Dualities Between Entropy Functions and Network Codes

Terence Chan[1] and Alex Grant

Institute for Telecommunications Research

University of South Australia, Australia

{terence.chan, alex.grant}@unisa.edu.au

Abstract

Characterization of the set of entropy functions Γ^* is an important open problem in information theory. The region Γ^* is central to the theory of information inequalities, and as such could be regarded as a key to the basic laws of information theory. Characterization of Γ^* has several important consequences. In probability theory, it would provide a solution for the implication problem of conditional independence. In communications networks, the capacity region of multi-source network coding is given in terms of Γ^*. More broadly, determination of Γ^* would have an impact on converse theorems for multi-terminal problems in information theory. This paper provides several new dualities between entropy functions and network codes. Given a function $g \geq 0$ defined on all proper subsets of N random variables, we provide a construction for a network multicast problem which is "solvable" if and only if g is the entropy function of a set of quasi-uniform random variables. The underlying network topology is fixed and the multicast problem depends on g only through link capacities and source rates. A corresponding duality is developed for linear networks codes, where the constructed multicast problem is linearly solvable if and only if g is linear group characterizable. Relaxing the requirement that the domain of g be subsets of random variables, we obtain a similar duality between polymatroids and the linear programming bound. These duality results provide an alternative proof of the insufficiency of linear (and abelian) network codes, and demonstrate the utility of non-Shannon inequalities to tighten outer bounds on network coding capacity regions.

[1]Terence Chan is also with the Department of Computer Science, University of Regina.

I. INTRODUCTION

Information inequalities are one of the central tools of information theory. An information inequality is a relation between information measures such as entropy and mutual information that holds regardless of the specific choice of joint probability distribution on the underlying random variables, see [1, Chapters 12–14]. Converse proofs involving chains of information inequalities are ubiquitous in the literature, extending back to Shannon. It is somewhat frustrating therefore, that a characterization of the complete set of information inequalities is lacking. Until the appearance of the Zhang-Yeung inequality [2], the only known inequalities were the so-called Shannon, or basic inequalities, being consequences of the non-negativity of conditional mutual information (which is a special case of non-negativity of information divergence). Starting with [3], large classes of conditional non-Shannon inequalities (e.g. contingent on imposition of certain Markov constraints) have been found [4]–[7]. A countably infinite class of unconstrained inequalities was reported in [8], indexed by the number of random variables N involved (one inequality for each N). More recently, additional unconstrained non-Shannon inequalities have been found [9]. Another countably infinite class of unconditional inequalities was recently found in [10]. This class differs from [8], in that a countably infinite number of inequalities were found for any fixed number of $N \geq 4$ random variables. As we shall see later, this result has profound implications.

An intimately related concept is the set of entropy functions Γ^*. Let $\mathcal{H}[\mathcal{L}]$ be a subset of a 2^N dimensional euclidean space. Each coordinate of this space will be indexed by a subset of a set $\mathcal{L}$ with N elements. Points $h \in \mathcal{H}[\mathcal{L}]$ can be regarded as functions, mapping from the set of all subsets of $\mathcal{L}$ onto $\mathbb{R}$ with $h(\emptyset) = 0$. Points in $\mathcal{H}[\mathcal{L}]$ belong to Γ^* if they correspond to a consistent choice of joint entropies for a set $\mathcal{L} = \{X_1, X_2, \ldots, X_N\}$ of N random variables. Members of Γ^* are called *entropic*, and members of the closure of Γ^*, denoted by $\bar{\Gamma}^*$, are called *almost entropic*.

Characterization of $\bar{\Gamma}^*$ is equivalent to determination of the set of all possible information inequalities [1, Section 12.3]. This characterization is lacking for $N > 3$. In contrast, we do know the set $\Gamma \supset \Gamma^*$ corresponding to the basic inequalities. This set contains some functions that obey the basic inequalities, but are not entropy functions and do not correspond to any joint distribution on N random variables. The basic inequalities are equivalent to the polymatroid

axioms, and hence Γ is simply the set of polymatroids, implying a polyhedral structure.

Characterization of Γ^* is an important open problem. It gives bounds for source coding problems [11]. As shown in [1], it would resolve the implication problem of conditional independence (determination of all additional conditional independence relations implied by a given set of conditional independence relationships). In other fields, information inequalities are also closely linked to group theory [12] and the theory of Kolmogorov complexity [13], [14]. The focus in this paper is however on the link between entropy functions and the capacity region of multi-source network coding.

The prevailing approach to data transport in communications networks is based on routing, in which intermediate nodes duplicate and forward packets towards their final destination. Although such a store-and-forward scheme is simple to implement, it does not guarantee efficient utilization of available transmission capacity. The network coding approach introduced in [15], [16] generalizes routing by allowing intermediate nodes to forward packets that are coded combinations of all received data packets. This seemingly simple change in approach yields many benefits. Not only can network coding increase throughput in multicast scenarios, it can also provide robustness to link failure [17], wiretap security [18], and minimal transmission cost [19]. Naturally, these advantages are obtained at the expense of increased node complexity.

One fundamental problem in network coding is to understand the capacity region and the classes of codes that achieve capacity. In the single session multicast scenario, the problem is well understood. In particular, the capacity region is characterized by max-flow/min-cut bounds and linear network codes are sufficient to achieve maximal throughput [16], [20].

Significant practical and theoretical complications arise in more general multicast scenarios, involving more than one session. It was recently proved that linear network codes are not sufficient for the multi-source problem [20]. Furthermore, the network coding capacity region is unknown. In fact, there are only a few tools in the literature for study the capacity region.

One powerful theoretical tool bounds the capacity region by the intersection of a set of hyperplanes (specified by the network topology and connection requirement) and the set of entropy functions Γ^* (inner bound), or its closure $\bar{\Gamma}^*$ (outer bound) [1], [21], [22]. Recently, these bounds have been tightened to obtain an exact expression for the capacity region, again in terms of Γ^* [23]. Unfortunately, the capacity region, or even the bounds cannot be computed in practice, due to the lack of an explicit characterization of the set of entropy functions for more

than three random variables. One way to resolve this difficulty is via relaxation of the bound, replacing the set of entropy functions with the set of polymatroids Γ. The resulting "linear programming" bound can be quite loose. Recent work [24] based on matroid theory showed that application of the Zhang-Yeung inequality [2] yields a tighter bound for the capacity region (by obtaining a better outer bound for the set of entropy functions).

The main results of this paper are new dualities between non-negative functions $g \in \mathcal{H}[\mathcal{L}]$ and network codes. These duality results are based on the construction of a special network multicast problem from functions g. The underlying network topology is fixed and the multicast problem depends on g only through the assignment of link capacities and source rates.

Three main kinds of duality are considered, corresponding to different restrictions on g and different kinds of network codes. First, we show in Theorem 1 that the constructed multicast problem is solvable (i.e. the constructed source rates and link capacities are in the capacity region) if and only if g is the entropy function of a set of quasi-uniform random variables. This duality is extended in Theorem 2 to show that the multicast problem is asymptotically solvable with ϵ error if and only if h is almost entropic.

The second duality restricts attention to linear network codes. We show that the multicast problem is linearly solvable if and only if g is linear group characterizable (i.e. g is an entropy function for random variables generated by vector spaces). A corresponding limiting form of this duality is also provided.

Finally, by relaxing the requirement that the domain of g be subsets of random variables, we obtain a duality between polymatroids and the linear programming bound.

These duality results yield several immediate implications. In particular, we provide an alternative proof to [20], [24] for the insufficiency of linear (and abelian) network codes, and demonstrate the utility of non-Shannon inequalities to tighten outer bounds on network coding capacity regions.

The paper is organized in the following way. Section II introduces some fundamentals of network coding. Section II-A focuses on network codes with algebraic structure, and random variables generated by groups with a variety of algebraic structures. We establish a relation between linear network codes and random variables generated by vector spaces and generalize this idea to define the concept of a group network code. A central theme of the paper is the trade-off between source rate and link capacity using network coding, i.e. determination of

the network coding capacity region. Section II-B introduces the definitions for admissibility and achievability in the network coding context. Section III introduces the concept of pseudo-variables, which generalize random variables in such a way that allows a notational unification of the linear programming bound with that of [21].

Section IV proves the duality results, Theorems 1 – 5. These results rely on the construction in Section IV-A of a special network and multicast problem from a function g. Section IV-B gives the duality between entropic functions and solvable multicast problems. Section IV-C provides the corresponding duality for linearly solvable multicast problems. These duality results are extended in Section IV-D to give a similar link between polymatroids and the linear programming bound, i.e. a function g is a polymatroid if and only if the constructed source rates and link capacities satisfy the bound. This result relies heavily on the notion of pseudo-variables introduced in Section III, and in particular on extension and adhesion of sets of pseudo-variables, discussed in Appendix I. Finally, in Section IV-E we give a one-way relation between the LP bound for linear codes, and polymatroids which also satisfy the Ingleton inequality.

Section V explores the implications of our results, which include the insufficiency of linear or even (abelian) group network codes, and the necessity for non-Shannon inequalities for determination of the network coding capacity region.

Notation: For a set $\mathcal{A}$, the power set $2^{\mathcal{A}} = \{\mathcal{B} : \mathcal{B} \subseteq \mathcal{A}\}$ denotes the set of all subsets of $\mathcal{A}$. Given a set of $|\mathcal{A}|$ variables $\{X_a, a \in \mathcal{A}\}$, and a subset $\mathcal{C} \subseteq \mathcal{A}$, the subscript $X_{\mathcal{C}}$ shall mean $\{X_c : c \in \mathcal{C}\}$. In contrast, the notation $Y_{[\mathcal{B}]}$ will be used to index a single variable out of a set of $2^{|\mathcal{A}|}$ variables $\{Y_{[\mathcal{B}]} : \mathcal{B} \in 2^{\mathcal{A}}\}$. Other notation will be introduced as necessary throughout the paper.

II. Networks, Codes and Capacity

A directed acyclic graph $\mathcal{G} = (\mathcal{P}, \mathcal{E})$ is commonly used as a simplified model of a communication network. The nodes $u \in \mathcal{P}$ and directed edges $e = (\mathrm{tail}(e), \mathrm{head}(e)) \in \mathcal{E}$ respectively model communication nodes and directed, error-free point-to-point communication links. The terms graph and network will be used interchangeably. For edges $e, f \in \mathcal{E}$, write $f \to e$ as shorthand for $\mathrm{head}(f) = \mathrm{tail}(e)$. Similarly, for an edge $f \in \mathcal{E}$ and a node $u \in \mathcal{P}$, the notations $f \to u$ and $u \to f$ respectively denote $\mathrm{head}(f) = u$ and $\mathrm{tail}(f) = u$. So far we have only specified the basic network topology. The communication problem is specified via imposition

of a connection requirement.

Definition 1 (Connection Requirement): For any network $\mathcal{G}$, a *connection requirement* $M = (\mathcal{S}, O, \mathcal{D})$ is specified by three components representing the sessions, originating nodes and destination nodes as follows. $\mathcal{S}$ is an index set of independent multicast sessions, each of which is a collection, or stream of data packets to be multicast to a prescribed set of destination nodes. $O : \mathcal{S} \mapsto \mathcal{P}$ is a source-location mapping, where $O(s)$ is the originating node for multicast session s. $\mathcal{D} : \mathcal{S} \mapsto 2^{\mathcal{P}}$ is a receiver-location mapping, where $\mathcal{D}(s) \subseteq \mathcal{P}$ is the set of nodes requiring the data of session s.

It should be noted that there is *no specified rate requirement*. The connection requirement differs from the usual concept of multicast requirement in that it only specifies *which* nodes require data from which other nodes, and not any particular desired information rate.

Given a connection requirement M, the goal of a network code is to efficiently multicast data for session s originating at node $O(s)$ to all receivers in the set $\mathcal{D}(s)$. Nodes are assumed to have sufficient computing power to implement any desired network coding scheme.

Let $\mathcal{F} = \mathcal{S} \cup \mathcal{E}$. For a network $\mathcal{G}$ and connection requirement M, a network code is specified by a set of source and edge alphabets $\{\mathcal{U}_f, f \in \mathcal{F}\}$ and a set of local coding functions

$$\Phi \triangleq \left\{ \phi_e : \prod_{f \in \mathcal{F}: f \to e} \mathcal{U}_f \mapsto \mathcal{U}_e : e \in \mathcal{E} \right\}$$

where for ease of notation, $s \to e$ indicates $O(s) \to e$, and $f \in \mathcal{F} : f \to e$ means any source or edge incident to edge e.

Data transmission takes place as follows. Session $s \in \mathcal{S}$ generates a source symbol U_s, which is assumed to be independent of other sessions and uniformly distributed over $\mathcal{U}_s$. The link symbol transmitted along $e \in \mathcal{E}$ is $U_e = \phi_e(U_f : f \in \mathcal{F}, f \to e)$. In other words, the symbol transmitted along an outgoing link of a node is a function of the available sources and incident link symbols.

We will refer to a network code by Φ, with the set of alphabets $\{\mathcal{U}_f, f \in \mathcal{F}\}$ implicitly defined. Since the input and link symbols are random variables, we can also refer to the code by the set of random variables $U_{\mathcal{F}}$, where their joint distribution is implied by Φ. Clearly,

$$H(U_{\mathcal{S}}) = \sum_{s \in \mathcal{S}} H(U_s) = \sum_{s \in \mathcal{S}} \log |\mathcal{U}_s| \quad \text{and}$$

$$H(U_e) \leq \log |\mathcal{U}_e|.$$

For a given network code Φ designed for a network $\mathcal{G}$ with connection requirement M, the error probability $P_e(\Phi)$ is defined as the probability that at least one receiver $d \in \bigcup_{s\in\mathcal{S}} \mathcal{D}(s)$ fails to correctly reconstruct one or more of its requested source messages $\{U_s : \mathcal{D}(s) = d\}$. A *zero-error* network code is one for which $P_e(\Phi) = 0$, implying that the source symbols U_s are deterministic functions of the corresponding receiver-incident edge symbols.

A. Algebraic network codes

The above formulation imposes no restriction on the choice of alphabets and local coding functions. However, in practice, it may be preferable to impose algebraic structure to reduce the complexity of encoding and decoding. The overwhelming majority of codes studied for the point-to-point channel are in fact linear, and linear codes are also of particular interest in the network coding context.

Definition 2 (Linear Network Code): A network code Φ is *linear* over a finite field $\mathbb{F}_q$ if all source and link alphabets $\mathcal{U}_f$ are vector spaces over some finite field $\mathbb{F}_q$, and all the local encoding functions ϕ_e are linear.

Clearly, for a linear network code, each source alphabet is a vector subspace and the symbol transmitted along link $e \in \mathcal{E}$ is a linear function of the inputs $U_{\mathcal{S}}$. As will be stated in Proposition 2, the set of all the kernels of these linear functions associated with all the links can be used to "construct" the set of source and link random variables defining the network code. To understand this relationship, we first review the construction of random variables from a finite group and its groups [12].

