Beyond Network Equivalence

Beyond Network Equivalence

Ralf Koetter, Michelle Effros, and Muriel Médard

Abstract— In earlier work, we described an equivalence result for network capacity. Roughly, that result is as follows. Given a network of noisy, memoryless, point-to-point channels, replace each channel by a noiseless, memoryless bit pipe of the same capacity yields a new network such that any collection of demands can be met on the network of noisy links if and only if the same demands can be met on the network of noiseless links. We here expand on these ideas to provide a framework for studying more general networks, including networks containing memoryless broadcast, multiple access, and interference channels. For each network in this broader class, we design two corresponding networks of noiseless, memoryless point-to-point links. These two networks provide upper and lower bounds in the following sense. Fix an arbitrary collection of demands. If the given demands can be met on the lower bounding network, then they can also be met on the original network. Likewise, if the given demands can be met on the original network, then they can also be met on the upper bounding network.

I. INTRODUCTION

In [1], we introduced a new tool for investigating the relationship between the two predominant approaches to studying communications through networks. In the first of these approaches, networks are treated as graphs comprised of nodes and the noise-free, capacitated links that connect them. A different approach is provided by multiterminal information theory, which deals with noisy channels, or rather the stochastic relationship of input and output variables at nodes in a network. As a first step towards establishing a connection between these seemingly disparate views, we established an equivalence relationship in the special case of a network of point-to-point links [1]. Roughly, that result is as follows. An arbitrary collection of demands can be met on a network of noisy, independent, memoryless links if and only if it can be met on another network where each noisy link is replaced by a noiseless bit pipe with capacity equal to the noisy link capacity.

We here apply the same approach to study more general stochastic components such as multiple access, broadcast, and interference channels. While there appears to be no obvious relationship between these canonical networks and

networks of error free bit pipes, our goal is to develop bit-pipe models for a broad library of components in such a way that the capacity of any network comprised of the given stochastic components can be bounded by the capacity of the network of noiseless bit pipes achieved by replacing each stochastic component by its corresponding bit-pipe model.

While existing achievability results for stochastic network models can be used to find bit-pipe models useful in bounding capacities from below, existing converses on network components like broadcast and multiple access channels fail to provide upper bounds on the behavior of those components when employed in larger networks. That is, it is sometimes necessary to operate component channels strictly above their individual capacities in order to obtain the capacities of the larger networks in which they are employed.

Given a library of bit-pipe models for stochastic components, bounding the capacity for any network comprised of these components is equivalent to finding the network coding capacity for a corresponding network of error-free bit pipes. In certain situations, most notably a multicast demand, the network capacity has nice and simple answers. Unfortunately, the general case is wide open. In fact, it is suspected that the central combinatorial network coding problem is hard, though NP hardness is only established for the class of linear network coding [2]. Nevertheless, we can solve this problem for small networks. This possibility to characterize, in principle, the rate region of a combinatorial network coding problem will be a corner stone for our investigations.

The situation is not unlike issues in complexity theory, where a lot of research is devoted to show that one problem is essentially as difficult as another one without being able to give precise expressions as to how difficult a problem is in absolute terms. Inspired by this analogy, we resort in this paper to characterizing the capacity of one network communication problem in terms of capacities of the central combinatorial network coding problem. This characterization lends insights to a variety of questions such as degree-of-freedom or high signal to noise ratio analysis.

The reduction of a network information theoretic question to its combinatorial essence is also at the heart of recent related publications, see e.g. [3]. While our approach and results are quite different, we believe it to be no coincidence that in both cases the reduction of a problem to its combinatorial essence is a central step.

II. THE SETUP

Our notation is similar to that of Cover and Thomas [4]. A multiterminal network comprises m nodes with associated

random variables $X^{(i)} \in \mathcal{X}^{(i)}$ which are transmitted from node i and $Y^{(i)} \in \mathcal{Y}^{(i)}$ which are received at node i. Both $\mathcal{X}^{(i)}$ and $\mathcal{Y}^{(i)}$ may be either discrete or continuous. The network is assumed to be memoryless, so it is characterized by a conditional probability distribution $p(\mathbf{y}|\mathbf{x}) = p(y^{(1)}, \ldots, y^{(m)}|x^{(1)}, \ldots, x^{(m)})$. A code of blocklength n operates the network over n time steps with the goal of communicating, for each distinct pair of nodes i and j, message $W^{(i \to j)} \in \mathcal{W}^{(i \to j)} \stackrel{\text{def}}{=} \{1, \ldots, 2^{nR^{(i \to j)}}\}$ from source node i to sink node j; $R^{(i \to j)}$ is called the rate of the transmission. The vector of rates $R^{(i \to j)}$ is denoted by $\mathcal{R}$.[1] We denote the random variables transmitted and received at node i at time-step t by $X_t^{(i)}$ and $Y_t^{(i)}$. A network is thus written as a triple $\left(\prod_{i=1}^{m} \mathcal{X}^{(i)}, p(\mathbf{y}|\mathbf{x}), \prod_{i=1}^{m} \mathcal{Y}^{(i)}\right)$ with the additional constraint that random variable $X_t^{(i)}$ is a function of random variables $\{Y_1^{(i)}, \ldots, Y_{t-1}^{(i)}, W^{(i \to 1)}, \ldots, W^{(i \to m)}\}$ alone.

While this characterization is very general, it does not exploit any information about the network's structure. The structure is given as a hypergraph G with node set $V = \{1, \ldots, m\}$ and hyperedge set E. Let $\mathscr{P}(V)$ denote the power set of V. Each directed hyperedge $e \in E$ takes the form $e = [V_1, V_2]$, $V_1, V_2 \in \mathscr{P}(V)$. Thus $E \subseteq \mathscr{P}(V) \times \mathscr{P}(V)$. For example, a point-to-point channel has a single transmitter $V_1 = \{i\}$ and a single receiver $V_2 = \{j\}$; a broadcast channel has a single transmitter $V_1 = \{i\}$ and multiple receivers $V_2 = \{j_1, \ldots, j_k\}$; and so on.

The indegree $d_{\text{in}}(i)$ and outdegree $d_{\text{out}}(i)$ of node i in hypergraph G are defined as $d_{\text{in}}(i) = |\{V_1|[V_1, V_2] \in E, i \in V_2\}|$ and $d_{\text{out}}(i) = |\{V_2|[V_1, V_2] \in E, i \in V_1\}|$. If a node has outdegree or indegree larger than one, then $\mathcal{X}^{(i)} = \prod_{d=1}^{d_{\text{out}}(i)} \mathcal{X}^{(i,d)}$ and $\mathcal{Y}^{(i)} = \prod_{d=1}^{d_{\text{in}}(i)} \mathcal{Y}^{(i,d)}$, and the inputs and outputs of node i at time t are given by $X_t^{(i)} = (X_t^{(i,1)}, \ldots, X_t^{(i,d_{\text{out}}(i))})$ and $Y_t^{(i)} = (Y_t^{(i,1)}, \ldots, Y_t^{(i,d_{\text{in}}(i))})$. Here $X_t^{(i,d)}$ denotes the input from node i at time t to the dth hyperedge for which node i is in the tail and $\mathcal{X}^{(i,d)}$ denotes its alphabet. Likewise, $Y_t^{(i,d)}$ denotes the output at node i at time t of the dth hyperedge for which node i is in the tail and $\mathcal{Y}^{(i,d)}$ denotes its alphabet. The orderings for the hyperedges outgoing from and incoming to node i are arbitrary but fixed. Given a hyperedge $e = [V_1, V_2]$, we use $V_1(e)$ and $V_2(e)$ to denote the input and output ports of hyperedge e and $X^{V_1(e)}$ and $Y^{V_2(e)}$ to denote the sets of input and output random variables of hyperedge e. For example, consider a link e from node i to node j. Then the input to hyperedge $e = [\{i\}, \{j\}]$ at node i is $X^{(i,s)}$ for some index $s \in \{1, \ldots, d_{\text{out}}(i)\}$ while the output from hyperedge $[\{i\}, \{j\}]$ at node j is $Y^{(j,r)}$ for some index $r \in \{1, \ldots, d_{\text{in}}(j)\}$. In this case, $V_1(e) = (i, s)$ and $V_2(e) = (j, r)$, giving $X^{V_1(e)} = X^{(i,s)}$ and $Y^{V_2(e)} = Y^{(j,r)}$ as desired. The collection of sets $\{V_1(e) : e \in E\}$ ($\{V_2(e) : e \in E\}$) forms a partition of the set of random variables $X^{(1,1)}, \ldots, X^{(m,d_{\text{out}}(m))}$ ($Y^{(1,1)}, \ldots, Y^{(m,d_{\text{in}}(m))}$).

