# Equivalence for Networks With Adversarial State

Oliver Kosut, *Member, IEEE*, and Jörg Kliewer, *Senior Member, IEEE*

***Abstract*— We address the problem of finding the capacity of noisy networks with either independent point-to-point compound channels (CC) or arbitrarily varying channels (AVC). These channels model the presence of a Byzantine adversary, which controls a subset of links or nodes in the network. We derive equivalence results showing that these point-to-point channels with state can be replaced by noiseless bit-pipes without changing the network capacity region. Exact equivalence results are found for the CC model, and for some instances of the AVC, including all nonsymmetrizable AVCs. These results show that a feedback path between the output and input of a CC can increase the equivalent capacity, and that if common randomness can be established between the terminals of an AVC (either by a feedback, a forward path, or via a third-party node), then again the equivalent capacity can increase. This leads to an observation that deleting an edge of arbitrarily small capacity can cause a significant change in network capacity. We also analyze an example involving an AVC for which no fixed-capacity bit-pipe is equivalent.**

***Index Terms*— Network security, equivalence, capacity, arbitrarily-varying channel, compound channel, active adversary.**

## I. Introduction

ONE fundamental problem in wireless and wireline networks is to achieve robustness against active adversaries. A common assumption is to consider Byzantine adversaries who observe all transmissions, messages, and channel noise values and interfere with the transmitted signals, i.e., by replacing a subset of the channel output values or by injecting additional noise to a specific subset of communication channels or nodes (the adversarial set) in the network. For example, for the adversarial noiseless case both in-network error correction approaches and capacity results under network coding have been presented, e.g., in [1]–[4].

The underlying uncertainty in the network due to the action of the adversary leads to channels with varying state in the adversarial set [5]. One possible model is to assume that the corresponding nodes have no knowledge about the exact channel state, but only that the state is selected from a finite set. In the case of a compound channel (CC) [6], [7] the selected state is fixed over the whole transmission of a codeword. In contrast, if the channel state varies from symbol to symbol in an unknown and arbitrary manner we have the case of an arbitrarily varying channel (AVC) [8]–[11].

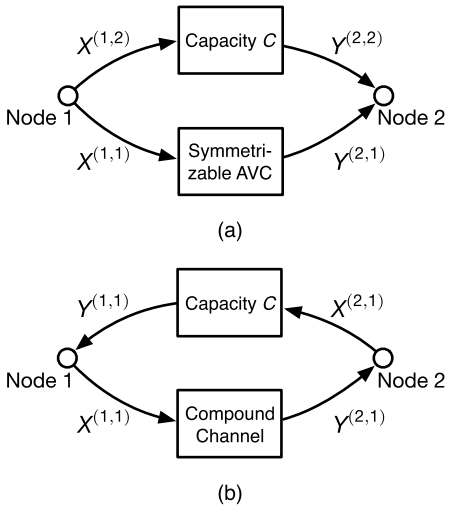


Fig. 1. Two-node networks with a capacity $C$ channel and (a) a symmetrizable AVC, (b) a CC. In general, the upper channel can be replaced with a single-source single-sink network having the same rate.

Note that the AVC has a (deterministic) capacity which is either zero or equals the random coding capacity [9]. The former case holds for a symmetrizable AVC, since such a channel can mimic a valid input sequence in such a way that it is impossible for the decoder to decide on the correct codeword. Even though transmission is not possible if such an AVC is considered in isolation, the situation changes in a network setting, as exemplarily depicted in Fig. 1(a). In this two-node network, source and destination nodes are connected via two parallel channels, a (fixed) channel with capacity $C$ and a symmetrizable AVC. Here, communication over the AVC is possible with a non-zero rate since common randomness with negligible rate $\epsilon > 0$ can be shared between both nodes [9]–[11] via the upper channel in Fig. 1(a). In a more general setup, in Fig. 1 this channel can be replaced with a single-source single-sink network of positive rate $C$.

In the following we consider the problem of reliable communication over a network of independent noisy point-to-point channels in the presence of active adversaries. A subset of the channels either consists of AVCs or CCs. This is in contrast to the model in [12], where the action of the adversary is directly modeled by injecting an arbitrary vector to the network edges in the adversarial set. By building on the results in [13] we identify cases where the adversarial capacity of the network equals the capacity of another network in which each channel is replaced by a noise-free bit-pipe. For a CC, the bit-pipe has capacity equal to the standard CC capacity if there is no

Manuscript received January 22, 2015; revised June 13, 2016 and March 20, 2017; accepted March 29, 2017. Date of publication May 5, 2017; date of current version June 14, 2017. This work was supported by the U.S. National Science Foundation under Grant CCF-1439465, Grant CCF-1440014, Grant CCF-1453718, and Grant CNS-1526547. This work was presented in part at the 2014 IEEE International Symposium on Information Theory. *(Corresponding author: Oliver Kosut.)*

O. Kosut is with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287 USA (e-mail: okosut@asu.edu).

J. Kliewer is with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ 07102 USA (email: jkliewer@njit.edu).

Communicated by S. S. Pradhan, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2017.2701804

feedback path from the output to the input; if there is, then the equivalent bit-pipe has higher capacity, because the state can be estimated at the output and relayed back to the input (see Fig. 1(b)). For an AVC, the equivalent bit-pipe has capacity equal to the random coding capacity if it is possible to establish common randomness between the input and output. This can be accomplished if any of the following hold: (i) the AVC is non-symmetrizable, (ii) there is a parallel forward path as in Fig. 1(a), (iii) there is a feedback path as for the CC in Fig. 1(b), or (iv) a third-party node can transmit to both the input and output nodes. If none of these hold, it appears to be difficult to obtain an equivalence result, as the strong converse does not hold for symmetrizable AVCs. Indeed, we illustrate in Sec. IX that there exist AVC networks in which no equivalent bit-pipe with fixed capacity exists.

These observations are related to the concept of super-activation [14] which for two channels $\mathcal{C}_1$ and $\mathcal{C}_2$ is defined by the observation that these channels can only be used for reliable communication if they are used jointly, but not in isolation. Super-activation has for example been studied in the context of arbitrarily-varying wiretap channels [15], [16], where it has been shown that there exist pairs of symmetrizable arbitrarily-varying wiretap channels which can be super-activated.

The structure of the paper is as follows. In Sec. II, we formally introduce the problem for both CC and AVC models. In Sec. III we describe the concept of stacked networks, introduced in [13], and state two preliminary lemmas. In Sec. IV, we introduce a lemma demonstrating that training sequences can be used for the CC model to reliably estimate the channel state. In Sec. V, we prove a lemma for the AVC model showing that having access to unlimited shared randomness among certain sets of nodes does not change the capacity region. In Sec. VI, given a channel model and a pair of nodes $u$ and $v$, we determine whether it is possible to transmit information at any positive rate from $u$ to $v$. These results will be used in the equivalence results for both state models: for the CC model, to determine whether feedback is possible, and for the AVC model, whether common randomness can be established (*cf.* Fig. 1). In Sec. VII we present our main equivalence results for the CC model, and in Sec. VIII for the AVC model. In Sec. IX we analyze an example AVC network that we show has no equivalent bit-pipe. In Sec. X we relate our results to the edge removal problem, which has proved difficult for state-less networks but we prove has a simple solution for both CC and AVC models. We conclude in Sec. XI.

## II. Model

Consider a network of nodes $\mathcal{V} := \{1, \ldots, m\}$ with state, given by

$$\mathcal{N} = \left( \prod_{v=1}^{m} \mathcal{X}^{(v)}, \mathcal{S}, p(\mathbf{y}|\mathbf{x}, s), \prod_{v=1}^{m} \mathcal{Y}^{(v)} \right). \tag{1}$$

Herein, $\mathcal{X}^{(v)}$ and $\mathcal{Y}^{(v)}$ denote the input and output alphabets of the node $v$ and $\mathcal{S}$ the set of network states, respectively. This network may represent either a CC or an AVC model. These both assume that the state is chosen not randomly but adversarially; in the CC model the adversary chooses a single state $s \in \mathcal{S}$ that remains constant throughout the code block, whereas in the AVC model the adversary chooses an arbitrary state sequence $s^n \in \mathcal{S}^n$. We assume that the adversary is blind, i.e., that it does not know the transmitted messages, but only the employed codebooks. In this paper we are interested in both CC and AVC problems, but only one at a time. Studying networks with both CC-type state and AVC-type state is beyond our scope.

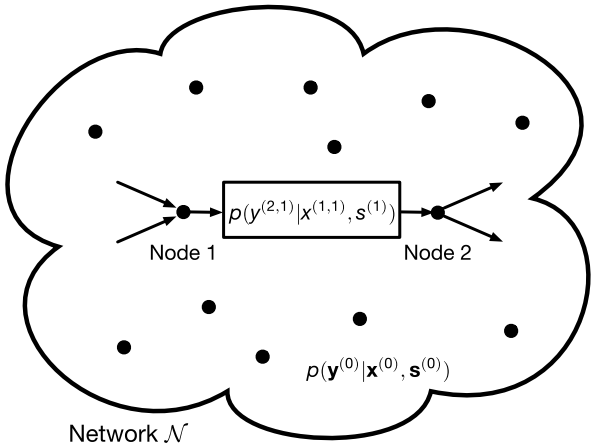


Fig. 2. Decomposition of a network $\mathcal{N}$ into a point-to-point channel between node 1 and 2 with conditional pmf $p(y^{(2,1)}|x^{(1,1)}, s^{(1)})$ and the channels specified by the rest of the network with pmf $p(\mathbf{y}^{(0)}|\mathbf{x}^{(0)}, s^{(0)})$.

We also assume that nodes may use private randomness, independently generated at each node, in their coding operations. This will be important for our results on the AVC achievability arguments, in which nodes generated random quantities and transmit them across the network. In our model we allow each node an unlimited amount of private randomness (in particular, a uniform random variable on the interval $[0, 1]$), although our achievability arguments require no more than $O(\log n)$ bits of private randomness are required at each node. Note that private randomness is quite different from shared randomness, which is a significant asset that trivializes many AVC problems; we do *not* assume that any shared randomness is available in this model. In Sec. V we show that certain forms of shared randomness have no effect on the capacity region of the AVC model, but this is not true for unrestricted shared randomness.

We further assume that there is an independent point-to-point channel from node 1 to node 2 with independent state. That is, $\mathcal{X}^{(1)} = \mathcal{X}^{(1,0)} \times \mathcal{X}^{(1,1)}$, $\mathcal{Y}^{(2)} = \mathcal{Y}^{(2,0)} \times \mathcal{Y}^{(2,1)}$, $\mathcal{S} = \mathcal{S}^{(0)} \times \mathcal{S}^{(1)}$, and

$$p(\mathbf{y}|\mathbf{x}, s) = p(\mathbf{y}^{(0)}|\mathbf{x}^{(0)}, s^{(0)}) p(y^{(2,1)}|x^{(1,1)}, s^{(1)}) \tag{2}$$

where $x^{(1,1)} \in \mathcal{X}^{(1,1)}$, $y^{(2,1)} \in \mathcal{Y}^{(2,1)}$, and $s^{(1)} \in \mathcal{S}^{(1)}$ represent the input, output, and state respectively for the point-to-point channel, and $\mathbf{x}^{(0)} \in \mathcal{X}^{(1,0)} \times \prod_{v \neq 1} \mathcal{X}^{(v)}$, $\mathbf{y}^{(0)} \in \mathcal{Y}^{(2,0)} \times \prod_{v \neq 2} \mathcal{Y}^{(v)}$, and $s^{(0)} \in \mathcal{S}^{(0)}$ represent the input, output, and state respectively for the remainder of the network. This decomposition is visualized in Fig. 2.

The point-to-point channel itself is given by

$$\mathcal{C} = (\mathcal{X}^{(1,1)}, \mathcal{S}^{(1)}, p(y^{(2,1)}|x^{(1,1)}, s^{(1)}), \mathcal{Y}^{(2,1)}). \tag{3}$$

Our main goal is to relate the capacity region of $\mathcal{N}$ to that when point-to-point channel $\mathcal{C}$ is replaced by a noiseless link of fixed capacity (i.e. a bit-pipe). In particular, for any $R \geq 0$, let $\mathcal{N}^R$ be the network in which $\mathcal{C}$ is replaced by a rate-$R$ noiseless (and state-less) bit-pipe $\mathcal{C}^R$ given by

$$\mathcal{C}^R = (\{0,1\}^R, \delta(y^{(2,1)} - x^{(1,1)}), \{0,1\}^R).$$

With other words, the noiseless bit-pipe of capacity $R$ transmits $\lfloor nR \rfloor$ bits over each block of $n$ channel uses with zero error probability for any integer $n \geq 1$.