Definition 3 (Construction of random variables from subgroups): Suppose that U is a random variable uniformly distributed over a group G. For any subgroup G_i, the set of left cosets of G_i forms a partition in G. Let $\mathcal{U}_i$ be an index set of the cosets of G_i in G. We can define a random variable U_i as a function of U such that U_i is the index of the coset of G_i that contains U, or simply that U_i is the coset of G_i that contains U. The resulting random variable is said to be *constructed* from G and G_i.

Definition 4 (Group characterizable random variables): A set of random variables $\{U_1, \ldots, U_N\}$

(and its induced entropy function) is called *group characterizable* if it is equivalent[1] to a set of random variables constructed from a finite group G and its subgroups $G_1, \cdots, G_N$.

If G is abelian, then $\{U_1, \cdots, U_N\}$ (and the entropy function) is called *abelian group characterizable*. If in addition G and $G_1, \cdots, G_N$ are all vector spaces, then the set of random variables (and the entropy function) is called *linear group characterizable*.

Denote the set of group characterizable entropy functions by $\Gamma^*_G \subset \Gamma^*$, the set of abelian group characterizable functions by Γ^*_{ab} and the set of linear (with respect to a finite field $\mathbb{F}_q$) group characterizable functions by $\Gamma^*_{L(q)}$. Then, it is clear that $\Gamma^*_{L(q)} \subset \Gamma^*_{\text{ab}} \subset \Gamma^*_G \subset \Gamma^*$.

Random variables constructed from subgroups have been shown to have many interesting properties. For example, suppose $\{U_1, \cdots, U_N\}$ is constructed from a finite group G and its subgroups $G_1, \cdots, G_N$. Then $H(U_\alpha) = \log |G|/|\bigcap_{i\in\alpha} G_i|$ for any non-empty subset $\alpha \subseteq \mathcal{N} \triangleq \{1, 2, \ldots, N\}$ [12]. It was also proved in [12] that a linear information inequality is valid if and only it is satisfied by all group characterizable random variables. Thus group characterizable random variables have an interesting role to play in the proof of information inequalities,

Before describing some additional properties of group characterizable random variables, we will need the concept of quasi-uniform random variables.

Definition 5 (Quasi-uniform random variable): A discrete finite random variable U defined on a sample space $\mathcal{U}$ is called *quasi-uniform* if and only if it is uniformly distributed over its support $\Omega(U)$. In other words, the probability distribution of U has the following form:

$$\Pr(U = u) = \begin{cases} 1/|\Omega(U)| & \text{if } u \in \Omega(U) \\ 0 & \text{otherwise} \end{cases}$$

Hence, $H(U) = \log |\Omega(U)|$.

Similarly, a set of random variables $U_1, U_2, \ldots, U_N$ (and its induced entropy function) is called quasi-uniform if and only if every subset of random variables $U_\alpha, \alpha \subseteq \{1, 2, \ldots, N\}$ is quasi-uniform, i.e. $H(U_\alpha) = \log |\Omega(U_\alpha)|$.

[1]Two sets of random variables $\{U_1, \cdots, U_N\}$ and $\{V_1, \cdots, V_N\}$ with probability distributions P_U and P_V respectively are "equivalent" if for each $i = 1, \cdots, N$, there is a one-to-one mapping τ_i from the support of U_i to the support of V_i such that $P_U(U_1, \cdots, U_N) = P_V(\tau_1(U_1), \cdots, \tau_N(U_N))$. In this paper, two sets of equivalent random variables will be regarded as identical.

Lemma 1 ([12], [25]): Random variables induced by groups and subgroups are quasi-uniform. Hence

$$\Gamma^*_{L(q)} \subset \Gamma^*_{\rm ab} \subset \Gamma^*_G \subset \Gamma^*_Q \subset \Gamma^*$$

where Γ^*_Q is the set of all quasi-uniform entropy functions.

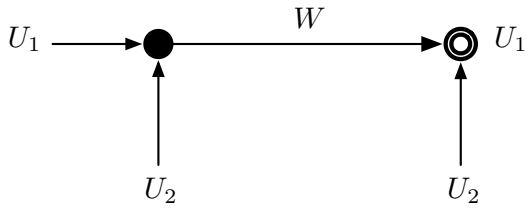


Fig. 1. The side-information network.

Lemma 2: With reference to Figure 1, consider a simple coding problem in which there is a transmitter (indicated by an open circle) and a receiver (indicated by a double circle) connected by a noiseless point-to-point link. A source U_1 is available at the transmitter, while correlated side-information U_2 is available at both transmitter and receiver. The coding problem is to encode U_1, U_2 into a symbol W defined on the sample space $\mathcal{W}$ such that U_1 can be constructed perfectly at receiver from W and U_2.

Suppose that $\{U_1, U_2\}$ is quasi-uniform. Then one can have a zero-error code with rate $\log|\Omega(U_1,U_2)|/|\Omega(U_2)| = H(U_1|U_2)$, where the code rate is defined as $\log|\mathcal{W}|$.

Proof: Since U_2 is available to both transmitter and receiver, U_1 can be reconstructed perfectly if the transmitter only sends the index of u_1 in the set $\{u_1 : (u_1,u_2) \in \Omega(U_1,U_2)\}$ for any given $u_2 \in \Omega(U_2)$. By the quasi-uniformity of $\{U_1, U_2\}$, the cardinality of the set $\{u_1 : (u_1,u_2) \in \Omega(U_1,U_2)\}$ is $|\Omega(U_1,U_2)|/|\Omega(U_2)|$ for any $u_2 \in \Omega(U_2)$. Hence, one can easily construct a zero-error code at a rate of $\log|\Omega(U_1,U_2)|/|\Omega(U_2)| = H(U_1|U_2)$ that solves the coding problem. ■

If the group and subgroups in question possess additional algebraic properties, the induced random variables may also satisfy certain additional properties. One interesting example, proved in [26], [27] is given as follows.

Proposition 1 (Ingleton's inequality): Suppose that the set of random variables $\{U_1, \ldots, U_N\}$ is abelian group characterizable. Let $\{V_1, V_2, V_3, V_4\} \subseteq \{U_1, \ldots, U_N\}$. Then

$$g(1,2)+g(1,3)+g(1,4)+g(2,3)+g(2,4) \geq g(1)+g(2)+g(3,4)+g(1,2,3)+g(1,2,4) \quad (1)$$

where $g(\alpha) \triangleq H(V_\alpha)$.

Proposition 2: Suppose that a set of random variables $\{U_f, f \in \mathcal{F}\}$ defines a zero-error linear network code. Then $\{U_f, f \in \mathcal{F}\}$ is linear group characterizable.

Proof: [Proof Sketch] Suppose that $\Phi = \{\phi_e, e \in \mathcal{E}\}$ is a zero-error linear network code with inputs $U_s \in \mathcal{U}_s$ for $s \in \mathcal{S}$ and link symbols $U_e \in \mathcal{U}_e$ for $e \in \mathcal{E}$. We will now construct a linear group characterization for the set of source/link random variables induced by Φ. Let

1) G be the vector space formed by the Cartesian product of $\prod_{s\in\mathcal{S}} \mathcal{U}_s$;
2) $\psi_s : G \mapsto \mathcal{U}_s$ be a linear function such that $\psi_s(U_s : s \in \mathcal{S}) = U_s$;
3) $\psi_e : G \mapsto \mathcal{U}_e$ be a linear function such that $U_e = \psi_e(U_s : s \in \mathcal{S})$; (*This is possible as all local coding functions ϕ_e are linear*)
4) G_f is the kernel of ψ_f, denoted by $\ker(\psi_f)$, for $f \in \mathcal{S} \cup \mathcal{E}$. Hence, G_f is a subspace of G.

Then it is straightforward to show that for any $(U_s : s \in \mathcal{S})$ and $f \in \mathcal{F}$, the value of $\psi_f(U_s : s \in \mathcal{S})$ can be uniquely determined from the index of the coset of G_f that contains $(U_s : s \in \mathcal{S})$ and vice versa. In other words, the link random variable U_f is equivalent to the one induced by the subspace G_f. ■

A natural interpretation of Proposition 2 is that linear network codes are those codes whose induced source and link random variables can be characterized by a vector space and its subspaces. Developing this line of thought more generally, we make the following definition.

Definition 6 (Group network code): A *group network code* is a network code $\{U_f, f \in \mathcal{F}\}$ whose source and link random variables are induced by a finite group G with subgroups $G_f, f \in \mathcal{F}$. Furthermore, a group network code is called abelian if G is abelian.

For a group network code $\Phi = \{U_f, f \in \mathcal{F}\}$, encoding at intermediate nodes works as follows. Suppose that the source and link random variables $\{U_f, f \in \mathcal{F}\}$ are characterized by a finite group and its subgroups G_f for $f \in \mathcal{F}$. For any $f \in \mathcal{F}$, let $\mathcal{U}_f$ be the index set for the set of left cosets of G_f in G. Each edge e receives symbols $\{U_f : f \to e\}$, which are indexes of cosets G_f in G. The symbol U_e to be transmitted along edge e is the index of the left coset G_e that contains the intersection of the cosets of G_f indexed by $\{U_f : f \to e\}$.

In fact, in the special case when the group and all its subgroups are vector spaces, we can index the coset of G_e as elements in a vector space such that U_e is indeed a linear function of $\{U_f : f \to e\}$.

Example 1: An R-module generalizes the concept of vector space, where the scalars are a members of a ring R, instead of a field. It consists of an abelian group K, and an operation of left multiplication by each element in R. In particular, for all $r, s \in R$ and $g, h \in K$,

$$rg \in K$$

$$(rs)g = r(sg)$$

$$(r+s)g = rg + sg$$

$$r(g+h) = rg + rh$$

$$0g = 0.$$

$R-module$ codes have been proposed as generalizations of linear network codes [20]. Messages to be transmitted along edges are elements in K. The only difference is that local encoding functions must be of the form

$$U_e = \sum_{f \in \mathcal{F}: f \to e} r_{fe} U_f$$

where $r_{fe} \in R$. As such, there exists elements $M_{es} \in R$ such that

$$U_e = \sum_{s \in \mathcal{S}} M_{es} U_s.$$

Let G be the $|\mathcal{S}|$-fold Cartesian product of K. For all $e \in \mathcal{E}$ and $s \in \mathcal{S}$, let

$$G_e = \left\{ (U_s \in K : s \in \mathcal{S}) : \sum_{s \in \mathcal{S}} M_{es} U_s = 0 \right\}$$

$$G_s = \left\{ (U_s \in K : s \in \mathcal{S}) : U_s = 0 \right\}.$$

Then it is straightforward to show that G_f is an abelian subgroup of G for $f \in \mathcal{F}$ and that the source and link random variables induced by the $R - module$ code is characterized by the subgroup G and its subgroups G_f, $f \in \mathcal{F}$.

B. The source rate-link capacity tradeoff

So far, we have only considered networks, and codes designed to meet particular connection requirements. Typically however, each link has limited capacity, and a fundamental design consideration is the tradeoff between supportable network throughput and link capacities. Of primary interest is determination of the minimal link capacities $\omega \triangleq (\omega_e : e \in \mathcal{E})$ required to transmit sources over a network at given rates $\lambda \triangleq (\lambda_s : s \in \mathcal{S})$ such that all receivers can reconstruct their desired messages with no, or arbitrarily small probability of error.

Definition 7 (Admissible rate-capacity tuple): Given a network $\mathcal{G} = (\mathcal{P}, \mathcal{E})$ and a connection requirement M, a rate-capacity tuple (λ, ω) is *admissible* if there exists a zero-error network code $\Phi = \{U_f, f \in \mathcal{S} \cup \mathcal{E}\}$, such that

$$H(U_e) \leq \log |\mathcal{U}_e| \leq \omega_e, \quad \forall e \in \mathcal{E},$$
$$H(U_s) = \log |\mathcal{U}_s| \geq \lambda_s, \quad \forall s \in \mathcal{S},$$

where U_e is the message symbol transmitted along link e and U_s is the input symbol generated at source s.

Coding over long block of symbols often improves the rate of point-to-point codes. Similarly, increased efficiency may be expected for network codes operating over a long block of source symbols. Therefore, we also consider the asymptotic tradeoff between source rates and link capacities.

Definition 8 (Asymptotically admissible): A rate-capacity tuple (λ, ω) is *asymptotically admissible* if there exists a sequence of zero-error network codes $\Phi^{(n)} = \{U_f^{(n)}, f \in \mathcal{S} \cup \mathcal{E}\}$ and positive normalizing constants $r(n)$ such that

$$\lim_{n\to\infty} \frac{1}{r(n)} H\left(U_e^{(n)}\right) \leq \lim_{n\to\infty} \frac{1}{r(n)} \log |\mathcal{U}_e^{(n)}| \leq \omega_e, \quad \forall e \in \mathcal{E},$$
$$\lim_{n\to\infty} \frac{1}{r(n)} H\left(U_s^{(n)}\right) = \lim_{n\to\infty} \frac{1}{r(n)} \log |\mathcal{U}_s^{(n)}| \geq \lambda_s, \quad \forall s \in \mathcal{S}.$$

The above two definitions consider zero-error network codes. Relaxing the requirement to allow arbitrarily small error probability prompts the following definition.

Definition 9 (Achievable rate-capacity tuple): A rate-capacity tuple (λ, ω) is *achievable* if there exists a sequence of network codes $\Phi^{(n)} \triangleq \{U_f^{(n)}, f \in \mathcal{S} \cup \mathcal{E}\}$ and positive normalizing

constants $r(n)$ such that

$$\lim_{n\to\infty}\frac{1}{r(n)}H\left(U_e^{(n)}\right)\leq\lim_{n\to\infty}\frac{1}{r(n)}\log|\mathcal{U}_e^{(n)}|\leq\omega_e,\quad\forall e\in\mathcal{E},$$
$$\lim_{n\to\infty}\frac{1}{r(n)}H\left(U_s^{(n)}\right)=\lim_{n\to\infty}\frac{1}{r(n)}\log|\mathcal{U}_s^{(n)}|\geq\lambda_s,\quad\forall s\in\mathcal{S},$$
$$\lim_{n\to\infty}P_e\left(\Phi^{(n)}\right)=0.$$

Assuming that the underlying network and connection requirement are known implicitly, the set of admissible, asymptotically admissible and achievable rate-capacity tuples will be denoted $\Upsilon^0,\Upsilon^\infty$ and Υ^ϵ respectively.

The preceding definitions place no restriction on the class of network codes under consideration. However, if a rate-capacity tuple is admissible/asymptotically admissible/achievable using a network code in a specific class $\mathcal{C}$ (e.g. the class of linear network codes), then that rate-capacity tuple is said to be admissible/asymptotically admissible/achievable by network codes in $\mathcal{C}$, and the corresponding sets are denoted $\Upsilon^0_\mathcal{C},\Upsilon^\infty_\mathcal{C}$ and $\Upsilon^\epsilon_\mathcal{C}$.

