness capacity

As mentioned in Section 1.1, the network of DMC's connecting the agents will be represented by the digraph $G = (V, E)$. 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 probabilities $\mathbf{Q}_e = (Q_e(y|x) : (x, y) \in \mathcal{X}_e \times \mathcal{Y}_e)$.

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

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

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

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

Thus $\delta^-(W)$ (resp. $\delta^+(W)$) is the set of edges that exit (resp. enter) a vertex not in W and enter (resp. exit) one in W, while $\sigma(W)$ is the set of edges that exit *and* enter vertices in W. 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 non-empty, contains exactly one edge, viz. a self-loop on v. It will also be convenient further to define

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

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

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

To generate common randomness, the agents communicate interactively over the network for a certain number, say n, of steps. This communication proceeds according to an agreed-upon set of rules that specifies each agent's channel inputs in each step, based on the channel outputs available to him from 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 exiting channel $e \in \delta_{out}(v)$, as a function $X_{e,k} = f_{e,k}(\mathbf{Y}_{e'}^{k-1} : e' \in \delta_{in}(v))$ of the sequences of outputs $\mathbf{Y}_{e'}^{k-1}$ received in the previous $k-1$ steps on all entering channels $e' \in \delta_{in}(v)$.
- He then transmits the symbols $X_{e,k}$ on their respective exiting channels.
- Finally, he receives the outputs $Y_{e',k}$ corresponding to the symbols transmitted in step k on all his entering 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 $[K] \stackrel{\text{def}}{=} \{1, 2, \ldots, K\}$ — 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 $\mathbf{Y}_{e'}^{n}$, $e' \in \delta_{in}(v)$, and takes values in the set $\{*\} \cup [K]$. Here, $S_v = *$ is supposed to indicate that v declared failure to generate common randomness.

Let $f_e = (f_{e,1}, \ldots, f_{e,n})$, and $\mathbf{f} = (f_e : e \in E)$. Let $\mathbf{S} = (S_v : v \in V)$. Then the pair $(\mathbf{f}, \mathbf{S})$, 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 $(\mathbf{f}, \mathbf{S})$ 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

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

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 make the following definition: $(\mathbf{f}, \mathbf{S})$ *is an* (n, K, λ) *protocol if*

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

To motivate this definition, suppose that $K = \exp\{nR - o(n)\}$ for some $R > 0$, and $\lambda = o(1)$ — what this means is that $(\mathbf{f}, \mathbf{S})$ is the n^{th} term in a *sequence* of (n, K_n, λ_n) protocols, with $\liminf_{n\to\infty}(1/n)\log K_n = R$ and $\lim_{n\to\infty} \lambda_n = 0$. Then, by (8),

$$Pr\{\exists\, l \in [K] \text{ such that } S_v = l \text{ for all } v \in V\} \geq 1 - \lambda. \tag{9}$$

This means that all the agents compute the *same* random output with high probability. In particular, the probability that some agent declares failure to generate common randomness is small.

Next, we will show that each S_v has an entropy of $nR - o(n)$, implying that it has an approximately uniform distribution. In what follows, if $W \subseteq V$, then $\mathbf{S}_W \stackrel{\text{def}}{=} (S_v : v \in W)$. First note that

$$\begin{aligned} H(\mathbf{S}_V) &\stackrel{\text{def}}{=} -\sum_{\mathbf{s}} Pr\{\mathbf{S}_V = \mathbf{s}\} \log Pr\{\mathbf{S}_V = \mathbf{s}\} \\ &\geq -\sum_{l=1}^{K} Pr\{S_v = l \text{ for all } v \in V\} \log Pr\{S_v = l \text{ for all } v \in V\} \\ &\geq \sum_{l=1}^{K} \left(\frac{1-\lambda}{K}\right) \log\left(\frac{K}{1+\lambda}\right) \\ &= \log K - [\lambda \log K + (1-\lambda)\log(1+\lambda)] \\ &= nR - o(n). \end{aligned} \tag{10}$$

Here, the first inequality holds because $-z \log z \geq 0$ for $z \in [0,1]$, and the second inequality is by (8). Now (9) implies $Pr\{S_u \neq S_w\} \leq \lambda$ for all u and w. Hence $H(S_u|S_w) \leq 1 + \lambda \log K$, by Fano's inequality. In fact, for all non-empty sets U and W of vertices, we have

$$\begin{aligned} H(\mathbf{S}_U|\mathbf{S}_W) &\leq \sum_{u \in U} H(S_u|\mathbf{S}_W) \\ &\leq |U|(1 + \lambda \log K) \\ &= o(n). \end{aligned} \tag{11}$$

The inequalities above are by the chain rule for entropy and the fact that conditioning cannot increase entropy. From (10) and (11), it follows that for any non-empty $W \subset V$,

$$\begin{aligned} H(\mathbf{S}_W) &= H(\mathbf{S}_V) - H(\mathbf{S}_{V-W}|\mathbf{S}_W) \\ &\geq nR - o(n). \end{aligned} \tag{12}$$

In particular, $H(S_v) \geq nR - o(n)$ for any v, as claimed earlier. For future reference, note also that by (11) and (12), with W replaced by U in (12),

$$\begin{aligned} I(\mathbf{S}_U; \mathbf{S}_W) &= H(\mathbf{S}_U) - H(\mathbf{S}_U|\mathbf{S}_W) \\ &\geq nR - o(n). \end{aligned} \tag{13}$$

The above considerations motivate the following definition of the common randomness capacity of the given network:

Definition 2.1 *R is an achievable rate of generating common randomness over the given network if there exists a sequence of (n, K_n, λ_n) protocols such that*

$$\lim_{n\to\infty} \lambda_n = 0 \qquad \textit{and} \qquad \liminf_{n\to\infty} \frac{\log K_n}{n} = R. \tag{14}$$

The common randomness capacity of the network is the supremum of all achievable rates.

Our main result is a "single-letter" characterization of the common randomness capacity, in terms of the topology of the network and the characteristics of the channels constituting it.

2.2 Characterization of the common randomness capacity

For ease of reference, we first record the definitions of some standard graph-theoretic concepts. All definitions are 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

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

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.

An *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; and (iii) T does not contain any circuits.

It can be verified easily that the above properties imply that (iv) $|T| = |V| - 1$; (v) for each $u \neq v$, there is a *unique* path in T from v to u; and (vi) the edges of T form a spanning tree in the undirected graph underlying G. We will denote the set of all arborescences rooted at v by $\mathcal{T}(v)$, and let $\mathcal{T} = \bigcup_{v \in V} \mathcal{T}(v)$.

If $(u, v) \in T$, then we will say that u is the *parent of* v *in* T, and v is a *child of* u *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 arborescence has at least one leaf).

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

$$a_v(\mathbf{P}) \stackrel{\text{def}}{=} \sum_{e \in \mathcal{E}_{in}(v)} H(\mathbf{Q}_e | P_e), \tag{15}$$

$$b_e(\mathbf{P}) \stackrel{\text{def}}{=} I(P_e; \mathbf{Q}_e), \tag{16}$$

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

$$\sum_{T \in \mathcal{T}(v)} r_T \leq a_v(\mathbf{P}), \quad \text{for each } v \in V; \tag{17}$$

$$\sum_{T : e \in T} r_T \leq b_e(\mathbf{P}), \quad \text{for each } e \in E. \tag{18}$$

Note that the constraint in (18) can be ignored if e is a self-loop, because the LHS is then a summation over an empty set (no arborescence contains a self-loop), and is therefore equal to zero.

The "achievability" part of our main result essentially states that for any $\mathbf{P}$ and any $\mathbf{r} \in \mathcal{R}(\mathbf{P})$, the rate $\sum_T r_T$ is achievable.

Theorem 2.1 (Achievability result) *Let*

$$C_*(\mathbf{P}) \stackrel{\text{def}}{=} \max_{\mathbf{r}\in\mathcal{R}(\mathbf{P})} \sum_{T\in\mathcal{T}} r_T, \tag{19}$$

$$C_* \stackrel{\text{def}}{=} \max_{\mathbf{P}} C_*(\mathbf{P}). \tag{20}$$

Then the common randomness capacity of the network is bounded from below by C_*.

To establish that the common randomness capacity is actually *equal* to C_*, we must also prove an appropriate "converse" result. For this purpose, it will be convenient first to derive different, more explicit, expressions for $C_*(\mathbf{P})$ and C_*.