In general, CCs and AVCs can be quite pathological, so we assume that alphabets $\mathcal{X}^{(v)}$, $\mathcal{S}$, and $\mathcal{Y}^{(v)}$ are all finite sets. Most of our results apply for more general alphabets under mild regularity conditions, but to avoid edge cases and complications we restrict ourselves to finite alphabets. We believe that the interesting consequences of the CC and AVC network models are captured with finite alphabets models, and that the complications that arise for general alphabets are unlikely to make a difference in practice.

*Notation:* Let $[k] = \{1, \ldots, k\}$. A rate vector $\mathcal{R}$ consists of multicast rates $R^{(\{v\}\to U)}$ from each source node $v$ to each destination set $U \subseteq \mathcal{V}$. With a singleton destination set $U = \{u\}$, we sometimes write simply $R^{(v\to u)}$. For each $(v, U)$ pair, there is a message $W^{(\{v\}\to U)} \in \mathcal{W}^{(\{v\}\to U)} = [2^{nR^{(\{v\}\to U)}}]$. Let $W^{(V\to *)}$ denote the vector of all messages originating at nodes $v \in V$, and let $\mathcal{W}^{(V\to *)}$ denote the corresponding message set. Also let $W$ denote the vector of all messages. For a set $\mathcal{A} \subset \mathcal{V}$, we write $X^{(\mathcal{A})} = (X^{(v)} : v \in \mathcal{A})$, and similarly for $Y^{(\mathcal{A})}$. We also write $\mathbf{x}$ for $X^{(\mathcal{V})}$ and $\mathbf{y}$ for $Y^{(\mathcal{V})}$, as in (1). For each node $v$, the private randomness generated at node $v$ is given by a random variable $Q_v$ drawn uniformly from the interval $[0, 1]$.

A blocklength-$n$ solution $\mathsf{S}(\mathcal{N})$ for network $\mathcal{N}$ is given by:

- for each $v \in \mathcal{V}$ and $t \in [n]$, an encoding function

$$X_t^{(v)} : (\mathcal{Y}^{(v)})^{t-1} \times \mathcal{W}^{(\{v\}\to *)} \times [0,1] \to \mathcal{X}^{(v)} \tag{4}$$

  by which node $v$ determines channel input symbol $X_t^{(v)}$ given previously received data $Y_{1:t-1}^{(v)}$, messages $W^{(\{v\}\to *)}$, and private randomness $Q_v$
- for each $(v, U)$ pair and each $u \in U$, a decoding function

$$\widehat{W}^{(\{v\}\to U),u} : (\mathcal{Y}^{(u)})^n \times \mathcal{W}^{(\{u\}\to *)} \times [0,1] \to \mathcal{W}^{(\{v\}\to U)} \cup \{e\} \tag{5}$$

  by which node $u$ determines an estimate $\widehat{W}^{(\{v\}\to U),u}$ of $W^{(\{v\}\to U)}$ given received signals $Y_{1:n}^{(u)}$, messages $W^{(\{u\}\to *)}$, and private randomness $Q_u$. Here, $e$ is a special symbol that denotes declaring an error.

Let $\widehat{W}$ be the complete vector of message estimates, and denote by $\{\widehat{W} \neq W\}$ the event that at least one message is incorrectly decoded. Note that the probability of this event depends on the state sequence $S^n$.

*Definition 1:* The CC-capacity region $\mathscr{R}_{\mathrm{CC}}(\mathcal{N})$ of network $\mathcal{N}$ is given by the closure of the set of rate vectors $\mathcal{R}$ for which there exists a sequence of blocklength-$n$ solutions for which

$$\max_{s\in\mathcal{S}} \Pr(\widehat{W} \neq W | S^n = (s, s, \ldots, s)) \to 0. \tag{6}$$

*Definition 2:* The AVC-capacity region $\mathscr{R}_{\mathrm{AVC}}(\mathcal{N})$ of network $\mathcal{N}$ is given by the closure of the set of rate vectors $\mathcal{R}$ for which there exists a sequence of blocklength-$n$ solutions for which

$$\max_{s^n\in\mathcal{S}^n} \Pr(\widehat{W} \neq W | S^n = s^n) \to 0. \tag{7}$$

It is easy to see that neither $\mathscr{R}_{\mathrm{CC}}(\mathcal{N})$ nor $\mathscr{R}_{\mathrm{AVC}}(\mathcal{N})$ change if the state is allowed to be randomized instead of deterministic, as long as this random choice is independent of the message and the operation of the channel, and for the CC model the state is fixed across the coding block.

Our goal is to prove achievability-type results of the form $\mathscr{R}(\mathcal{N}^R) \subseteq \mathscr{R}(\mathcal{N})$ and converse-type results of the form $\mathscr{R}(\mathcal{N}) \subseteq \mathscr{R}(\mathcal{N}^R)$ for both CC and AVC models.

## III. Stacked Networks

We adopt the notion from [13] of *stacked networks*, wherein we denote by $\underline{\mathcal{N}}$ a network with $N$ independent copies of the network $\mathcal{N}$. Each copy (layer) contains an instance of every channel input and every channel output, all operating independently.[1] Underlines denote stacked variables and vectors, and the argument $\ell$ refers to layer $\ell$, where $\ell \in [N]$. That is, $\underline{X}^{(v)}(\ell)$ is the symbol transmitted by node $v$ in layer $\ell$, and $\underline{Y}^{(v)}(\ell)$ is the symbol received by node $v$ in layer $\ell$. Moreover, we denote $\underline{X}^{(v)} = (\underline{X}^{(v)}(\ell) : \ell \in [N])$ and similarly for $\underline{Y}^{(v)}$. The corresponding alphabets are given by $\underline{\mathcal{X}}^{(v)}$, etc. Message sets are correspondingly increased by a factor of $N$; that is, $\underline{\mathcal{W}}^{(\{v\}\to U)} = (\mathcal{W}^{(\{v\}\to U)})^N$. Rates are therefore defined by $\underline{R}^{(\{v\}\to U)} = |\underline{\mathcal{W}}^{(\{v\}\to U)}|/(nN)$.

We need to differentiate between the CC and AVC models for stacked networks, because for the CC model the state remains constant across time and across layers, whereas for the AVC model the state may vary between layers. For the CC model, the distribution of channel outputs $\underline{\mathbf{Y}} = (\underline{Y}^{(v)} : v \in \mathcal{V})$ given channel inputs $\underline{\mathbf{X}} = (\underline{X}^{(v)} : v \in \mathcal{V})$ and state $s \in \mathcal{S}$ is

$$p(\underline{\mathbf{y}}|\underline{\mathbf{x}}, s) = \prod_{\ell=1}^{N} p(\underline{\mathbf{y}}(\ell)|\underline{\mathbf{x}}(\ell), s) \tag{8}$$

where $\underline{\mathbf{X}}(\ell)$ and $\underline{\mathbf{Y}}(\ell)$ are the vectors of transmitted and received symbols respectively in layer $\ell$. For the AVC model, there is a different state in each layer, denoted as $\underline{S}(\ell)$ for layer $\ell$. The distribution of $\underline{\mathbf{Y}}$ given $\underline{\mathbf{X}}$ and state vector $\underline{S} = (\underline{S}(\ell) : \ell \in [N])$ is

$$p(\underline{\mathbf{y}}|\underline{\mathbf{x}}, \underline{s}) = \prod_{\ell=1}^{N} p(\underline{\mathbf{y}}(\ell)|\underline{\mathbf{x}}(\ell), \underline{s}(\ell)). \tag{9}$$

Solutions for stacked networks are defined similarly to those for unstacked networks, the only difference being that each coding function has access to all stacks from prior time

[1] With the exception that in the CC model, the state is constant across all layers of the network and all time.

instances. In particular, the transmitted symbols for all layers at node $v$ and time $t$ are determined by the causal encoding function

$$\underline{X}_t^{(v)} : (\underline{\mathcal{Y}}^{(v)})^{t-1} \times \underline{\mathcal{W}}^{(\{v\}\to *)} \to \underline{\mathcal{X}}^{(v)} \tag{10}$$

and the decoding function for message $\underline{W}^{(\{v\}\to U)}$ at node $u \in U$ is given by

$$\underline{\widehat{W}}^{(\{v\}\to U),u} : (\underline{\mathcal{Y}}^{(u)})^{n} \times \underline{\mathcal{W}}^{(\{u\}\to *)} \to \underline{\mathcal{W}}^{(\{u\}\to U)} \cup \{e\}. \tag{11}$$

Note that node $v$ has access to its received symbols and messages in all layers when deciding its transmissions. The capacity regions for the stacked networks $\mathscr{R}_{\mathrm{CC}}(\underline{\mathcal{N}})$ and $\mathscr{R}_{\mathrm{AVC}}(\underline{\mathcal{N}})$ are defined analogously as above for unstacked networks.

The following two preliminary lemmas are simple extensions of Lemmas 1 and 4 respectively from [13] to include state.

*Lemma 1:* For any network $\mathcal{N}$, $\mathscr{R}_{\mathrm{CC}}(\mathcal{N}) = \mathscr{R}_{\mathrm{CC}}(\underline{\mathcal{N}})$ and $\mathscr{R}_{\mathrm{AVC}}(\mathcal{N}) = \mathscr{R}_{\mathrm{AVC}}(\underline{\mathcal{N}})$.

*Proof:* The proof for the two state models are largely the same, so we describe them both simultaneously and discuss differences only when they arise. We first prove $\mathscr{R}_{\mathrm{CC}}(\mathcal{N}) \subseteq \mathscr{R}_{\mathrm{CC}}(\underline{\mathcal{N}})$ and $\mathscr{R}_{\mathrm{AVC}}(\mathcal{N}) \subseteq \mathscr{R}_{\mathrm{AVC}}(\underline{\mathcal{N}})$. Consider any rate $\mathcal{R}$ in the interior of the capacity region for $\mathcal{N}$, and we prove that $\mathcal{R}$ is achievable for $\underline{\mathcal{N}}$. This is sufficient because of the closure operation in the definition of the capacity regions. Given any $\lambda > 0$, for $n$ sufficiently large there exists a blocklength-$n$ solution $\mathsf{S}(\mathcal{N})$ on network $\mathcal{N}$ with rate $\mathcal{R}$ and probability of error $\lambda/N$. We construct a solution for stacked network $\underline{\mathcal{N}}$ by repeating $\mathsf{S}(\mathcal{N})$ identically and independently on each layer of $\underline{\mathcal{N}}$. Note that the independence refers to the encoding operations at the nodes, wherein layers are independent of each other, but not necessarily to the input and output random variables at different layers, which may be made dependent via the adversarial state. However, it is still the case that the probability of error for each layer is at most $\lambda/N$, because each layer looks like an ordinary CC or AVC-type model.[2] Thus, by the union bound, the probability of error for the stacked solution is at most $\lambda$.

We now prove $\mathscr{R}_{\mathrm{CC}}(\mathcal{N}) \supseteq \mathscr{R}_{\mathrm{CC}}(\underline{\mathcal{N}})$ and $\mathscr{R}_{\mathrm{AVC}}(\mathcal{N}) \supseteq \mathscr{R}_{\mathrm{AVC}}(\underline{\mathcal{N}})$. Given any blocklength-$n$ solution on $\underline{\mathcal{N}}$, it may be "unraveled" to form a blocklength-$nN$ solution on $\mathcal{N}$ with identical rate and probability of error. In particular, the symbols transmitted at time $t$ by the $N$ layers of $\underline{\mathcal{N}}$ are transmitted at times $(t-1)N+1, \ldots, tN$ on $\mathcal{N}$. Thus causality is maintained at each node. For the AVC model, the same unraveling operation forms an equivalence between state sequences selections for the length-$n$ solution on $\underline{\mathcal{N}}$ and the length-$nN$ solution on $\mathcal{N}$. Thus the worst case probability of error is unchanged. For the CC model, since the state is required to be constant across layers in $\underline{\mathcal{N}}$, the state selection is unchanged and fixed over the blocklength-$nN$ solution. This means that again the state selections are equivalent between the two models, so the probability of error is unchanged. ∎

*Lemma 2:* The capacity regions $\mathscr{R}_{\mathrm{CC}}(\mathcal{N}^R)$ and $\mathscr{R}_{\mathrm{AVC}}(\mathcal{N}^R)$ are continuous in $R$ for all $R > 0$.