In this paper, we are interested in two special classes of network codes, (i) linear network codes (with respect to an underlying finite field $\mathbb{F}_q$) and (ii) abelian group network codes. The sets of admissible/asymptotically admissible/achievable rate-capacity tuples by linear network codes are respectively denoted by $\Upsilon^0_{L(q)},\Upsilon^\infty_{L(q)}$ and $\Upsilon^\epsilon_{L(q)}$. Similarly, the set of admissible/asymptotically admissible/achievable rate-capacity tuples by abelian group network codes are respectively denoted by $\Upsilon^0_{\mathrm{ab}},\Upsilon^\infty_{\mathrm{ab}}$ and $\Upsilon^\epsilon_{\mathrm{ab}}$.

Discovering the hidden structure of these sets of rate-capacity tuples is the key to understanding the tradeoff between source rates and edge capacities. In the following, we list some basic structural properties of $\Upsilon^0_\mathcal{C},\Upsilon^\infty_\mathcal{C}$ and $\Upsilon^\epsilon_\mathcal{C}$ when $\mathcal{C}$ is either the class of all network codes, linear network codes or abelian group network codes.

P1) The sets $\Upsilon^0_\mathcal{C},\Upsilon^\infty_\mathcal{C}$ and $\Upsilon^\epsilon_\mathcal{C}$ are closed under addition. In other words, if tuples (λ,ω) and (λ',ω') are in $\Upsilon^0_\mathcal{C}$ (or respectively in $\Upsilon^\infty_\mathcal{C}$ and $\Upsilon^\epsilon_\mathcal{C}$), then the element-wise addition of the two tuples will still be in the same set.

P2) $\Upsilon^\infty_\mathcal{C}$ and $\Upsilon^\epsilon_\mathcal{C}$ are closed convex cones, and $\overline{\mathrm{con}}(\Upsilon^0_\mathcal{C})=\Upsilon^\infty_\mathcal{C}$ where $\overline{\mathrm{con}}(\Upsilon^0_\mathcal{C})$ is the minimal closed convex cone containing $\Upsilon^0_\mathcal{C}$.

P3) Admissibility implies asymptotic admissibility which further implies achievability, $\Upsilon^0_\mathcal{C}\subseteq\Upsilon^\infty_\mathcal{C}\subseteq\Upsilon^\epsilon_\mathcal{C}$.

III. Pseudo-variables and bounds

The sets of admissible/achievable rate-capacity tuples are difficult to characterize explicitly. In fact, we will show later that finding these sets is at least as hard as determining the set of entropy functions Γ^*. Due to the difficulty of the problem, results on characterizing the set of achievable rate-capacity tuples are quite limited [21], [24], [28], [29]. While inner bounds and outer bounds constructed with entropic/almost entropic functions exist [1], these bounds are not computable and hence are of limited practical use. The only known computable outer bound is the Linear Programming (LP) bound, which is constructed using polymatroids [1]. The remainder of this section provides a brief review of these bounds. We use the opportunity to introduce notation (differing slightly from the original manuscripts), facilitating later discussion.

Let $\mathcal{L}$ be a nonempty finite set. Recall that $\mathcal{H}[\mathcal{L}]$ (or simply $\mathcal{H}$) is a real euclidean space which has $2^{|\mathcal{L}|}$ dimensions and coordinates indexed by the set of all subsets of $\mathcal{L}$ and that $g(\emptyset) = 0$ for all $g \in \mathcal{H}[\mathcal{L}]$. Specifically, if $g \in \mathcal{H}$, then its coordinates will be denoted by $(g(\mathcal{A}) : \mathcal{A} \subseteq \mathcal{L})$. We call $\mathcal{L}$ a ground set. Each $g \in \mathcal{H}$ can also be viewed as a real-valued function $g : 2^{\mathcal{L}} \mapsto \mathbb{R}$ defined on each subset of $\mathcal{L}$.

Definition 10 (Polymatroid): A function $g \in \mathcal{H}[\mathcal{L}]$ is a *polymatroid* if it satisfies

$$g(\emptyset) = 0 \tag{2}$$

$$g(\mathcal{A}) \geq g(\mathcal{B}), \quad \text{if } \mathcal{B} \subseteq \mathcal{A} \qquad \text{non-decreasing} \tag{3}$$

$$g(\mathcal{A}) + g(\mathcal{B}) \geq g(\mathcal{A} \cup \mathcal{B}) + g(\mathcal{A} \cap \mathcal{B}) \qquad \text{submodular} \tag{4}$$

Note (2) and (3) imply non-negativity of a polymatroid. Let $\mathcal{L}$ be a set of discrete random variables with finite entropies. Note that $\mathcal{L}$ contains random variables rather than indexes for a set of random variables. This induces a function $g \in \mathcal{H}$ where $g(\mathcal{A})$ is the joint entropy of the set of random variables $\emptyset \neq \mathcal{A} \subseteq \mathcal{L}$. Functions so-defined will be called *entropy functions*.

It is well-known that entropy functions are polymatroids over the ground set $\mathcal{L}$. In fact, in the context of entropy functions, the polymatroid axioms are completely equivalent to the basic information inequalities (i.e. non-negativity of conditional mutual information) [1, p. 297]. It is by now well-known however that there are other information inequalities that are not implied by the polymatroid axioms. The set of entropy functions is denoted Γ^*, while the set of polymatroids is Γ.

While an entropy function takes a subset of random variables as argument, a polymatroid g

more generally takes a subset of the ground set $\mathcal{L}$ as argument, where the elements of $\mathcal{L}$ may or may not be random variables. For simplicity, we shall call the elements of the ground set of a polymatroid *pseudo-variables*. They differ from random variables in that they do not necessarily take values, and there may be no associated joint probability distribution function.

It must be emphasized that pseudo-variables are only defined in the context of a polymatroid g defined on the ground set $\mathcal{L}$. The elements of $\mathcal{L}$ are not pseudo-variables by themselves in the absence of an associated polymatroid.

Carrying these ideas further, we will call $g(\mathcal{A})$ the *pseudo-entropy* of the set of pseudo-variables $\mathcal{A}$, and g is a *pseudo-entropy function*. Treating pseudo-variables as a set of basic objects associated with a polymatroid yields notational simplification. For example, random variables are simply pseudo-variables possessing a probability distribution such that their pseudo-entropy function is the same as the entropy function. As such, we extend the use of $H(\mathcal{A})$ to refer to the pseudo-entropy of a set of pseudo-variables $\mathcal{A}$.

Definition 11 (Entropic function): A set of pseudo-variables (and its associated pseudo-entropy function) is called *entropic* if its pseudo-entropy function is the same as an entropy function of a set of random variables.

Similarly, a set of pseudo-variables (and their pseudo-entropy function) is called *linear group characterizable* if its pseudo-entropy function is the same as an entropy function of a set of linear group characterizable random variables.

The following two definitions generalize concepts of functional dependence and independence to pseudo-variables.

Definition 12 (Functional dependence): Let $\mathcal{L}$ be a set of pseudo-variables. A pseudo-variable $X \in \mathcal{L}$ is said to be a *function* of a set of pseudo-variables $\mathcal{A} \subseteq \mathcal{L}$ if $H(\{X\} \cup \mathcal{A}) = H(\mathcal{A})$. This relation will be denoted by $H(X|\mathcal{A}) = 0$.

Definition 13 (Independence): Two subsets of pseudo-variables $\mathcal{A}$ and $\mathcal{B}$ are called *independent* if $H(\mathcal{A} \cup \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B})$, and this relationship will be denoted by $\mathcal{A} \perp \mathcal{B}$. Similarly, if $H(\bigcup_{j\in\mathcal{J}} \mathcal{A}_j) = \sum_{j\in\mathcal{J}} H(\mathcal{A}_j)$, write $\perp_{j\in\mathcal{J}} \mathcal{A}_j$.

Clearly, these definitions are consistent with the usual ones used for random variables. The following bound re-states the linear programming bound [1, Section 15.6] in terms of pseudo-variables.

Definition 14 (LP bound): Given a network $\mathcal{G}$ and a connection requirement M, the LP bound

is the set of rate-capacity tuples (λ, ω) such that there exists a set of pseudo-variables $\{U_s : s \in \mathcal{S}, U_e : e \in \mathcal{E}\}$ satisfying the following "connection constraint":

$$\begin{aligned} H\left(U_e \mid U_f : f \to e\right) &= 0, \quad e \in \mathcal{E} \\ H\left(U_s \mid U_f : f \to u\right) &= 0, \quad u \in \mathcal{D}(s) \\ \perp_{s\in\mathcal{S}} U_s & \\ H(U_s) &\geq \lambda_s, \quad s \in \mathcal{S} \\ H(U_e) &\leq \omega_e, \quad e \in \mathcal{E}. \end{aligned} \tag{5}$$

Denote the set of rate-capacity tuples that satisfy the LP bound by Υ_{LP}. From [1] it is known that $\Upsilon_{LP} \supseteq \Upsilon^{\epsilon}$. It is interesting to notice that the use of pseudo-variables gives a notational unification of an inner bound and an outer bound given in [1] as follows:

Proposition 3 (Inner and Outer bounds): Given a network $\mathcal{G}$ and a connection requirement M, let Υ_{in} resp. Υ_{out} be the set of rate-capacity tuples (λ, ω) such that there exists a set of *entropic* resp. *almost entropic* pseudo-variables $\{U_s : s \in \mathcal{S}, U_e : e \in \mathcal{E}\}$ satisfying (5). Then $\Upsilon_{\text{in}} \subseteq \Upsilon^{\epsilon} \subseteq \Upsilon_{\text{out}} \subseteq \Upsilon_{LP}$.

Proof: The proof is straightforward by rewriting the bounds obtained in [1]. ■

Similar to the LP bound, we define the following bound for abelian group network codes (including linear network codes) as follows.

Definition 15 (LP-Ingleton bound): Given a network $\mathcal{G}$ and a connection requirement M, the LP-Ingleton bound is the set of rate-capacity tuples (λ, ω) such that there exists a set of pseudo-variables $\{U_s : s \in \mathcal{S}, U_e : e \in \mathcal{E}\}$ satisfying the Ingleton inequalities (1) and the connection constraint (5).

Proposition 4: Denote the set of rate-capacity tuples that satisfy the LP-Ingleton bound by $\Upsilon_{LP,I}$. Then $\Upsilon_{LP,I}$ contains $\Upsilon^{\epsilon}_{\text{ab}}$.

Proof: First notice that all source and link random variables of an abelian group network code must satisfy the Ingleton inequalities. The proposition then follows by using a similar argument as in [1] that proves $\Upsilon_{LP} \supseteq \Upsilon^{\epsilon}$. ■

Since the LP and LP-Ingleton bounds are defined by intersections of several linear half-spaces and hyperplanes, these bounds are polyhedral. Together with the following duality results, this implies that LP bounds are not generally tight (this is proved Section V).

IV. Entropy functions, network codes and duality

Given a network, a connection requirement and a rate-capacity tuple, the *multicast problem* is to determine whether or not the rate-capacity tuple is admissible or achievable (perhaps even restricted to codes in a particular class). In this section, we construct multicast problems from non-negative functions. This construction yields several dualities between properties of the generating function and the solubility of the multicast problem. We establish three main dualities. The first duality relates entropy functions and network codes. It can be paraphrased as follows.

A function is quasi-uniform if and only if its induced rate-capacity tuple is admissible.

This is shown in Theorem 1. Theorem 2 provides an extension which implies

A function is almost entropic if and only if its induced rate-capacity tuple is achievable.

The second duality proves similar results for linear network codes.

An entropy function is linear group characterizable if and only if its induced rate-capacity tuple is admissible by linear network codes.

This is Theorem 3. Again, Theorem 4 extends the result, relating almost linear group characterizable functions and achievable rate-capacity tuples with linear network codes.

The third duality, Theorem 5 relates polymatroids and the linear programming bound.

A function is a polymatroid if and only if its induced rate-capacity tuple satisfies the LP bound.

We also give a partial result for an extension to polymatroids that also satisfy the Ingleton inequality.

Despite their apparent simplicity, these results leads to many interesting corollaries: linear network codes (or more generally, abelian group network codes) are suboptimal, the LP bound is not tight, and in general the network coding capacity region is not a polytope. These consequences will be described in more detail in Section V.

A. Constructing multicast problems

Let $h \in \mathcal{H}[\mathcal{N}]$, be a given non-negative function over the ground set $\mathcal{N} = \{1, 2, \ldots, N\}$. The proof for the main result relies on the construction of a special network $\mathcal{G}^\dagger$, a connection requirement $M^\dagger$ and a rate-capacity tuple $\mathsf{T}(h) \triangleq (\lambda(h), \omega(h))$.

Figure 2 defines the network topology, connection requirement and edge capacities. For convenience, the network is divided into several subnetworks. To differentiate the roles of network nodes, source nodes are indicated by open circles, destination nodes are double circles, and intermediate nodes are solid circles. By construction, each node takes only one role. The label beside a source node is the input message available to that source node (this defines the source location mapping O). The label beside a receiver node indicates the desired source message to be reconstructed at that destination node (this defines the destination location mapping $\mathcal{D}$). To simplify notation, each capacitated edge is labeled with a pair of symbols denoting the edge message (and corresponding random variable), and the edge capacity. Unlabelled edges are assumed to be uncapacitated, or to have a finite but sufficiently large capacity (such as $\sum_\alpha h(\alpha)$) to losslessly forward all received messages.

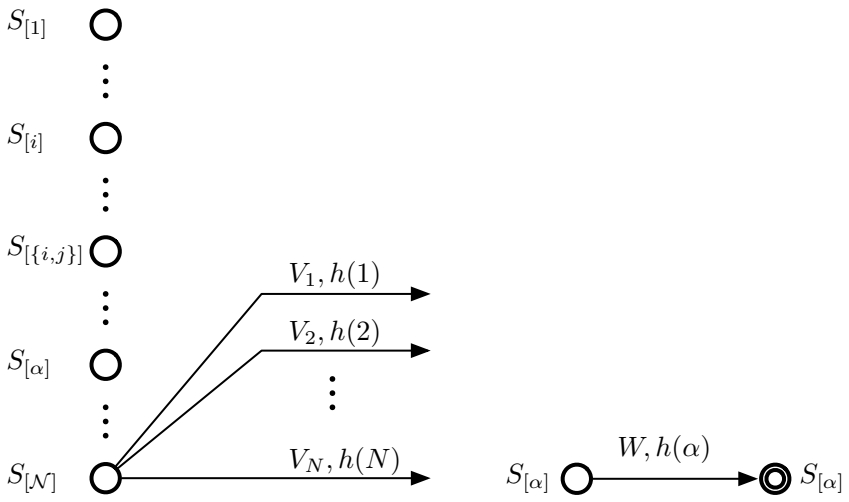
The first part of the network, shown in Figure 2(a), contains the sources. There are $2^N - 1$ independent sessions, $\mathcal{S} = \left\{S_{[\alpha]} : \emptyset \neq \alpha \subseteq 2^{\mathcal{N}}\right\}$[2]. The desired source rate associated with session α is $h(\alpha)$. Singletons $\{i\} \in 2^{\mathcal{N}}$ will be denoted without brackets, e.g. $h(i)$ and $S_{[i]}$. There are N specific edge messages that are of particular interest. Rather than naming all edge variables $U_e, e \in \mathcal{E}$, we label these N particular edge variables V_j, $j = 1, \ldots, N$. Remaining edge variables will be labelled with generic symbols W, W', W'', W^* and W^{**}. Source $S_{[\mathcal{N}]}$ generates the network coded messages $V_1, V_2, \ldots, V_N$ which are duplicated as required and forwarded to the rest of the network. The remaining part of the network is divided into subnetworks of three types, shown in Figures 2(b), 2(c) and 2(d).