When the characterization corresponds to network $G =$ (V, E), we factor $p(\mathbf{y}|\mathbf{x})$ to give the characterization:

$$\left(\prod_{i=1}^{m} \mathcal{X}^{(i)}, \prod_{e \in E} p(y^{V_2(e)}|x^{V_1(e)}), \prod_{i=1}^{m} \mathcal{Y}^{(i)}\right), \quad (1)$$

again with the additional property that random variable $X_t^{(i)}$ is a function of random variables $\{Y_1^{(i)}, \ldots, Y_{t-1}^{(i)}, W^{(i \to 1)}, \ldots, W^{(i \to m)}\}$ alone. As in [1], we investigate some information theoretic aspects of replacing factors in the factorization of $p(\mathbf{y}|\mathbf{x})$.

Remark 1: In some situations it is important to be able to embed the transmissions of various nodes in a schedule which may or may not depend on the messages to be sent and the symbols that were received in the network. It is straightforward to model such a situation in the above setup by including in the input and output alphabets symbols for the case when nothing was sent on a particular link. In this way we can assume that at each time t random variables $X_t^{(i)}$ and $Y_t^{(i)}$ are given.

Definition 1: Let a network

$$\mathcal{N} \stackrel{\text{def}}{=} \left(\prod_{i=1}^{m} \mathcal{X}^{(i)}, \prod_{e \in E} p(y^{V_2(e)}|x^{V_1(e)}), \prod_{i=1}^{m} \mathcal{Y}^{(i)}\right)$$

be given corresponding to a hypergraph $G = (V, E)$. A blocklength-n solution $\mathcal{S}(\mathcal{N})$ to this network is defined as a set of encoding and decoding functions:

$$X_t^{(i)} : \left(\mathcal{Y}^{(i)}\right)^{t-1} \times \prod_{k=1}^{m} \mathcal{W}^{(i \to k)} \to \mathcal{X}^{(i)}$$
$$\hat{W}^{(j \to i)} : \left(\mathcal{Y}^{(i)}\right)^{n} \times \prod_{k=1}^{m} \mathcal{W}^{(i \to k)} \to \mathcal{W}^{(j \to i)}$$

mapping $(Y_1^{(i)}, \ldots, Y_{t-1}^{(i)}, W^{(i \to 1)}, \ldots, W^{(i \to m)})$ to $X_t^{(i)}$ for each $i \in V$ and $t \in \{1, \ldots, n\}$ and mapping $(Y_1^{(i)}, \ldots, Y_n^{(i)}, W^{(i \to 1)}, \ldots, W^{(i \to m)})$ to $\hat{W}^{(j \to i)}$ for each $i, j \in V$. The solution $\mathcal{S}(\mathcal{N})$ is called a $(\lambda, \mathcal{R})$-solution, denoted $(\lambda, \mathcal{R}) - \mathcal{S}(\mathcal{N})$, if the specified encoding and decoding functions imply $\Pr(W^{(i \to j)} \neq \hat{W}^{(i \to j)}) < \lambda$ for all source and sink pairs i, j.

Definition 2: The rate region $\mathscr{R}(\mathcal{N}) \subset \mathbb{R}_+^{m(m-1)}$ of a network $\mathcal{N}$ is the closure of all rate vectors $\mathcal{R}$ such that for any $\lambda > 0$, a $(\lambda, \mathcal{R}) - \mathcal{S}(\mathcal{N})$ solution exists.[2]

Since we do not know the properties of an optimal solution for a network $\mathcal{N}$, we cannot eliminate the possibility that such solutions vary with time (due, for example, to feedback). As a result, we cannot apply the usual typicality arguments across time. Instead, we create multiple, independent copies of a network – here viewed as layers in a stack of network copies – and apply typicality arguments across layers of the stack.

Given a network $\mathcal{N}$ defined on hypergraph G, the N-fold stacked network $\underline{\mathcal{N}}$ is defined on N copies of G. Since the vertex and edge sets of $\underline{\mathcal{N}}$ are multisets and not sets, the stacked network is not a network, and new definitions are required. We carry over notation and variable

---

[1] Nodes do not transmit messages to themselves. For the sake of brevity, we do not make this explicit in our notation.

[2] In this and later definitions of rate regions, the blocklength n required to achieve a solution with error probability λ need not be a constant. It suffices that each error probability can be achieved with some (possibly distinct) blocklength.

definitions from the network $\mathcal{N}$ to the stacked network $\underline{\mathcal{N}}$ by underlining the variable names. So for any distinct $i, j \in V$, $\underline{W}^{(i \to j)} \in \underline{\mathcal{W}}^{(i \to j)} \stackrel{\text{def}}{=} (\mathcal{W}^{(i \to j)})^N$ is the N-dimensional vector of messages that the copies of node i send to the corresponding copies of node j, and $\underline{X}_t^{(i)} \in \underline{\mathcal{X}}^{(i)} \stackrel{\text{def}}{=} \mathcal{X}^N$ and $\underline{Y}_t^{(i)} \in \underline{\mathcal{Y}}^{(i)} \stackrel{\text{def}}{=} \mathcal{Y}^N$ are the N-dimensional vectors of channel inputs and channel outputs, respectively, for node i at time t. The variables in the ℓ-th layer of the stack are denoted by an argument ℓ, for example $\underline{W}^{(i \to j)}(\ell)$ is the message from node i to node j in the ℓ-th layer of the stack and $\underline{X}_t^{(i)}(\ell)$ is the layer-ℓ channel input from node i at time t. Since $\underline{W}^{(i \to j)}$ is an N-dimensional vector of messages, when $W^{(i \to j)} \in \mathcal{W}^{(i \to j)} \stackrel{\text{def}}{=} \{1, \ldots, 2^{nR}\}$, $\underline{W}^{(i \to j)} \in \underline{\mathcal{W}}^{(i \to j)} \stackrel{\text{def}}{=} \{1, \ldots, 2^{nR}\}^N$. We therefore define the rate $R^{(i \to j)}$ for a stacked network to be $(\log |\underline{\mathcal{W}}^{(i \to j)}|)/(nN)$; this normalization makes the rate of a network and its corresponding stacked network comparable.

Definition 3: Let a network

$$\mathcal{N} \stackrel{\text{def}}{=} \left(\prod_{i=1}^m \mathcal{X}^{(i)}, \prod_{e \in E} p(y^{V_2(e)} | x^{V_1(e)}), \prod_{i=1}^m \mathcal{Y}^{(i)} \right)$$

be given corresponding to a hypergraph $G = (V, E)$. Let $\underline{\mathcal{N}}$ be the N-fold stacked network for $\mathcal{N}$. A blocklength-n solution $\mathcal{S}(\underline{\mathcal{N}})$ to this network is defined as a set of encoding and decoding functions

$$\underline{X}_t^{(i)} : (\underline{\mathcal{Y}}^{(i)})^{t-1} \times \prod_{k=1}^m \underline{\mathcal{W}}^{(i \to k)} \to \underline{\mathcal{X}}^{(i)}$$
$$\underline{\hat{W}}^{(j \to i)} : (\underline{\mathcal{Y}}^{(i)})^n \times \prod_{k=1}^m \underline{\mathcal{W}}^{(i \to k)} \to \underline{\mathcal{W}}^{(j \to i)}$$

mapping

$$(\underline{Y}_1^{(i)}, \ldots, \underline{Y}_{t-1}^{(i)}, \underline{W}^{(i \to 1)}, \ldots, \underline{W}^{(i \to m)})$$

to $\underline{X}_t^{(i)}$ for each $t \in \{1, \ldots, n\}$ and $i \in V$ and mapping $(\underline{Y}_1^{(i)}, \ldots, \underline{Y}_n^{(i)}, \underline{W}^{(i \to 1)}, \ldots, \underline{W}^{(i \to m)})$ to $\underline{\hat{W}}^{(j \to i)}$ for each $i, j \in V$. The solution $\mathcal{S}(\underline{\mathcal{N}})$ is called a $(\lambda, \mathcal{R})$-solution for $\underline{\mathcal{N}}$, denoted $(\lambda, \mathcal{R}) - \mathcal{S}(\underline{\mathcal{N}})$, if the encoding and decoding functions imply

$$\Pr(\underline{W}^{(i \to j)} \neq \underline{\hat{W}}^{(i \to j)}) < \lambda$$

for all source and sink pairs i, j.

Definition 4: The rate region $\mathscr{R}(\underline{\mathcal{N}}) \subset \mathbb{R}_+^{m(m-1)}$ of a stacked network $\underline{\mathcal{N}}$ is the closure of all rate vectors $\mathcal{R}$ such that a $(\lambda, \mathcal{R}) - \mathcal{S}(\underline{\mathcal{N}})$ solution exists for any $\lambda > 0$ and all N sufficiently large.