[2]In the CC model, the adversary is restricted to maintain a constant state across layers, but this assumption is not necessary for this direction of proof.

*Proof:* We employ a very similar proof technique as that of [13, Lemma 4]. By Lemma 1, it is equivalent to prove continuity for $\mathscr{R}_{\mathrm{CC}}(\underline{\mathcal{N}}^R)$ and $\mathscr{R}_{\mathrm{AVC}}(\underline{\mathcal{N}}^R)$. Fix any $\delta \in (0, R)$ and rate vector $\mathcal{R} \in \mathrm{int}(\mathscr{R}_{\mathrm{CC}}(\underline{\mathcal{N}}^{R+\delta}))$ (resp. $\mathcal{R} \in \mathrm{int}(\mathscr{R}_{\mathrm{AVC}}(\underline{\mathcal{N}}^{R+\delta}))$). Assume that $\underline{\mathcal{N}}^{R+\delta}$ has $N$ layers. Let $\underline{\mathcal{N}}^{R-\delta}$ be an $N'$-fold stacked network with

$$N'(R-\delta) \geq N(R+\delta). \tag{12}$$

For all $\lambda > 0$, there exists solution $\mathsf{S}(\underline{\mathcal{N}}^{R+\delta})$ with rate vector $\mathcal{R}$ and probability of error $\lambda$. We define a solution $\mathsf{S}(\underline{\mathcal{N}}^{R-\delta})$ based on $\mathsf{S}(\underline{\mathcal{N}}^{R+\delta})$ as follows. Use precisely the same coding operations aside from the bit-pipe $\underline{\mathcal{C}}^{R+\delta}$ for the first $N$ layers of the stack, and send the $\lfloor N(R+\delta) \rfloor$ bits to be sent across $\underline{\mathcal{C}}^{R+\delta}$ instead across the bit-pipe $\underline{\mathcal{C}}^{R-\delta}$. This can be done because of (12). Note that the resulting rate vector for $\mathsf{S}(\underline{\mathcal{N}}^{R-\delta})$ is

$$\mathcal{R}' = \frac{\mathcal{R}N}{N'} > \mathcal{R}\frac{N}{N(R+\delta)/(R-\delta)+1}. \tag{13}$$

Thus the difference between $\mathcal{R}$ and $\mathcal{R}'$ vanishes as $N \to \infty$ and $\delta \to 0$.

Recall that for the CC-model (resp. AVC-model), the state does not affect operation of the bit-pipes. Meanwhile, as the rest of the network is operated identically in the two solutions—aside from the $N'-N$ unused layers in the solution on $\underline{\mathcal{N}}^{R-\delta}$—the effect of the state is precisely the same. Thus the modified solution on $\underline{\mathcal{N}}^{R-\delta}$ has precisely the same probability of error $\lambda$. Therefore $\mathcal{R}' \in \mathscr{R}_{\mathrm{CC}}(\underline{\mathcal{N}}^{R-\delta})$ (resp. $\mathcal{R}' \in \mathscr{R}_{\mathrm{AVC}}(\underline{\mathcal{N}}^{R-\delta})$). ∎

## IV. Compound Channel Training Lemma

The following lemma will be used several times in CC results. It asserts that CC states can be estimated using training sequences.

*Lemma 3:* Fix a point-to-point CC $(\mathcal{X}, \mathcal{S}, p(y|x,s), \mathcal{Y})$. For any input sequence $x_{1:n} \in \mathcal{X}^n$ and output sequence $y_{1:n} \in \mathcal{Y}^n$, define the set of maximum likelihood state estimates as

$$\hat{\mathcal{S}}(x_{1:n}, y_{1:n}) = \left\{\hat{s} \in \mathcal{S} : p(y_{1:n}|x_{1:n}, \hat{s}) = \max_{s' \in \mathcal{S}} p(y_{1:n}|x_{1:n}, s')\right\}. \tag{14}$$

For any state $s \in \mathcal{S}$, let the set of states equivalent to $s$ be[3]

$$\bar{\mathcal{S}}(s) = \{\bar{s} \in \mathcal{S} : p(y|x,\bar{s}) = p(y|x,s) \text{ for all } x \in \mathcal{X}, y \in \mathcal{Y}\}. \tag{15}$$

Then, for any $s \in \mathcal{S}$,

$$\lim_{n\to\infty} \Pr\left(\hat{\mathcal{S}}(\alpha_{1:n}, Y_{1:n}) \neq \bar{\mathcal{S}}(s)\right) = 0 \tag{16}$$

[3]In many cases, each state induces a distinct channel distribution, so we would have $\bar{\mathcal{S}}(s) = \{s\}$. However, there are important scenarios when this is not the case, such as when $S$ is the full network channel state, and the channel from $X$ to $Y$ represents just part of the overall network channel model. Different states might induce the same behavior from $X$ to $Y$ but different behaviors elsewhere in the network. For example, consider two BSCs and a ternary network state $S$ such that the crossover probabilities of the two channels are $(0,0)$ if $S=0$, $(0,1)$ if $S=1$, or $(1,0)$ if $S=2$. Thus $S=0$ and $S=1$ induce exactly the same behavior in the first channel, but are materially different when considering the entire network.

where $\alpha_{1:n}$ is a random training sequence drawn uniformly i.i.d. from $\mathcal{X}^n$, and $Y_{1:n} \sim p(y_{1:n}|\alpha_{1:n}, s)$.

*Proof:* Fix $s \in \mathcal{S}$. Note that if $\bar{s} \in \bar{\mathcal{S}}(s)$, then by definition all probabilities for $\bar{s}$ are identical to those for $s$, so $\bar{s} \in \hat{\mathcal{S}}(x_{1:n}, y_{1:n})$ if and only if $s \in \hat{\mathcal{S}}(x_{1:n}, y_{1:n})$. Thus, to prove (16) we need to show that with probability approaching 1,

$$p(Y_{1:n}|\alpha_{1:n}, s) > p(Y_{1:n}|\alpha_{1:n}, s') \text{ for all } s' \in \bar{\mathcal{S}}(s)^c \tag{17}$$

where $\bar{\mathcal{S}}(s)^c = \mathcal{S} \setminus \bar{\mathcal{S}}(s)$.

Note that $\hat{\mathcal{S}}(\alpha_{1:n}, Y_{1:n})$ consists of the set of $\hat{s}$ that minimize

$$-\frac{1}{n}\sum_{t=1}^{n} \log p(Y_t|\alpha_t, \hat{s}). \tag{18}$$

For any $s' \in \mathcal{S}$, the quantities $-\log p(Y_t|\alpha_t, s')$ are i.i.d. with expected value

$$\frac{1}{|\mathcal{X}|}\sum_{x,y} -p(y|x,s)\log p(y|x,s') = H(Y|X, S=s) + \Delta_{s,s'} \tag{19}$$

where

$$\Delta_{s,s'} := \sum_{x\in\mathcal{X}} \frac{1}{|\mathcal{X}|} D\big(p(y|x,s)\|p(y|x,s')\big). \tag{20}$$

Note that $\Delta_{s,s'} = 0$ if and only if $s' \in \bar{\mathcal{S}}(s)$. Let $\delta_s = \min\{\Delta_{s,s'} : \Delta_{s,s'} > 0\}$. We have $\delta_s > 0$ since $\mathcal{S}$ is finite. Let $\tau_s = H(Y|X, S=s) + \delta_s/2$. Hence

$$\mathbb{E}\big[-\log p(Y_t|\alpha_t, s)\big] > \tau_s, \tag{21}$$

$$\mathbb{E}\big[-\log p(Y_t|\alpha_t, s')\big] < \tau_s \text{ for any } s' \in \bar{\mathcal{S}}(s)^c. \tag{22}$$

Thus, by the Law of Large Numbers,

$$\Pr\left(-\frac{1}{n}\sum_{t=1}^{n} \log p(Y_t|\alpha_t, s) > \tau_s\right) \to 1, \tag{23}$$

$$\Pr\left(-\frac{1}{n}\sum_{t=1}^{n} \log p(Y_t|\alpha_t, s') < \tau_s \text{ for all } s' \in \bar{\mathcal{S}}(s)^c\right) \to 1. \tag{24}$$

Therefore (17) holds with probability approaching 1. ∎

## V. Arbitrarily Varying Shared Randomness Lemma

A key element of proving AVC equivalence results, and indeed of many existing AVC results, is the role of shared randomness between nodes. This is the essence of the difference between the classical deterministic and random coding models for the point-to-point AVC, and so one may ask exactly when does having shared randomness between nodes change or not change the capacity region. In this section, we prove a generic lemma stating that the capacity region of an AVC network does not change if certain groups of nodes have access to shared randomness. This lemma will be used several times in proving our equivalence results. For the no adversary model, it was shown in [17] that the capacity region for average probability of error does not change even if all nodes have access to a single infinite entropy source of common randomness (modeled as a uniform random variable on the unit interval). This strong result does not hold for the AVC model, but, as stated below, for a given node $v$, if node $v$ is allowed to share an infinite entropy source of randomness with all other nodes to which it can communicate at any positive rate, then the capacity region does not change.

The proof is a generalization of the random code reduction [11, Lemma 12.8]. This lemma proves that for the point-to-point AVC an arbitrary amount of shared randomness between encoder and decoder can be reduced to an asymptotically negligible amount (in particular, $O(\log n)$ bits). This leads to [11, Th. 12.11], stating that the capacity of an AVC is either 0 or the random coding capacity, because if the capacity is positive, than a small amount of shared randomness can be set up, and thus full shared randomness can be simulated. We use essentially the same technique here.

To be precise, we define the following variant on our coding model.

*Definition 3:* Let $\tilde{\mathscr{R}}_{\text{AVC}}(\mathcal{N})$ be the capacity region for the AVC network $\mathcal{N}$ for the following shared randomness coding model. For each node $v$, let $\tilde{Q}_v$ be a uniform random variable on the interval $[0, 1]$, independent from each other, from the messages, from channel noise, and from the state sequence. Assume $\tilde{Q}_v$ is available at node $v$ and at all nodes $u$ for which there exists a rate vector $\mathcal{R} \in \mathscr{R}_{\text{AVC}}(\mathcal{N})$ with $R^{(v\to u)} > 0$.

*Lemma 4:* For any network $\mathcal{N}$, $\tilde{\mathscr{R}}_{\text{AVC}}(\mathcal{N}) = \mathscr{R}_{\text{AVC}}(\mathcal{N})$.

Before proving Lemma 4, we need the following lemma, which is the essence of the random code reduction.