With reference to Figure 2(b), type 0 subnetworks connect a single source to one receiver. There are $2^N - 1$ type 0 subnetworks, indexed by the choice of $\emptyset \neq \alpha \in 2^{\mathcal{N}}$.

Referring to Figure 2(c), there are $2^N - 1$ type 1 subnetworks, one for each nonempty $\alpha \in 2^{\mathcal{N}}$. These subnetworks introduce an edge of capacity $h(\mathcal{N}) - h(\alpha)$ between source $S_{[\mathcal{N}]}$ and a sink requiring $S_{[\mathcal{N}]}$. There is an intermediate node which has another $|\alpha|$ incident edges (from Figure 2(a)), carrying the messages $V_\alpha = \{V_j, j \in \alpha\}$. The intermediate node then has an edge of capacity $h(\alpha)$ to the sink.

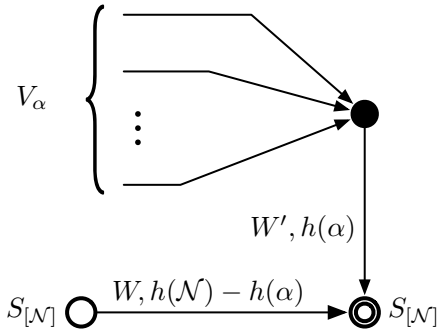
Finally, Figure 2(d) shows the structure of the type 2 subnetworks. Type 2 subnetworks are indexed by a set α, where $\emptyset \neq \alpha \subset \mathcal{N}$ and an element $i \in \alpha, i \notin \mathcal{N}$. Each type 2 subnetwork

[2]For simplicity, we use the same symbol to denote the index of a multicast session and the associated source random variable.

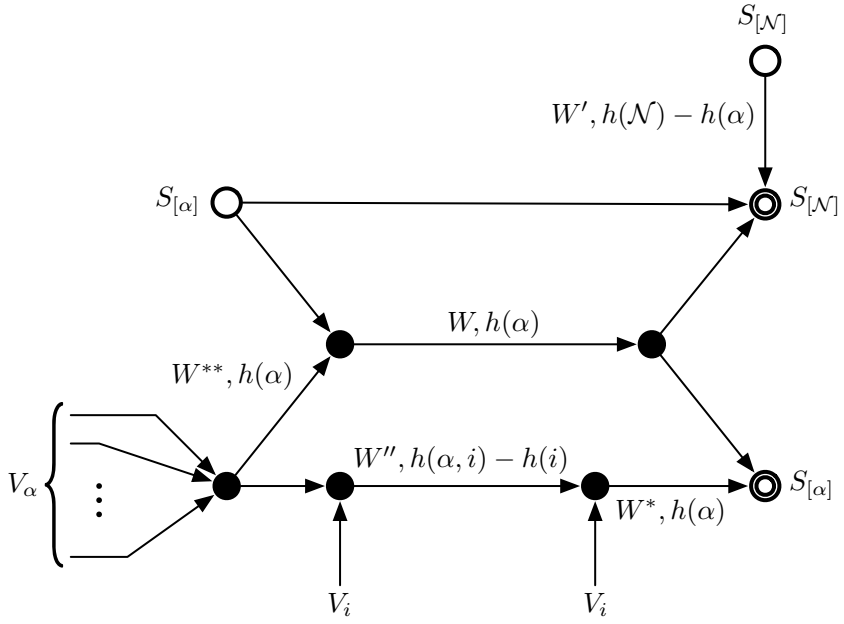


(a) The sources.

(b) Type 0 subnetworks



(c) Type 1 subnetworks



(d) Type 2 subnetworks

Fig. 2. The network $\mathcal{G}^\dagger$.

connects two sources $S_{[\alpha]}$ and $S_{[\mathcal{N}]}$ and two receivers respectively requiring $S_{[\alpha]}$ and $S_{[\mathcal{N}]}$. In addition, there are $|\alpha|+2$ other incident edges from Part 1 of the network, carrying V_α and two copies of V_i. For notational simplicity, we have written $h\left(\alpha \cup \{i\}\right) \triangleq h(\alpha, i)$.

So far, we have described a network $\mathcal{G}^\dagger$, a connection requirement $M^\dagger$ and have assigned rates to sources and capacities to links. Clearly $M^\dagger$ depends only on N, and not in any other way on h. Similarly, the topology of the network $\mathcal{G}^\dagger$ depends only on N. The choice of h affects only the source rates and edge capacities, which are collected into the rate-capacity tuple $\mathsf{T}(h)$. Also, we can assume without loss of generality that $\mathsf{T}(h)$ is a linear function of h.

Example 2: Figure 3 shows the topology of the network $\mathcal{G}^\dagger$ when $N = 2$. Edge labels are omitted for clarity.

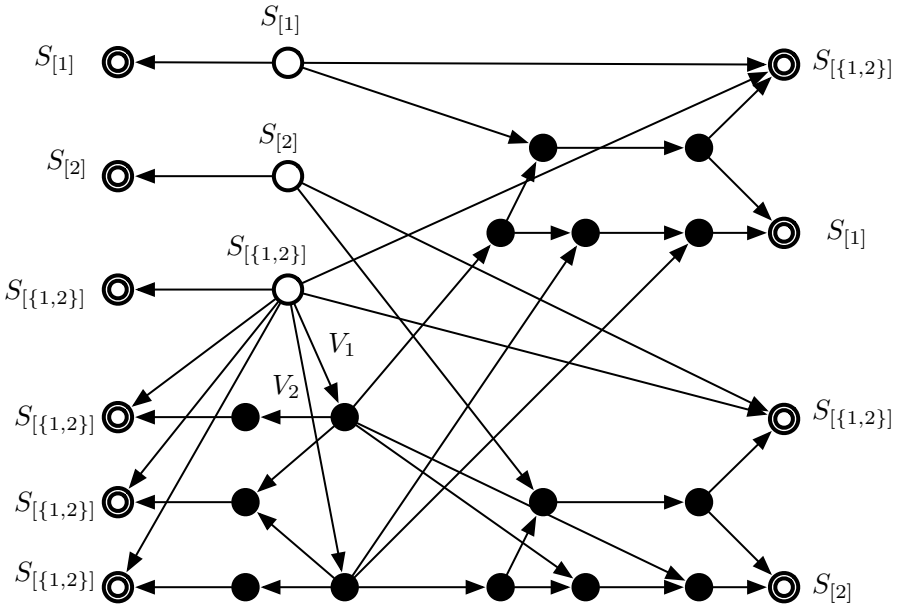


Fig. 3. The network $\mathcal{G}^\dagger$ when $N = 2$.

B. First Duality: Entropy functions and network codes

Theorem 1: Let h be in $\mathcal{H}[\mathcal{N}]$ for $\mathcal{N} = \{1, 2, \ldots, N\}$. The induced rate-capacity tuple $\mathsf{T}(h)$ is admissible on the network $\mathcal{G}^\dagger$ and connection requirement $M^\dagger$, if and only if h is quasi-uniform, i.e.,

$$h \in \Gamma_Q^* \iff \mathsf{T}(h) \in \Upsilon^0.$$

We begin with a proof of the only-if statement, i.e. starting with the assumption of admissibility, we must demonstrate that the function is quasi-uniform. By Definition 7, admissibility of $\mathsf{T}(h)$ on $\mathcal{G}^\dagger, M^\dagger$ requires existence of a zero-error network code Φ with source messages $S_{[\alpha]}$, $\emptyset \neq \alpha \subseteq \mathcal{N}$ and a subset of its coded messages $V_{\mathcal{N}}$ satisfying

$$H\left(S_{[\alpha]}\right) \geq h(\alpha), \quad \alpha \subseteq \mathcal{N} \tag{6}$$

$$H\left(S_{[\alpha]} : \alpha \subseteq \mathcal{N}\right) = \sum_{\alpha \subseteq \mathcal{N}} H(S_{[\alpha]}) \tag{7}$$

$$H\left(V_i\right) \leq h(i), \quad i \in \mathcal{N}. \tag{8}$$

The remaining goal is to prove $H(V_\alpha) = h(\alpha)$ for every $\alpha \subseteq \mathcal{N}$. To this end, we prove the following series of Lemmas 3–8, each predicated on admissibility of $\mathsf{T}(h)$ on $\mathcal{G}^\dagger, M^\dagger$.

Lemma 3: $H\left(S_{[\alpha]}\right) = h(\alpha)$ for all $\emptyset \neq \alpha \subseteq \mathcal{N}$.

Proof: Consider the type 0 subnetworks of Figure 2(b). Admissibility implies that each receiver can correctly reconstruct its required source message. This is not possible unless $H(S_{[\alpha]}) \leq H(W) \leq h(\alpha)$, which together with (6) proves the lemma. ■

Lemma 4: $h(\alpha) \leq H(V_\alpha)$ for all $\emptyset \neq \alpha \subseteq \mathcal{N}$.

Proof: Consider type 1 subnetworks in Figure 2(c). In order for the receiver to correctly determine the requested source message $S_{[\mathcal{N}]}$, it must be true that $H(V_\alpha) + H(W) \geq H(S_{[\mathcal{N}]})$. Furthermore, $H(W) \leq h(\mathcal{N}) - h(\alpha)$. Hence,

$$\begin{aligned} H(V_\alpha) + h(\mathcal{N}) - h(\alpha) &\geq H(V_\alpha) + H(W) \\ &\geq H(S_{[\mathcal{N}]}) \\ &\geq h(\mathcal{N}), \end{aligned}$$

where the last line follows from (6). As a result, $H(V_\alpha) \geq h(\alpha)$. ■

Lemma 5: $H(V_j) = h(j)$ for all $j \in \mathcal{N}$.

Proof: A direct consequence of Lemma 4 and (8). ■

By Lemma 5 we have taken a small step towards our goal, establishing $H(V_\alpha) = h(\alpha)$ for $|\alpha| = 1$. Extension to all α will be achieved by induction on $|\alpha|$. To this end, the remaining lemmas take the hypothesis $H(V_\alpha) = h(\alpha)$ for $|\alpha| = k < N$, and are proved in the context of type 2 subnetworks indexed by α and an element $i \in \mathcal{N}$, $i \notin \alpha$, as shown in Figure 2(d).

Lemma 6: In type 2 subnetworks, $W \perp S_{[\alpha]}$. Furthermore, if $V_\alpha = h(\alpha)$, then $H(V_\alpha|W, S_{[\alpha]}) = 0$.

Proof: By (7), $S_{[\alpha]} \perp S_{[\mathcal{N}]}$ and hence

$$\begin{aligned} H\left(S_{[\alpha]}\right) + H\left(S_{[\mathcal{N}]}\right) &= H\left(S_{[\alpha]}, S_{[\mathcal{N}]}\right) \\ &\leq H\left(S_{[\alpha]}, S_{[\mathcal{N}]}, W, W'\right) \\ &\overset{(i)}{\leq} H(W, S_{[\alpha]}, W') \\ &= H(W, S_{[\alpha]}) + H(W' \mid W, S_{[\alpha]}) \\ &\overset{(ii)}{\leq} H(W, S_{[\alpha]}) + H(W') \\ &\leq H(W) + H(S_{[\alpha]}) + H(W') \\ &\overset{(iii)}{\leq} h(\alpha) + H(S_{[\alpha]}) + H(W') \\ &\overset{(iv)}{\leq} h(\alpha) + H(S_{[\alpha]}) + h(\mathcal{N}) - h(\alpha) \\ &\overset{(v)}{=} H(S_{[\alpha]}) + H(S_{[\mathcal{N}]}). \end{aligned}$$

The inequality (i) follows from the fact that $S_{[\mathcal{N}]}$ is determined from $W, S_{[\alpha]}, W'$ at the upper receiver in Figure 2(d). Inequality (ii) is by discarding conditioning (note that both W and W' depend on $S_{[\mathcal{N}]}$, so this is indeed only an inequality). Inequalities (iii) and (iv) follow from the type 2 subnetwork capacity constraints,

$$H(W) \leq h(\alpha) \tag{9}$$

$$H(W') \leq h(\mathcal{N}) - h(\alpha) \tag{10}$$

and from Lemma 3. Finally, (v) is by Lemma 3. Thus the series of inequalities is actually a series of identities, and as a result,

$$H(W) = h(\alpha) \tag{11}$$

$$H(W, S_{[\alpha]}) = H(W) + H(S_{[\alpha]}) = 2h(\alpha) \tag{12}$$

which proves $W \perp S_{[\alpha]}$. Now consider

$$\begin{aligned} H(V_\alpha|W, S_{[\alpha]}) &= H(V_\alpha, W, S_{[\alpha]}) - H(W, S_{[\alpha]}) \\ &\overset{(i)}{=} H(V_\alpha, S_{[\alpha]}) - H(W, S_{[\alpha]}) \\ &\leq H(V_\alpha) + H(S_{[\alpha]}) - H(W, S_{[\alpha]}) \\ &\overset{(ii)}{=} H(V_\alpha) - h(\alpha) \\ &= 0 \text{ if } H(V_\alpha) = h(\alpha) \end{aligned}$$

where (i) holds since W is a function of $V_\alpha, S_{[\alpha]}$ and (ii) is by (11) and (12). ■

Lemma 7: In type 2 subnetworks, $H(W|V_\alpha, W^*) = H(W|W^*) = H(W)$, or equivalently, $I(W; V_\alpha, W^*) = 0$.