Theorem 1 from [1] shows that the rate regions for a network $\mathcal{N}$ and its corresponding stacked network $\underline{\mathcal{N}}$ are identical. That result further demonstrates that the error probability for the stacked network can be made to decay exponentially in the number of layers N and that for any rate $\mathcal{R} \in \text{int}(\mathscr{R}(\underline{\mathcal{N}}))$ there exists a $(2^{-N\delta}, \mathcal{R}) - \mathcal{S}(\underline{\mathcal{N}})$ solution for $\underline{\mathcal{N}}$ that first channel codes each message $\underline{W}^{(i \to j)}$ and then sends the channel coded description $\underline{\tilde{W}}^{(i \to j)}$ of $\underline{W}^{(i \to j)}$ through the stacked network *using the same solution $\mathcal{S}(\mathcal{N})$ independently in each layer of $\underline{\mathcal{N}}$*. We call a solution of this type a stacked solution.

Definition 5: Let a network

$$\mathcal{N} \stackrel{\text{def}}{=} \left(\prod_{i=1}^m \mathcal{X}^{(i)}, \prod_{e \in E} p(y^{V_2(e)} | x^{V_1(e)}), \prod_{i=1}^m \mathcal{Y}^{(i)} \right)$$

be given corresponding to a hypergraph $G = (V, E)$. Let $\underline{\mathcal{N}}$ be the N-fold stacked network for $\mathcal{N}$. A blocklength-n stacked solution $\mathcal{S}(\underline{\mathcal{N}})$ to this network is defined as a set of mappings:

$$\tilde{\underline{W}}^{(i \to j)} : \underline{\mathcal{W}}^{(i \to j)} \to \tilde{\underline{\mathcal{W}}}^{(i \to j)}$$
$$X_t^{(i)} : (\mathcal{Y}^{(i)})^{t-1} \times \prod_{k=1}^m \tilde{\mathcal{W}}^{(i \to k)} \to \mathcal{X}^{(i)}$$
$$\hat{\tilde{W}}^{(j \to i)} : (\mathcal{Y}^{(i)})^n \times \prod_{k=1}^m \tilde{\mathcal{W}}^{(i \to k)} \to \tilde{\mathcal{W}}^{(j \to i)}$$
$$\underline{\hat{W}}^{(i \to j)} : \tilde{\underline{\mathcal{W}}}^{(i \to j)} \to \underline{\mathcal{W}}^{(i \to j)}$$

such that

$$\tilde{\underline{W}}^{(i \to j)} = \tilde{\underline{W}}^{(i \to j)}(\underline{W}^{(i \to j)})$$
$$\underline{X}_t^{(i)}(\ell) = X_t^{(i)} \left(\underline{Y}_1^{(i)}(\ell), \ldots, \underline{Y}_{t-1}^{(i)}(\ell), \right.$$
$$\left. \tilde{\underline{W}}^{(i \to 1)}(\ell), \ldots, \tilde{\underline{W}}^{(i \to m)}(\ell) \right)$$
$$\hat{\tilde{\underline{W}}}^{(j \to i)}(\ell) = \hat{\tilde{W}}^{(j \to i)} \left(\underline{Y}_1^{(i)}(\ell), \ldots, \underline{Y}_n^{(i)}(\ell), \right.$$
$$\left. \tilde{\underline{W}}^{(i \to 1)}(\ell), \ldots, \tilde{\underline{W}}^{(i \to m)}(\ell) \right)$$
$$\underline{\hat{W}}^{(i \to j)} = \underline{W}^{(i \to j)}(\hat{\tilde{\underline{W}}}^{(i \to j)})$$

for each $i, j \in V$, $t \in \{1, \ldots, n\}$, and $\ell \in \{1, \ldots, N\}$. The solution $\underline{\mathcal{S}}(\underline{\mathcal{N}})$ is called a stacked $(\lambda, \mathcal{R})$-solution, denoted $(\lambda, \mathcal{R}) - \underline{\mathcal{S}}(\underline{\mathcal{N}})$, if the specified mappings imply

$$\Pr(\underline{W}^{(i \to j)} \neq \underline{\hat{W}}^{(i \to j)}) < \lambda$$

for all source and sink pairs i, j.

Theorem 1: [1, Theorem 1] The rate regions $\mathscr{R}(\mathcal{N})$ and $\mathscr{R}(\underline{\mathcal{N}})$ are identical. Further, there exist stacked solutions of type $(2^{-N\delta}, \mathcal{R}) - \underline{\mathcal{S}}(\underline{\mathcal{N}})$ for each $\mathcal{R} \in \text{int}(\mathscr{R}(\underline{\mathcal{N}}))$.

Since the proof of Theorem 1 shows that stacked solutions can obtain all rates in the interior of $\mathscr{R}(\underline{\mathcal{N}})$, we restrict our attention to stacked solutions going forward; there is no loss of generality in this restriction.

Using a stacked solution, if $\tilde{\underline{W}}^{(i \to j)}(1), \ldots, \tilde{\underline{W}}^{(i \to j)}(N)$ are independent and identically distributed (i.i.d.), then for each time t

$$(\underline{X}_t^{(1)}(1), \ldots, \underline{X}_t^{(m)}(1), \underline{Y}_t^{(1)}(1), \ldots, \underline{Y}_t^{(m)}(1)), \ldots,$$
$$(\underline{X}_t^{(1)}(N), \ldots, \underline{X}_t^{(m)}(N), \underline{Y}_t^{(1)}(N), \ldots, \underline{Y}_t^{(m)}(N))$$

are also i.i.d. since the solutions in the layers of $\underline{\mathcal{N}}$ are independent and identical.

Our message inputs $\tilde{\underline{W}}^{(i \to j)}(1), \ldots, \tilde{\underline{W}}^{(i \to j)}(N)$ are not the messages $\underline{W}^{(i \to j)}(1), \ldots, \underline{W}^{(i \to j)}(N)$ (which are independent and uniformly distributed over $\underline{\mathcal{W}}^{(i \to j)}(\ell)$ by assumption) but the channel coded versions of those messages. To maintain the desired i.i.d. structure across

$\underline{\tilde{W}}^{(i \to j)}(1), \ldots, \underline{\tilde{W}}^{(i \to j)}(N)$, we employ a random channel code design that draws each of the $2^{NnR^{(i \to j)}}$ codewords in the codebook for message $\underline{W}^{(i \to j)}$ independently and uniformly at random from the message set $\tilde{\mathcal{W}}^{(i \to j)}$. Maintaining the distribution over random codes, $\underline{\tilde{W}}^{(i \to j)}(1), \ldots, \underline{\tilde{W}}^{(i \to j)}(N)$ are i.i.d. as desired. We put off the choice of a particular instance of this code until the entire code is in place. Our argument employs several random codes, and a single instance of all codes is chosen jointly.

III. INTUITION AND SUMMARY OF RESULTS

As in [1], our goal is not to give the capacity regions of networks with respect to various demands, which is an intractable problem owing to its combinatorial nature. Rather, we wish to develop bounds on the capacity of stochastic components in terms of capacities of networks of point-to-point links. Given the existence of a solution $(\lambda, \mathcal{R})$-$\mathcal{S}(\mathcal{N})$ for a network $\mathcal{N}$, we will try to imply statements for the existence of a solution $(\lambda', \mathcal{R}')$-$\mathcal{S}(\mathcal{N}')$ for a network $\mathcal{N}'$.

Assume, for example, a network contains a hyperedge $\bar{e} = [\{i_1, i_2\}, \{j\}]$ which is a multiple access channel describing the interference that results at node j when nodes i_1 and i_2 transmit. The input and output random variables are $X^{V_1(\bar{e})} = (X^{(i_1, s_1)}, X^{(i_2, s_2)})$ and $Y^{V_2(\bar{e})} = Y^{(j,r)}$, where $i_1, i_2, j \in \{1, \ldots, m\}$, $s_1 \in \{1, \ldots, d_{\text{out}}(i_1)\}$, $s_2 \in \{1, \ldots, d_{\text{out}}(i_2)\}$, and $r \in \{1, \ldots, d_{\text{in}}(j)\}$. The transition probability for the network thus factors as:

$$p(y^{(j,r)}|x^{(i_1,s_1)}, x^{(i_2,s_2)}) \prod_{e \in E \setminus \{\bar{e}\}} p(y^{V_2(e)}|x^{V_1(e)}).$$

Let another network $\hat{\mathcal{N}}$ be given with random variables $\hat{X}^{(i_1,s_1)}$, $\hat{X}^{(i_2,s_2)}$, and $\hat{Y}^{(j,r)}$ replacing $X^{(i_1,s_1)}$, $X^{(i_2,s_2)}$, and $Y^{(j,r)}$ in $\mathcal{N}$. We have replaced the multiple access channel $p(y^{(j,r)}|x^{(i_1,s_1)}, x^{(i_2,s_2)})$ with another multiple access channel $\hat{p}(\hat{y}^{(j,r)}|\hat{x}^{(i_1,s_1)}, \hat{x}^{(i_2,s_2)})$. We want $\hat{p}(\hat{y}^{(j,r)}|\hat{x}^{(i_1,s_1)}, \hat{x}^{(i_2,s_2)})$ to be the distribution for a collection of bit pipes and for the existence of a $(\lambda, \mathcal{R})$-$\mathcal{S}(\mathcal{N})$ solution to imply the existence of a $(\hat{\lambda}, \mathcal{R})$-$\mathcal{S}(\hat{\mathcal{N}})$ solution, where $\hat{\lambda}$ can be made arbitrarily small if λ can. Since node j need not decode, multiple access channel capacity is not necessarily a relevant characterization of the channel's behavior. For example in an additive Gaussian multiple access channel, node j might contribute a real-valued estimation of the sum $X^{(i_1,s_1)} + X^{(i_2,s_2)}$; a collection of bit pipes cannot immediately deliver the same functionality.