*Lemma 5:* Let $Q$ and $Z$ be independent random variables with (not necessarily finite) alphabets $\mathcal{Q}$ and $\mathcal{Z}$ respectively. Let $f(q, z, s^n) \in [0, 1]$ be a function defined for $q \in \mathcal{Q}$, $z \in \mathcal{Z}$ and $s^n \in \mathcal{S}^n$. Suppose for some $\eta > 0$,

$$\mathbb{E} f(Q, Z, s^n) \le \eta \text{ for all } s^n \in \mathcal{S}^n. \tag{25}$$

Then for sufficiently large $n$, there exist $q_1, \ldots, q_{n^2} \in \mathcal{Q}$ such that

$$\frac{1}{n^2}\sum_{j=1}^{n^2} \mathbb{E} f(q_j, Z, s^n) \le 2\eta \text{ for all } s^n. \tag{26}$$

*Proof:* Let $Q_1, \ldots, Q_{n^2}$ be i.i.d. random variables with the same distribution as $Q$, all independent of $Z$. We have

$$\mathbb{P}\left(\frac{1}{n^2}\sum_{j=1}^{n^2} \mathbb{E}[f(Q_j, Z, s^n)|Q_j] > 2\eta \text{ for any } s^n\right) \tag{27}$$

$$\le \sum_{s^n} \mathbb{P}\left(\frac{1}{n^2}\sum_{j=1}^{n^2} \mathbb{E}[f(Q_j, Z, s^n)|Q_j] > 2\eta\right) \tag{28}$$

$$= \sum_{s^n} \mathbb{P}\left(2^{\sum_{j=1}^{n^2} \mathbb{E}[f(Q_j,Z,s^n)|Q_j]} > 2^{n^2 2\eta}\right) \tag{29}$$

$$\le \sum_{s^n} 2^{-n^2 2\eta}\, \mathbb{E} 2^{\sum_{j=1}^{n^2} \mathbb{E}[f(Q_j,Z,s^n)|Q_j]} \tag{30}$$

$$= \sum_{s^n} 2^{-n^2 2\eta} \left( \mathbb{E} 2^{\mathbb{E}[f(Q,Z,s^n|Q)]} \right)^{n^2} \tag{31}$$

$$\leq \sum_{s^n} 2^{-n^2 2\eta} (1 + \mathbb{E} f(Q, Z, s^n))^{n^2} \tag{32}$$

$$\leq |\mathcal{S}|^n 2^{-n^2 2\eta} (1+\eta)^{n^2} \tag{33}$$

$$\leq |\mathcal{S}|^n 2^{-n^2 \eta (2 - \log e)} \tag{34}$$

where (28) follows from the union bound, (30) from Markov's inequality, (31) from the fact that $Q_j$ for $j \in [n^2]$ are i.i.d. with the same distribution as $Q$, (32) follows from the fact that $f(q, z, s^n) \in [0, 1]$ and $2^x \leq 1 + x$ for any $x \in [0, 1]$, (33) follows from the assumption in (25), and (34) follows because $1 + \eta \leq e^\eta$. As $2 > \log e$, the quantity in (34) is vanishing in $n$. Thus, for sufficiently large $n$ the probability in (27) is strictly less than 1, meaning there exists at least one set of constants $\{q_j\}_{j \in [n^2]}$ satisfying (26). ∎

*Proof of Lemma 4:* It is obvious that $\mathscr{R}_{\text{AVC}}(\mathcal{N}) \subseteq \tilde{\mathscr{R}}_{\text{AVC}}(\mathcal{N})$. To prove $\tilde{\mathscr{R}}_{\text{AVC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{AVC}}(\mathcal{N})$, let $\mathcal{R}$ be a rate vector in the interior of $\tilde{\mathscr{R}}_{\text{AVC}}(\mathcal{N})$, and we prove that $\mathcal{R} \in \mathscr{R}_{\text{AVC}}(\mathcal{N})$. For sufficiently large $n$ there exists an $n$-length solution $\mathsf{S}(\mathcal{N})$ for the random coding model with rate $\mathcal{R}$ and probability of error $2^{-m}\epsilon$. Given $q_1, \ldots, q_m \in [0, 1]$ and $s^n \in \mathcal{S}^n$, let $e(q_1, \ldots, q_m, s^n)$ be the probability of error for $\mathsf{S}(\mathcal{N})$ conditioned on $\tilde{Q}_i = q_i$ for $i \in \mathcal{V}$, and $S^n = s^n$. Note that this quantity is averaged over the random choice of messages, the random channel noise, and any private randomness. Thus

$$\mathbb{E} e(\tilde{Q}_1, \ldots, \tilde{Q}_m, s^n) \leq 2^{-m}\epsilon \text{ for all } s^n. \tag{35}$$

We next prove that there exist $q_{ij} \in [0, 1]$ for $i \in \mathcal{V}$ and $j \in [n^2]$ such that

$$\frac{1}{n^{2m}} \sum_{j_1, \ldots, j_m \in [n^2]} e(q_{1j_1}, \ldots, q_{mj_m}, s^n) \leq \epsilon \text{ for all } s^n. \tag{36}$$

We now apply Lemma 5 $m$ times to the initial random coding probability of error in (35). In particular, by (35), applying Lemma 5 with particularizations $e \to f$, $\tilde{Q}_1 \to Q$, and $(\tilde{Q}_2, \ldots, \tilde{Q}_m) \to Z$, there exists $q_{1j} \in [0, 1]$ for $j \in [n^2]$ where

$$\frac{1}{n^2} \sum_{j=1}^{n^2} \mathbb{E} e(q_{1j}, \tilde{Q}_2, \ldots, \tilde{Q}_m) \leq 2^{-m+1}\epsilon \text{ for all } s^n. \tag{37}$$

Let $A_1$ be uniformly distributed on $\{q_{11}, \ldots, q_{1n^2}\}$. Thus (37) may be rewritten

$$\mathbb{E} e(A_1, \tilde{Q}_2, \ldots, \tilde{Q}_m) \leq 2^{-m+1}\epsilon \text{ for all } s^n. \tag{38}$$

Now applying Lemma 5 again with particularizations $e \to f$, $\tilde{Q}_2 \to Q$, and $(A_1, \tilde{Q}_3, \ldots, \tilde{Q}_m) \to Z$ allows us to conclude that there exist $q_{2j} \in [0, 1]$ for $j \in [n^2]$ such that

$$\frac{1}{n^2} \sum_{j=1}^{n^2} \mathbb{E} e(A_1, q_{2j}, \tilde{Q}_3, \ldots, \tilde{Q}_m) \leq 2^{-m+2}\epsilon \text{ for all } s^n. \tag{39}$$

Repeating this argument $m$ times proves (36).

We now construct a solution on network $\mathcal{N}$ using only private randomness as follows. At each node $v$, from private randomness $Q_v$ generate a random variable $J_v$ uniformly distributed in $[n^2]$, independent of all messages and received signals. By definition, for each node $u$ for which there exists $\mathcal{R} \in \mathscr{R}_{\text{AVC}}(\mathcal{N})$ with $R^{(v \to u)} > 0$, there is a positive rate solution with arbitrarily small probability of error that conveys data from node $v$ to node $u$. Using these positive rate solutions, $J_v$ may be transmitted to all such nodes $u$ essentially for free, because $\log(n^2)$ bits is sub-linear in $n$. Subsequently, all nodes proceed with the original code as if $\tilde{Q}_v = q_{vJ_v}$. Since in the shared randomness coding model, $\tilde{Q}_v$ is only available at these nodes $u$, $J_v$ has been successfully delivered to all the nodes that require it. For state sequence $s^n$, the resulting probability of error is given by

$$\frac{1}{n^{2m}} \sum_{j_1, \ldots, j_m \in [n^2]} e(q_{1j_1}, \ldots, q_{mj_m}, s^n) \tag{40}$$

which is at most $\epsilon$ by (36). This proves that the probability of error for code using only private randomness can be made arbitrarily small. ∎

Note that the above argument works equally well for stacked networks; therefore we also have $\tilde{\mathscr{R}}_{\text{AVC}}(\underline{\mathcal{N}}) = \mathscr{R}_{\text{AVC}}(\underline{\mathcal{N}})$.

## VI. Positive Rate Conditions

For both CC and AVC models, it will be important to know whether any information at all can be sent between nodes. This positive (but arbitrarily small) rate will be used for feedback in the CC model and generating shared randomness in the AVC model (see Fig. 1). Thus in this section we investigate the set of node pairs $(u, v)$ for which positive rate can be sent from $u$ to $v$. We do this first without state, and then extend it for the CC and AVC models.

### A. Positive Rate Without State

Assume for now that $\mathcal{S}$ contains only a single element, in which case $\mathscr{R}_{\text{CC}}(\mathcal{N}) = \mathscr{R}_{\text{AVC}}(\mathcal{N})$, and we denote both by $\mathscr{R}(\mathcal{N})$. We form a set $\mathcal{P} \subset \mathcal{V} \times \mathcal{V}$ and subsequently show that $\mathcal{P}$ is precisely the set of node pairs that can sustain positive rate. For the CC model, we will be interested in whether $(2, 1) \in \mathcal{P}$; *i.e.* whether feedback is possible with respect to the point-to-point channel from node 1 to node 2. On the other hand, for the AVC model, we care whether there exists a node $u$ such that $(u, 1), (u, 2) \in \mathcal{P}$.

The set $\mathcal{P}$ is formed via the following steps:

1) Initialize $\mathcal{P}$ as $\{(u, u) : u \in \mathcal{V}\}$.
2) If there is a pair of nodes $(u, v) \notin \mathcal{P}$, and a set $\mathcal{A} \subset \mathcal{V}$ such that $(j, v) \in \mathcal{P}$ for all $j \in \mathcal{A}$, and
$$\max_{p(x^{(u)}), x^{(\{u\}^c)}} I(X^{(u)}; Y^{(\mathcal{A})} | X^{(\{u\}^c)} = x^{(\{u\}^c)}) > 0, \tag{41}$$
then add $(u, v)$ to $\mathcal{P}$.
3) Repeat step 2 until there are no additional such pairs $(u, v)$.

Note that the condition in step (2) on a pair of nodes $(u, v)$ is monotonic in the sense that if it holds at any point in the procedure, adding other pairs to $\mathcal{P}$ cannot cause it to cease holding. Thus, no matter the order in which pairs are added to $\mathcal{P}$, any pair that satisfies the condition at any point will

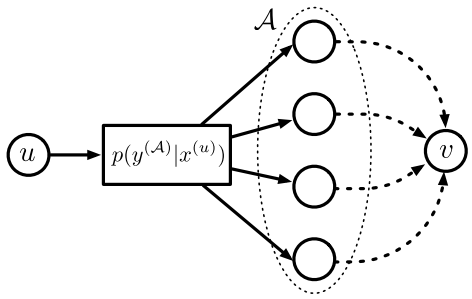


Fig. 3. Positive rate can be established between the pair of nodes $(u, v)$ if (41) is satisfied and for all $j \in \mathcal{A}$, positive rate can be sent from $j$ to $v$.

eventually be added. Thus, the above procedure defines $\mathcal{P}$ uniquely.

The mutual information in (41) represents the capacity of a point-to-point channel with input $X^{(u)}$ and output $Y^{(\mathcal{A})}$, even though $Y^{(\mathcal{A})}$ represents all received values by nodes in $\mathcal{A}$, which are not available at any single receiver. Additionally, we maximize over constants $x^{(\{u\}^c)}$ in case the channel from $X^{(u)}$ to $Y^{(\mathcal{A})}$ only has positive capacity for certain transmissions by the other nodes.

*Theorem 6:* If $(u, v) \in \mathcal{P}$, then there exists an $\mathcal{R} \in \mathscr{R}(\mathcal{N})$ with $R^{(u\to v)} > 0$.

*Proof:* A detailed proof is given by the proof of the stronger result Lemma 9, to be stated below. Roughly, the solution is visualized in Fig. 3 and derived as follows. A node may trivially send arbitrary amounts of information to itself; thus $R^{(u\to u)} > 0$ is achievable for any $u \in \mathcal{V}$. We proceed by induction to prove the theorem for pairs $(u, v) \in \mathcal{P}$ with $u \neq v$. Consider the specific step in the construction of $\mathcal{P}$ at which $(u, v)$ is added, and let $\mathcal{A}$ satisfy (41). We assume that for all $j \in \mathcal{A}$, positive rate can be sent from $j$ to $v$. To send positive rate from $u$ to $v$, we employ a point-to-point channel code from $X^{(u)}$ to $Y^{(\mathcal{A})}$. A message is chosen at node $u$, and the corresponding codeword is transmitted by node $u$ and received by nodes in $\mathcal{A}$. Next, the received sequences are transmitted from nodes in $\mathcal{A}$ to node $v$ using positive-rate solutions that are assumed to exist by the induction hypothesis and since by construction $(j, v) \in \mathcal{P}$ for all $j \in \mathcal{A}$. Finally, node $v$ decodes the point-to-point code. ∎

The following theorem gives the converse result, stating that if $(u, v) \notin \mathcal{P}$, then values received at node $v$ are conditionally independent of values sent from node $u$ given messages that originate outside node $u$. This indicates that all information known at node $v$ originates outside of node $u$; i.e., the input at node $u$ cannot influence the output at node $v$. This is a much stronger statement than a simple converse, and indeed even stronger than a usual "strong" converse, but it is necessary to prove equivalence results.

*Theorem 7:* If $(u, v) \notin \mathcal{P}$, then for any solution $\mathcal{S}(\mathcal{N})$, $X_{1:n}^{(u)} \to W^{(\{u\}^c\to *)} \to Y_{1:n}^{(v)}$ forms a Markov chain.