Proof: Recalling that $i \notin \alpha \subset \mathcal{N}$,

$$\begin{aligned} H(W|V_\alpha, W^*) &\geq H(W|V_\alpha, W^*, V_i) \\ &\overset{(i)}{=} H(W|V_\alpha, V_i) \\ &\overset{(ii)}{=} H(W|V_\alpha, V_i) + H(S_{[\alpha]}|V_\alpha, V_i, W) \\ &= H(W, S_{[\alpha]}|V_\alpha, V_i) \\ &\geq H(S_{[\alpha]}|V_\alpha, V_i) \\ &\overset{(iii)}{=} H(S_{[\alpha]}) \\ &\overset{(iv)}{=} h(\alpha) \\ &\overset{(v)}{\geq} H(W) \\ &\geq H(W|W^*) \\ &\geq H(W|V_\alpha, W^*) \end{aligned}$$

where (i) follows from the fact that W^* is a function of V_α, V_i, (ii) follows from that $S_{[\alpha]}$ can be reconstructed at the lower receiver, and (iii) follows from independence of $S_{[\alpha]}$ and (V_α, V_i), since by (7) $S_{[\alpha]} \perp S_{[\mathcal{N}]}$ and all the $V_j, j \in \mathcal{N}$ depend only on $S_{[\mathcal{N}]}$. Finally, (iv) is by Lemma 3, (v) is by the capacity constraint (9) and the remaining inequalities simply add extra conditioning. Thus the chain of inequalities is actually a chain of identities, the last three proving the lemma.

■

Lemma 8: In type 2 subnetworks, assuming $H(V_\alpha) = h(\alpha)$, $H(W^*|V_\alpha) = H(V_\alpha|W^*) = 0$.

Proof:

$$\begin{aligned} H(V_\alpha|W^*) &= H(V_\alpha|W^*, W) + I(V_\alpha; W|W^*) \\ &\overset{(i)}{=} H(V_\alpha|W^*, W) \\ &\leq H(V_\alpha, S_{[\alpha]}|W^*, W) \\ &= H(V_\alpha|W^*, W, S_{[\alpha]}) + H(S_{[\alpha]}|W^*, W) \\ &\overset{(ii)}{=} H(V_\alpha|W^*, W, S_{[\alpha]}) \\ &\leq H(V_\alpha|W, S_{[\alpha]}) \\ &\overset{(iii)}{=} 0. \end{aligned}$$

where (i) follows from Lemma 7, (ii) is because $S_{[\alpha]}$ can be reconstructed at the lower receiver, and (iii) is by Lemma 6, assuming $H(V_\alpha) = h(\alpha)$. Since conditional entropies are non-negative

$$H(V_\alpha|W^*) = 0. \tag{13}$$

On the other hand,

$$\begin{aligned} H(W^*|V_\alpha) &= H(W^*, V_\alpha) - H(V_\alpha) \\ &= H(W^*) + H(V_\alpha|W^*) - H(V_\alpha) \\ &\leq h(\alpha) - h(\alpha) = 0 \end{aligned}$$

where the last inequality uses (13), the type 2 subnetwork capacity bound $H(W^*) \leq h(\alpha)$ and the assumption $H(V_\alpha) = h(\alpha)$. Non-negativity of conditional entropy yields $H(W^*|V_\alpha) = 0$. ■

We are now ready to assemble the preceding lemmas into a proof for the only-if part of Theorem 1. *Proof:* [Proof: only-if part of Theorem 1] The goal is to prove $H(V_\alpha) = h(\alpha)$ for all non-empty subsets $\alpha \subseteq \mathcal{N}$. This was already shown for $|\alpha| = 1$ in Lemma 5. Extension to all α will be achieved using induction. First, assume the hypothesis is true for all $\alpha \subset \mathcal{N}$ with $1 \leq |\alpha| \leq k < N$. For any $i \in \mathcal{N}$ and $\alpha \subset \mathcal{N}$ such that $i \notin \alpha$ and $|\alpha| = k$, consider the type 2 subnetwork of Figure 2(d). We must show that $H(V_\alpha, V_i) = h(\alpha \cup \{i\}) \triangleq h(\alpha, i)$. By Lemma 4 we already know that $H(V_\alpha, V_i) \geq h(\alpha, i)$. Therefore it remains only to prove

$H(V_\alpha, V_i) \leq h(\alpha, i)$. Now

$$\begin{aligned} H(V_i, V_\alpha) &\leq H(V_i, V_\alpha, W^*) \\ &\overset{(i)}{=} H(V_i, W^*) \\ &\leq H(V_i, W^*, W'') \\ &\overset{(ii)}{=} H(V_i, W'') \\ &\leq H(V_i) + H(W'') \\ &\overset{(iii)}{\leq} H(V_i) + h(\alpha, i) - h(i) \\ &\overset{(iv)}{=} h(i) + h(\alpha, i) - h(i) \\ &= h(\alpha, i) \end{aligned}$$

where (i) follows from Lemma 8 (which holds under the induction hypothesis), (ii) is due to the fact that W^* is a function of W'', V_i and (iii) is from the subnetwork 2 capacity bound $H(W'') \leq h(\alpha, i) - h(i)$. Finally, (iv) is by Lemma 5.

Up to this point, we have proved that h is the entropy function of a set of random variables $\{V_1, \ldots, V_N\}$. To show that h is indeed quasi-uniform, it suffices to prove that for any subset α of $\mathcal{N}$, the set of random variables V_α is quasi-uniform. Since we have just showed that $H(V_\alpha) = h(\alpha)$, if the receiver in the type 1 subnetwork can decode $S_{[\mathcal{N}]}$, then $H(V_\alpha|W') = H(W'|V_\alpha) = 0$. Hence, $H(W') = h(\alpha)$. Now according to the link capacity constraint, W' is defined on an alphabet set of size $2^{h(\alpha)}$, and W' (and hence V_α) must be quasi-uniform. ■

It remains to prove the "if" statement in the theorem, i.e. to show that quasi-uniform random variables imply admissibility. *Proof:* [Proof: if part of Theorem 1] It suffices to show that one can construct a network code (defined by input variables, and message variables) meeting the connection requirement subject to the individual capacity constraint on each link.

The construction for the input variables is simple. For any $\emptyset \neq \alpha \subseteq \mathcal{N}$, define $S_{[\alpha]}$ to be a quasi-uniform random variable with entropy $h(\alpha)$. These input variables are also assumed to be independent. It remains to show that we can construct edge variables satisfying the capacity constraints, and which allow each receiver to reconstruct the requested messages perfectly.

By the quasi-uniformity of $S_{[\alpha]}$, it is clear that all receivers in type 0 subnetworks can reconstruct their requested message simply by having the source transmit the uncoded message,

$W = S_{[\alpha]}$.

Let $\{V_j : j \in \mathcal{N}\}$ be a set of quasi-uniform random variables whose entropy function is h. Since $H(V_{\mathcal{N}}) = H(S_{[\mathcal{N}]})$, there is a one-to-one mapping between $\Omega(V_{\mathcal{N}})$ and $\Omega(S_{[\mathcal{N}]})$. As they are both quasi-uniform, $S_{[\mathcal{N}]}$ and $(V_j : j \in \mathcal{N})$ can be regarded as the same.

For type 1 networks, by quasi-uniformity of V_α, one can send V_α unencoded as W'. Then the receivers see V_α and an auxiliary message W defined on a sample space of size at most $2^{h(\mathcal{N})-h(\alpha)}$. Reconstructing $S_{[\mathcal{N}]}$ at the receiver is equivalent to reconstructing $V_{\mathcal{N}\setminus\alpha}$ at the receiver.

By the quasi-uniformity of $S_{[\alpha]}$ and Lemma 2, $V_{\mathcal{N}\setminus\alpha}$ can be compressed to a symbol W of size $2^{h(\mathcal{N})-h(\alpha)}$ such that $V_{\mathcal{N}\setminus\alpha}$ can be losslessly reconstructed from W and V_α.

It remains to verify that receivers in type 2 subnetworks can reconstruct all requested messages. Recall that both $S_{[\alpha]}$ and V_α are quasi-uniform. Assume without loss of generality that their supports are $\{0, 1, 2, \ldots, 2^{h(\alpha)} - 1\}$. Then we can define $W \triangleq V_\alpha + S_{[\alpha]} \mod 2^{h(\alpha)}$. It is easy to verify the following properties:

$$H\left(W \mid V_\alpha, S_{[\alpha]}\right) = H\left(S_{[\alpha]} \mid W, V_\alpha\right) = H\left(V_\alpha \mid W, S_{[\alpha]}\right) = 0, \tag{14}$$

$$\log|\Omega(W)| = h(\alpha). \tag{15}$$

By (14), the upper receiver can correctly reconstruct V_α from $S_{[\alpha]}$ and W. Using a similar compression scheme as used in type 1 subnetworks, source $S_{[\mathcal{N}]}$ is compressed to $h(\mathcal{N}) - h(\alpha)$ bits, allowing lossless reconstruction of $S_{[\mathcal{N}]}$ at the upper receiver.

On the other hand, it is easy to see that $\{V_\alpha, V_i\}$ is quasi-uniform. Hence V_α can be compressed into W'' with a support of size $|\Omega(W'')| = 2^{h(\alpha,i)-h(i)}$ such that V_α can be reconstructed by using W'' and V_i. As a result, W^* may be transmitted as V_α without any encoding. The lower receiver can then recover $S_{[\alpha]}$ from V_α and W.

Since all receivers can reconstruct their requested source messages with properly constructed message random variables satisfying the capacity constraints, the rate-capacity tuple $\mathsf{T}(h)$ is admissible. ■

Definition 16: A polymatroid h is called *almost entropic* if there exists a sequence of entropic pseudo-entropy functions $h^{(k)}$ and positive constants $r(k)$ such that $\lim_{k\to\infty} h^{(k)}/r(k) = h$.

As $\bar{\Gamma}^*$ is a closed and convex cone [30], the set of all almost entropic functions is $\bar{\Gamma}^*$. Theorem 1 establishes a duality, or equivalence between the quasi-uniformity of h and admissibility of

$\mathsf{T}(h)$. The following theorem extends this result to a duality between almost entropic h and asymptotically admissible (and achievable) $\mathsf{T}(h)$.

Theorem 2: Let $h \in \mathcal{H}[\mathcal{N}]$ for $\mathcal{N} = \{1, 2, \ldots, N\}$ and let $\mathsf{T}(h)$ be an induced rate-capacity tuple. Then we have,

$$h \in \bar{\Gamma}^* \iff \mathsf{T}(h) \in \Upsilon^\infty \iff \mathsf{T}(h) \in \Upsilon^\epsilon.$$

In other words, the rate-capacity tuple $\mathsf{T}(h)$ is asymptotically admissible (or achievable) on the network $\mathcal{G}^\dagger$ and connection requirement $M^\dagger$ if and only if h is almost entropic.

Proof: Suppose that h is almost entropic. We will first show that $\mathsf{T}(h) \in \Upsilon^\infty$. By [12], [26], one can construct a sequence of quasi-uniform entropic functions $h^{(n)}$ and normalizing constants $r(n)$ that $\lim_{n\to\infty} h^{(n)}(\alpha)/r(n) = h(\alpha)$. By Theorem 1, each $\mathsf{T}(h^{(n)})$ is admissible. By property P2, the set Υ^∞ of asymptotically admissible rate-capacity tuples is a closed and convex cone and hence $\mathsf{T}(h) \in \Upsilon^\infty$.

Clearly, $\mathsf{T}(h) \in \Upsilon^\infty$ implies that $\mathsf{T}(h) \in \Upsilon^\epsilon$. It remains to show that $\mathsf{T}(h)$ is achievable implying that h is almost entropic. Suppose that $\mathsf{T}(h) \in \Upsilon^\epsilon$. According to Definition 9, one can construct a sequence of normalizing constants $r(n)$ and network codes $\Phi^{(n)}$ with source messages $\{S^{(k)}_{[\alpha]}, \alpha \subseteq \mathcal{N}\}$ and edge messages $V^{(k)}_N$ such that[3]

$$\lim_{k\to\infty} \frac{1}{r(n)} H\left(S^{(n)}_{[\alpha]}\right) \geq h(\alpha) \tag{16}$$

$$\lim_{k\to\infty} \frac{1}{r(n)} H\left(V^{(n)}_i\right) \leq h(i) \tag{17}$$

$$\lim_{n\to\infty} P_e\left(\Phi^{(n)}\right) = 0. \tag{18}$$

For each value of the sequence index n, consider the network $\mathcal{G}^\dagger$ and connection requirement $M^\dagger$ of Figure 2 with sources $\mathcal{S} = \left\{S^{(n)}_{[\alpha]}, \emptyset \neq \alpha \in 2^{\mathcal{N}}\right\}$ and edge messages $V^{(n)}_{\mathcal{N}}$. By the Fano inequality, the entropy of any source $s \in \mathcal{S}$ conditioned on the edge variables incident to any node in $\mathcal{D}(s)$ can be made as small as desired by increasing n. Following a similar procedure as in the proof for Theorem 1, it can be proved that for any non-empty subset $\emptyset \neq \alpha \subseteq \mathcal{N}$,

$$\lim_{k\to\infty} \frac{1}{r(n)} H\left(V^{(k)}_\alpha\right) = h(\alpha).$$

In other words, h is almost entropic. ■

[3]By the Bolzano-Wierstrass Theorem which says that any sequence in a closed and bounded interval has a convergent subsequence, we can safely assume that $\lim_{k\to\infty} \frac{1}{r(k)} H(S^{(k)}_{[\alpha]}, V^{(k)}_\beta)$ exists for any nonempty subsets α, β of $\mathcal{N}$.

C. Second Duality: Linear group characterizable functions and linear network codes

The first duality shows that h is quasi-uniform (almost entropic) if and only if $\mathsf{T}(h)$ is admissible (achievable). We will now prove a similar result, restricting the network codes to be linear.

Theorem 3: Let $h \in \mathcal{H}[\mathcal{N}]$ for $\mathcal{N} = \{1, 2, \ldots, N\}$. The induced rate-capacity tuple $\mathsf{T}(h)$ is admissible using linear network codes on the network $\mathcal{G}^\dagger$ and connection requirement $M^\dagger$, if and only if h is linear group characterizable, i.e.,

$$h \in \Gamma^*_{L(q)} \iff \mathsf{T}(h) \in \Upsilon^0_{L(q)}$$

Proof: [Proof: only-if part of Theorem 3] The proof of the only-if part is very similar to the one given in Theorem 1. Suppose that $\mathsf{T}(h) \in \Upsilon^0_{L(q)}$, i.e., it is admissible using a linear network code Φ on the network $\mathcal{G}^\dagger$ and connection requirement $M^\dagger$. By Proposition 2, the set of induced source and link random variables by Φ is linear group characterizable. Using the same argument as in the proof for Theorem 1, h is the entropy function of a subset of these linear group characterizable random variables. Hence, h is linear group characterizable.

In fact, using the same argument, we can show that if the induced rate-capacity tuple $\mathsf{T}(h)$ is admissible using abelian network codes on the network $\mathcal{G}^\dagger$ and connection requirement $M^\dagger$, then h is abelian group characterizable. ■

Before we prove the if part of Theorem 3, we need the following lemma which serves a similar role as Lemma 2 in the proof of Theorem 1 by justifying the feasibility of certain "compression" scheme.