Our proofs adopt the operational perspective from [1]. Intuitively, we prove that the rate region for network $\mathcal{N}$ is a subset of the rate region for network $\hat{\mathcal{N}}$ by proving that if anyone shows us a way to operate network $\mathcal{N}$ at one rate point, then we can find a way to operate network $\hat{\mathcal{N}}$ at the same rate point. Roughly, we do this by showing that we can emulate the solution designed for $\mathcal{N}$ on network $\hat{\mathcal{N}}$ to sufficient accuracy that if the error probability can be made arbitrarily small on $\mathcal{N}$ so too can the error probability on $\hat{\mathcal{N}}$. Since we do not know what optimal solutions look like, we must show that we can emulate all possible solutions.

The proofs of our main results follow the same outline as the proof of [1, Theorem 3]. An outline of that argument appears in Section IV. The results for multiple access, broadcast, and interference channels then follow in Section V.

IV. OUTLINE OF PROOFS

Given a network $\mathcal{N}$, let $\hat{\mathcal{N}}^-$ and $\hat{\mathcal{N}}^+$ be the proposed lower and upper bounding networks, respectively. We wish to prove

$$\mathscr{R}(\hat{\mathcal{N}}^-) \subseteq \mathscr{R}(\hat{\mathcal{N}}) \subseteq \mathscr{R}(\hat{\mathcal{N}}^+).$$

By Theorem 1, it suffices to prove

$$\mathscr{R}(\underline{\hat{\mathcal{N}}}^-) \subseteq \mathscr{R}(\underline{\hat{\mathcal{N}}}) \subseteq \mathscr{R}(\underline{\hat{\mathcal{N}}}^+).$$

Proof that $\mathscr{R}(\underline{\hat{\mathcal{N}}}^-) \subseteq \mathscr{R}(\underline{\hat{\mathcal{N}}})$: A new lower bound can be derived for any achievability result on each of the channel components. Suppose that rate vector $R^{(V_1(\bar{e}), V_2(\bar{e}))}$ is achievable on channel $p(y^{V_2(\bar{e})}|x^{V_1(\bar{e})})$. Then the lower bounding network replaces channel $p(y^{V_2(\bar{e})}|x^{V_1(\bar{e})})$ by a collection of noiseless links that deliver precisely the rates promised by $R^{(V_1(\bar{e}), V_2(\bar{e}))}$. For any $\mathcal{R} \in \text{int}(\mathscr{R}(\underline{\hat{\mathcal{N}}}^-))$, there exists a $(2^{-N\delta}, \mathcal{R})$-$\underline{\mathcal{S}}(\underline{\hat{\mathcal{N}}}^-)$ stacked solution by Theorem 1. We run the solution $\underline{\mathcal{S}}(\underline{\hat{\mathcal{N}}}^-)$ on $\underline{\mathcal{N}}$ by using a dimension-n, rate-$R^{(V_1(\bar{e}), V_2(\bar{e}))}$ channel code across the N copies $p(y^{V_2(\bar{e})}|x^{V_1(\bar{e})})$ at each time instant.

Proof that $\mathscr{R}(\underline{\hat{\mathcal{N}}}) \subseteq \mathscr{R}(\underline{\hat{\mathcal{N}}}^+)$: For any $\mathcal{R} \in \text{int}(\mathscr{R}(\underline{\mathcal{N}}))$, there exists a $(2^{-N\delta}, \mathcal{R})$-$\underline{\mathcal{S}}(\underline{\mathcal{N}})$ stacked solution by Theorem 1. Let $\mathcal{S}(\mathcal{N})$ be the solution applied independently in each layer of $\underline{\mathcal{S}}(\underline{\mathcal{N}})$, and let n be the blocklength of solution $\mathcal{S}(\mathcal{N})$. For each $t \in \{1, \ldots, n\}$, we use

$$p_t(x^{V_1(\bar{e})}, y^{V_2(\bar{e})}) = p_t(x^{V_1(\bar{e})})p(y^{V_2(\bar{e})}|x^{V_1(\bar{e})})$$

to designate the distribution established across the channel described by hyperedge $\bar{e}$ at time t when we run solution $\mathcal{S}(\mathcal{N})$ on the network $\mathcal{N}$. Then

$$p_t(\underline{x}^{V_1(\bar{e})}, \underline{y}^{V_2(\bar{e})}) = \prod_{\ell=1}^N p_t(\underline{x}^{V_1(\bar{e})}(\ell), \underline{y}^{V_2(\bar{e})}(\ell)).$$

For any $\epsilon = (\epsilon(1), \ldots, \epsilon(n))$ with $\epsilon(t) > 0$ for all t and any $\bar{\lambda} > 0$, let $A_{\epsilon,t}^{(N)}$ designate the jointly typical set of channel inputs and outputs with respect to $p_t(x^{V_1(\bar{e})}, y^{V_2(\bar{e})})$ and parameter $\epsilon(t)$. Our typical set is slightly modified from its usual definition to enforce not only that each subset of the elements of $(\underline{x}^{V_1(\bar{e})}, \underline{y}^{V_2(\bar{e})})$ are jointly typical but also that for each typical input $\underline{x}^{V_1(\bar{e})}$, the conditional probability of observing a jointly typical output $\underline{y}^{V_2(\bar{e})}$ across the channel is sufficiently high. We then define $B_{\epsilon,t}^{(N)}(\bar{\lambda})$ to designate the subset of vectors $(\underline{x}^{V_1(\bar{e})}, \underline{y}^{V_2(\bar{e})}) \in A_{\epsilon,t}^{(N)}$ for which conditional probability of observing an error in decoding one or more message is at least $\bar{\lambda}$ given $(\underline{X}^{V_1(\bar{e})}, \underline{Y}^{V_2(\bar{e})}) = (\underline{x}^{V_1(\bar{e})}, \underline{y}^{V_2(\bar{e})})$. Lemma 2 bounds the size of $B_{\epsilon,t}^{(N)}(\bar{\lambda})$ as a fraction of $A_{\epsilon,t}^{(N)}$ and the probability of this set under distribution $p_t(\underline{x}^{V_1(\bar{e})}, \underline{y}^{V_2(\bar{e})})$.

Lemma 2: If there exists a $(2^{-N\delta}, \mathcal{R})$-$\underline{\mathcal{S}}(\underline{\mathcal{N}})$ stacked solution for a stacked network $\underline{\mathcal{N}}$, then

$$\frac{|B_{\epsilon,t}^{(N)}(\bar{\lambda})|}{|A_{\epsilon,t}^{(N)}|} \leq \frac{m^2 2^{-N(\delta - f_1(\epsilon,t))}}{\bar{\lambda}} \text{ and}$$

$$\Pr(B_{\epsilon,t}^{(N)}(\bar{\lambda})) \leq \frac{m^2 2^{-N(\delta - f_2(\epsilon,t))}}{\bar{\lambda}}$$

for N sufficiently large.

The functions $f_1(\epsilon,t)$ and $f_2(\epsilon,t)$ vary with the distribution $p_t(x^{V_1(\bar{e})}, y^{V_2(\bar{e})})$ and the coding instance, but all go to zero as $\epsilon(t)$ goes to zero. The proof of Lemma 2 is essentially a counting argument. Since the total error probability is $2^{-N\delta}$ and typical vectors have roughly equal probability, there can't be very many with high conditional error probability.

The next task in each proof is to design a source code that approximates the typical behavior of the channel that we plan to replace. The descriptions of the source codes in all three cases appear below. The source encoders map each channel input $\underline{X}^{V_1(\bar{e})}$ in $\hat{\mathcal{N}}$ to one or more binary descriptions. The binary descriptions are sent across the noiseless bit pipes. The source decoders map the binary descriptions to the emulated channel outputs $\underline{Y}^{V_2(\bar{e})}$, which proceed through the solution as if they were the usual channel outputs.