Note that $C_*(\mathbf{P})$ is defined in (19) to be the optimal value of a certain linear program (LP). We will now write down the *dual* to this LP. From now on, we will refer to the LP in (19) as the *primal*. The dual LP has a variable $x_v \in \mathbf{R}_+$ for each $v \in V$, and a variable $y_e \in \mathbf{R}_+$ for each $e \in E$. The dual constraints are

$$x_v + \sum_{e\in T} y_e \geq 1, \quad \text{for each } v \in V \text{ and } T \in \mathcal{T}(v). \tag{21}$$

Let $\mathcal{D}$ denote the dual feasible region (which does not depend on $\mathbf{P}$). Thus $\mathcal{D}$ is the polyhedron of all vectors $(\mathbf{x},\mathbf{y}) \in \mathbf{R}_+^V \times \mathbf{R}_+^E$ satisfying (21).

The dual objective is to minimize $\sum_{v\in V} a_v(\mathbf{P})x_v + \sum_{e\in E} b_e(\mathbf{P})y_e$. By linear programming duality, the optimal values of the primal and dual problems are equal, i.e.,

$$C_*(\mathbf{P}) = \min_{(\mathbf{x},\mathbf{y})\in\mathcal{D}} \left[\sum_{v\in V} a_v(\mathbf{P})x_v + \sum_{e\in E} b_e(\mathbf{P})y_e\right]. \tag{22}$$

The key step in obtaining more convenient expressions for $C_*(\mathbf{P})$ and C_* is to decompose the polyhedron $\mathcal{D}$ as the vector sum of the convex hull of its extreme points and the cone generated by its extreme directions. (Every polyhedron of non-negative vectors can be so decomposed.) Now, the cone generated by the extreme directions of $\mathcal{D}$ equals all of $\mathbf{R}_+^V \times \mathbf{R}_+^E$ because $\mathcal{D}$ has the following property: if $(\mathbf{x},\mathbf{y}) \in \mathcal{D}$, and $\mathbf{x}' \geq \mathbf{x}$, $\mathbf{y}' \geq \mathbf{y}$, then $(\mathbf{x}',\mathbf{y}') \in \mathcal{D}$. As for the extreme points of $\mathcal{D}$, the following result identifies a finite set $D_0 \subseteq \mathcal{D}$ that contains all of them.

Theorem 2.2 (Combinatorial result) *For each non-empty subset* W *of* V, *let* $\mathbf{x}(W) \in \mathbf{R}_+^V$ *and* $\mathbf{y}(W) \in \mathbf{R}_+^E$ *be given by*

$$x_v(W) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } v \in W; \\ 0 & \text{otherwise;} \end{cases} \qquad \text{and} \qquad y_e(W) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } e \in \delta^-(W); \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

For each $1 \leq m < |V|$, *and each collection of non-empty and pairwise disjoint subsets* $V_1, \ldots, V_{m+1}$ *of* V, *let* $\mathbf{y}(V_1, \ldots, V_{m+1}) \in \mathbf{R}_+^E$ *be given by*

$$y_e(V_1, \ldots, V_{m+1}) \stackrel{\text{def}}{=} \begin{cases} 1/m & \text{if } e \in \bigcup_{i=1}^{m+1} \delta^-(V_i); \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Let D_0 *be the set consisting of the vectors* $(\mathbf{x}(W),\mathbf{y}(W))$ *and* $(\mathbf{0},\mathbf{y}(V_1, \ldots, V_{m+1}))$ *defined above. Then* $D_0 \subseteq \mathcal{D}$; *in fact,*

$$\mathcal{D} = conv(D_0) + \mathbf{R}_+^V \times \mathbf{R}_+^E. \tag{25}$$

Here, $conv(D_0)$ *denotes the convex hull of all the vectors in* D_0.

The desired decomposition of $\mathcal{D}$ is given by (25). We are now in a position to state the "converse" part of the main result.

Theorem 2.3 (Converse result) *Let*

$$C^*(\mathbf{P}) \stackrel{\text{def}}{=} \min_{(\mathbf{x},\mathbf{y})\in D_0} \left[\sum_{v\in V} a_v(\mathbf{P})x_v + \sum_{e\in E} b_e(\mathbf{P})y_e\right], \tag{26}$$

$$C^* \stackrel{\text{def}}{=} \max_{\mathbf{P}} C^*(\mathbf{P}). \tag{27}$$

Then the common randomness capacity of the network is bounded from above by C^*.

We claim now that $C_*(\mathbf{P}) = C^*(\mathbf{P})$ for every $\mathbf{P}$, and hence $C_* = C^*$. To see this, note that by (25), and the non-negativity of the vectors $\mathbf{a}(\mathbf{P})$ and $\mathbf{b}(\mathbf{P})$, the optimal value of the dual LP equals the minimum of the dual objective function over $\text{conv}(D_0)$, which in turn equals the minimum over D_0, by linearity. But the dual optimal value also equals $C_*(\mathbf{P})$ by LP duality (see (22)), and the minimum of the dual objective over D_0 equals $C^*(\mathbf{P})$ by definition (see (26)). Hence $C_*(\mathbf{P}) = C^*(\mathbf{P})$. It follows now that both C_* and C^* equal the common randomness capacity of the given network.

Theorem 2.3 essentially states that if $\lim_{n\to\infty} \lambda_n = 0$, then there does not exist a sequence of (n, K_n, λ_n) protocols for generating common randomness with $\liminf_{n\to\infty}(1/n)\log K_n > C^*$. (We can actually prove a slightly stronger result, with liminf replaced by limsup.) In the usual terminology, this is a "weak" converse to Theorem 2.1. A "strong" converse would state that even if only $\limsup_{n\to\infty} \lambda_n < 1$ is assumed (instead of $\lim_{n\to\infty} \lambda_n = 0$), there does not exist a sequence of (n, K_n, λ_n) protocols with $\limsup_{n\to\infty}(1/n)\log K_n > C^*$. Such a result was proved in [17] for the two-agent case. By similar methods, it is indeed possible to prove a "strong" converse in the general case, too. However, we have omitted the details of the extension.

3 Proof of the achievability result

The intuition behind the achievability result can be summed up as follows: 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 $H(\mathbf{Q}_e|P_e)$ bits of noise. If the agents always use codewords from reliable block codes for communication, then they can recover this randomness due to noise from their respective incoming channels. This way, agent v can extract $\sum_{e\in\delta_{in}(v)} H(\mathbf{Q}_e|P_e)$ 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 arborescence comes in. An 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 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 through each arborescence $T \in \mathcal{T}(v)$. Since v gets only $\sum_{e \in \delta_{in}(v)} H(\mathbf{Q}_e|P_e)$ bits of randomness per step from noise, we have the constraint $\sum_{T \in \mathcal{T}(v)} r_T \leq \sum_{e \in \delta_{in}(v)} H(\mathbf{Q}_e|P_e)$ for each $v \in V$. Also, the channel e can reliably convey only $I(P_e; \mathbf{Q}_e)$ bits per step if its input symbols are used according to the distribution P_e. Therefore, we also have the capacity constraint $\sum_{T \ni e} r_T \leq I(P_e; \mathbf{Q}_e)$ for each $e \in E$. These are just the conditions (17) and (18) defining the polyhedron $\mathcal{R}(\mathbf{P})$.

So, as long as the vector of rates $\mathbf{r}$ is in $\mathcal{R}(\mathbf{P})$, 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 to all the other agents.

In Section 3.1, we will make the above reasoning precise using the following lemma, which was the basis for the achievability proof in [17] also.