Lemma 9: Consider a special case of the network depicted in Figure 1 where the left node receives $T_1(a)$ and $T_2(a)$ as inputs, where T_1 and T_2 are two linear functions defined on a vector space $\mathbf{A}$ over $\mathbb{F}_q$. Let the kernels of T_1 and T_2 be respectively $\mathbf{B}_1$ and $\mathbf{B}_2$. Then, there exists a linear function W of $T_1(a)$ and $T_2(a)$ such that (1) $T_1(a)$ is uniquely determined from W and $T_2(a)$, and (2) W takes at most $q^{\dim \mathbf{B}_2 - \dim \mathbf{B}_1 \cap \mathbf{B}_2}$ different values.

Proof: From $\mathbf{B}_1$ and $\mathbf{B}_2$, we can construct three subspaces $\mathbf{W}_1$, $\mathbf{W}_2$ and $\mathbf{W}_0$ such that

$$\dim \mathbf{W}_0 + \dim \mathbf{W}_1 + \dim \mathbf{W}_2 + \dim \mathbf{B}_1 \cap \mathbf{B}_2 = \dim \mathbf{A}$$

and that for each $i = 1, 2$, the subspace $\mathbf{B}_i$ is equal to the linear span of $\mathbf{W}_i$ and $\mathbf{B}_1 \cap \mathbf{B}_2$. Hence any $a \in \mathbf{A}$ can be written uniquely as $a = a_0 + a_1 + a_2 + b$ where $a_i \in \mathbf{W}_i$ for $i = 1, 2, 3$ and $b \in \mathbf{B}_1 \cap \mathbf{B}_2$.

Since $\ker(T_1) = \mathbf{B}_1$, we have $T_1(a_0 + a_1 + a_2 + b) = T_1(a_2) + T_1(b)$. Furthermore, one can easily construct a linear function T_1^* such that $T_1^*(T_1(a)) = (a_2, b)$. Similarly, there exists a linear function T_2^* such that $T_2^*(T_2(a)) = (a_1, b)$.

To compute $T_1(a)$ at node 2, it suffices to compute a_2 as b can be computed directly from $T_2(a)$. A simple counting argument shows that a_2 lies in a vector subspace of dimension $\dim \mathbf{B}_2 - \dim \mathbf{B}_1 \cap \mathbf{B}_2$. Therefore, we can set $W = a_2$ over the network and it takes at most $q^{\dim \mathbf{B}_2 - \dim \mathbf{B}_1 \cap \mathbf{B}_2}$ different values. ■

Now we may continue our proof for Theorem 3. *Proof:* [Proof: if part of Theorem 3] To prove the direct part of Theorem 3, we need to show that if h is linear group characterizable, then one can construct a linear network code (defined by the induced source and link random variables) meeting the connection requirement subject to the individual capacity constraint on each link.

Suppose that h is linear group characterizable by a vector space $\mathbf{V}$ and its subspaces $\mathbf{V}_1, \ldots, \mathbf{V}_N$, defined over a field $\mathbb{F}_q$. Assume without loss of generality that the subspaces intersect only at the zero vector, $\bigcap_{j=1}^{N} \mathbf{V}_j = \{\mathbf{0}\}$. As such, $h(\mathcal{N}) = \log q \cdot (\dim \mathbf{V})$ and for any $\alpha \subseteq \mathcal{N}$, we have $h(\alpha) = \log q \cdot (\dim \mathbf{V} - \dim \bigcap_{j \in \alpha} \mathbf{V}_j)$.

For $j = 1, \ldots, N$, construct linear functions f_j over $\mathbf{V}$ such that $\ker(f_j) = \mathbf{V}_j$. The source random variable $S_{[\mathcal{N}]}$ is uniformly distributed over $\mathbf{V}$ such that the link symbols transmitted in Figure 2(a) are $V_j = f_j(S_{[\mathcal{N}]})$. For any other $\emptyset \neq \alpha \subset \mathcal{N}$, define $S_{[\alpha]}$ to be a random variable, uniformly distributed over a vector space of dimension $\log_q 2 \cdot h(\alpha)$ (hence, $H(S_{[\alpha]}) = h(\alpha)$). All these source random variables are assumed to be independent.

Up to this point, we have described how source and link random variables are defined in Figure 2(a). It remains to show that we can construct a linear network code, consisting of a set of link random variables which are linear functions of the incident source/link random variables, satisfying the capacity constraints, and which allow each receiver to reconstruct the requested messages perfectly.

For type 0 subnetworks, all receivers can reconstruct their requested message simply by having the source transmit the uncoded message, $W = S_{[\alpha]}$. Clearly, the associated link random variables in these subnetworks are linear functions of the incident ones and meet the capacity constraint.

For type 1 subnetworks, let $W' = (V_i : i \in \alpha) = (f_i(S_{[\mathcal{N}]}) : i \in \alpha)$, which depends linearly on $S_{[\mathcal{N}]}$. Note that $(f_i(a) : i \in \alpha) = \mathbf{0}$ if and only if $f_i(a) = \mathbf{0}$ for all $i \in \alpha$, or equivalently,

when $a \in \bigcap_{i\in\alpha} \mathbf{V}_i$. By the rank-nullity theorem, W' can take at most $|\mathbf{V}|/|\bigcap_{i\in\alpha} \mathbf{V}_i|$ different values. We can thus treat W' as a vector in space of dimension $\dim \mathbf{V} - \dim \bigcap_{i\in\alpha} \mathbf{V}_i$.

As a result, the subnetwork can now be treated as a special case of Lemma 9 such that $T_1(a) = a$ and $T_2(a) = (f_i(a) : i \in \alpha)$. The dimensions of the kernels of T_1 and T_2 are respectively 0 and $\dim \bigcap_{i\in\alpha} \mathbf{V}_i$. By Lemma 9, the required rate is thus $\log q \cdot (\dim \bigcap_{i\in\alpha} \mathbf{V}_i) = h(\mathcal{N}) - h(\alpha)$.

Similarly, for type 2 subnetworks, let $W^{**} = (f_i(S_{[\mathcal{N}]}) : i \in \alpha)$. As before, we can treat W^{**} as a vector of length $\dim \mathbf{V} - \dim \bigcap_{i\in\alpha} \mathbf{V}_i$. Similarly, $S_{[\alpha]}$ can also be regarded as a vector of the same length. We can therefore define W by vector addition, $W = S_{[\alpha]} + W^{**}$. Consequently, the receiver in the upper branch can reconstruct V_α by subtracting $S_{[\alpha]}$ from W. As before, one can find W' as a linear function of $S_{[\mathcal{N}]}$ and this function allows $S_{[\mathcal{N}]}$ to be reconstructed from W' and V_α.

For the lower branch, we can identify a special case of Figure 1 with $T_1(a) = V_\alpha$ and $T_2(a) = V_i$. One can construct W'' such that (1) W'' is a linear function of $T_1(a)$ and $T_2(a)$, (2) the kernel $\ker(T_1) = \bigcap_{j\in\alpha} \mathbf{V}_j$ and $\ker(T_2) = \mathbf{V}_i$, and (3) the rate required is $\dim \bigcap_{j\in\alpha} \mathbf{V}_j - \dim \mathbf{V}_i \bigcap_{j\in\alpha} \mathbf{V}_j$. Therefore, we can reconstruct V_α from W'' and $T_2(a)$ where $T_1(a) = V_\alpha$. Again, treating V_α as a vector of length $\dim \mathbf{V} - \dim \bigcap_{i\in\alpha} \mathbf{V}_i$, the receiver at the lower branch can reconstruct $S_{[\alpha]}$ by subtracting V_α from W. ■

So far, we have proved that h is linear group characterizable if and only if the rate-capacity tuple $\mathsf{T}(h)$ is admissible with a linear network code. As before, we can further generalize the result to include the case when h is almost linear group characterizable according to the following definition.

Definition 17: A polymatroid h is called *almost linear group characterizable* if there exists a sequence of linear group characterizable entropy functions $h^{(k)}$ and positive constants $r(k)$ such that $\lim_{k\to\infty} h^{(k)}/r(k) = h$.

It is easy to prove that the set of all almost linear group characterizable polymatroids is $\overline{\mathrm{con}}(\Gamma^*_{L(q)})$, the minimal closed and convex cone containing $\Gamma^*_{L(q)}$.

Theorem 4: Let $h \in \mathcal{H}[\mathcal{N}]$ for $\mathcal{N} = \{1, 2, \ldots, N\}$ and let $\mathsf{T}(h)$ be an induced rate-capacity tuple. Then we have

$$h \in \overline{\mathrm{con}}(\Gamma^*_{L(q)}) \iff \mathsf{T}(h) \in \Upsilon^{\infty}_{L(q)} \iff \mathsf{T}(h) \in \Upsilon^{\epsilon}_{L(q)}.$$

In other words, the rate-capacity tuple $\mathsf{T}(h)$ is asymptotically admissible (or achievable) by linear

network codes on the network $\mathcal{G}^\dagger$ and connection requirement $M^\dagger$ if and only if h is is almost linear group characterizable.

Proof: Suppose that $h \in \overline{\text{con}}(\Gamma^*_{L(q)})$. By Definition 17, one can construct a sequence of linear group characterizable entropy functions $h^{(k)}$ and positive constants $r(k)$ such that $\lim_{k\to\infty} h^{(k)}/r(k) = h$. By Theorem 3, each $\mathsf{T}(h^{(n)})$ is admissible by linear network codes. By property P2, the set $\Upsilon^\infty_{L(q)}$ of asymptotically admissible rate-capacity tuples is a closed and convex cone and hence $\mathsf{T}(h) \in \Upsilon^\infty_{L(q)}$.

Clearly, $\mathsf{T}(h) \in \Upsilon^\infty_{L(q)}$ implies that $\mathsf{T}(h) \in \Upsilon^\epsilon_{L(q)}$. It remains to prove that $\mathsf{T}(h) \in \Upsilon^\epsilon_{L(q)}$ implies $h \in \overline{\text{con}}(\Gamma^*_{L(q)})$.

Suppose that $\mathsf{T}(h)$ is achievable by linear network codes. Then one can construct a sequence of normalizing constants $r(n)$ and linear network codes $\Phi^{(n)}$ with source messages $(S^{(k)}_{[\alpha]}, \alpha \subseteq \mathcal{N})$ and edge messages $(V^{(k)}_j, j \in \mathcal{N})$ such that

$$\lim_{k\to\infty} \frac{1}{r(n)} H\left(S^{(n)}_{[\alpha]}\right) \geq h(\alpha) \tag{19}$$

$$\lim_{k\to\infty} \frac{1}{r(n)} H\left(V^{(n)}_j\right) \leq h(j) \tag{20}$$

$$\lim_{n\to\infty} P_e\left(\Phi^{(n)}\right) = 0. \tag{21}$$

Similar to the proof given in Theorem 2, it can be proved that for any non-empty subset $\emptyset \neq \alpha \subseteq \mathcal{N}$, $\lim_{k\to\infty} \frac{1}{r(n)} H\left(V^{(k)}_\alpha\right) = h(\alpha)$. In addition, as $(V^{(k)}_j, j \in \mathcal{N})$ is linear group characterizable, h is almost linear group characterizable. ■

D. Third Duality: Polymatroids and the LP bound

Theorem 2 provides a duality between entropy functions and network codes, namely that a function $h \in \mathcal{H}[\mathcal{N}]$ is almost entropic if and only if $\mathsf{T}(h)$ is achievable on $\mathcal{G}^\dagger$, $M^\dagger$. As the set of almost entropic functions $\bar{\Gamma}^*$ has no explicit characterization for four or more variables, the sets of admissible or achievable rate-capacity tuples are unknown. Therefore computable bounds such as the linear programming bound are of great interest.

Let Γ be the set of all polymatroids. Definition 14 writes the LP bound in terms of constraints on pseudo-variables. The following theorem provides a direct generalization of the ideas of the previous sections to pseudo-variables.

Theorem 5: Suppose $h \in \mathcal{H}[\mathcal{N}]$. A rate-capacity tuple $(\lambda(h), \omega(h))$ satisfies the LP bound if and only if h is a polymatroid,

$$h \in \Gamma \iff \mathsf{T}(h) \in \Upsilon_{LP}.$$

Proof: The "only if" part of the proof is a direct generalization of the proof of Theorem 1. Suppose $(\lambda(h), \omega(h))$ satisfies the LP bound. By Definition 14 there exists a set of pseudo-variables satisfying the set of (in)equalities in (5). In particular, there are pseudo-variables $\{S_{[\alpha]}, \emptyset \neq \alpha \subseteq \mathcal{N}\}$ and $V_{\mathcal{N}}$ such that

$$H(S_{[\alpha]}) \geq h(\alpha), \quad \alpha \subseteq \mathcal{N}, \tag{22}$$

$$H(S_{[\alpha]} : \alpha \subseteq \mathcal{N}) = \sum_{\alpha \subseteq \mathcal{N}} H(S_{[\alpha]}) \tag{23}$$

$$H(V_i) \leq h(i). \tag{24}$$

Following the same steps as in the proof for Theorem 1 (translating random variables to pseudo-variables), shows that h is the pseudo-entropy function of $V_{\mathcal{N}}$. Hence, h is a polymatroid.

To prove the direct part, suppose h is a polymatroid over the ground set $\mathcal{L} = \{V_1, V_2, \ldots, V_N\}$ (i.e. h is the pseudo-entropy function of $V_{\mathcal{N}}$). We must exhibit a set of pseudo-variables satisfying the set of (in)equalities (5). Whereas the proof for Theorem 1 constructs auxiliary random variables via data compression, we need to show how to analogously adhere auxiliary pseudo-variables W, W'' etc. to the set of pseudo-variables $V_{\mathcal{N}}$. In contrast to random variables, we cannot rely on coding theorems, or other probabilistic constructions that assume the existence of an underlying probability distribution. Nevertheless, it is possible to adhere pseudo-variables. This is accomplished in Appendix I, where proof of the direct part is also completed. ■

E. Fourth Duality: Ingleton polymatroids and the LP bound for linear codes?

Finally, we can consider rate-capacity tuples which satisfy the LP-Ingleton bound of Definition 15. The following theorem establishes a relation to Ingleton polymatroids (i.e., a polymatroid satisfying Ingleton inequalities). This is shown in one direction only. Let $\Gamma_{LP,I}$ be the set of all Ingleton polymatroids.