The final step of each proof is to bound the change in the solution's error probability caused by replacing channels by source codes across noiseless bit pipes. In each case, we show that the difference between the conditional distribution on channel outputs given channel inputs created by the source codes and the conditional distribution established by the channel decays exponentially in the blocklength N. Careful choice of parameters then allows us to prove that the cumulative impact of this small difference on the solution's error probability can be made asymptotically small.

The source code descriptions follow.

A. Multiple Access Channels

For the multiple access channel, we design two encoders

$$\alpha_{N,t}^{(1)}: \quad \underline{\mathcal{X}}^{(i_1,s_1)} \to \{1,\ldots,2^{NR_1}\}$$
$$\alpha_{N,t}^{(2)}: \quad \underline{\mathcal{X}}^{(i_1,s_1)} \times \underline{\mathcal{X}}^{(i_2,s_2)} \to \{1,\ldots,2^{NR_2}\}$$

and a single decoder

$$\beta_{N,t}: \{1,\ldots,2^{NR_1}\} \times \{1,\ldots,2^{NR_2}\} \to \underline{\mathcal{Y}}^{(j,r)}.$$

Given $p(u|x^{(i_1,s_1)}, x^{(i_2,s_2)}) = p(u|x^{(i_1,s_1)})$ and distribution $p_t(x^{(i_1,s_1)}, x^{(i_2,s_2)}, y^{(j,r)})$ from Section IV, the random decoder design draws $\{\gamma_{N,t}(1),\ldots,\gamma_{N,t}(2^{NR_1})\}$ i.i.d. from the distribution $\prod_{\ell=1}^{N} p_t(u_\ell)$. For each $w_1 \in \{1,\ldots,2^{NR_1}\}$, draw $\{\beta_{N,t}(w_1,1),\ldots,\beta_{N,t}(w_1,2^{NR_2})\}$ i.i.d. from distribution $p_t(y^{(j,r)}|\gamma_N(w_1)) = \prod_{\ell=1}^{N} p_t(y^{(j,r)}|\gamma_N(w_1,\ell))$, where $\gamma_{N,t}(w_1,\ell)$ is the ℓth component of N-vector $\gamma_{N,t}(w_1)$.

The random design of encoders $\alpha_{N,t}^{(1)}$ and $\alpha_{N,t}^{(2)}$ proceeds as follows. For each $\underline{x}^{(i_1,s_1)} \in \underline{\mathcal{X}}^{(i_1,s_1)}$, encoder $\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)})$ is chosen uniformly at random from the indices $w_1 \in \{1,\ldots,2^{NR_1}\}$ for which $(\gamma_{N,t}(w_1), \underline{x}^{(i_1,s_1)})$ is typical. If there are no such indices, then $\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)}) = 1$.

For each $(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}) \in \underline{\mathcal{X}}^{(i_1,s_1)} \times \underline{\mathcal{X}}^{(i_2,s_2)}$, encoder $\alpha_{N,t}^{(2)}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)})$ is chosen uniformly at random from the indices w_2 for which

$$\Big(\gamma_{N,t}(\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)})), \underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)},$$
$$\beta_{N,t}(\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)}), w_2)\Big)$$

is jointly typical and

$$(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}, \beta_{N,t}(\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)}), w_2)) \notin B_{\epsilon,t}^{(N)}(\bar{\lambda}).$$

If there is no such index, then $\alpha_{N,t}^{(2)}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}) = 1$.

The rates in the upper bound are chosen so that the probability of finding jointly typical codewords is high.

B. Broadcast Channels

For the broadcast channel, we design one source encoder

$$\alpha_{N,t}: \underline{\mathcal{X}}^{(i,s)} \to \{1,\ldots,2^{NR_1}\} \times \{1,\ldots,2^{nR_2}\}$$

and two decoders

$$\beta_{N,t}^{(1)}: \quad \{1,\ldots,2^{NR_1}\} \to \underline{\mathcal{Y}}^{(j_1,r_1)}$$
$$\beta_{N,t}^{(2)}: \quad \{1,\ldots,2^{NR_1}\} \times \{1,\ldots,2^{NR_2}\} \to \underline{\mathcal{Y}}^{(j_2,r_2)}.$$

To design the decoder, draw $\{\beta_{N,t}^{(1)}(1),\ldots,\beta_{N,t}^{(1)}(2^{NR_1})\}$ i.i.d. from distribution $\prod_{\ell=1}^{N} p_t(y^{(j_1,r_1)}(\ell))$, and for each w_1 draw $\{\beta_{N,t}^{(2)}(w_1,1),\ldots,\beta_{N,t}^{(2)}(w_1,2^{NR_2})\}$ i.i.d. from distribution $\prod_{\ell=1}^{N} p_t(y^{(j_2,r_2)}(\ell)|\beta_{N,t}^{(1)}(w_1,\ell))$, where $\beta_{N,\ell}^{(1)}(w_1,\ell)$ is the ℓth component of $\beta_{N,t}^{(1)}(w_1)$.

The encoder output $\alpha_{N,t}(x^N)$ is chosen uniformly from the collection of indices (w_1, w_2) for which the triple $(x^N, \beta_{N,t}^{(1)}(w_1), \beta_{N,t}^{(2)}(w_2))$ is in $A_{\epsilon,t}^{(N)} \setminus B_{\epsilon,t}^{(N)}(\bar{\lambda})$; $(\alpha_{N,t}^{(1)}(x^N), \alpha_{N,t}^{(2)}(x^N)) = 1$ if no such pair exists.

C. Interference Channels

For the interference channel, we design a source code with two encoders and two decoders. We consider two mechanisms for this design. These two mechanisms result in four possible scenarios by reversing the roles of $(X^{(i_1,s_1)}, Y^{(j_1,r_1)})$ and $(X^{(i_2,s_2)}, Y^{(j_2,r_2)})$.

Mechanism 1: We design a source code with encoders

$$\alpha_{N,t}^{(1)}: \quad \underline{\mathcal{X}}^{(i_1,s_1)} \to \{1,\ldots,2^{NR_{11}}\} \times \{1,\ldots,2^{nR_{12}}\}$$
$$\alpha_{N,t}^{(2)}: \quad \underline{\mathcal{X}}^{(i_1,s_1)} \times \underline{\mathcal{X}}^{(i_2,s_2)} \to \prod_{j=1}^{2} \{1,\ldots,2^{NR_{2j}}\}$$

and decoders

$$\beta_{N,t}^{(1)}: \quad \prod_{j=1}^{2}\prod_{i=1}^{2} \{1,\ldots,2^{NR_{ij}}\} \to \underline{\mathcal{Y}}^{(j_1,r_1)}$$
$$\beta_{N,t}^{(2)}: \quad \prod_{i=1}^{2} \{1,\ldots,2^{NR_{i2}}\} \to \underline{\mathcal{Y}}^{(j_2,r_2)}.$$

Given any distribution

$$p(u, v|x^{(i_1,s_1)}, x^{(i_2,s_2)}, y^{(j_1,r_1)}, y^{(j_2,r_2)})$$
$$= p(v|x^{(i_1,s_1)}) p(u|x^{(i_1,s_1)}, v),$$

and the distribution $p_t(x^{(i_1,s_1)}, x^{(i_2,s_2)}, y^{(j_1,r_1)}, y^{(j_2,r_2)})$ from Section IV, the random decoder design begins by drawing

$$\{\underline{V}(1), \ldots, \underline{V}(2^{NR_{12}})\}$$

i.i.d. from the distribution $\prod_{\ell=1}^{N} p(v(\ell))$ and for each $w_{12} \in \{1, \ldots, 2^{NR_{12}}\}$ drawing

$$\{\underline{U}(1, w_{12}), \ldots, \underline{U}(2^{NR_{11}}, w_{12})\}$$

i.i.d. from the distribution $\prod_{\ell=1}^{N} p(u(\ell)|\underline{V}(w_{12}, \ell))$, where $\underline{V}(w_{12}, \ell)$ is the ℓth component of N-vector $\underline{V}(w_{12})$. We then proceed to draw

$$\{\beta_{N,t}^{(2)}(1, w_{12}), \ldots, \beta_{N,t}^{(2)}(2^{nR_{22}}, w_{12})\}$$

i.i.d. from the distribution $\prod_{\ell=1}^{N} p(y^{(j_2,r_2)}(\ell)|\underline{V}(w_{12}, \ell))$. Finally, for each

$$(w_{11}, w_{12}, w_{22}) \in \prod_{(i,j) \in \{1,2\}^2 \setminus \{(2,1)\}} \{1, \ldots, 2^{nR_{ij}}\},$$

draw

$$\{\beta_{N,t}^{(1)}(w_{11}, w_{12}, 1, w_{22}), \ldots, \beta_{N,t}^{(1)}(w_{11}, w_{12}, 2^{NR_{21}}, w_{22})\}$$

i.i.d. from the distribution

$$\prod_{\ell=1}^{N} p(y^{(j_2,r_2)}(\ell)|\underline{U}(w_{11}, w_{12}, \ell), \underline{V}(w_{12}, \ell), \beta_{N,t}^{(2)}(w_{12}, w_{22}, \ell)).$$

The random design of encoders $\alpha_{N,t}^{(1)}$ and $\alpha_{N,t}^{(2)}$ proceeds as follows. For each $\underline{x}^{(i_1,s_1)} \in \mathcal{X}^{(i_1,s_1)}$, encoder output $\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)}) = (\alpha^{11}(\underline{x}^{(i_1,s_1)}), \alpha^{12}(\underline{x}^{(i_1,s_1)}))$, is chosen uniformly from the family of indices of codewords that is jointly typical and not in $B_{\epsilon,t}^{(N)}(\bar{\lambda})$. If there are no such codewords, then $\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)}) = (1,1)$. Similarly, for each $(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}) \in \mathcal{X}^{(i_1,s_1)} \times \mathcal{X}^{(i_2,s_2)}$, encoder output

$$\alpha_{N,t}^{(2)}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)})$$
$$= (\alpha_{N,t}^{21}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}), \alpha_{N,t}^{22}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}))$$

is chosen uniformly at random from the family of indices of codewords that is jointly typical and not in $B_{\epsilon,t}^{(N)}(\bar{\lambda})$. $\alpha_{N,t}^{(2)}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}) = (1,1)$ otherwise.