*Proof:* Fix $(u, v) \notin \mathcal{P}$. Let $\mathcal{A} := \{i : (i, v) \in \mathcal{P}\}$. By the definition of $\mathcal{P}$, for any $i \notin \mathcal{A}$,

$$\max_{p(x^{(\{i\})}), x^{(\{i\}^c)}} I(X^{(\{i\})}; Y^{(\mathcal{A})} | X^{(\{i\}^c)} = x^{(\{i\}^c)}) = 0. \quad (42)$$

In other words, the conditional distribution $p(y^{(\mathcal{A})}|\mathbf{x})$ does not depend on $x^{(\{i\})}$. As this holds for all $i \notin \mathcal{A}$, it must be that $p(y^{(\mathcal{A})}|\mathbf{x}) = p(y^{(\mathcal{A})}|x^{(\mathcal{A})})$. Hence, for any solution $\mathcal{S}(\mathcal{N})$, we have the Markov chain

$$X_t^{(\mathcal{A}^c)} \to X_t^{(\mathcal{A})} \to Y_t^{(\mathcal{A})} \quad (43)$$

for each time $t$. We may now write

$$p\left(y_{1:n}^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}\right) \quad (44)$$

$$= \prod_{t=1}^{n} p\left(y_t^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}, y_{1:t-1}^{(\mathcal{A})}\right) \quad (45)$$

$$= \prod_{t=1}^{n} \sum_{x_t^{(\mathcal{V})}} p\left(x_t^{(\mathcal{V})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}, y_{1:t-1}^{(\mathcal{A})}\right) p\left(y_t^{(\mathcal{A})} \middle| x_t^{(\mathcal{V})}\right) \quad (46)$$

$$= \prod_{t=1}^{n} \sum_{x_t^{(\mathcal{V})}} p\left(x_t^{(\mathcal{V})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}, y_{1:t-1}^{(\mathcal{A})}\right) p\left(y_t^{(\mathcal{A})} \middle| x_t^{(\mathcal{A})}\right) \quad (47)$$

$$= \prod_{t=1}^{n} \sum_{x_t^{(\mathcal{A})}} p\left(x_t^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, x_{1:n}^{(\mathcal{A}^c)}, y_{1:t-1}^{(\mathcal{A})}\right) p\left(y_t^{(\mathcal{A})} \middle| x_t^{(\mathcal{A})}\right) \quad (48)$$

$$= \prod_{t=1}^{n} \sum_{x_t^{(\mathcal{A})}} p\left(x_t^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, y_{1:t-1}^{(\mathcal{A})}\right) p\left(y_t^{(\mathcal{A})} \middle| x_t^{(\mathcal{A})}\right) \quad (49)$$

$$= \prod_{t=1}^{n} p\left(y_t^{(\mathcal{A})} \middle| w^{(\mathcal{A})}, y_{1:t-1}^{(\mathcal{A})}\right) = p\left(y_{1:n}^{(\mathcal{A})} \middle| w^{(\mathcal{A})}\right) \quad (50)$$

where (47) follows from (43), and (49) follows by the dependency requirements of the coding at nodes in $\mathcal{A}$. From this derivation, we conclude that $X_{1:n}^{(\mathcal{A}^c)} \to W^{(\mathcal{A})} \to Y_{1:n}^{(\mathcal{A})}$ forms a Markov chain. This completes the proof since $v \in \mathcal{A}$ and $u \in \mathcal{A}^c$. ∎

Theorems 6 and 7 completely determine when any positive rate is achievable, as stated in the following corollary.

*Corollary 8:* There exists a rate vector $\mathcal{R} \in \mathscr{R}(\mathcal{N})$ with $R^{(\{v\}\to U)} > 0$ if and only if $(v, i) \in \mathcal{P}$ for all $i \in U$.

Note that the "only if" direction of Corollary 8 is weaker than Theorem 7, because even if $R^{(v\to u)}$ cannot be positive, it does not mean that the strong statement of Theorem 7 holds.

### *B. Positive Rate for the CC Model*

We now extend the above results for CC-type state. For each $s \in \mathcal{S}$, define $\mathcal{P}_s$ as above for $\mathcal{P}$, but with fixed state $S = s$. Let $\mathcal{P}_{\text{CC}} = \bigcap_{s\in\mathcal{S}} \mathcal{P}_s$.

For any state $s$ such that $(u, v) \in \mathcal{P}_s$, the following lemma establishes the existence of solutions for the CC model with positive rate from $u$ to $v$ such that (i) if the state is $s$, node $v$ can reliably decode the message; and (ii) if the state is *not* $s$, node $v$ either decodes correctly or declares an error. Recall that we use the symbol $e$ to signify a decoder declaring an error. We construct these solutions using training sequences (*cf.* Lemma 3), wherein node $v$ only decodes if $s$ is among the most likely states. Thus if the true state is not $s$, either node $v$ will discover this and declare an error, or the channel is indistinguishable from that with state $s$, so node $v$ will decode reliably. The solutions from this lemma will be used

to prove that positive rate can be transmitted from $u$ to $v$ for $(u, v) \in \mathcal{P}_{\mathrm{CC}}$.

*Lemma 9:* For any state $s \in \mathcal{S}$, and all $(u, v) \in \mathcal{P}_s$, there exist a sequence of solutions $\mathsf{S}^{(n)}_{u,v,s}(\mathcal{N})$ with rate $R^{(u\to v)} > 0$ such that

1) if $S = s$ then the probability of error vanishes with $n$, and
2) if $S \neq s$ then the probability of making an error without declaring an error (*i.e.* that $\widehat{W}^{(u\to v)} \notin \{W^{(u\to v)}, e\}$) vanishes with $n$.

*Proof:* We adopt the convention that a node may send arbitrary amounts of information to itself; thus the lemma is immediate if $u = v$. We proceed by induction to prove the theorem for pairs $(u, v) \in \mathcal{P}_s$ with $u \neq v$. Consider the specific step in the construction of $\mathcal{P}_s$ at which $(u, v)$ was added. There is a set $\mathcal{A} \subset \mathcal{V}$ such that for some distribution $p(x^{(u)})$ and constant $x^{(\{u\}^c)}$,

$$I(X^{(u)}; Y^{(\mathcal{A})} | X^{(\{u\}^c)} = x^{(\{u\}^c)}, S = s) > 0, \tag{51}$$

and $(j, v)$ for all $j \in \mathcal{A}$ has already been added to $\mathcal{P}_s$. We assume there exist sequences of solutions $\mathsf{S}^{(n_j)}_{j,v,s}(\mathcal{N})$ for all $j \in \mathcal{A}$, with rates $R^{(j\to v)} > 0$, satisfying the probability of error constraints in the statement of the lemma. Fix a length $n$ to be determined later.

We now describe the coding procedure. Initially node $u$ chooses a message $W^{(u\to v)} \in \mathcal{W}^{(u\to v)} = [2^{n\tilde{R}^{(u\to v)}}]$, where $\tilde{R}^{(u\to v)}$ is any positive number strictly smaller than the mutual information in (51). Coding proceeds in three sessions, described as follows. The lengths of the first two sessions are $n$, and that of the third session is $\sum_{j\in\mathcal{A}} n_j$. Thus the quantity $\tilde{R}^{(u\to v)}$ is not the rate achieved by the code, because the overall blocklength is longer than $n$.

*Session 1:* Node $u$ transmits a training sequence $\alpha_{1:n}$ drawn randomly and uniformly from $(\mathcal{X}^{(u)})^n$ while other nodes transmit the constant $x^{(\{u\}^c)}$. The training sequence constitutes part of the codebook and is revealed to all nodes prior to coding. For each $j \in \mathcal{A}$, let $Y^{(j)}_{1:n}$ be the received sequence at node $j$ for each $j \in \mathcal{A}$.

*Session 2:* Node $u$ transmits $W^{(u\to v)}$ via an $n$-length point-to-point channel code from $X^{(u)}$ to $Y^{(\mathcal{A})}$ with input distribution $p(x^{(u)})$ and distribution conditioned on $X^{(\{u\}^c)} = x^{(\{u\}^c)}$ and $S = s$, while all other nodes transmit the constant $x^{(\{u\}^c)}$. Let $Y^{(j)}_{n+1:2n}$ be the received sequence at node $j$ at each $j \in \mathcal{A}$.

*Session 3:* Dividing into $|\mathcal{A}|$ sub-sessions, we run one sub-session for each $j \in \mathcal{A}$, in which $\mathsf{S}^{(n_j)}_{j,v,s}(\mathcal{N})$ is employed to transmit $Y^{(j)}_{1:2n}$ from $j$ to $v$, where the blocklength is given by

$$n_j = \left\lceil \frac{2n \log |\mathcal{Y}^{(j)}|}{R^{j\to v}} \right\rceil \tag{52}$$

so that $2^{n_j R^{(j\to v)}} \geq |\mathcal{Y}^{(j)}|^{2n}$. Let $\hat{Y}^{(j)}_{1:2n}$ be the decoded sequence at node $v$.

*Decoding:* If any of the solutions $\mathsf{S}^{(n_j)}_{j,v,s}(\mathcal{N})$ declares an error, then node $v$ declares an error. Otherwise, given $\hat{Y}^{(\mathcal{A})}_{1:n}$ node $v$ determines whether $s$ is among the most likely states given the training sequence; that is

$$p\big(\hat{Y}^{(\mathcal{A})}_{1:n} | \alpha_{1:n}, X^{(\{u\}^c)}_{1:n} = x^{(\{u\}^c)}_{1:n}, S = s\big) = \max_{s'} p\big(\hat{Y}^{(\mathcal{A})}_{1:n} | \alpha_{1:n}, X^{(\{u\}^c)}_{1:n} = x^{(\{u\}^c)}_{1:n}, S = s'\big). \tag{53}$$

If (53) does not hold, then node $v$ declares an error. If it does, then node $v$ decodes the message from $\hat{Y}^{(\mathcal{A})}_{n+1:2n}$ using the point-to-point channel decoder. Let $\widehat{W}^{(u\to v)}$ be the decoded message.

*Achieved Rate:* Recall that all we need to show is that the achieved rate $R^{(u\to v)}$ is positive. The total blocklength for the code is $2n + \sum_{j\in\mathcal{A}} n_j$, so the overall rate is given by

$$R^{(u\to v)} = \frac{n\tilde{R}^{(u\to v)}}{2n + \sum_{j\in\mathcal{A}} n_j} \tag{54}$$

$$\geq \frac{\tilde{R}^{(u\to v)}}{2 + \sum_{j\in\mathcal{A}} \frac{2\log|\mathcal{Y}^{(j)}|}{R^{j\to v}} + \frac{|\mathcal{A}|}{n}}. \tag{55}$$

Note that $R^{(u\to v)}$ is bounded above 0 for sufficiently large $n$.

*Probability of Error Analysis:* First consider the case that $S = s$. We need to show that $\Pr(\widehat{W}^{(u\to v)} \neq W^{(u\to v)})$ can be made arbitrarily small. Define the error events

$$\mathcal{E}_1 := \left\{\hat{Y}^{(\mathcal{A})}_{1:2n} \neq Y^{(\mathcal{A})}_{1:2n}\right\}, \tag{56}$$

$$\mathcal{E}_2 := \left\{s \notin \arg\max_{s'} p(\hat{Y}^{(\mathcal{A})}_{1:n} | \alpha_{1:n}, X^{(\{u\}^c)} = x^{(\{u\}^c)}, S = s')\right\}, \tag{57}$$

$$\mathcal{E}_3 := \left\{\widehat{W}^{(u\to v)} \neq W^{(u\to v)}\right\}. \tag{58}$$

The overall error event is $\mathcal{E}_3$, and we can upper bound its probability by

$$\Pr(\mathcal{E}_3 | S = s) \leq \Pr(\mathcal{E}_1 | S = s) + \Pr(\mathcal{E}_1^c \cap \mathcal{E}_2 | S = s) + \Pr(\mathcal{E}_3 | \mathcal{E}_1^c, \mathcal{E}_2^c, S = s). \tag{59}$$

By the inductive assumptions that $\mathsf{S}^{(n_j)}_{j,v,s}(\mathcal{N})$ have vanishing probability of error given state $s$ for each $j \in \mathcal{A}$, $\Pr(\mathcal{E}_1) \to 0$. By Lemma 3, $\Pr(\mathcal{E}_1^c \cap \mathcal{E}_2 | S = s) \to 0$. Finally, $\Pr(\mathcal{E}_3 | \mathcal{E}_1^c, \mathcal{E}_2^c, S = s)$ is merely the probability of error of the point to point code from $u$ to $\mathcal{A}$, so it vanishes as $n \to \infty$. Thus the overall probability of error may be made arbitrarily small.