Lemma 3.1 *Consider a DMC with finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and transition probabilities $\mathbf{Q} = (Q(y|x) : (x, y) \in \mathcal{X} \times \mathcal{Y})$. Suppose P is an n-type on $\mathcal{X}$ (i.e., P is a probability distribution on $\mathcal{X}$ such that $nP(x)$ is an integer for all $x \in \mathcal{X}$). Then:*

(1) For any $R' \geq 0$ and $L \leq \exp(nR')$, there exist sequences $\mathbf{c}(1), \ldots, \mathbf{c}(L)$ in $\mathcal{X}^n$, all of type P, and a partition of $\mathcal{Y}^n$ into subsets $\mathcal{C}(1), \ldots, \mathcal{C}(L)$, such that

$$1 - Q^n(\mathcal{C}(i)|\mathbf{c}(i)) \leq \exp\{-n\alpha(R', P) + o(n)\}, \quad \text{for all } i \in [L]. \tag{28}$$

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

(2) If $R'' \geq 0$ and $M \leq \exp(nR'')$ then, for any $\mathbf{c} \in \mathcal{X}^n$ of type P and any $\mathcal{C} \subseteq \mathcal{Y}^n$, there exists a partition of $\mathcal{C}$ into subsets $\mathcal{C}(*), \mathcal{C}(1), \ldots, \mathcal{C}(M)$, such that $Q^n(\mathcal{C}(j)|\mathbf{c})$ is the same for all $j \in [M]$, and

$$Q^n(\mathcal{C}(*)|\mathbf{c}) \leq \exp\{-n\beta(R'', P) + o(n)\}. \tag{29}$$

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

Proof: The first part is a standard result — see Theorem 5.2 on p. 165 of [7] for a proof. The second part is Lemma 3.2 in [17]. □

3.1 A protocol for generating common randomness

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 the first part of Lemma 3.1 says that we can find a "constant-composition" block code for channel e, with blocklength n and $\exp\{\lfloor n(I(P_e; \mathbf{Q}_e) - \epsilon) \rfloor\}$ codewords — all of type P_e — and a maximal error probability bounded by $\exp(-\alpha n)$, for

some $\alpha > 0$. Using this code, it is possible to convey one of $\exp\{\lfloor n(I(P_e; \mathbf{Q}_e) - \epsilon)\rfloor\}$ messages reliably across the channel e.

By the second part of Lemma 3.1, each decoding set in the above code can further be partitioned into $\exp\{\lfloor n(H(\mathbf{Q}_e|P_e) - \epsilon)\rfloor\}$ 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 bitstring of length $\lfloor n(H(\mathbf{Q}_e|P_e) - \epsilon)\rfloor$, or the failure symbol $*$. The entropy of B_e is close to $n(H(\mathbf{Q}_e|P_e) - \epsilon)$ because $Pr\{B_e = *\} \leq \exp(-\beta n)$ and, conditional on $B_e \neq *$, all the bitstrings are equiprobable.

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

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

1. First, for each 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 $\lfloor nr_T \rfloor$, or the failure symbol $*$ (we will describe in Stage 3 how it is determined).

   Observe that the channel e must therefore carry one of $\exp(\lfloor nr_T \rfloor) + 1$ messages for each arborescence T to which it belongs, so that the total number of possible messages is $\prod_{T \ni e}(\exp(\lfloor nr_T \rfloor) + 1)$. If ϵ is small enough, this number is indeed smaller than $\exp\{\lfloor n(I(P_e; \mathbf{Q}_e) - \epsilon)\rfloor\}$ — the number of codewords in the code for channel e — because $\mathbf{r}$ was assumed to be in the interior of $\mathcal{R}(\mathbf{P})$.

2. Next, for each 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 $\lfloor n(H(\mathbf{Q}_e|P_e) - \epsilon)\rfloor$, or the failure symbol $*$.

3. Finally, agent v determines the messages $M_{T,v}(k+1)$ to be transmitted in the next round to his children in T. There are two cases to consider here:

   (a) For each arborescence T that is *not* rooted at v, $M_{T,v}(k+1)$ equals $\hat{M}_{T,v}(k)$, the estimate that v made of the message sent in round k by his parent in T.

   (b) For the arborescences T that *are* rooted at v, $M_{T,v}(k+1)$ is 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 $\lfloor nr_T \rfloor$.

      In Case ii above, observe that we need a total of $\sum_{T \in \mathcal{T}(v)} \lfloor nr_T \rfloor$ bits to define the messages $M_{T,v}(k+1)$, $T \in \mathcal{T}(v)$. If ϵ is small enough, this number is indeed smaller than $\sum_{e \in \delta_{in}(v)} \lfloor n(H(\mathbf{Q}_e|P_e) - \epsilon) \rfloor$ — the length of the concatenation of all the $B_e(k)$ — because $\mathbf{r}$ was assumed to be in the interior of $\mathcal{R}(\mathbf{P})$.

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. Observe that in round k, $1 \leq k < n$, the root of each arborescence transmits a message to all his children, which is derived in round $k-1$ from the noise variables on the root's incoming channels. Each of these children forms an estimate of that message in round k, and in turn conveys that estimate to his own children in round $k+1$, and so on. Eventually, every agent forms an estimate of the message. All these estimates coincide with the original message if no decoding errors occur. 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 the messages are bitstrings contributing to common randomness.

Let $d_T(v)$ denote the *depth of v in T*, i.e., the number of edges in the unique path in T from the root to v (we will say that the root is itself at depth 0). Then, by the remarks made in the preceding paragraph, the message $M_{T,v}(k+d_T(v))$ transmitted by v in round $k+d_T(v)$ to his children in T is v's own estimate of $M_{T,u}(k)$, where u is the root of T. Let

$$\mathbf{M}_v \stackrel{\text{def}}{=} \Big(M_{T,v}(k + d_T(v)) : T \in \mathcal{T},\ 1 \leq k \leq n - d\Big), \tag{30}$$

where $d \stackrel{\text{def}}{=} \max_{T,v} d_T(v)$ is the maximum depth of any arborescence. Then, $\mathbf{M}_v$ is the vector of v's estimates of the messages sent by the roots of all arborescences in rounds $1, 2, \ldots, n-d$ (for arborescences rooted at v, these estimates are of course the same as the actual messages themselves). Clearly, if no decoding errors occur in any round on any of the channels, then the $\mathbf{M}_v$'s will all be equal.

Here is how agent v determines his decision random variable S_v. If some component of $\mathbf{M}_v$ equals $*$, i.e., if $M_{T,v}(k+d_T(v)) = *$ for some $T \in \mathcal{T}$ and $1 \leq k \leq n-d$, then agent v declares failure to generate common randomness, and sets $S_v = *$. On the other hand, if no component of $\mathbf{M}_v$ equals $*$, then 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 bitstrings $M_{T,v}(k+d_T(v))$, $T \in \mathcal{T}$, $1 \leq k \leq n-d$. Thus, the S_v's all take values in $\{*\} \cup [K]$, where

$$\begin{aligned} K &= \exp\left\{(n-d)\sum_{T\in\mathcal{T}} \lfloor nr_T \rfloor\right\} \\ &= \exp\left\{n^2\left(\sum_{T\in\mathcal{T}} r_T\right) - o(n^2)\right\}. \end{aligned} \tag{31}$$

It can be shown that the protocol satisfies (8) with

$$\lambda = n|E|\left[\exp(-\alpha n) + \exp(-\beta n)\right] \tag{32}$$

$$= o(1). \tag{33}$$

The RHS of (32) is 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. The proof of (32) is quite similar to that of Claim 3.1 in [17], and will not be repeated here (the main idea is to use the "Markovian" nature of the protocol to bound the probabilities involved).

From (31) and (33), it follows that the protocol achieves the rate $\sum_{T\in\mathcal{T}} r_T$. This completes the achievability proof.