Mechanism 2: In this case, we design a source code with encoders

$$\alpha_{N,t}^{(1)}: \quad (\mathcal{X}^{(i_1,s_1)})^N \to \{1, \ldots, 2^{NR_{11}}\} \times \{1, \ldots, 2^{nR_{12}}\}$$

$$\alpha_{N,t}^{(2)}: \quad (\mathcal{X}^{(i_1,s_1)})^N \times (\mathcal{X}^{(i_2,s_2)})^N \to \prod_{j=1}^{2} \{1, \ldots, 2^{NR_{2j}}\}$$

and decoders

$$\beta_{N,t}^{(1)}: \quad \prod_{i=1}^{2} \{1, \ldots, 2^{NR_{i2}}\} \to (\mathcal{Y}^{(j_1,r_1)})^N$$

$$\beta_{N,t}^{(2)}: \quad \prod_{j=1}^{2} \prod_{i=1}^{2} \{1, \ldots, 2^{NR_{ij}}\} \to (\mathcal{Y}^{(j_2,r_2)})^N.$$

While the encoder arguments have not changed, this formulation differs from the previous one in that roles of the decoder

has changed – in this case decoder 2 has access to both sets of descriptions while decoder 1 has access to only one set.

Given any distribution

$$p(u, v|x^{(i_1,s_1)}, x^{(i_2,s_2)}, y^{(j_1,r_1)}, y^{(j_2,r_2)})$$
$$= p(u|x^{(i_1,s_1)})p(v|x^{(i_1,s_1)}, u),$$

and the distribution $p_t(x^{(i_1,s_1)}, x^{(i_2,s_2)}, y^{(j_1,r_1)}, y^{(j_2,r_2)})$ from Section IV, the random decoder design begins by drawing

$$\{\underline{U}(1), \ldots, \underline{U}(2^{NR_{11}})\}$$

i.i.d. from the distribution $\prod_{\ell=1}^{N} p(u(\ell))$ and for each $w_{11} \in \{1, \ldots, 2^{NR_{11}}\}$ drawing

$$\{\underline{V}(w_{11}, 1), \ldots, \underline{V}(w_{11}, 2^{NR_{12}})\}$$

i.i.d. from the distribution $\prod_{\ell=1}^{N} p(v(\ell)|\underline{U}(w_{11}, \ell))$, where $\underline{U}(w_{11}, \ell)$ is the ℓth component of N-vector $\underline{U}(w_{11})$. We then proceed to draw

$$\{\beta_{N,t}^{(1)}(w_{11}, 1), \ldots, \beta_{N,t}^{(1)}(w_{11}, 2^{nR_{21}})\}$$

i.i.d. from the distribution $\prod_{\ell=1}^{N} p(y_{1,\ell}|\underline{U}(w_{11}, \ell))$. Finally, for each

$$(w_{11}, w_{12}, w_{21}) \in \prod_{(i,j) \in \{1,2\}^2 \setminus \{(2,2)\}} \{1, \ldots, 2^{nR_{ij}}\},$$

draw

$$\{\beta_{N,t}^{(2)}(w_{11}, w_{12}, w_{21}, 1), \ldots, \beta_{N,t}^{(2)}(w_{11}, w_{12}, w_{21}, 2^{NR_{22}})\}$$

i.i.d. from the distribution

$$\prod_{\ell=1}^{N} p(y^{(j_2,r_2)}(\ell)|$$
$$\underline{U}(w_{11}, \ell), \underline{V}(w_{11}, w_{12}, \ell), \beta_{N,t}^{(1)}(w_{11}, w_{21}, \ell)).$$

The random design of encoders $\alpha_{N,t}^{(1)}$ and $\alpha_{N,t}^{(2)}$ proceeds as follows. For each $\underline{x}^{(i_1,s_1)}$, encoder output $\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)}) = (\alpha^{11}(\underline{x}^{(i_1,s_1)}), \alpha^{12}(\underline{x}^{(i_1,s_1)}))$, is chosen uniformly from the set of indices of jointly typical codewords. If there are no such codewords, then $\alpha_{N,t}^{(1)}(\underline{x}^{(i_1,s_1)}) = (1,1)$. Similarly, for each $(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)})$, encoder output

$$\alpha_{N,t}^{(2)}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)})$$
$$= (\alpha_{N,t}^{21}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}), \alpha_{N,t}^{22}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}))$$

is chosen uniformly at random from the set of indices of jointly typical codewords for which the channel input and output are not in $B_{\epsilon,t}^{(N)}(\bar{\lambda})$.

$\mathcal{W}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)})$ if it is non-empty and from $\{1, \ldots, 2^{NR_{21}}\} \times \{1, \ldots, 2^{NR_{22}}\}$ otherwise. If there are no such codewords, then $\alpha_{N,t}^{(2)}(\underline{x}^{(i_1,s_1)}, \underline{x}^{(i_2,s_2)}) = (1,1)$.

V. RESULTS

The discussion that follows presents our main results for three canonical networks. Our proof of the lower bound allows us to turn any existing achievability result into a lower bounding network. Only one example is given for the broadcast channel and none are given for the interference channel due to space constraints.

A. Multiple Access Channels

Consider a network $\mathcal{N}$ with transition probabilities (1), where $p(y^{V_2(\bar{e})}|x^{V_1(\bar{e})}) = p(y^{(j,r)}|x^{(i_1,s_1)}, x^{(i_2,s_2)})$ corresponds to a multiple access channel with transmitter nodes i_1 and i_2 and receiver node j. We wish to derive upper and lower bounds on the capacity of the given network as a function of the capacity of corresponding networks where the multiple access channel is replaced by a collection of noiseless, point-to-point bit pipes.

Our main result for multiple access channels shows that we can bound the capacity of a network containing a multiple access channel by the capacities of two distinct networks in which the multiple access channel is replaced by a collection of point-to-point noiseless links.

Theorem 3: Given a network

$$
\begin{aligned}
\mathcal{N} = & \Big(\mathcal{X}^{(1,1)} \times \cdots \times \mathcal{X}^{(i_1,s_1)} \times \cdots \times \mathcal{X}^{(i_2,s_2)} \\
& \times \cdots \times \mathcal{X}^{(m,d_{\text{out}}(m))}, p(y^{(j,r)}|x^{(i_1,s_1)}, x^{(i_2,s_2)}) \\
& \cdot \prod_{e \in E \setminus \{\bar{e}\}} p(y^{V_2(e)}|x^{V_1(e)}), \mathcal{Y}^{(1,1)} \times \cdots \\
& \times \mathcal{Y}^{(j,r)} \times \cdots \times \mathcal{Y}^{(m,d_{\text{in}}(m))} \Big)
\end{aligned}
$$

with node indegree $d_{\text{in}}(i)$ and outdegree $d_{\text{out}}(i)$ for each $i \in \{1, \ldots, m\}$, define a new network $\hat{\mathcal{N}}$