Now consider the case that $S = \bar{s} \neq s$. Define the additional error events

$$\mathcal{E}_4 := \Big\{\text{solution } \mathsf{S}^{(n_j)}_{j,v,s}(\mathcal{N}) \text{ declares an error for some } j \in \mathcal{A}\Big\}, \tag{60}$$

$$\mathcal{E}_5 := \left\{\widehat{W}^{(u\to v)} \notin \{W^{(u\to v)}, e\}\right\}. \tag{61}$$

We need to show $\Pr(\mathcal{E}_5) \to 0$ as $n \to \infty$. If either $\mathcal{E}_2$ or $\mathcal{E}_4$ occurs, then node $v$ declares an error, so $\mathcal{E}_5 \subset \mathcal{E}_2^c \cap \mathcal{E}_4^c$. In addition, $\mathcal{E}_5 \subset \mathcal{E}_3$, so

$$\Pr(\mathcal{E}_5 | S = \bar{s}) \leq \Pr(\mathcal{E}_3 \cap \mathcal{E}_2^c \cap \mathcal{E}_4^c | S = \bar{s}) \tag{62}$$

$$\leq \Pr(\mathcal{E}_1 \cap \mathcal{E}_4^c | S = \bar{s}) + \Pr(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_4^c | S = \bar{s}). \tag{63}$$

The first term in (63) vanishes by the inductive assumption on $\mathsf{S}_{j,v,s}^{(n_j)}(\mathcal{N})$ for all $j \in \mathcal{A}$. To bound the second term, we consider two cases. First, that $p(y^{(\mathcal{A})}|x^{(u)}, x^{(\{u\}^c)}, \bar{s}) \neq p(y^{(\mathcal{A})}|x^{(u)}, x^{(\{u\}^c)}, s)$ for any $x^{(u)} \in \mathcal{X}^{(u)}$ and $y^{(\mathcal{A})} \in \mathcal{Y}^{(\mathcal{A})}$. Then $\Pr(\mathcal{E}_1^c \cap \mathcal{E}_2^c | S = \bar{s}) \to 0$ by Lemma 3. Otherwise, the channel from $x^{(u)}$ to $Y^{(\mathcal{A})}$ conditioned on $X^{(\{u\}^c)} = x^{(\{u\}^c)}$ is identical for $S = \bar{s}$ and $S = s$. Hence the operation of the point-to-point code from $X^{(u)}$ to $Y^{(\mathcal{A})}$ works just as well for $S = \bar{s}$ as for $S = s$, so $\Pr(\mathcal{E}_3 \cap \mathcal{E}_1^c \cap \mathcal{E}_2^c | S = \bar{s}) \to 0$. ∎

The following theorem gives the positive rate result (equivalent to Theorems 6 and 7) for the CC model.

*Theorem 10:* If $(u, v) \in \mathcal{P}_{\mathrm{CC}}$, then there exists a rate vector $\mathcal{R} \in \mathscr{R}_{\mathrm{CC}}(\mathcal{N})$ with $R^{(u\to v)} > 0$. Conversely, if $(u, v) \notin \mathcal{P}_{\mathrm{CC}}$, then for any solution $\mathcal{S}(\mathcal{N})$ there exists $s \in \mathcal{S}$ such that with $S^n = (s, s, \ldots, s)$, $X_{1:n}^{(u)} \to W^{(\{u\}^c\to *)} \to Y_{1:n}^{(v)}$ forms a Markov chain.

*Proof:* To prove the converse, note that if $(u, v) \notin \mathcal{P}_{\mathrm{CC}}$ then $(u, v) \notin \mathcal{P}_s$ for some $s \in \mathcal{S}$. With this fixed state, the proof follows exactly as that of Theorem 7.

Now we prove achievability. Suppose $(u, v) \in \mathcal{P}_{\mathrm{CC}}$. Thus $(u, v) \in \mathcal{P}_s$ for all $s \in \mathcal{S}$. Let $\mathsf{S}_{u,v,s}^{(n)}(\mathcal{N})$ be the sequence of solutions asserted by Lemma 9. Let $R_s^{(u\to v)} > 0$ be the rate for code $\mathsf{S}_{u,v,s}^{(n)}(\mathcal{N})$. Let $\tilde{R}^{(u\to v)} = \min_{s\in\mathcal{S}} R_s^{(u\to v)}$.

We construct a solution to send positive rate from $u$ to $v$ as follows. First node $u$ chooses a message $W^{(u\to v)} \in [2^{n\tilde{R}^{(u\to v)}}]$. Coding proceeds in $|\mathcal{S}|$ sessions. In the session associated with $s \in \mathcal{S}$, we employ $\mathsf{S}_{u,v,s}^{(n)}(\mathcal{N})$ to send $W^{(u\to v)}$ from $u$ to $v$. After all sessions are complete, node $v$ decodes by choosing $\widehat{W}^{(u\to v)}$ to be the output of the first solution that did not declare an error. By Lemma 9, with high probability the solution associated with the true state will not make an error, and any solution associated with a false state will not make an error without declaring an error. Thus the probability of error is small. As the total blocklength for the code is $n|\mathcal{S}|$, the achieved rate is $\tilde{R}^{(u\to v)}/|\mathcal{S}| > 0$. ∎

### C. Positive Rate for the AVC Model

Recall that, as defined in [10], an AVC $p(y|x, s)$ is *symmetrizable* if there exists a probability transition matrix $p(s|x)$ such that

$$\sum_{s\in\mathcal{S}} p(y|x, s)p(s|x') = \sum_{s\in\mathcal{S}} p(y|x', s)p(s|x), \quad \text{for all } x, x' \in \mathcal{X}, y \in \mathcal{Y}. \tag{64}$$

As shown in [10], a point-to-point AVC has positive capacity if and only if it is non-symmetrizable. Now define $\mathcal{P}_{\mathrm{AVC}}$ using the same procedure as above for $\mathcal{P}$, but replace (41) with the condition that there exists $x^{(\{u\}^c)} \in \mathcal{X}^{(\{u\}^c)}$ such that the channel from $X^{(u)}$ to $Y^{(\mathcal{A})}$, conditioned on $X^{(\{u\}^c)} = x^{(\{u\}^c)}$, is non-symmetrizable.

*Theorem 11:* If $(u, v) \in \mathcal{P}_{\mathrm{AVC}}$, then there exists a rate vector $\mathcal{R} \in \mathscr{R}_{\mathrm{AVC}}(\mathcal{N})$ with $R^{(u\to v)} > 0$.

*Proof:* The proof follows from the same argument as for Theorem 6, except that we replace the point-to-point channel code from $X^{(u)}$ to $Y^{(\mathcal{A})}$ with an AVC code. By the assumption that this channel is non-symmetrizable, positive rate can be achieved by the results in [10]. ∎

## VII. Compound Channel Equivalence

In this section and the next we simplify notation by writing $X$ for $X^{(1,1)}$, $Y$ for $Y^{(2,1)}$, and $S$ for $S^{(1)}$. Since we are primarily interested in the independent channel $\mathcal{C}$, there should be no confusion.

There are two relevant capacities for the compound channel: first, the standard capacity expression for a compound channel

$$\underline{C} = \max_{p(x)} \min_{s\in\mathcal{S}} I(X; Y|S = s), \tag{65}$$

and second, the capacity of a compound channel if the state is known at the encoder and the decoder, wherein the min and max are reversed:

$$\bar{C} = \min_{s\in\mathcal{S}} \max_{p(x)} I(X; Y|S = s). \tag{66}$$

In other words, $\bar{C}$ and $\underline{C}$ represent the capacities of the independent channel $\mathcal{C}$ depending on whether compound state knowledge is available at the encoder or not.

Of course, $\underline{C} \leq \bar{C}$. Let $\mathcal{P}_{\mathrm{CC}}$ be defined as above for $\mathcal{N}$. As stated in the following theorem, the compound channel is equivalent to a bit-pipe with rate either $\underline{C}$ or $\bar{C}$, depending on whether the rest of the network can sustain any positive feedback rate from node 2 to node 1.

*Theorem 12:*

$$\mathscr{R}_{\mathrm{CC}}(\mathcal{N}) = \begin{cases} \mathscr{R}_{\mathrm{CC}}(\mathcal{N}^{\bar{C}}) & \text{if } (2,1) \in \mathcal{P}_{\mathrm{CC}} \\ \mathscr{R}_{\mathrm{CC}}(\mathcal{N}^{\underline{C}}) & \text{if } (2,1) \notin \mathcal{P}_{\mathrm{CC}}. \end{cases} \tag{67}$$

We prove this theorem in several lemmas, which in combination with continuity from Lemma 2 prove the theorem.

*Lemma 13:* For all networks with links $(2,1) \notin \mathcal{P}_{\mathrm{CC}}$ if $R < \underline{C}$, then $\mathscr{R}_{\mathrm{CC}}(\mathcal{N}^R) \subseteq \mathscr{R}_{\mathrm{CC}}(\mathcal{N})$.

*Proof:* The proof follows an almost identical argument as that of [13, Lemma 5], which proved that a bit-pipe may simulate a point-to-point noisy channel via a traditional channel code. Recalling that $\underline{C}$ is the usual compound channel capacity, $R < \underline{C}$ implies the existence of a reliable compound channel code at rate $R$. Replacing the channel code in the proof of [13, Lemma 5] with such a compound channel code proves our result. ∎

*Lemma 14:* For all networks with links $(2,1) \in \mathcal{P}_{\mathrm{CC}}$ if $R > \bar{C}$, then $\mathscr{R}_{\mathrm{CC}}(\mathcal{N}) \subseteq \mathscr{R}_{\mathrm{CC}}(\mathcal{N}^R)$.

*Proof:* Let $s^* = \operatorname{argmin}_s \max_{p(x)} I(X; Y)$. We may use [13, Th. 6], which proves that a bit-pipe can simulate a noisy channel with less capacity, to simulate the channel $p(y|x, s^*)$ over the bit-pipe of rate $R$, since $R > I(X; Y)$ for this channel and any input distribution. ∎

*Lemma 15:* For all networks with links $(2,1) \in \mathcal{P}_{\mathrm{CC}}$ if $R < \bar{C}$, then $\mathscr{R}_{\mathrm{CC}}(\mathcal{N}^R) \subseteq \mathscr{R}_{\mathrm{CC}}(\mathcal{N})$.

*Proof:* By Theorem 6, since $(2,1) \in \mathcal{P}_{\mathrm{CC}}$, there exists a solution $\mathsf{S}_0(\mathcal{N})$ such that $R^{(2\to 1)} > 0$. Given a solution $\mathsf{S}(\mathcal{N}^R)$, we construct a solution $\mathsf{S}(\mathcal{N})$ with three sessions. In session 1, node 1 sends a training sequence so that node 2 can learn the state. In session 2, this estimated state is transmitted back to node 1 using $\mathsf{S}_0(\mathcal{N})$. In session 3, node 1

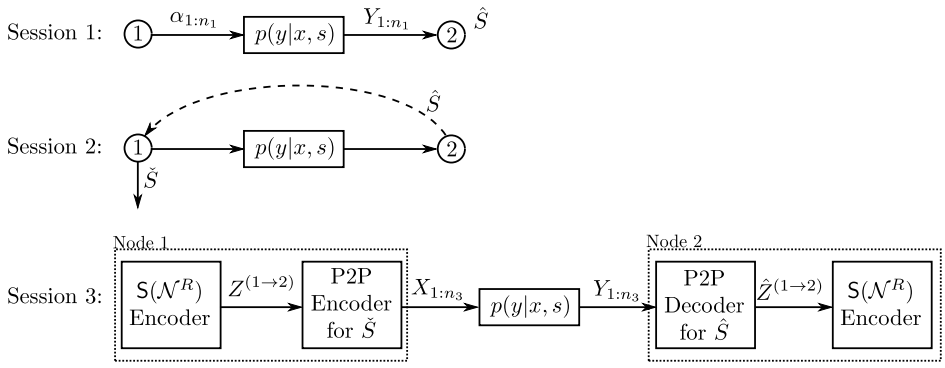


Fig. 4. The structure of the proof of Lemma 15. Training is used in Session 1 to learn the state; in Session 2 the estimated state is sent back to the transmitter; in Session 3 a point-to-point channel code is used based on the estimated state.

uses this estimated state to transmit a message across $\mathcal{C}$ while the rest of $\mathsf{S}(\mathcal{N}^R)$ is conducted. This technique is illustrated in Fig. 4. We give more details as follows.