4 Proof of the combinatorial result

For each $T \in \mathcal{T}$, define the vectors $\boldsymbol{\mu}(T) \in \mathbf{R}_+^V$ and $\boldsymbol{\nu}(T) \in \mathbf{R}_+^E$ as follows:

$$\mu_v(T) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } T \in \mathcal{T}(v); \\ 0 & \text{otherwise;} \end{cases} \quad \text{and} \quad \nu_e(T) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } e \in T; \\ 0 & \text{otherwise.} \end{cases} \tag{34}$$

$\boldsymbol{\mu}(T)$ and $\boldsymbol{\nu}(T)$ may be regarded as incidence vectors for the arborescence T — the former indicates the vertex at which T is rooted, and the latter which edges belong to T. Let

$$C_0 \stackrel{\text{def}}{=} \{(\boldsymbol{\mu}(T), \boldsymbol{\nu}(T)) : T \in \mathcal{T}\}. \tag{35}$$

Then, observe that the polyhedron $\mathcal{D}$ (the dual feasible region) is precisely the set of vectors $(\mathbf{x}, \mathbf{y}) \in \mathbf{R}_+^V \times \mathbf{R}_+^E$ such that $\sum_{v\in V} \mu_v x_v + \sum_{e\in E} \nu_e y_e \geq 1$ for all $(\boldsymbol{\mu}, \boldsymbol{\nu}) \in C_0$; i.e.,

$$\mathcal{D} = \left\{(\mathbf{x}, \mathbf{y}) \in \mathbf{R}_+^V \times \mathbf{R}_+^E : \boldsymbol{\mu}\cdot\mathbf{x} + \boldsymbol{\nu}\cdot\mathbf{y} \geq 1 \text{ for all } (\boldsymbol{\mu}, \boldsymbol{\nu}) \in C_0\right\}. \tag{36}$$

Now define a new polyhedron $\mathcal{C}$ by substituting D_0 for C_0 in the above definition; i.e., let

$$\mathcal{C} = \left\{(\boldsymbol{\mu}, \boldsymbol{\nu}) \in \mathbf{R}_+^V \times \mathbf{R}_+^E : \mathbf{x}\cdot\boldsymbol{\mu} + \mathbf{y}\cdot\boldsymbol{\nu} \geq 1 \text{ for all } (\mathbf{x}, \mathbf{y}) \in D_0\right\}. \tag{37}$$

In the terminology of polyhedral combinatorics ([11], Section 6 of Chapter 30), $\mathcal{D}$ (resp. $\mathcal{C}$) is the *blocker* of C_0 (resp. D_0).

Our goal is to prove (25). The starting point of the proof is the following fundamental result in the theory of blocking polyhedra: if $\mathcal{D}$ and $\mathcal{C}$ are the blockers of finite sets C_0 and D_0, respectively, then (25) holds if and only if

$$\mathcal{C} \;=\; \mathrm{conv}\,(C_0) \;+\; \mathbf{R}_+^V \times \mathbf{R}_+^E \tag{38}$$

holds. See [11], Theorem 6.3 in Chapter 30, for a proof of this equivalence, which is based on Farkas's lemma. If either (25) or (38) holds, then $\mathcal{D}$ and $\mathcal{C}$ are said to be a *blocking pair of polyhedra*, and the matrices whose rows are the vectors in C_0 and D_0, respectively, are said to be a *blocking pair of matrices.*

We will therefore prove (25) by proving (38). In doing so, it will be convenient to be able to treat vertices and edges on an equal footing. For this purpose, we will first extend G in the following way: let $\tilde{G} = (\tilde{V}, \tilde{E})$, where $\tilde{V} = V \cup \{r\}$, and $\tilde{E} = E \cup \{(r, v) : v \in V\}$. Thus $\tilde{G}$ is obtained from G by adding a new vertex r, and adding new edges from r to all the vertices in G. We may then identify the edge (r, v) in $\tilde{G}$ with the vertex v in G.

We will prove the statement (38) about G by proving an equivalent statement about $\tilde{G}$. Some notation will be useful here. Let

$$\delta^+(r) \;\stackrel{\mathrm{def}}{=}\; \{(r, v) : v \in V\}\,. \tag{39}$$

For $W \subseteq V$, continue to define $\delta^-(W)$ as in (2). Further, let

$$\tilde{\delta}^-(W) \;\stackrel{\mathrm{def}}{=}\; \delta^-(W) \cup \{(r, v) : v \in W\}\,. \tag{40}$$

We will denote the set of all arborescences in $\tilde{G}$ by $\tilde{\mathcal{T}}$. Note that *all* arborescences in $\tilde{G}$ must be rooted at r, since no edge in $\tilde{G}$ enters r. Further, there is a one-to-one correspondence between the arborescences in the original digraph G, and the elements of $\tilde{\mathcal{T}}$ that have exactly one edge exiting r; if $T \in \tilde{\mathcal{T}}$ satisfies $T \cap \delta^+(r) = \{(r, v)\}$, then T corresponds to the arborescence $T - \{(r, v)\}$ in G, which is rooted at v.

To derive the equivalent of (38) in $\tilde{G}$, we must define the equivalents in $\tilde{G}$ of the finite sets C_0 and D_0, and the polyhedron $\mathcal{C}$. First, for each $T \in \tilde{\mathcal{T}}$, define an incidence vector $\boldsymbol{\xi}(T) \in \mathbf{R}_+^{\tilde{E}}$ as follows:

$$\xi_e(T) \;\stackrel{\mathrm{def}}{=}\; \begin{cases} 1 & \text{if } e \in T; \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

Observe that if $T \cap \delta^+(r) = \{(r, v)\}$ then $\boldsymbol{\xi}(T)$ contains all the information in the vectors $\boldsymbol{\mu}(T')$ and $\boldsymbol{\nu}(T')$, where $T' \stackrel{\mathrm{def}}{=} T - \{(r, v)\}$ is the arborescence in G (rooted at v) to which T corresponds. Therefore, the appropriate equivalent in $\tilde{G}$ of the set C_0 is

$$\tilde{C}_0 \;\stackrel{\mathrm{def}}{=}\; \left\{\boldsymbol{\xi}(T) :\; T \in \tilde{\mathcal{T}},\; |T \cap \delta^+(r)| = 1\right\}. \tag{42}$$

Next, for each non-empty subset W of V, let $\mathbf{z}(W) \in \mathbf{R}_+^{\tilde{E}}$ be given by

$$z_e(W) \;\stackrel{\mathrm{def}}{=}\; \begin{cases} 1 & \text{if } e \in \tilde{\delta}^-(W); \\ 0 & \text{otherwise;} \end{cases} \tag{43}$$

and, for each collection of two or more non-empty and pairwise disjoint subsets $V_1, \ldots, V_{m+1}$ of V, let $\mathbf{z}(V_1, \ldots, V_{m+1}) \in \mathbf{R}_+^{\tilde{E}}$ be given by

$$z_e(V_1, \ldots, V_{m+1}) \stackrel{\text{def}}{=} \begin{cases} 1/m & \text{if } e \in \bigcup_{i=1}^{m+1} \delta^-(V_i); \\ 0 & \text{otherwise.} \end{cases} \tag{44}$$

Observe that $\mathbf{z}(W)$ and $\mathbf{z}(V_1, \ldots, V_{m+1})$ contain all the information in $(\mathbf{x}(W), \mathbf{y}(W))$ and $(\mathbf{0}, \mathbf{y}(V_1, \ldots, V_{m+1}))$, respectively. Therefore, the appropriate equivalent of D_0 in $\tilde{G}$ is the set $\tilde{D}_0$ consisting of all the vectors $\mathbf{z}(W)$ and $\mathbf{z}(V_1, \ldots, V_{m+1})$ defined above. Finally, it is obvious that the equivalent of the polyhedron $\mathcal{C}$ is

$$\tilde{\mathcal{C}} \stackrel{\text{def}}{=} \left\{ \boldsymbol{\xi} \in \mathbf{R}_+^{\tilde{E}} : \mathbf{z} \cdot \boldsymbol{\xi} \geq 1 \text{ for all } \mathbf{z} \in \tilde{D}_0 \right\}. \tag{45}$$

The statement (38) about G is then clearly equivalent to the following statement about $\tilde{G}$:

$$\tilde{\mathcal{C}} = \text{conv}(\tilde{C}_0) + \mathbf{R}_+^{\tilde{E}}. \tag{46}$$

The RHS of (46) is the polyhedron of all vectors in $\mathbf{R}_+^{\tilde{E}}$ that are greater than or equal to convex combinations of the incidence vectors of those arborescences in $\tilde{G}$ that use exactly one edge exiting r. Essentially, (46) says that this polyhedron can also be described as the set of vectors in $\mathbf{R}_+^{\tilde{E}}$ that satisfy a certain system of linear inequalities. We will derive (46) from the following corollary of Fulkerson's *optimum arborescence theorem*, which describes in a similar way the polyhedron of vectors in $\mathbf{R}_+^{\tilde{E}}$ that are greater than or equal to convex combinations of the incidence vectors of *all* arborescences in $\tilde{G}$ (without any degree constraints on the root r).