$$
\begin{aligned}
\hat{\mathcal{N}} = & \Big(\mathcal{X}^{(1,1)} \times \cdots \times \hat{\mathcal{X}}^{(i_1,s_1)} \times \cdots \times \hat{\mathcal{X}}^{(i_1,\hat{d}_{\text{out}}(s_1)} \times \\
& \cdots \times \hat{\mathcal{X}}^{(i_2,s_2)} \times \cdots \times \mathcal{X}^{(m,\hat{d}_{\text{out}}(m))}, \\
& \delta(\hat{x}^{(i_1,s_1)} - \hat{y}^{(j,r)}) \delta(\hat{x}^{(i_1,\hat{d}_{\text{out}}(i_1))} - \hat{x}^{(i_2,\hat{d}_{\text{in}}(i_2))}) \\
& \cdot \delta(\hat{x}^{(i_2,s_2)} - \hat{y}^{(j,\hat{d}_{\text{in}}(j))}) \prod_{e \in E \setminus \{\bar{e}\}} p(y^{V_2(e)}|x^{V_1(e)}), \\
& \mathcal{Y}^{(1,1)} \times \cdots \times \hat{\mathcal{Y}}^{(i_2,\hat{d}_{\text{in}}(i_2))} \times \cdots \times \hat{\mathcal{Y}}^{(j,r)} \times \cdots \\
& \times \hat{\mathcal{Y}}^{(j,\hat{d}_{\text{in}}(j))} \times \cdots \times \mathcal{Y}^{(m,\hat{d}_{\text{in}}(m))} \Big),
\end{aligned}
$$

with node indegrees $\hat{d}_{\text{in}}(i_2) = d_{\text{in}}(i_2) + 1$, $\hat{d}_{\text{in}}(j) = d_{\text{in}}(j) + 1$, and $\hat{d}_{\text{in}}(i) = d_{\text{in}}(i)$ for all $i \in \{1, \ldots, m\} \setminus \{i_2, j\}$ and node outdegrees $\hat{d}_{\text{out}}(i_1) = d_{\text{out}}(i_1) + 1$ and $\hat{d}_{\text{out}}(i) = d_{\text{out}}(i)$ for all $i \in \{1, \ldots, m\} \setminus \{i_1\}$. This replaces the single hyperedge $\bar{e} = [\{i_1, i_2\}, \{j\}]$ by new edges $\hat{e}_1 = [\{i_1\}, \{j\}]$, $\hat{e}_2 = [\{i_2\}, \{j\}]$, $\hat{e}_1 = [\{i_1\}, \{i_2\}]$ which are bit pipes of capacities R_1, R_2, and R_3, respectively.

   1) If

$$
\begin{aligned}
R_1 & < I(X^{(i_1,s_1)}; Y^{(j,r)}|X^{(i_2,s_2)}) \\
R_2 & < I(X^{(i_2,s_2)}; Y^{(j,r)}|X^{(i_1,s_1)}) \\
R_1 + R_2 & < I(X^{(i_1,s_1)}, X^{(i_2,s_2)}; Y^{(j,r)}) \\
R_3 & = 0,
\end{aligned}
$$

for some input distribution $p(x^{(i_1,s_1)}, x^{(i_2,s_2)}) = p(x^{(i_1,s_1)})p(x^{(i_2,s_2)})$ on alphabet $\mathcal{X}^{(i_1,s_1)} \times \mathcal{X}^{(i_2,s_2)}$, then $\mathscr{R}(\hat{\mathcal{N}}) \subseteq \mathscr{R}(\mathcal{N})$.

   2) If there exists a conditional distribution

$p(u|x^{(i_1,s_1)}, x^{(i_2,s_2)}) = p(u|x^{(i_1,s_1)})$ for which

$$
\begin{aligned}
R_1 & > I(X^{(i_1,s_1)}; U) \\
R_2 & > I(X^{(i_1,s_1)}, X^{(i_2,s_2)}; Y^{(j,r)}|U) \\
R_3 & = \infty,
\end{aligned}
$$

for all distributions $p(x^{(i_1,s_1)}, x^{(i_2,s_2)})$ on alphabet $\mathcal{X}^{(i_1,s_1)} \times \mathcal{X}^{(i_2,s_2)}$, then $\mathscr{R}(\mathcal{N}) \subseteq \mathscr{R}(\hat{\mathcal{N}})$.

B. Broadcast Channels

Consider a network $\mathcal{N}$ with transition probabilities (1), where $p(y^{V_2(\bar{e})}|x^{V_1(\bar{e})}) = p(y^{(j_1,r_1)}, y^{(j_2,r_2)}|x^{(i,s)})$ corresponds to a broadcast channel with transmitter node i and receiver nodes j_1 and j_2. We next derive upper and lower bounds on the capacity of this network as a function of the capacity of corresponding networks where the broadcast channel has been replaced by noiseless, point-to-point bit pipes.

Theorem 4: Given a network

$$
\begin{aligned}
\mathcal{N} = & \Big(\mathcal{X}^{(1,1)} \times \cdots \times \mathcal{X}^{(i,s)} \times \cdots \times \mathcal{X}^{(m,d_{\text{out}}(m))}, \\
& p(y^{(j_1,r_1)}, y^{(j_2,r_2)}|x^{(i,s)}) \prod_{e \in E \setminus \{\bar{e}\}} p(y^{V_2(e)}|x^{V_1(e)}), \\
& \mathcal{Y}^{(1,1)} \times \cdots \times \mathcal{Y}^{(j_1,r_1)} \times \cdots \times \mathcal{Y}^{(j_2,r_2)} \times \cdots \\
& \times \mathcal{Y}^{(m,d_{\text{in}}(m))} \Big)
\end{aligned}
$$

with node indegree $d_{\text{in}}(j)$ and outdegree $d_{\text{out}}(j)$ for each $j \in \{1, \ldots, m\}$, define a new network $\hat{\mathcal{N}}$

$$
\begin{aligned}
\hat{\mathcal{N}} \\
= & \Big(\mathcal{X}^{(1,1)} \times \cdots \times \hat{\mathcal{X}}^{(i,s)} \times \cdots \times \hat{\mathcal{X}}^{(i,\hat{d}_{\text{out}}(s))} \times \cdots \\
& \times \mathcal{X}^{(m,\hat{d}_{\text{out}}(m))}, \delta(\hat{x}^{(i,s)} - \hat{y}^{(j_1,r_1)}) \\
& \cdot \delta(\hat{x}^{(i,\hat{d}_{\text{out}}(i))} - \hat{y}^{(j_2,r_2)}) \delta(\hat{x}^{(j_1,\hat{d}_{\text{out}}(j_1))} - \hat{y}^{(j_2,\hat{d}_{\text{in}}(j_2))}) \\
& \prod_{e \in E \setminus \{\bar{e}\}} p(y^{V_2(e)}|x^{V_1(e)}), \mathcal{Y}^{(1,1)} \times \cdots \times \hat{\mathcal{Y}}^{(j_1,r_1)} \times \\
& \cdots \times \hat{\mathcal{Y}}^{(j_2,r_2)} \times \cdots \times \hat{\mathcal{Y}}^{(j_2,\hat{d}_{\text{in}}(j_2))} \times \cdots \\
& \times \mathcal{Y}^{(m,\hat{d}_{\text{in}}(m))} \Big),
\end{aligned}
$$

with node outdegrees $\hat{d}_{\text{out}}(i) = d_{\text{out}}(i) + 1$, $\hat{d}_{\text{out}}(j_1) = d_{\text{out}}(j_1) + 1$, and $\hat{d}_{\text{out}}(j) = d_{\text{out}}(j)$ for all $j \in \{1, \ldots, m\} \setminus \{i, j_1\}$ and node indegrees $\hat{d}_{\text{in}}(j_2) = d_{\text{in}}(j_2) + 1$ and $\hat{d}_{\text{in}}(j) = d_{\text{in}}(j)$ for all $j \in \{1, \ldots, m\} \setminus \{i_1\}$. This replaces hyperedge $\bar{e} = [\{i\}, \{j_1, j_2\}]$ by three new edges $\hat{e}_1 = [\{i\}, \{j_1\}]$, $\hat{e}_2 = [\{i\}, \{j_2\}]$, and $\hat{e}_3 = [\{j_1\}, \{j_2\}]$ which are bit pipes of capacities R_1, R_2, and R_0, respectively.

   1) Suppose $p(y^{(j_1,r_1)}, y^{(j_2,r_2)}|x^{(i,s)})$ describes a physically or stochastically degraded broadcast channel with $Y^{(j_1,r_1)}$ as the more degraded channel output. Then

$$
\begin{aligned}
R_1 & < I(U; Y^{(j_1,r_1)}) \\
R_2 & < I(X^{(i,s)}; Y^{(j_2,r_2)}|U) \\
R_0 & < I(U; Y^{(j_1,r_1)})
\end{aligned}
$$

for some $p(u)p(x^{(i,s)}|u)$, then $\mathscr{R}(\hat{\mathcal{N}}) \subseteq \mathscr{R}(\mathcal{N})$.

2) If

$$
\begin{aligned}
R_1 &> I(X^{(i,s)}; Y^{(j_1,r_1)}) \\
R_2 &> I(X^{(i,s)}; Y^{(j_2,r_2)}|Y^{(j_1,r_1)}) \\
R_0 &= I(X^{(i,s)}; Y^{(j_1,r_1)})
\end{aligned}
$$

for all $p(x^{(i_1,s_1)}, x^{(i_2,s_2)})$, then $\mathscr{R}(\mathcal{N}) \subseteq \mathscr{R}(\hat{\mathcal{N}})$.