*Session 1:* We employ a random coding argument wherein we choose a training sequence $\alpha_{1:n_1}$ randomly and uniformly from $\mathcal{X}^{n_1}$. This sequence forms the codebook for session 1, and it is revealed to nodes 1 and 2. Node 1 transmits $\alpha_{1:n_1}$ into $\mathcal{C}$ while the inputs to all other channels are arbitrary. Let $Y_{1:n_1}$ be the output of $\mathcal{C}$. Node 2 forms a state estimate by choosing $\hat{S}$ arbitrarily from the set of $\hat{s} \in \mathcal{S}$ such that $p(Y_{1:n_1}|\alpha_{1:n_1}, \hat{s}) = \max_{s'} p(Y_{1:n_1}|\alpha_{1:n_1}, s')$.

*Session 2:* Employ $\mathsf{S}_0(\mathcal{N})$ with blocklength $n_2$ to transmit $\hat{S}$ from node 2 to node 1. Let $\check{S}$ be the recovered value at node 1. Assume $n_2$ is large enough such that $2^{n_2 R^{(2\to 1)}} \geq |\mathcal{S}|$.

*Session 3:* The network conducts $\mathsf{S}(\mathcal{N}^R)$, but signals to be sent along the bit-pipe $\mathcal{C}^R$ are instead transmitted across the noisy link $\mathcal{C}$ by encoding them at node 1 using an encoder point-to-point channel with state $\check{S}$, while node 2 employs a decoder for the channel with state $\hat{S}$. Let $n_3$ be the blocklength of this session. Denote by $Z^{(1\to 2)} \in [2^{n_3 R}]$ the signal to be sent across bit-pipe $\mathcal{C}^R$, and $\hat{Z}^{(1\to 2)}$ the estimate at node 2.

*Probability of Error Analysis:* Assume the state is $s$. Define the following error events:

$$\mathcal{E}_1 := \{p(y|x,s) \neq p(y|x,\hat{S}) \text{ for any } x, y\}, \tag{68}$$

$$\mathcal{E}_2 := \{\check{S} \neq \hat{S}\}, \tag{69}$$

$$\mathcal{E}_3 := \{\hat{Z}^{(1\to 2)} \neq Z^{(1\to 2)}\}. \tag{70}$$

We may bound the probability of error by

$$\Pr(\mathcal{E}_1) + \Pr(\mathcal{E}_2) + \Pr(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3). \tag{71}$$

By Lemma 3, $\Pr(\mathcal{E}_1) \to 0$ as $n_1 \to \infty$. By Theorem 6, $\Pr(\mathcal{E}_2) \to 0$ as $n_2 \to \infty$. The effective rate of the point-to-point code in Session 3 is $\frac{nR}{n_3}$, where the total blocklength is $n = n_1 + n_2 + n_3$. Since by assumption $R < \bar{C}$, for sufficiently large $n_3/(n_1 + n_2)$ the effective rate is bounded below $\bar{C}$. Moreover, $\bar{C} \leq \max_{p(x)} I(X; Y|S = s)$, so the effective rate is bounded below the capacity of the point-to-point channel with state $s$. As long as $\mathcal{E}_1$ and $\mathcal{E}_2$ do not hold, then $\hat{S} = \check{S}$ are a state for which the operation of the channel is identical to that of $s$, so the channel with this state has the same capacity as with $s$. Hence $\Pr(\mathcal{E}_1^c \cap \mathcal{E}_2^c \cap \mathcal{E}_3) \to 0$ as $n_3 \to \infty$. ■

The following theorem is essentially equivalent to [13, Th. 4], but with a compound channel instead of a standard channel without state.

*Lemma 16:* For all networks with links $(2, 1) \notin \mathcal{P}_{\text{CC}}$ if $R > \underline{C}$, then $\mathscr{R}_{\text{CC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{CC}}(\mathcal{N}^R)$.

*Proof:* By Lemma 1 it suffices to show that $\mathscr{R}_{\text{CC}}(\underline{\mathcal{N}}) \subseteq \mathscr{R}_{\text{CC}}(\underline{\mathcal{N}}^R)$. Fix any $\mathcal{R} \in \text{int}(\mathscr{R}_{\text{CC}}(\mathcal{N}))$ and $\lambda > 0$.

*Choose Code and Define Distributions:* Let $\mathsf{S}(\mathcal{N})$ be a rate-$\mathcal{R}$ solution on network $\mathcal{N}$ for some blocklength $n$. By Theorem 10, for solution $\mathsf{S}(\mathcal{N})$, $X_{1:n}^{(2)} \to W^{(\{2\}^c\to *)} \to Y_{1:n}^{(1)}$ forms a Markov chain. Moreover, the state $S$ only has direct impact on $Y_{1:n}^{(2)}$, which in turn only has direct impact on $X_{1:n}^{(2)}$. Thus $S \to X_{1:n}^{(2)} \to (W^{(\{2\}^c\to *)}, Y_{1:n}^{(1)})$ forms a Markov chain.[4] Combining these two chains yields

$$S \to X_{1:n}^{(2)} \to W^{(\{2\}^c\to *)} \to Y_{1:n}^{(1)}. \tag{72}$$

Since $W^{(\{2\}^c\to *)}$ is drawn uniformly from $\mathcal{W}^{(\{2\}^c\to *)}$ and independently from $S$, the distribution of $(W^{(\{2\}^c\to *)}, Y_{1:n}^{(1)})$ does not depend on $S$. Thus the distribution of $X_{1:n}^{(1)}$ also does not depend on $S$, as it is a function of $(W^{(\{1\}\to *)}, Y_{1:n}^{(1)})$. Therefore, for each time $t$ we may define $p_t(x)$ to be the distribution of $X_t^{(1)}$ independent of $S$. Let $p(x) = \frac{1}{n}\sum_{t=1}^n p_t(x)$ and let

$$s^* = \operatorname*{argmin}_{s\in\mathcal{S}} I(X; Y|S = s), \tag{73}$$

where $X$ is drawn from $p(x)$. Let $p_t(x, y) = p_t(x)p(y|x, s^*)$.

*Typical Set:* Define $\widehat{A}_{\epsilon,t}^{(N)}$ to be the $N$-length typical set according to distribution $p_t(x, y)$ as in [13, Appendix II].

*Design of Channel Emulators:* By concavity of mutual information with respect to the input variable,

$$\frac{1}{n}\sum_{t=1}^n I(X_t; Y_t|S = s^*) \leq I(X; Y|S = s^*) \tag{74}$$

$$= \min_s I(X; Y|S = s) \tag{75}$$

$$\leq \underline{C} < R. \tag{76}$$

Let $R_t := I(X_t; Y_t|S = s^*) + \Delta$ where $\Delta > 0$ is chosen so that $\frac{1}{n}\sum_{t=1}^n R_t = R$.

[4] We have written $S$ as a random variable even though it is arbitrary rather than random. By $S \to A \to B$ we mean that $p(b|a, s) = p(b|a)$.

Randomly design decoder $\beta_{N,t} : [2^{NR_t}] \to \underline{\mathcal{Y}}$ by drawing codewords $\beta_{N,t}(1), \ldots, \beta_{N,t}(2^{NR_t})$ from the i.i.d. distribution with marginal $p_t(\underline{y})$. Define encoder $\alpha_{N,t} : \underline{\mathcal{X}} \to [2^{NR_t}]$ as

$$\alpha_{N,t}(\underline{x}) = \begin{cases} k & \text{if } (\underline{x}, \beta_{N,t}(k)) \in \widehat{A}_{\epsilon,t}^{(N)} \\ 1 & \text{if } \nexists k \text{ s.t. } (\underline{x}, \beta_{N,t}(k)) \in \widehat{A}_{\epsilon,t}^{(N)}. \end{cases} \tag{77}$$

Note that the number of bits required to send $(\alpha_{N,t}(\underline{X}))_{t=1}^n$ is $\sum_{t=1}^n NR_t = nNR$, so we may send all these encoded functions via a bit-pipe of rate $R$.

The rest of the proof follows essentially that of [13, Th. 6]. This involves creating a stacked solution for $\underline{\mathcal{N}}$ with exponentially decreasing probability of error, and then converting it into a solution for $\underline{\mathcal{N}}^R$ by employing the channel emulators at nodes 1 and 2 to simulate the noisy channel over the rate-$R$ bit-pipe. Finally, the error probability can be bounded provided correct parameters are chosen for the typical set $\widehat{A}_{\epsilon,t}^{(N)}$, which can be done for our problem by virtue of the fact that $R_t - I(X_t; Y_t|S = s^*) = \Delta > 0$. ∎

## VIII. Arbitrarily Varying Channel Equivalence

The random coding capacity of a point-to-point AVC is defined as the maximum rate that can be achieved if the encoder and decoder have access to shared randomness (inaccessible to the adversary). It is given by

$$C_r = \max_{p(x)} \min_{p(s)} I(X; Y). \tag{78}$$

Moreover, the max and min may be interchanged without changing the quantity, because of the convexity properties of the mutual information. Without shared randomness, as shown in [10], the capacity of an AVC is 0 if the channel is symmetrizable, and $C_r$ if not. Thus, in all cases, $C_r$ is an upper bound on the capacity. The following theorem provides the corresponding network-level converse.

*Theorem 17:* $\mathscr{R}_{\text{AVC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{AVC}}(\mathcal{N}^{C_r})$.

The proof of this theorem requires a slightly different approach to network equivalence than that of [13]. In particular, we use the following Universal Channel Simulation lemma; a version of this result was stated in [17] and used for an alternative proof of the network equivalence result. The advantage of this result is that it shows that the difference in distribution (as measured by total variational distance) between a DMC and a simulated channel over a noiseless bit-pipe may be arbitrarily small for any input sequence. That is, no assumptions need to be made on the distribution of the input, which is important because in the AVC setting, this input distribution may be influenced by the adversary, and hence unknown. While [17] did not give a complete proof of this lemma, we have provided a proof in Appendix.[5]

*Lemma 18:* Consider a DMC $(\mathcal{X}, q(y|x), \mathcal{Y})$ with capacity $C$. Given a rate $R > C$, a noiseless channel simulation code $(f, g)$ consists of

- $f : \mathcal{X}^n \times [0, 1] \to \{0, 1\}^{nR}$,
- $g : \{0, 1\}^{nR} \times [0, 1] \to \mathcal{Y}^n$.