Theorem 4.1 (Fulkerson) *Let*

$$\tilde{A}_0 \stackrel{\text{def}}{=} \left\{ \boldsymbol{\xi}(T) : T \in \tilde{\mathcal{T}} \right\}, \tag{47}$$

and let

$$\tilde{\mathcal{A}} \stackrel{\text{def}}{=} \left\{ \boldsymbol{\xi} \in \mathbf{R}_+^{\tilde{E}} : \mathbf{z} \cdot \boldsymbol{\xi} \geq 1 \textit{ for all } \mathbf{z} \in \tilde{B}_0 \right\}, \tag{48}$$

where $\tilde{B}_0 \subseteq \tilde{D}_0$ is the set consisting only of the vectors $\mathbf{z}(W)$ defined in (43). Then

$$\tilde{\mathcal{A}} = \textit{conv}(\tilde{A}_0) + \mathbf{R}_+^{\tilde{E}}. \tag{49}$$

Proof: See [10]. □

Observe that $\tilde{A}_0 \supseteq \tilde{C}_0$ and $\tilde{\mathcal{A}} \supseteq \tilde{\mathcal{C}}$, so that (46) does not follow by a "sandwich" argument from (49). The following curious fact is worth noting: although $\tilde{\mathcal{C}}$ is obtained by adding many more constraints to those defining $\tilde{\mathcal{A}}$, it actually has far fewer extreme points than $\tilde{\mathcal{A}}$.

We will prove (46) by proving containment in both directions, beginning with the easier direction $\tilde{\mathcal{C}} \supseteq \text{conv}(\tilde{C}_0) + \mathbf{R}_+^{\tilde{E}}$. (This part of the proof does not require Theorem 4.1.) Since $\tilde{\mathcal{C}}$ is a convex set with the property that $\boldsymbol{\xi}' \in \tilde{\mathcal{C}}$ whenever $\boldsymbol{\xi}' \geq \boldsymbol{\xi}$ for some $\boldsymbol{\xi} \in \tilde{\mathcal{C}}$, the desired result will follow if we just show that $\tilde{\mathcal{C}} \supseteq \tilde{C}_0$, i.e., $\mathbf{z} \cdot \boldsymbol{\xi} \geq 1$ for all $\mathbf{z} \in \tilde{D}_0$ and $\boldsymbol{\xi} \in \tilde{C}_0$. This is the content of Parts (a) and (b) of the following lemma. Part (c) of the lemma states that much more is true: each vector in $\tilde{C}_0$ is actually an extreme point of $\tilde{\mathcal{C}}$.

Lemma 4.1 *Let $T \in \tilde{\mathcal{T}}$. Then:*

(a) for any non-empty subset W of V, $\mathbf{z}(W) \cdot \boldsymbol{\xi}(T) \geq 1$, i.e., T contains at least one edge from $\tilde{\delta}^-(W)$.

(b) if $|T \cap \delta^+(r)| = 1$ then, for any collection of two or more non-empty and pairwise disjoint subsets $V_1, \ldots, V_{m+1}$ of V, $\mathbf{z}(V_1, \ldots, V_{m+1}) \cdot \boldsymbol{\xi}(T) \geq 1$, i.e., T contains at least m edges from $\bigcup_{i=1}^{m+1} \delta^-(V_i)$.

(c) if $|T \cap \delta^+(r)| = 1$ then $\boldsymbol{\xi}(T)$ is an extreme point of $\tilde{\mathcal{C}}$.

Proof: Appendix. □

To prove containment in the other direction, viz. $\tilde{\mathcal{C}} \subseteq \text{conv}(\tilde{C}_0) + \mathbf{R}_+^{\tilde{E}}$, it suffices to prove that every extreme point of $\tilde{\mathcal{C}}$ is in $\tilde{C}_0$, because $\tilde{\mathcal{C}}$ is clearly equal to the set of all vectors that are greater than or equal to convex combinations of its extreme points. We will actually prove that

$$\min_{\boldsymbol{\xi} \in \tilde{\mathcal{C}}} \mathbf{q} \cdot \boldsymbol{\xi} \geq \min_{\boldsymbol{\xi} \in \tilde{C}_0} \mathbf{q} \cdot \boldsymbol{\xi} \quad \text{for } \textit{any} \text{ vector } \mathbf{q} \in \mathbf{R}_+^{\tilde{E}}. \tag{50}$$

Since the minimum of any non-negative linear functional over $\tilde{\mathcal{C}}$ must occur at one of its extreme points, and since $\tilde{\mathcal{C}} \supseteq \tilde{C}_0$ is already known, it will follow from (50) that $\tilde{C}_0$ contains every extreme point of $\tilde{\mathcal{C}}$. (Together with Part (c) of Lemma 4.1, this will actually imply that $\tilde{C}_0$ is *precisely* the set of extreme points of $\tilde{\mathcal{C}}$.)

The main idea in the proof of (50), besides Theorem 4.1, is the following lemma:

Lemma 4.2 *Every extreme point $\boldsymbol{\xi}$ of the polyhedron $\tilde{\mathcal{C}}$ satisfies $\sum_{e \in \delta^+(r)} \xi_e = 1$.*

Proof: Appendix. □

Let any $\mathbf{q} \in \mathbf{R}_+^{\tilde{E}}$ be given, and define a vector $\mathbf{q}' \in \mathbf{R}_+^{\tilde{E}}$ as follows:

$$q'_e \stackrel{\text{def}}{=} \begin{cases} q_e + M & \text{if } e \in \delta^+(r); \\ q_e & \text{otherwise.} \end{cases} \tag{51}$$

Later, we will take M to be a large positive number. Now, since $\sum_{e \in \delta^+(r)} \xi_e = 1$ for any $\boldsymbol{\xi} \in \tilde{C}_0$, we have $\mathbf{q} \cdot \boldsymbol{\xi} = \mathbf{q}' \cdot \boldsymbol{\xi} - M$ for any $\boldsymbol{\xi} \in \tilde{C}_0$, and hence

$$\min_{\boldsymbol{\xi} \in \tilde{C}_0} \mathbf{q} \cdot \boldsymbol{\xi} = \min_{\boldsymbol{\xi} \in \tilde{C}_0} \mathbf{q}' \cdot \boldsymbol{\xi} - M. \tag{52}$$

By Lemma 4.2, $\mathbf{q} \cdot \boldsymbol{\xi} = \mathbf{q}' \cdot \boldsymbol{\xi} - M$ for any extreme point $\boldsymbol{\xi}$ of $\tilde{\mathcal{C}}$, so that

$$\min_{\boldsymbol{\xi} \in \tilde{\mathcal{C}}} \mathbf{q} \cdot \boldsymbol{\xi} = \min_{\boldsymbol{\xi} \in \tilde{\mathcal{C}}} \mathbf{q}' \cdot \boldsymbol{\xi} - M. \tag{53}$$

Because of (52) and (53), (50) will be proved if we show that

$$\min_{\boldsymbol{\xi} \in \tilde{\mathcal{C}}} \mathbf{q}' \cdot \boldsymbol{\xi} \geq \min_{\boldsymbol{\xi} \in \tilde{C}_0} \mathbf{q}' \cdot \boldsymbol{\xi}. \tag{54}$$

To prove (54), we must bring Theorem 4.1 into play. The key to doing this is the following observation: if M is very large, then

$$\min_{\boldsymbol{\xi} \in \tilde{C}_0} \mathbf{q}' \cdot \boldsymbol{\xi} = \min_{\boldsymbol{\xi} \in \tilde{A}_0} \mathbf{q}' \cdot \boldsymbol{\xi}. \tag{55}$$

The reason is this: the RHS of (55) is the minimum of $\sum_{e \in T} q'_e$ over all arborescences T in $\tilde{G}$. If M is very large — say, $M > \sum_{e \in \tilde{E}} q_e$ — then any arborescence T with more than one edge exiting r would have a much larger value of $\sum_{e \in T} q'_e$ than one with exactly one edge exiting r (each T must use at least one edge exiting r). Therefore, the RHS of (55) must actually equal the the minimum of $\sum_{e \in T} q'_e$ restricted to arborescences T with exactly one edge exiting r, which is just the LHS.

By the conclusion of Theorem 4.1, viz. (49), and the fact that $\tilde{\mathcal{A}} \supseteq \tilde{\mathcal{C}}$, we have

$$\begin{aligned} \min_{\boldsymbol{\xi} \in \tilde{A}_0} \mathbf{q}' \cdot \boldsymbol{\xi} &= \min_{\boldsymbol{\xi} \in \tilde{\mathcal{A}}} \mathbf{q}' \cdot \boldsymbol{\xi} \\ &\leq \min_{\boldsymbol{\xi} \in \tilde{\mathcal{C}}} \mathbf{q}' \cdot \boldsymbol{\xi}. \end{aligned} \tag{56}$$

From (55) and (56), we obtain the desired conclusion (54). This completes the proof.