Remark 2: Given a non-degraded channel and any achievable rate triple (R_0, R_1, R_2), a corresponding result can be derived with the same network topology. (Here R_0 represents the rate for the common information.)

C. Interference Channels

Finally, consider a network $\mathcal{N}$ with transition probabilities (1), where $p(y^{V_2(\bar{e})}|x^{V_1(\bar{e})}) = p(y^{(j_1,r_1)}, y^{(j_2,r_2)}|x^{(i_1,s_1)}, x^{(i_2,s_2)})$ corresponds to an interference channel with transmitter nodes i_1 and i_2 and receiver nodes j_1 and j_2. We next describe an upper bound on the capacity of this network as a function of the capacity of two networks where the interference channel has been replaced by noiseless, point-to-point bit pipes. We omit the lower bounds due to space constraints. As in all cases, any known achievability bound for the given network can be applied to create a lower bounding network.

Theorem 5: Given a network

$$
\begin{aligned}
\mathcal{N} = \Big(&\mathcal{X}^{(1,1)} \times \cdots \times \mathcal{X}^{(i_1,s_1)} \times \cdots \times \mathcal{X}^{(i_2,s_2)} \times \cdots \\
&\times \mathcal{X}^{(m,d_{\text{out}}(m))}, \\
&p(y^{(j_1,r_1)}, y^{(j_2,r_2)}|x^{(i_1,s_1)}, x^{(i_2,s_2)}) \\
&\cdot \prod_{e \in E \setminus \{\bar{e}\}} p(y^{V_2(e)}|x^{V_1(e)}), \\
&\mathcal{Y}^{(1,1)} \times \cdots \times \mathcal{Y}^{(j_1,r_1)} \times \cdots \times \mathcal{Y}^{(j_2,r_2)} \times \cdots \\
&\times \mathcal{Y}^{(m,d_{\text{in}}(m))} \Big)
\end{aligned}
$$

with node indegree $d_{\text{in}}(j)$ and outdegree $d_{\text{out}}(j)$ for each $j \in \{1, \ldots, m\}$. For the sake of notational brevity, let $d_1 = d_{\text{out}}(i_1) + 1$, $d_2 = d_{\text{out}}(i_1) + 2$, $d_3 = d_{\text{out}}(i_2) + 1$, $d_4 = d_{\text{out}}(j_1) + 1$, $d_5 = d_{\text{out}}(j_2) + 1$, $d_6 = d_{\text{in}}(j_1) + 1$, $d_7 = d_{\text{in}}(j_1) + 2$, $d_8 = \text{in}(j_2) + 1$, $d_9 = d_{\text{in}}(j_2) + 2$, and $d_{10} = d_{\text{in}}(i_2) + 1$. Define a new network $\hat{\mathcal{N}}$

$$
\begin{aligned}
\hat{\mathcal{N}} = \Big(&\mathcal{X}^{(1,1)} \times \cdots \times \hat{\mathcal{X}}^{(i_1,s_1)} \times \cdots \times \hat{\mathcal{X}}^{(i_1,d_1)} \times \cdots \\
&\times \hat{\mathcal{X}}^{(i_1,d_2)} \times \cdots \times \hat{\mathcal{X}}^{(i_2,s_2)} \times \cdots \times \hat{\mathcal{X}}^{(i_2,d_3)} \times \cdots \\
&\times \hat{\mathcal{X}}^{(j_1,d_4)} \times \cdots \times \hat{\mathcal{X}}^{(j_2,d_5)} \times \cdots \times \mathcal{X}^{(m,\hat{d}_{\text{out}}(m))}, \\
&\cdot \delta(\hat{x}^{(i_1,s_1)} - \hat{y}^{(j_1,r_1)}) \delta(\hat{x}^{(i_2,s_2)} - \hat{y}^{(j_2,r_2)}) \\
&\cdot \delta(\hat{x}^{(i_1,d_1)} - \hat{y}^{(j_2,d_8)}) \delta(\hat{x}^{(i_2,d_3)} - \hat{y}^{(j_1,d_6)})
\end{aligned}
$$

$$
\begin{aligned}
&\cdot \delta(\hat{x}^{(j_1,d_4)} - \hat{y}^{(j_2,d_9)}) \delta(\hat{x}^{(j_2,d_5)} - \hat{y}^{(j_1,d_7)}) \\
&\cdot \delta(\hat{x}^{(i_1,d_2)} - \hat{y}^{(i_2,d_10)}) \prod_{e \in E \setminus \{\bar{e}\}} p(y^{V_2(e)}|x^{V_1(e)}), \\
&\mathcal{Y}^{(1,1)} \times \cdots \times \hat{\mathcal{Y}}^{(i_2,d_{10})} \times \cdots \times \hat{\mathcal{Y}}^{(j_1,r_1)} \times \cdots \\
&\times \hat{\mathcal{Y}}^{(j_1,d_6)} \times \cdots \times \hat{\mathcal{Y}}^{(j_1,d_7)} \times \cdots \times \hat{\mathcal{Y}}^{(j_2,r_2)} \times \cdots \\
&\times \hat{\mathcal{Y}}^{(j_2,d_8)} \times \cdots \times \hat{\mathcal{Y}}^{(j_2,d_9)} \times \cdots \times \mathcal{Y}^{(m,d_{\text{in}}(m))} \Big).
\end{aligned}
$$

This replaces hyperedge $\bar{e} = [\{i_1, i_2\}, \{j_1, j_2\}]$ by edges $\hat{e}_1 = [\{i_1\}, \{j_1\}]$, $\hat{e}_2 = [\{i_2\}, \{j_2\}]$, $\hat{e}_3 = [\{i_1\}, \{j_2\}]$, $\hat{e}_4 = [\{i_2\}, \{j_1\}]$, $\hat{e}_5 = [\{j_1\}, \{j_2\}]$, $\hat{e}_6 = [\{j_2\}, \{j_1\}]$, and $\hat{e}_7 = [\{i_1\}, \{i_2\}]$ which are bit pipes of capacities R_{11}, R_{22}, R_{12} R_{21}, R_5, R_6, and R_7, respectively.

1) If there exists a distribution $p(u, v|x^{(i_1,s_1)}, x^{(i_2,s_2)}) = p(v|x^{(i_1,s_1)})p(u|x^{(i_1,s_1)}, v)$ for which

$$
\begin{aligned}
R_{11} &> I(X^{(i_1,s_1)}; U|V) \\
R_{12} &> I(X^{(i_1,s_1)}; V) \\
R_{21} &> I(X^{(i_1,s_1)}, X^{(i_2,s_2)}; Y^{(j_1,r_1)}|U, V, Y^{(j_2,r_2)}) \\
R_{22} &> I(X^{(i_1,s_1)}, X^{(i_2,s_2)}; Y^{(j_2,r_2)}|V)
\end{aligned}
$$

for all $p(x^{(i_1,s_1)}, x^{(i_2,s_2)})$ and $R_5 = 0$, $R_6 = R_{12} + R_{22}$, and $R_7 = \infty$, then $\mathscr{R}(\mathcal{N}) \subseteq \mathscr{R}(\hat{\mathcal{N}})$.

2) If there exists a distribution $p(u, v|x^{(i_1,s_1)}, x^{(i_2,s_2)}) = p(u|x^{(i_1,s_1)})p(v|x^{(i_1,s_1)}, u)$ for which

$$
\begin{aligned}
R_{11} &> I(X^{(i_1,s_1)}; U) \\
R_{12} &> I(X^{(i_1,s_1)}; V|U) \\
R_{21} &> I(X^{(i_1,s_1)}, X^{(i_2,s_2)}; Y^{(j_1,r_1)}|U) \\
R_{22} &> I(X^{(i_1,s_1)}, X^{(i_2,s_2)}; Y^{(j_2,r_2)}|U, V, Y_1)
\end{aligned}
$$

for all $p(x^{(i_1,s_1)}, x^{(i_2,s_2)})$ and $R_5 = R_{11} + R_{21}$, $R_6 = 0$, and $R_7 = \infty$, then $\mathscr{R}(\mathcal{N}) \subseteq \mathscr{R}(\hat{\mathcal{N}})$.

REFERENCES

[1] R. Koetter, M. Effros, and M. Médard, "An equivalence theory of network capacity," in *Proceedings of the IEEE Information Theory Workshop*, (Volos, Greece), pp. 326–330, June 2009.

[2] M. Langberg, M. Sprintson, and J. Bruck, "Network coding: a computational perspective," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 145–157, 2008.

[3] A. Avestimehr, S. Diggavi, and D. Tse, "Approximate capacity of Gaussian relay networks," in *Proceedings of the IEEE International Symposium on Information Theory*, pp. 474–478, July 2008.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, second ed., 2006.