Theorem 6: Suppose $h \in \mathcal{H}[\mathcal{N}]$. If a rate-capacity tuple $(\lambda(h), \omega(h))$ satisfies the LP bound for linear codes, then h is an Ingleton polymatroid, i.e.,

$$\mathsf{T}(h) \in \Upsilon_{LP,I} \Rightarrow h \in \Gamma_{LP,I}.$$

Proof: Suppose $(\lambda(h), \omega(h))$ satisfies the LP-Ingleton bound. By Definition 15 there exists a set of Ingleton pseudo-variables satisfying the set of (in)equalities in (5). In particular, there are pseudo-variables $\{S_{[\alpha]}, \emptyset \neq \alpha \subseteq \mathcal{N}\}$ and $V_{\mathcal{N}}$ such that

$$H(S_{[\alpha]}) \geq h(\alpha), \quad \alpha \subseteq \mathcal{N}, \tag{25}$$

$$H(S_{[\alpha]} : \alpha \subseteq \mathcal{N}) = \sum_{\alpha \subseteq \mathcal{N}} H(S_{[\alpha]}) \tag{26}$$

$$H(V_i) \leq h(i). \tag{27}$$

Following the same steps as in the proof for Theorem 1 (translating random variables to pseudo-variables), shows that h is the pseudo-entropy function of $V_{\mathcal{N}}$. Hence, h is an Ingleton polymatroid. ■

We conjecture that the converse of the fourth duality should also hold. In fact, it can be proved that if the converse fails to hold, then there exists a polymatroid satisfying Ingleton inequalities but which is not almost linear group characterizable. Therefore determination of whether the converse of the fourth duality holds is a very interesting open question.

V. Implications

The results of Section IV while interesting in their own right, have several consequential applications. First, in Section V-A we consider implications to the determination of the network coding capacity region (in the absence of any restriction on the class of network codes). Secondly, we discuss the sub-optimality of linear network codes in Section V-B.

A. The capacity region

Implication 1 (Hardness of a multicast problem): Determination of the set of achievable source rate-link capacity tuples Υ^{ϵ} is at least as hard as the problem of determining the set of all almost entropic functions.

Similarly, determination of the set of source rate-link capacity tuples achieved by linear network codes $\Upsilon^{\epsilon}_{L(q)}$ is at least as hard as the problem of determining the set of all almost linear group characterizable entropy functions.

Proof: By Theorem 2, a polymatroid h is almost entropic (and almost linear group characterizable) if and only if the induced rate-capacity tuple $(\lambda(h), \omega(h))$ is achievable (with linear

network codes). In other words, the problem of determining the set of all almost entropic (and almost linear group characterizable) functions can be reduced to the solubility of a corresponding multicast problem. ■

In [24], a network, called the Vámos network, was constructed from the Vámos matroid. This was later used to prove that the LP bound is not tight and the bound can be tightened by applying a non-Shannon information inequality proved in [2].

In the following, we will use the duality results obtained in Section IV to provide another proof for the looseness of LP bound.

Implication 2 (Looseness of LP bound): The LP outer bound can be tightened by any non-Shannon information inequality.

Proof: Theorem 5 shows that the rate-capacity tuple $(\lambda(h), \omega(h))$ is in the LP bound if h is a polymatroid. Yet, Theorem 2 proves that $(\lambda(h), \omega(h))$ is achievable if and only if h is almost entropic. Consider the function h defined as follows [2]:

$$\begin{aligned} h(1) = h(2) = h(3) = (4) &= 2a > 0 \\ h(1,2) &= 3a \\ h(3,4) &= 4a \\ h(1,3) = h(1,4) = h(2,3) &= h(2,4) = 3a \\ h(i,j,k) = 4a = h(1,2,3,4),\ &\forall \text{ distinct } i,j,k. \end{aligned}$$

It can be verified directly that $h \in \Gamma_4$. However, the non-Shannon information inequality obtained in [2] shows that $h \notin \bar{\Gamma}_4^*$. While the rate-capacity tuple $\mathsf{T}(h)$ satisfies the LP bound, it is not achievable, as it is not almost entropic. ■

Using the same argument, any non-Shannon information inequality [2], [9], [10] will remove some polymatroids which are not almost entropic. The corresponding tuples in the LP bound will not be achievable. In other words, *any set of non-Shannon information inequalities can be used to tighten the LP bound*.

In fact, together with the fact that $\bar{\Gamma}^*$ is not a polyhedron when the number of random variables is at least four [10], our duality results lead to very interesting consequences.

First, we show that the set of achievable rate-capacity tuples is not a polyhedron in general. Second, the LP bound is not only loose, but it remains loose even when tightened via application

of any finite number of linear non-Shannon information inequalities.

Proposition 5: The set of almost entropic functions is not a polytope.

Proof: [Proof sketch] The following is a sketch of the proof given by Matúš [10]. Matúš constructed a convergent sequence of entropic functions $g_t \to g_0$ with one-side tangent $\dot{g}_{0+} \triangleq \lim_{t\to 0^+}(g_t - g_0)/t$. Clearly, if $\bar{\Gamma}_n^*$ is polyhedral, there exists $\epsilon > 0$ such that $g_0 + \epsilon \dot{g}_{0+} \in \bar{\Gamma}_n^*$. This was shown not to be the case, since $g_0 + \epsilon \dot{g}_{0+}$ violates some of the information inequalities proved in [10]. Therefore, $\bar{\Gamma}_n^*$ is not polyhedral. Furthermore, there are infinitely many information inequalities. ■

Implication 3 (Set of achievable rate-capacity tuples): The sets of achievable rate-capacity tuples Υ^∞ and Υ^ϵ for the network $\mathcal{G}^\dagger$ and connection requirement $M^\dagger$ are not polytopes (when $N \geq 4$).

Proof: Consider the sequence $g_t \to g_0$ from the proof of Proposition 5. By Theorem 2, $\mathsf{T}(g_t)$ and $\mathsf{T}(g_0)$ are asymptotically admissible. As $\mathsf{T}(h)$ is a linear function of h, we have

$$\dot{T} \triangleq \lim_{t\to 0^+} (\mathsf{T}(g_t) - \mathsf{T}(g_0))/t = \mathsf{T}(\dot{g}_{0+}). \tag{28}$$

For any $\epsilon > 0$,

$$\mathsf{T}(g_0) + \epsilon \dot{T} = \mathsf{T}(g_0 + \epsilon \dot{g}_{0+}). \tag{29}$$

As $g_0 + \epsilon \dot{g}_{0+}$ is not almost entropic, $\mathsf{T}(g_0) + \epsilon \dot{T}$ is not achievable. In other words, Υ^∞ and Υ^ϵ are not polytope. ■

Now the LP bound is a polytope, while the capacity region is not. Furthermore, the introduction of any finite number of additional linear inequalities in the LP bound simply results in another polytope. Hence

Implication 4 (Looseness of polyhedral bounds): The LP bound is not tight. Furthermore, any finite number of linear information inequalities cannot tighten the LP bound Υ_{LP} to the set of achievable rate-capacity tuples Υ^ϵ. In fact, any polyhedral outer bound for Υ^ϵ is not tight.

Proof: A direct consequence of Theorem 3 and Proposition 5. ■

B. Suboptimality of linear network codes

As discussed in Section II-A, it may be practically desirable to use network codes with nice algebraic properties that simplify encoding and decoding operations. Most algebraic network

codes considered in the literature are linear, and these were shown in [16] to be optimal for single session multicast.

Since the appearance of [16], it has been an open question as to whether linear network codes are in general optimal. This question was recently answered in the negative by Dougherty et. al [20]. Their proof constructs a special network containing two subnetworks such that the base fields required for optimality by each of the subnetworks have different characteristics, establishing a contradiction.

The following provides an alternative proof using a completely different approach, making use of the duality between entropy functions and achievability established in Section IV. The proof is an immediate consequence of the duality results and that some entropic functions are not almost linear group characterizable.

Implication 5 (Suboptimality of linear network codes): There is a network and a connection requirement such that the use of abelian network codes is suboptimal, including linear network codes, R–module codes, and time-sharing of such.

Proof: Consider a set of four random variables U_1, U_2, U_3, U_4 constructed using the projective plane described in [2]. The entropy function of these random variables is

$$\begin{aligned} h(1) = h(2) = h(3) = (4) &= \log 13 \\ h(1,2) &= \log 6 + \log 13 \\ h(3,4) &= \log 13 + \log 12 \\ h(1,3) = h(1,4) = h(2,3) = h(2,4) &= \log 13 + \log 4 \\ h(i,j,k) = \log 13 + \log 12 = h(1,2,3,4),\ &\forall \text{ distinct } i,j,k. \end{aligned}$$

Since h is the entropy function of a set of random variables, $\mathsf{T}(h)$ is achievable, by Theorem 2. Since h does not satisfy the Ingleton inequality

$$\begin{aligned} h(1,2) + h(1,3) + h(1,4) + h(2,3) + h(2,4) \geq \\ h(1) + h(2) + h(3,4) + h(1,2,3) + h(1,2,4), \end{aligned} \quad (30)$$

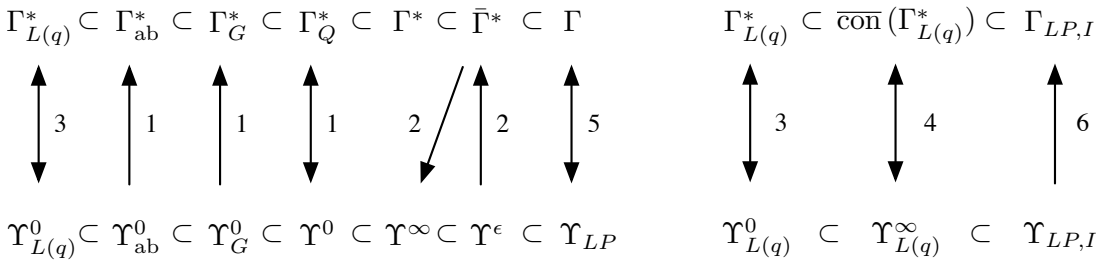
h is not almost linear group characterizable. By Theorem 4, $\mathsf{T}(h)$ is not achievable by linear network codes. ■

Implication 6 (Suboptimality of abelian group network codes): There is a network and a multicast requirement for which abelian codes are (asymptotically) suboptimal.

Proof: All abelian group characterizable entropy function must satisfy the Ingleton inequality. The corollary then follows. ■

VI. Conclusion

Entropy functions and network coding are already closely connected, through the network coding capacity region which is expressed in terms of Γ^*. The main results of this paper, summarized in Figure 4, further strengthens this connection. Figure 4 shows the inclusion relationships of the various sets of interest, as well as the implications between set membership of h and $\mathsf{T}(h)$ established by the theorems. Each arrow is labeled by the Theorem number which establishes the relation. Note that the relation of $\overline{\text{con}}(\Gamma^*_{L(q)})$ to sets other than $\Gamma^*_{L(q)}$ shown in Figure 4(a) is unknown, hence the linear code relationships are shown separately in Figure 4(b).



(a) (b) Linear codes.

Fig. 4. Summary of the duality results.

Given a non-negative real function g whose domain consists of all non-empty subsets of N random variables, we have provided a construction for a network and a connection requirement such that a rate-capacity tuple is achievable if and only if g is almost entropic (i.e. satisfies every information inequality). The network topology depends only on the number of random variables, and not on the function g, which affects the construction only through the assignment of source rates and link capacities.

An extension of this result shows that a rate-capacity tuple for the constructed multicast problem is achievable by linear network codes if and only if the entropy function g is almost

linear group characterizable. A further extension shows that the induced rate-capacity tuple satisfies the linear programming bound if and only if the function g is a polymatroid (i.e. satisfies all Shannon-type inequalities). This extension is obtained using the concept of pseudo-variables, which replace random variables in the domain of g. These pseudo-variables are abstract objects that do not take any values, and are not associated with any probability distribution. The key is that polymatroids defined over set of pseudo-variables behave very similar to entropy functions, except that they lie in Γ rather than Γ^*. This definition of pseudo-variables is not just a matter of terminology. It is a non-trivial matter to generalize notions of extension and adhesion of random variables (which rely on the existence of a probability distribution) to pseudo-variables. We provided some examples of such extensions and adhesions, which leaves the proof of the main theorem intact under a substitution of pseudo-variables for random variables. We anticipate that this concept of pseudo-variables, and their differences from random variables, may yet bear more fruit in uncovering the structure of Γ^*

The seemingly simple duality between entropy vectors and network codes has a number of powerful implications. It renders the problems of network code solubility is at least as hard as determination of $\bar{\Gamma}^*$. We also obtain alternate proofs that the LP bound is not tight, and that non-Shannon inequalities such as the Zhang-Yeung inequality indeed tighten the LP bound. However no additional finite number of inequalities can improve the LP bound to the capacity region. Finally, we have proved the suboptimality of abelian network codes, including linear codes, R-module codes and any scheme that time-shares between such codes. The duality result also provides a tool to compare different classes of network codes. Rather than comparing the codes directly, one can now compare the sets of entropy functions induced by the codes.

ACKNOWLEDGEMENT

This work was supported in part by the Australian Government under ARC grant DP0557310, and by the Defence Science and Technology Organisation under contracts 4500485167 and 4500550654.

APPENDIX I
PROOF FOR CONVERSE OF THEOREM 5

Before we prove the direct part of Theorem 5, we will prove some intermediate results which show how to *extend* sets of pseudo-variables (build new pseudo-variables from old ones), and how to *adhere* additional pseudo-variables to a given set of pseudo-variables (consistently join two sets of pseudo-variables). These results are provided in Section I-A. The proof of Theorem 5 follows in Section I-B.

A. Adhesion and extension for pseudo-variables

For random variables, adhesion or extension is facilitated by the existence of an underlying probability distribution. For example, consider two sets of random variables $\mathcal{L} = \{X, U\}$ and $\mathcal{L}^* = \{X, W\}$ with respective underlying distributions P_{XU} and P^*_{XW}. Suppose that the marginals over X coincide, $P_X = P^*_X$. We can then easily adhere P_{XU} and P^*_{XW} to obtain a new distribution Q_{XUW} such that its marginals over $\mathcal{L}$ and $\mathcal{L}^*$ coincide, $Q_{XU} = P_{XU}$ and $Q_{XW} = P^*_{XW}$. One possibility is $Q_{XUV} = P_{XU}P^*_{XW}/P_X$. In general, for any sets of random variables $\mathcal{L}$ and $\mathcal{L}^*$ with respective distributions P and P^* coinciding on $\mathcal{L} \cap \mathcal{L}^*$, we can construct a new distribution over $\mathcal{L} \cup \mathcal{L}^*$ such that its marginals over $\mathcal{L}$ and $\mathcal{L}^*$ are P and P^*. Clearly, the entropy function for $\mathcal{L} \cup \mathcal{L}^*$ is an extension of those belonging to $\mathcal{L}$ and $\mathcal{L}^*$.

Consider another simple example. Let $\mathcal{A} \subset \mathcal{L}$ be a subset of the random variables $\mathcal{L}$. Then we can define a new random variable $W \triangleq \mathcal{A}$. By doing so, we have constructed a new variable, and extended both the distribution and entropy function. Clearly there are various ways to adhere or extend sets of random variables. Doing this for pseudo-variables is not so straightforward. The following results provide several adhesion and extension methods for pseudo-variables.