5 Proof of the converse result

Let $(\mathbf{f}, \mathbf{S})$ be any (n, K, λ) protocol for generating common randomness, with $K = \exp\{nR - o(n)\}$ and $\lambda = o(1)$; as in Section 2.1, this means that $(\mathbf{f}, \mathbf{S})$ is the n^{th} term in a sequence of (n, K_n, λ_n) protocols, with $\liminf_{n \to \infty} (1/n) \log K_n = R$ and $\lim_{n \to \infty} \lambda_n = 0$. We will then prove that $R \leq C^*$.

Let $X_{e,k}$ and $Y_{e,k}$ be the random variables representing the input and output, respectively, on edge e in step k, when the agents communicate according to the protocol $(\mathbf{f}, \mathbf{S})$. For each $e \in E$, define a probability distribution P_e on $\mathcal{X}_e$ as follows:

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

Thus P_e is the average of the distributions of $X_{e,1}, \ldots, X_{e,n}$. It will be convenient to define random variables X_e and Y_e, for each $e \in E$, with the following joint distribution:

$$Pr\{X_e = x, Y_e = y\} = P_e(x) Q_e(y|x), \quad (x, y) \in \mathcal{X}_e \times \mathcal{Y}_e. \tag{57}$$

Let $\mathbf{P} = (P_e : e \in E)$. We will actually prove that $R \leq C^*(\mathbf{P}) + o(1)$. This will obviously imply that $R \leq C^* + o(1)$, whence we will get the desired result $R \leq C^*$ by letting n tend to ∞. In the proof, the following notation will be useful: if $E' \subseteq V \times V$ then

$$\begin{aligned} \mathbf{X}_{E',k} &\stackrel{\text{def}}{=} (X_{e,k} : e \in E' \cap E) \quad \text{and} \quad \mathbf{X}_{E'}^{k} \stackrel{\text{def}}{=} (\mathbf{X}_{E',i} : 1 \leq i \leq k); \\ \mathbf{Y}_{E',k} &\stackrel{\text{def}}{=} (Y_{e,k} : e \in E' \cap E) \quad \text{and} \quad \mathbf{Y}_{E'}^{k} \stackrel{\text{def}}{=} (\mathbf{Y}_{E',i} : 1 \leq i \leq k). \end{aligned}$$

Recall also that $\mathbf{S}_W \stackrel{\text{def}}{=} (S_v : v \in W)$, if $W \subseteq V$.

Now, $C^*(\mathbf{P})$ was defined in (26) to be the minimum of the dual objective function $\sum_{v\in V} a_v(\mathbf{P})x_v + \sum_{e\in E} b_e(\mathbf{P})y_e$ over all vectors $(\mathbf{x},\mathbf{y})$ in D_0. To show that $R \leq C^*(\mathbf{P}) + o(1)$, it suffices therefore to show that R is bounded above by the value of the dual objective function at each $(\mathbf{x},\mathbf{y})$ in D_0 (neglecting $o(1)$ terms). The finite set D_0 consists of a vector $(\mathbf{x}(W),\mathbf{y}(W))$ for each non-empty set W of vertices — call these vectors of the first type — and a vector $(\mathbf{0},\mathbf{y}(V_1,\ldots,V_{m+1}))$ for each choice of two or more non-empty and pairwise disjoint sets $V_1,\ldots,V_{m+1}$ of vertices — call these vectors of the second type. For vectors of the first type, we must prove that, up to $o(1)$ terms,

$$\begin{aligned} R &\leq \sum_{v\in V} a_v(\mathbf{P})x_v(W) + \sum_{e\in E} b_e(\mathbf{P})y_e(W) \\ &= \sum_{v\in W} a_v(\mathbf{P}) + \sum_{e\in\delta^-(W)} b_e(\mathbf{P}) \qquad (58) \\ &= \sum_{v\in W}\sum_{e\in\delta^-(v)} H(Y_e|X_e) + \sum_{e\in\delta^-(W)} I(X_e;Y_e). \qquad (59) \end{aligned}$$

Here, (58) is by (23), and (59) is by (15), (16), and (57). For vectors of the second type, we must prove that, up to $o(1)$ terms,

$$\begin{aligned} R &\leq \sum_{e\in E} b_e(\mathbf{P})y_e(V_1,\ldots,V_{m+1}) \\ &= \frac{1}{m}\sum_{j=1}^{m+1}\sum_{e\in\delta^-(V_j)} b_e(\mathbf{P}) \qquad (60) \\ &= \frac{1}{m}\sum_{j=1}^{m+1}\sum_{e\in\delta^-(V_j)} I(X_e;Y_e). \qquad (61) \end{aligned}$$

Here, (60) is by (24), and (61) is again by (16) and (57).

5.1 Vectors of the first type

Let W be a non-empty set of vertices. Then $nR - o(n) \leq H(\mathbf{S}_W)$, by (12). So (59) will be proved if we show that $H(\mathbf{S}_W)$ is bounded above by n times the RHS of (59), which we proceed to do now:

$$\begin{aligned} H(\mathbf{S}_W) &\leq H\left(\mathbf{Y}^n_{\delta_{in}(W)}\right) \qquad (62) \\ &= \sum_{k=1}^{n} H\left(\mathbf{Y}_{\delta_{in}(W),k}|\mathbf{Y}^{k-1}_{\delta_{in}(W)}\right) \qquad (63) \\ &= \sum_{k=1}^{n} H\left(\mathbf{Y}_{\sigma(W),k},\mathbf{Y}_{\delta^-(W),k}|\mathbf{Y}^{k-1}_{\delta_{in}(W)}\right) \qquad (64) \\ &= \sum_{k=1}^{n} H\left(\mathbf{Y}_{\sigma(W),k},\mathbf{Y}_{\delta^-(W),k}|\mathbf{Y}^{k-1}_{\delta_{in}(W)},\mathbf{X}_{\sigma(W),k}\right) \qquad (65) \\ &\leq \sum_{k=1}^{n}\left[H\left(\mathbf{Y}_{\sigma(W),k}|\mathbf{X}_{\sigma(W),k}\right) + H\left(\mathbf{Y}_{\delta^-(W),k}\right)\right] \qquad (66) \end{aligned}$$

$$\leq \sum_{k=1}^{n}\left[\sum_{e\in\sigma(W)} H(Y_{e,k}|X_{e,k}) + \sum_{e\in\delta^-(W)} H(Y_{e,k})\right] \tag{67}$$

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

$$\leq n\left[\sum_{e\in\sigma(W)} H(Y_e|X_e) + \sum_{e\in\delta^-(W)} H(Y_e)\right] \tag{68}$$

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

$$= n\left[\sum_{v\in W}\sum_{e\in\delta^-(v)} H(Y_e|X_e) + \sum_{e\in\delta^-(W)} I(X_e;Y_e)\right]. \tag{69}$$

Here, (62) is because $\mathbf{S}_W$ is a function of $\mathbf{Y}^n_{\delta_{in}(W)}$; (63) is by the chain rule for entropy; (64) is by (5); (65) is because $\mathbf{X}_{\sigma(W),k}$ is a function of $\mathbf{Y}^{k-1}_{\delta_{in}(W)}$; (66) and (67) are by the chain rule and the fact that conditioning cannot increase entropy; (68) is by the concavity of the entropy and conditional entropy functions; and (69) is because $\sigma(W)\cup\delta^-(W) = \bigcup_{v\in W}\delta^-(v)$.

5.2 Vectors of the second type

Let $V_1,\ldots,V_{m+1}$ be non-empty and pairwise disjoint sets of vertices. Let

$$U \stackrel{\text{def}}{=} V - \bigcup_{i=1}^{m+1} V_i, \tag{70}$$

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

$$W_j \stackrel{\text{def}}{=} \bigcup_{i=1}^{j} V_i. \tag{71}$$

Then, by (13), $nR - o(n) \leq I(\mathbf{S}_{W_j};\mathbf{S}_{V_{j+1}})$, $j=1,2,\ldots,m$. To prove (61), it will therefore suffice to prove that $(1/m)\sum_{j=1}^{m} I(\mathbf{S}_{W_j};\mathbf{S}_{V_{j+1}})$ is bounded above by n times the RHS of (61). To this end, note that

$$\sum_{j=1}^{m} I\left(\mathbf{S}_{W_j};\mathbf{S}_{V_{j+1}}\right) \leq \sum_{j=1}^{m} I\left(\mathbf{Y}^n_{\delta_{in}(W_j)};\mathbf{Y}^n_{\delta_{in}(V_{j+1})}\right) \tag{72}$$

$$\leq \sum_{j=1}^{m} I\left(\mathbf{Y}^n_{\delta_{in}(W_j)},\mathbf{Y}^n_{W_j\times U};\mathbf{Y}^n_{\delta_{in}(V_{j+1})},\mathbf{Y}^n_{V_{j+1}\times U}\right) \tag{73}$$

$$= \sum_{j=1}^{m}\left[H\left(\mathbf{Y}^n_{\delta_{in}(W_j)},\mathbf{Y}^n_{W_j\times U}\right) + H\left(\mathbf{Y}^n_{\delta_{in}(V_{j+1})},\mathbf{Y}^n_{V_{j+1}\times U}\right) - H\left(\mathbf{Y}^n_{\delta_{in}(W_{j+1})},\mathbf{Y}^n_{W_{j+1}\times U}\right)\right] \tag{74}$$

$$= H\left(\mathbf{Y}^n_{\delta_{in}(W_1)},\mathbf{Y}^n_{W_1\times U}\right) + \sum_{j=1}^{m} H\left(\mathbf{Y}^n_{\delta_{in}(V_{j+1})},\mathbf{Y}^n_{V_{j+1}\times U}\right)$$

$$- H\left(\mathbf{Y}^n_{\delta_{in}(W_{m+1})}, \mathbf{Y}^n_{W_{m+1}\times U}\right) \tag{75}$$

$$= \sum_{j=1}^{m+1} H\left(\mathbf{Y}^n_{\delta_{in}(V_j)}, \mathbf{Y}^n_{V_j\times U}\right) - H\left(\mathbf{Y}^n_{\delta_{in}(W_{m+1})}, \mathbf{Y}^n_{W_{m+1}\times U}\right). \tag{76}$$

Here, (72) is because $\mathbf{S}_{W_j}$ is a function of $\mathbf{Y}^n_{\delta_{in}(W_j)}$, and $\mathbf{S}_{V_{j+1}}$ is a function of $\mathbf{Y}^n_{\delta_{in}(V_{j+1})}$; (73) is because $I(Z_1; Z_2) \leq I(Z_1, Z_3; Z_2, Z_4)$; (74) is obtained using the formula $I(Z_1; Z_2) = H(Z_1) + H(Z_2) - H(Z_1, Z_2)$, and the definition (71); (75) is obtained by "telescoping" the sum in (74); and in (76) we have used the fact that $W_1 = V_1$, by the definition in (71). We will bound each of the terms in (76) separately. First,

$$H\left(\mathbf{Y}^n_{\delta_{in}(V_j)}, \mathbf{Y}^n_{V_j\times U}\right)$$

$$= \sum_{k=1}^{n} H\left(\mathbf{Y}_{\delta_{in}(V_j),k}, \mathbf{Y}_{V_j\times U,k} | \mathbf{Y}^{k-1}_{\delta_{in}(V_j)}, \mathbf{Y}^{k-1}_{V_j\times U}\right) \tag{77}$$

$$= \sum_{k=1}^{n} H\left(\mathbf{Y}_{\delta^-(V_j),k}, \mathbf{Y}_{\sigma(V_j),k}, \mathbf{Y}_{V_j\times U,k} | \mathbf{Y}^{k-1}_{\delta_{in}(V_j)}, \mathbf{Y}^{k-1}_{V_j\times U}\right) \tag{78}$$

$$\leq \sum_{k=1}^{n} H\left(\mathbf{Y}_{\delta^-(V_j),k}, \mathbf{Y}_{\sigma(V_j),k}, \mathbf{Y}_{V_j\times U,k} | \mathbf{Y}^{k-1}_{\delta_{in}(V_j)}, \mathbf{X}_{\sigma(V_j),k}, \mathbf{X}_{V_j\times U,k}\right) \tag{79}$$

$$\leq \sum_{k=1}^{n} \Big[H\left(\mathbf{Y}_{\delta^-(V_j),k}\right) + H\left(\mathbf{Y}_{\sigma(V_j),k} | \mathbf{X}_{\sigma(V_j),k}\right) + H\left(\mathbf{Y}_{V_j\times U,k} | \mathbf{X}_{V_j\times U,k}\right) \Big] \tag{80}$$

$$\leq \sum_{k=1}^{n} \Big[\sum_{e\in\delta^-(V_j)} H(Y_{e,k}) + \sum_{e\in\sigma(V_j)} H(Y_{e,k}|X_{e,k}) + \sum_{e\in(V_j\times U)\cap E} H(Y_{e,k}|X_{e,k}) \Big]. \tag{81}$$

Here, (77) is by the chain rule; (78) is by (5); (79) is obtained from (78) by first conditioning on $\mathbf{X}_{\sigma(V_j),k}$ and $\mathbf{X}_{V_j\times U,k}$ also (this does not change (78) since these are both functions of $\mathbf{Y}^{k-1}_{\delta_{in}(V_j)}$), and then dropping the conditioning on $\mathbf{Y}^{k-1}_{V_j\times U}$ (conditioning cannot increase entropy); and (80) and (81) are by the chain rule and the fact that conditioning cannot increase entropy.

Next, we will bound $H(\mathbf{Y}^n_{\delta_{in}(W_{m+1})}, \mathbf{Y}^n_{W_{m+1}\times U})$ from below. For convenience, let $E' \stackrel{\text{def}}{=} \delta_{in}(W_{m+1}) \cup (W_{m+1} \times U)$. Then,

$$H\left(\mathbf{Y}^n_{\delta_{in}(W_{m+1})}, \mathbf{Y}^n_{W_{m+1}\times U}\right) \tag{82}$$

$$= H(\mathbf{Y}^n_{E'})$$

$$= \sum_{k=1}^{n} H\left(\mathbf{Y}_{E',k} | \mathbf{Y}^{k-1}_{E'}\right) \tag{83}$$

$$\geq \sum_{k=1}^{n} H\left(\mathbf{Y}_{E',k} | \mathbf{Y}^{k-1}_{E}\right) \tag{84}$$

$$= \sum_{k=1}^{n} H\left(\mathbf{Y}_{E',k}|\mathbf{Y}_E^{k-1}, \mathbf{X}_{E',k}\right) \quad (85)$$

$$= \sum_{k=1}^{n} \sum_{e\in E'} H\left(Y_{e,k}|X_{e,k}\right) \quad (86)$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{m+1} \left[\sum_{e\in\delta_{in}(V_j)} H\left(Y_{e,k}|X_{e,k}\right) + \sum_{e\in(V_j\times U)\cap E} H\left(Y_{e,k}|X_{e,k}\right) \right] \quad (87)$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{m+1} \left[\sum_{e\in\delta^-(V_j)} H\left(Y_{e,k}|X_{e,k}\right) + \sum_{e\in\sigma(V_j)} H\left(Y_{e,k}|X_{e,k}\right) + \sum_{e\in(V_j\times U)\cap E} H\left(Y_{e,k}|X_{e,k}\right) \right]. \quad (88)$$

Here, (83) is again by the chain rule; (84) is because conditioning cannot increase entropy; (85) is because $\mathbf{X}_{E',k}$ is a function of $\mathbf{Y}_E^{k-1}$; (86) is by the chain rule and the fact that, given $X_{e,k}$, $Y_{e,k}$ is conditionally independent of $\mathbf{Y}_E^{k-1}$, $\mathbf{X}_{E-\{e\},k}$, and $\mathbf{Y}_{E-\{e\},k}$; (87) is by the definition $E' = \delta_{in}(W_{m+1}) \cup (W_{m+1} \times U)$ and (71); and (88) is by (5).