Lemma 10 (Functional extension): Let $\mathcal{L}$ be a set of pseudo-variables. For any given $\mathcal{A} \subseteq \mathcal{L}$, one can adhere a new pseudo-variable Y to $\mathcal{L}$ such that $H(Y|\mathcal{A}) = H(\mathcal{A}|Y) = 0$. In other words, there exists a polymatroid g over $\mathcal{L} \cup \{Y\}$ satisfying

$$g(\mathcal{B}) = H(\mathcal{B}) \quad \forall \mathcal{B} \subseteq \mathcal{L} \tag{31}$$

$$g(Y) = g(\mathcal{A}) = g(\{Y\} \cup \mathcal{A}). \tag{32}$$

Proof: Define g over $\mathcal{L} \cup \{Y\}$ such that for all $\mathcal{B} \subseteq \mathcal{L}$,

$$g(\mathcal{B}) = H(\mathcal{B}) \text{ and } g(\{Y\} \cup \mathcal{B}) = H(\mathcal{B} \cup \mathcal{A}). \tag{33}$$

It is straightforward to show that g is a polymatroid satisfying (31) and (32). ■

In light of Definition 12, we shall refer to (33) as functional extension and denote the new variable as $J_{\mathcal{A}}$. Clearly, any subset of pseudo-variables in $\mathcal{A}$ is a function of $J_{\mathcal{A}}$.

Lemma 11 (Sum extension): Let $\{X, Y\}$ be a set of pseudo-variables such that $H(X) = H(Y)$ and $X \perp Y$. Then one can adhere a new pseudo-variable Z to $\{X, Y\}$ such that $H(Z) = H(X)$ and $H(Z|X, Y) = H(X|Y, Z) = H(Y|X, Z) = 0$.

Proof: Let g be the pseudo-entropy function for $\{X, Y\}$. Extend g such that $g(Z) = g(X)$ and $g(X, Z) = g(Y, Z) = g(X, Y, Z) = g(X, Y)$. The resulting extended g is still a polymatroid. ■

Lemma 11 shows that for any independent pseudo-variables X and Y of equal pseudo-entropies, one can construct a pseudo-variable Z, denoted $Z = X \oplus Y$ such that its pseudo-entropy is the same as X and Y, and any single pseudo-variable is a function of the two others. Structurally, this mimics the modulo-2 addition of two i.i.d binary random variables.

Lemma 12 (SW extension): Let $\{X, Y\}$ be two pseudo-variables. Then one can adhere a new pseudo-variable Z to $\{X, Y\}$ such that

$$H(Z) = H(X|Y),$$

$$H(X|Z, Y) = 0,$$

$$H(Z|X) = 0.$$

Proof: Let g be the pseudo-entropy of $\{X, Y\}$ and extend it as follows: $g(Z) = g(X, Y) - g(Y)$, $g(Z, Y) = g(X, Y, Z) = g(X, Y)$, and $g(X, Z) = g(X)$. The resulting extended g is still a polymatroid. ■

Lemma 12 shows that starting with pseudo-variables X, Y, one can construct another pseudo-variable Z with pseudo-entropy $H(X, Y) - H(Y)$ such that X is a function of Y, Z and Z is a function of X. For simplicity, we use the symbol $J_{X|Y}$ to denote the new pseudo-variable Z.

Lemmas 10–12 show that sets of pseudo-variables can be explicitly extended to obtain new pseudo-variables. In the following, we study adhesion of existing sets of pseudo-variables.

Lemma 13 (Independent adhesion): Let $\mathcal{L}$ and $\mathcal{L}^*$ be two disjoint sets of pseudo-variables. Then they can adhere to each other *independently* such that for any $\mathcal{A} \subseteq \mathcal{L} \cup \mathcal{L}^*$,

$$H(\mathcal{A}) = H(\mathcal{A} \cap \mathcal{L}) + H(\mathcal{A} \cap \mathcal{L}^*). \tag{34}$$

Proof: Let g and g^* be the pseudo-entropies of $\mathcal{A}$ and $\mathcal{A}^*$, and for each $\mathcal{A} \subseteq \mathcal{L} \cup \mathcal{L}^*$ set $g(\mathcal{A}) = g(\mathcal{A} \cap \mathcal{L}) + g^*(\mathcal{A} \cap \mathcal{L}^*)$. It can be verified that g is a polymatroid. ■

Any subsets $\mathcal{A} \subseteq \mathcal{L}$ and $\mathcal{B} \subseteq \mathcal{L}^*$ are independent, $\mathcal{A} \perp \mathcal{B}$ under the independent adhesion of $\mathcal{L}$ and $\mathcal{L}^*$ in Lemma 13. Before we continue with more complicated adhesions, we need the following proposition from [31].

Proposition 6: Let $\mathcal{L}$ and $\mathcal{L}^*$ be two sets of pseudo-variables coinciding over $\mathcal{L}' \triangleq \mathcal{L} \cap \mathcal{L}^*$, i.e. for all $\mathcal{A} \subseteq \mathcal{L}'$, the pseudo-entropy of $\mathcal{A}$ is the same with respect $\mathcal{L}$ and $\mathcal{L}^*$. Further, suppose

$$\Delta(\mathcal{A}, \mathcal{B}) \geq \Delta(\mathcal{L}' \cap \mathcal{A}, \mathcal{L}' \cap \mathcal{B}), \tag{35}$$

for all flats[4] $\mathcal{A}, \mathcal{B}$ of $\mathcal{L}$ where $\Delta(\mathcal{A}, \mathcal{B}) \triangleq H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A} \cup \mathcal{B}) - H(\mathcal{A} \cap \mathcal{B})$. Then $\mathcal{L}$ and $\mathcal{L}^*$ can adhere to each other.

Proof: See Theorem 1 in [31]. ■

Corollary 1: Let $\mathcal{L} = \{X, Y, Z\}$ be a set of pseudo-variables, such that Z is a function of X, Y and X is a function of Y, Z. Let $\mathcal{L}^*$ be another set of pseudo-variables such that $\mathcal{L}$ and $\mathcal{L}^*$ coincide over $\mathcal{L} \bigcap \mathcal{L}^* = \{X, Y\}$. Then $\mathcal{L}^*$ and $\mathcal{L}$ can adhere to each other.

Proof: It is easy to verify that $\{X, Y\}$ and $\{Y, Z\}$ cannot be flats of $\mathcal{L}$. To prove the corollary, it suffices to prove that (35) is satisfied for all flats of $\mathcal{L}$.

Suppose that $\mathcal{A}$ and $\mathcal{B}$ are flats of $\mathcal{L}$. If either $\mathcal{A}$ or $\mathcal{B}$ is the empty set, $\{Z\}$ or $\{X, Y, Z\}$, then either $\mathcal{L}' \cap \mathcal{A} \subseteq \mathcal{L}' \cap \mathcal{B}$ or $\mathcal{L}' \cap \mathcal{B} \subseteq \mathcal{L}' \cap \mathcal{A}$. As a result, $\Delta(\mathcal{L}' \cap \mathcal{A}, \mathcal{L}' \cap \mathcal{B}) = 0$ and (35) holds. On the other hand, if both $\mathcal{A}$ and $\mathcal{B}$ are subsets of $\{X, Y\}$, then it is obvious that (35) remains true. Now, suppose $\mathcal{A} = \{X, Z\}$. Then (35) holds for $\mathcal{B} = \{X\}$ or $\{X, Z\}$. Finally, when $\mathcal{A} = \{X, Z\}$ and $\mathcal{B} = \{Y\}$, by direct verification, (35) still holds. Combining all the cases, we see that (35) indeed holds for all flats of $\mathcal{L}$. ■

Corollary 1 directly leads to the following result.

Theorem 7: Let $\mathcal{L}^* \supseteq \{X, Y\}$. Then one can adhere the pseudo-variable $Z = J_{X|Y}$ to $\mathcal{L}^*$.

If in addition $H(X) = H(Y)$, it is possible adhere a pseudo-variable $Z = X \oplus Y$ to $\mathcal{L}^*$.

B. Proof for direct part of Theorem 5

Proof: To prove the direct part, we must exhibit a set of pseudo-variables satisfying the set of (in)equalities (5). Our construction works as follows:

[4]A subset $\mathcal{A}$ of the ground set $\mathcal{L}$ is a *flat* if $H(\mathcal{A}') > H(\mathcal{A})$ for all proper supersets $\mathcal{A}'$ containing $\mathcal{A}$.

- Let $V_1, \ldots, V_N$ be pseudo-variables whose pseudo-entropy function is h.
- By Lemma 10, we can adhere $S_{[\mathcal{N}]} \triangleq J_{\mathcal{L}}$ to $\mathcal{L} = \{V_1, \ldots, V_n\}$.
- For any non-empty subset α of $\mathcal{N}$, let $S_{[\mathcal{N}]}$ be a pseudo-variable whose pseudo-entropy is $H(V_\alpha)$.
- By Lemma 13, we adhere independent pseudo-variables $S_{[\alpha]}$ to the current set of pseudo-variables $\{V_1, \ldots, V_N, S_{[\mathcal{N}]}\}$.
- By Theorem 7, we can further adhere auxiliary pseudo-variables such as J_{V_α}, $J_{S_{[\mathcal{N}]}|J_{V_\alpha}}$, $J_{V_\alpha} \oplus S_{[\alpha]}$ etc.

Now, we will show how to associate pseudo-variables to edges. If the edge is uncapacitated, then the associated pseudo-variable is the join of the set of pseudo-variables incident to that edge. It remains to show that for the three subnetworks, we can adhere pseudo-variables meeting all the constraints of the LP bound.

Consider type 0 subnetworks. Let $W = S_{[\alpha]}$. Then, (5) clearly holds. In type 1 subnetworks let $W = J_{S_{[\mathcal{N}]}|J_{V_\alpha}}$ and $W' = J_{V_\alpha}$. Again, (5) holds. Finally, for type 2 subnetworks, let $W = S_{[\alpha]} \oplus J_{V_\alpha}$, $W' = J_{S_{[\mathcal{N}]}|J_{V_\alpha}}$, $W'' = J_{J_{V_\alpha}|V_i}$, and $W^* = W^{**} = J_{V_\alpha}$. By direct verification, the set of (in)equalities (5) holds. ■

REFERENCES

[1] R. Yeung, *A First Course in Information Theory*. Kluwer Academic/Plenum Publishers, 2002.

[2] Z. Zhang and R. W. Yeung, "On the characterization of entropy function via information inequalities," *IEEE Trans. Inform. Theory*, vol. 44, pp. pp. 1440–1452, 1998.

[3] ——, "A non-Shannon-type conditional information inequality of information quantities," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1982–1986, Nov. 1997.

[4] R. W. Yeung and Z. Zhang, "A class of non-Shannon-type information inequalities and their applications," *Communications in Information and Systems*, vol. 1, pp. 87–100, 2001.

[5] ——, "A class of non-Shannon-type information inequalities and their applications," in *IEEE Int. Symp. Inform. Theory*, Washington, DC, 2001, p. 231.

[6] I. Sason, "Identification of new classes of non-Shannon type constrained information inequalities and their relation to finite groups," in *IEEE Int. Symp. Inform. Theory*, Lausanne, Switzerland, 2002, p. 236.

[7] F. Matúš, "Piecewise linear conditional information inequality," *IEEE Trans. Inform. Theory*, vol. 52, pp. 236–238, Jan. 2006.

[8] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin, "A new class of non-Shannon-type inequalities for entropies," *Communications in Information and Systems*, vol. 2, no. 2, pp. 147–165, Dec. 2002.

[9] R. Dougherty, C. Freiling, and K. Zeger, "Six new non-Shannon information inequalities," in *IEEE Int. Symp. Inform. Theory*, July 2006, pp. 233–236.

[10] F. Matúš, "Infinitely many information inequalities," in *IEEE Int. Symp. Inform. Theory*, 2007.

[11] R. W. Yeung and Z. Zhang, "Distributed source coding for satellite communications," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1111–1120, May 1999.

[12] T. H. Chan and R. W. Yeung, "On a relation between information inequalities and group theory," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1992–1995, 2002.

[13] D. Hammer, A. E. Romashchenko, A. Shen, and N. K. Vereshchagin, "Inequalities for Shannon entropy and Kolmogorov complexity," *J. Comp. Syst. Sci.*, vol. 60, pp. 442–464, 2000.

[14] A. Romashchenko, N. Vereshchagin, and A. Shen, "Combinatorial interpretation of Kolmogorov complexity," in *15th Annual IEEE Conf. Computational Complexity*, Florence, Italy, July 2000, pp. 131–137.

[15] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[16] S.-Y. R. Li, R. Yeung, and N. Cai, "Linear network coding," *IEEE Trans. Inform. Theory*, vol. 49, no. 2, pp. 371–381, Feb. 2003.

[17] A. F. Dana, R. Gowaikar, R. Palanki, B. Hassibi, and M. Effros, "Capacity of wireless erasure networks," *IEEE Trans. Inform. Theory*, vol. 52, pp. 789–804, March 2006.

[18] N. Cai and R. W. Yeung, "Secure network coding," in *IEEE Int. Symp. Inform. Theory*, Lausanne, Switzerland, 2002, p. 323.

[19] D. S. Lun, N. Ratanakar, M. Médard, R. Koetter, D. R. Karger, T. Ho, and E. Ahmed, "Minimum-cost multicast over coded packet networks," *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2608–2623, Jun. 2006.

[20] R. Dougherty, C. Freiling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2745–2759, Aug. 2005.

[21] L. Song, R. Yeung, and N. Cai, "Zero-error network coding for acyclic networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3129–3139, Dec. 2003.

[22] R. W. Yeung, S.-Y. R. Li, N. Cai, and Z. Zhang, *Network Coding Theory*, ser. Foundations and Trends in Communications and Information Theory. Now Publishers, 2006.

[23] X. Yan, R. W. Yeung, and Z. Zhang, "The capacity region for multi-source multi-sink network coding," in *IEEE Int. Symp. Inform. Theory*, Nice, France, Jun. 2007, pp. 116–120.

[24] R. Dougherty, C. Freiling, and K. Zeger, "Matroids, networks, and non-Shannon information inequalities," *IEEE Trans. Inform. Theory*, 2007.

[25] T. H. Chan, "A combinatorial approach to information inequalities," *Communications in Information and Systems*, vol. 1, pp. 1–14, 2001.

[26] ——, "Aspects of information inequalities and its applications," Master's thesis, The Chinese University of Hong Kong, 1998.

[27] ——, "Group characterizable entropy functions," 2007, arxiv.org/cs.IT/0702064.

[28] ——, "Capacity regions for linear and abelian network code," in *NETCOD*, San Diego, USA, 2007.

[29] ——, "Capacity region of probabilistic network codes," in *Canadian Workshop Inform. Theory*, Montreal, Canada, June 2005, pp. 167–170.

[30] R. Yeung, "A framework for linear information inequalities," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1924–1934, Nov. 1997.

[31] F. Matúš, "Adhesivity of polymatroids," *Discrete Math.*, 2007.