From (76), (81), and (88), we have, after cancelling common terms,

$$\begin{aligned} \frac{1}{m}\sum_{j=1}^{m} I\left(\mathbf{S}_{W_j};\mathbf{S}_{V_{j+1}}\right) &\leq \frac{1}{m}\sum_{k=1}^{n}\sum_{j=1}^{m+1}\sum_{e\in\delta^-(V_j)} [H(Y_{e,k}) - H(Y_{e,k}|X_{e,k})] \\ &= \frac{1}{m}\sum_{j=1}^{m+1}\sum_{e\in\delta^-(V_j)}\left[\sum_{k=1}^{n} I(X_{e,k};Y_{e,k})\right] \\ &\leq n\left[\frac{1}{m}\sum_{j=1}^{m+1}\sum_{e\in\delta^-(V_j)} I(X_e;Y_e)\right]. \end{aligned} \quad (89)$$

The inequality in (89) is by the concavity of the mutual information function. This completes the proof of the converse.

Appendix

Proof of Lemma 4.1:

(a) Pick any $v \in W$. Then there is a path in T from r to v. Since $r \notin W$, there must be an edge in this path that exits a vertex not in W and enters one in W.

(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 $|T\cap\delta^+(r)| = 1$, then $\boldsymbol{\xi}(T) \in \tilde{C}$. To prove that $\boldsymbol{\xi}(T)$ is actually an extreme point of $\tilde{C}$, we must find a subset of the inequalities defining $\tilde{C}$, whose *unique* solution, when converted to equalities, is $\boldsymbol{\xi}(T)$. Consider the following

set of $|\tilde{E}|$ inequalities:

$$\xi_e \geq 0, \quad e \in \tilde{E} - T; \tag{90}$$

$$\mathbf{z}(W_v) \cdot \boldsymbol{\xi}(T) \geq 1, \quad v \in V. \tag{91}$$

Here, for each $v \in V$, $W_v \subseteq V$ is the set of vertices in the sub-arborescence of T that is rooted at v (including v); in other words, W_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 $\tilde{\delta}^-(W_v)$, viz., the edge that enters v. Using this fact, it can be verified that $\boldsymbol{\xi}(T)$ is the unique solution of the equations obtained from (90) and (91). □

Proof of Lemma 4.2:

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

Let $\boldsymbol{\xi} \in \tilde{C}$ be a given minimal vector. Now $\tilde{C}$ is defined by a set of constraints, one for each vector in $\tilde{D}_0$. From the constraint $\mathbf{z}(V) \cdot \boldsymbol{\xi} \geq 1$, we have $\sum_{e \in \delta^+(r)} \xi_e \geq 1$. Let $\xi_{r,v_1}, \ldots, \xi_{r,v_k}$ 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 $\tilde{C}$. 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 $W_1, \ldots, W_k$ of V, such that $v_i \in W_i$ and

$$\mathbf{z}(W_i) \cdot \boldsymbol{\xi} = 1. \tag{92}$$

We claim that

$$\mathbf{z}(W_j \cup W_{j'}) \cdot \boldsymbol{\xi} = 1 \quad \text{if } W_j \cap W_{j'} \text{ is non-empty.} \tag{93}$$

The proof of this claim is by a neat trick adapted from [10], which uses the max-flow min-cut theorem. Let $u \in W_j \cap W_{j'}$. 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 $\mathbf{z}(W) \cdot \boldsymbol{\xi} \geq 1$ then imply that every cut $(\tilde{V} - W, W)$, $u \in W \subseteq V$, separating r and u has capacity at least 1. So, by (92) applied to W_j and $W_{j'}$, both $(\tilde{V} - W_j, W_j)$ and $(\tilde{V} - W_{j'}, W_{j'})$ are min-cuts. But by [9], Corollary I.5.4, this means that $(\tilde{V} - (W_j \cup W_{j'}), W_j \cup W_{j'})$ is also a min-cut, which is the same as saying $\mathbf{z}(W_j \cup W_{j'}) \cdot \boldsymbol{\xi} = 1$.

Now, by repeatedly combining pairs of intersecting sets, we can express $W_1 \cup \cdots \cup W_k$ as the union of pairwise disjoint sets $V_1, \ldots, V_{l+1}$, with $0 \leq l < k$. Here, each V_j is the union of certain of the W_i's. By applying (93) 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.,

$$\mathbf{z}(V_j) \cdot \boldsymbol{\xi} = 1, \quad 1 \leq j \leq l+1. \tag{94}$$

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,

$$\begin{aligned} \sum_{e \in \delta^+(r)} \xi_e &= \sum_{i=1}^{k} \xi_{r,v_i} \\ &\leq \sum_{j=1}^{l+1} \sum_{v \in V_j} \xi_{r,v}. \end{aligned} \tag{95}$$

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

$$\begin{aligned} \sum_{j=1}^{l+1} \sum_{v \in V_j} \xi_{r,v} &= \sum_{v \in V_1} \xi_{r,v} \\ &\leq \sum_{e \in \tilde{\delta}^-(V_1)} \xi_e \\ &= \mathbf{z}(V_1) \cdot \boldsymbol{\xi} \\ &= 1. \end{aligned} \tag{96}$$

The last equality above is by (94). On the other hand, if $l > 0$, then

$$\begin{aligned} \sum_{j=1}^{l+1} \sum_{v \in V_j} \xi_{r,v} &= \sum_{j=1}^{l+1} \left[\sum_{e \in \tilde{\delta}^-(V_j)} \xi_e - \sum_{e \in \delta^-(V_j)} \xi_e \right] \\ &= \sum_{j=1}^{l+1} \mathbf{z}(V_j) \cdot \boldsymbol{\xi} \; - \; l \left[\mathbf{z}(V_1, \ldots, V_{l+1}) \cdot \boldsymbol{\xi} \right] \\ &\leq (l+1) - l \\ &= 1. \end{aligned} \tag{97}$$

Here, the second equality is by the definitions (43) and (44). The inequality is by (94), and the constraint $\mathbf{z}(V_1, \ldots, V_{l+1}) \cdot \boldsymbol{\xi} \geq 1$, which holds because $\boldsymbol{\xi} \in \tilde{C}$ and the V_j's are non-empty and pairwise disjoint.

The desired result follows from (95), (96), and (97). □

References

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

[2] R. Ahlswede and I. Csiszàr. Common randomness in information theory and cryptography, Part 1: Secret sharing. *IEEE Transactions on Information Theory*, Vol. 39(No. 4), July 1993.

[3] R. Ahlswede and I. Csiszar. Common randomness in information theory and cryptography, Part 2: CR capacity. Technical Report 95-101, Universitat Bielefeld, Discrete Strukturen in der Mathematik, 1995.

[4] R. Ahlswede and G. Dueck. Identification in the presence of feedback - a discovery of new capacity formulas. *IEEE Transactions on Information Theory*, Vol. 35(No. 1), January 1989.

[5] R. Ahlswede and G. Dueck. Identification via channels. *IEEE Transactions on Information Theory*. Vol. 35(No. 1), January 1989.

[6] R. Ahlswede and B. Verboven. On identification via multiway channels with feedback. *IEEE Transactions on Information Theory*, Vol. 37(No. 5), September 1991.

[7] I. Csiszar and J. Korner. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1981.

[8] I. Csiszar and P. Narayan. The capacity of the arbitrarily varying channel revisited: Positivity constraints. *IEEE Transactions on Information Theory*, 34:181–193, March 1988.

[9] L.R. Ford and D.R. Fulkerson. *Flows in Networks*. Princeton University Press, 1962.

[10] D.R. Fulkerson. Packing rooted directed cuts in a weighted directed graph. *Mathematical Programming*, 6:1–13, 1974.

[11] R.L. Graham, M. Grotschel, and L. Lovasz. *Handbook of Combinatorics*. MIT Press, 1995.

[12] L. Lovász. Communication complexity: A survey. In B.H. Korte et al., editors, *Paths, Flows and VLSI layout*. Springer-Verlag, 1990.

[13] U.M. Maurer. Perfect cryptographic security from partially independent channels. *Proc. of the 23rd Annual ACM Symposium on the Theory of Computing*, 1991.

[14] U.M. Maurer. Secret key agreement by public discussion from common information. *IEEE Transactions on Information Theory*, Vol. 39(No. 3), May 1993.

[15] A. Orlitsky and A. El Gamal. Communication complexity. In Y. Abu-Mostafa, editor, *Complexity in Information Theory*. Springer-Verlag, 1988.

[16] A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley, 1986.

[17] S. Venkatesan and V. Anantharam. The common randomness capacity of a pair of independent discrete memoryless channels. Technical Report UCB/ERL M95/85, Electronics Research Laboratory, Univ. of California, Berkeley, September 1995.