Let $p(y^n|x^n)$ be the conditional pmf of $Y^n$ given $X^n$ where $Q \sim \text{Unif}[0, 1]$ and

$$Y^n = g(f(X^n, Q), Q).$$

Let $d_{\text{TV}}(p, q)$ be the total variational distance between two distributions $p$ and $q$. There exists a sequence of length-$n$ channel simulation codes where

$$\lim_{n\to\infty} \max_{x^n} d_{\text{TV}}(p(y^n|x^n), q(y^n|x^n)) = 0. \tag{79}$$

*Proof of Theorem 17:* By the continuity property from Lemma 2, it will be enough to show that $\mathscr{R}_{\text{AVC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{AVC}}(\mathcal{N}^R)$ for all $R > C_r$. Let

$$p^\star(s) := \operatorname*{argmin}_{p(s)} \max_{p(x)} I(X; Y). \tag{80}$$

Let $p^\star(y|x) = \sum_s p^\star(s)p(y|x, s)$. Note that $C_r$ is the capacity of the ordinary channel with transition matrix $p^\star(y|x)$. Since the probability of error for the AVC model is maximized over all choices for $S^n$, it cannot increase if we assume $S^n$ is drawn i.i.d. from $p^\star(s)$. Thus, the capacity region can only enlarge if the AVC is replaced by the ordinary channel $p^\star(y|x)$ in $\mathcal{N}$. In particular, if we let $\tilde{\mathcal{N}}$ be the network in which the AVC is replaced by this channel, we have $\mathscr{R}_{\text{AVC}}(\mathcal{N}) \subseteq \mathscr{R}_{\text{AVC}}(\tilde{\mathcal{N}})$. Thus it will be enough to show $\mathscr{R}_{\text{AVC}}(\tilde{\mathcal{N}}) \subseteq \mathscr{R}_{\text{AVC}}(\mathcal{N}^R)$. Moreover, by Lemmas 1 and 4 it will be enough to show $\mathscr{R}_{\text{AVC}}(\underline{\tilde{\mathcal{N}}}) \subseteq \tilde{\mathscr{R}}_{\text{AVC}}(\underline{\mathcal{N}}^R)$, where as in Sec. V $\tilde{\mathscr{R}}$ refers to the capacity region under the shared randomness model from Definition 3. Take any rate vector $\mathcal{R}$ in the interior of $\mathscr{R}_{\text{AVC}}(\underline{\tilde{\mathcal{N}}})$, and let $\mathsf{S}(\underline{\tilde{\mathcal{N}}})$ be a solution with rate vector $\mathcal{R}$ and probability of error at most $\lambda$. We convert this to a randomized solution on $\underline{\mathcal{N}}^R$ as follows. By assumption $R > C_r \geq 0$, so it is certainly possible to transmit data at some positive rate from node 1 to node 2 on network $\underline{\mathcal{N}}^R$; thus, by Definition 3, the shared randomness coding model allows arbitrary shared randomness between nodes 1 and 2.

By Lemma 18, for sufficiently large $N$, there exists a length-$N$ channel simulation code $(f, g)$ with rate $R$ where the induced distribution $p(\underline{y}|\underline{x})$ satisfies

$$\max_{\underline{x}} d_{\text{TV}}\left(p(\underline{y}|\underline{x}), \prod_{\ell=1}^N p^\star(\underline{y}(\ell)|\underline{x}(\ell))\right) \leq \lambda/n. \tag{81}$$

Note that in the network $\underline{\tilde{\mathcal{N}}}$, $\underline{X}_{1:n}$ and $\underline{Y}_{1:n}$ are related by

$$p(\underline{y}_{1:n}|\underline{x}_{1:n}) = \prod_{t=1}^n \prod_{\ell=1}^N p^\star(\underline{y}_t(\ell)|\underline{x}_t(\ell)). \tag{82}$$

We form a randomized code on network $\underline{\mathcal{N}}^R$ by replacing the noisy channel $p^\star(y|x)$ with the channel simulation code used across the layers and repeated $n$ times, once for each time $t \in [n]$. This causes $\underline{X}^n$ and $\underline{Y}^n$ to be related by

$$\prod_{t=1}^n p(\underline{y}_t|\underline{x}_t). \tag{83}$$

While we have eliminated the state for the channel from node 1 to node 2, the state $s^{(0)}$ for the rest of the network remains.

[5] In fact, the result stated in [17] is slightly different: it states that one DMC can be simulated by another; here we only show that a DMC can be simulated by a noiseless bit-pipe. The result of [17] can be recovered from ours by concatenating an ordinary channel code for the DMC to be simulated to the simulation code.

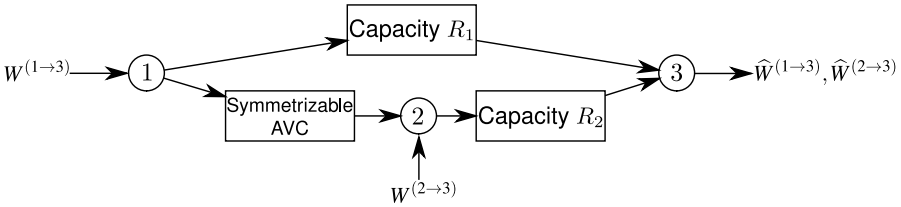


Fig. 5. Example network with a symmetrizable AVC from node 1 to node 2 that does not satisfy the conditions of Corollary 20. The network also contains a rate $R_1$ bit-pipe between nodes 1 and 3 and a rate $R_2$ bit-pipe between nodes 2 and 3. Proposition 21 gives the complete capacity region for this network, which cannot be equated to the capacity region of a network in which the AVC is replaced by any bit-pipe of fixed capacity.

Fix a complete state sequence $\underline{s}_{1:n}^{(0)}$, and consider the distribution of random variables

$$\underline{W}, \widehat{\underline{W}}, \underline{\mathbf{X}}_{1:n}^{(0)}, \underline{\mathbf{Y}}_{1:n}^{(0)}, \underline{X}_{1:n}, \underline{Y}_{1:n}. \tag{84}$$

conditioned on $\underline{s}_{1:n}^{(0)}$. In particular, we wish to bound the total variational distance between the above distribution for the original code on $\underline{\tilde{\mathcal{N}}}$, and that for the randomized code on $\underline{\mathcal{N}}^R$. Let $\mathsf{P}_0$ be the probability law for the distribution of the original code, and for each $t \in [n]$, let $\mathsf{P}_t$ be the probability law in which the original noisy channel distribution is replaced by the induced distribution of the channel simulation code for all times $t' \le t$. Thus $\mathsf{P}_n$ is the probability law for the code on $\underline{\mathcal{N}}^R$, and the difference between $\mathsf{P}_{t-1}$ and $\mathsf{P}_t$ is only the distribution at time $t$. Using the generic fact about total variational distance that

$$d_{\mathrm{TV}}\big(p(a,b)p(c|b)p(d|a,b,c),\ p(a,b)q(c|b)p(d|a,b,c)\big) \le \max_b d_{\mathrm{TV}}(p(c|b), q(c|b)). \tag{85}$$

we have, for any $t \in [n]$,

$$d_{\mathrm{TV}}(\mathsf{P}_{t-1}, \mathsf{P}_t) \le \max_{\underline{x}_t} d_{\mathrm{TV}}\left(p(\underline{y}_t|\underline{x}_t), \prod_{\ell=1}^{N} p^\star(\underline{y}_t(\ell)|\underline{x}_t(\ell))\right) \tag{86}$$

$$\le \lambda/n \tag{87}$$

where we have applied (81). By the triangle inequality,

$$d_{\mathrm{TV}}(\mathsf{P}_0, \mathsf{P}_n) \le \lambda. \tag{88}$$

In particular,

$$d_{\mathrm{TV}}(\mathsf{P}_0(\underline{w}, \underline{\hat{w}}), \mathsf{P}_n(\underline{w}, \underline{\hat{w}})) \le \lambda \tag{89}$$

meaning the probability of error for the randomized code on $\underline{\mathcal{N}}^R$ is at most $\lambda$ more than the original probability of error for the code on $\underline{\tilde{\mathcal{N}}}$. Note that the state sequence $\underline{s}_{1:n}^{(0)}$ affects channel outputs, and thus, via coding operations, may subsequently affect channel inputs. However, because the total variation bound in (81) holds for all input sequences, the effect of the state sequence on the distribution of the channel inputs is irrelevant. Therefore, for any state sequence the resulting randomized code on $\underline{\mathcal{N}}^R$ has probability of error at most $2\lambda$. Since $\lambda$ may be arbitrarily small, this implies $\mathcal{R} \in \tilde{\mathscr{R}}(\underline{\mathcal{N}}^R)$. ∎

[11, Th. 12.11] states that the capacity of a point-to-point AVC is either 0 or $C_r$. This is shown by proving that a small header can be transmitted from encoder to decoder that allows the encoder and decoder to simulate common randomness. This small header can be sent using any code that achieves positive rate. The following is an extension of this result to the network setting wherein the header may originate at any node and be transmitted to both nodes 1 and 2.

*Theorem 19:* If for some node $u$, there exists a rate vector $\mathcal{R}_1 \in \mathscr{R}_{\mathrm{AVC}}(\mathcal{N})$ with $R_1^{(u\to1)} > 0$ and a rate vector $\mathcal{R}_2 \in \mathscr{R}_{\mathrm{AVC}}(\mathcal{N})$ with $R_2^{(u\to2)} > 0$, then $\mathscr{R}_{\mathrm{AVC}}(\mathcal{N}) = \mathscr{R}_{\mathrm{AVC}}(\mathcal{N}^{C_r})$.

*Proof:* In light of Theorem 17, we have only to prove that $\mathscr{R}(\mathcal{N}^{C_r}) \subseteq \mathscr{R}(\mathcal{N})$. Applying Lemmas 1, 2, and 4, it is enough to prove $\mathscr{R}(\underline{\mathcal{N}}^R) \subseteq \tilde{\mathscr{R}}(\underline{\mathcal{N}})$ for all $R < C_r$. By the assumption of the theorem, there exists node $u$ that can transmit data to both nodes 1 and 2; thus by Definition 3, in the shared randomness coding model, $Q_u$ is available at both nodes 1 and 2. By [11, Lemma 12.10], there exists a randomized point-to-point AVC code achieving any rate $R < C_r$ with arbitrarily small probability of error. Given any solution on $\mathscr{R}(\underline{\mathcal{N}}^R)$, we adapt it into a randomized code on $\underline{\mathcal{N}}$ be employing an $N$-length randomized point-to-point channel code across layers, once for each time $t \in [n]$, using the shared randomness $Q_u$. Since the probability of error of the AVC code is vanishing, for sufficiently large $N$ the probability of the overall code is also vanishing. ∎

The following corollary provides a sufficient condition for equivalence for the AVC. It follows immediately from Theorem 11 and Theorem 19.

*Corollary 20:* If there exists a node $u$ such that $(u,1) \in \mathcal{P}_{\mathrm{AVC}}$ and $(u,2) \in \mathcal{P}_{\mathrm{AVC}}$, then $\mathscr{R}_{\mathrm{AVC}}(\mathcal{N}) = \mathscr{R}_{\mathrm{AVC}}(\mathcal{N}^{C_r})$.

## IX. AVC Example Network

This section examines the example network shown in Fig. 5. This network illustrates that when a point-to-point AVC does not satisfy the condition of Corollary 20, it is not necessarily equivalent to a zero-capacity bit-pipe, or indeed any bit-pipe with fixed capacity. The channel from node 1 to node 2 is a symmetrizable AVC given by $p(y|x,s)$, with random code capacity $C_r$. The channel from node 1 to node 3 is a bit-pipe with capacity $R_1$, where we assume $R_1 > 0$, and that from node 2 to node 3 is a bit-pipe with capacity $R_2$. We first determine the capacity region of this network, and then find the capacity region if the AVC were replaced by a bit-pipe of capacity fixed capacity $\tilde{R}$; these two regions do not coincide for any $\tilde{R}$. Roughly, equivalence cannot hold because the symmetrizable AVC leads to a situation in which node 2 can determine that the data sent by node 1 is one of a small number of possibilities. All of these possibilities can be sent along

link $(2,3)$, where node 3 can determine which is the correct one using side information from link $(1,3)$. Thus, as long as $R_2$ is not too large, each bit sent on link $(2,3)$ for message $W^{(1\to 3)}$ contributes only a fraction of a bit of useful data; no such phenomenon can occur with a fixed-capacity bit-pipe, since an additional bit would add either a full bit or zero bits to the overall capacity.

It was shown in [18] that with list decoding—even for quite short lists—the capacity of a symmetrizable AVC is given by its random code capacity. In particular, [18] defines the *symmetrizability* of an AVC $p(y|x,s)$ as the largest integer $M$ for which there exists a stochastic matrix $p(s|x_1,\ldots,x_M)$ such that

$$\sum_{s\in\mathcal{S}} p(y|x,s)p(s|x_1,\ldots,x_M) \tag{90}$$

is symmetric in $x, x_1, \ldots, x_M$. A channel is symmetrizable, in the sense formulated in [10] and discussed above in (64), if and only if $M \geq 1$. It is shown in [18] that for an AVC with symmetrizability $M$, the decoder can reliably list-decode at rate $C_r$ with list size $M+1$. This result will be instrumental in our examination of the example network.

For the network shown in Fig. 5, the only positive achievable rates for this network are $R^{(1\to 3)}$ and $R^{(2\to 3)}$. The following proposition characterizes the capacity region for this network.

*Proposition 21:* The capacity region for the network shown in Fig. 5 is given by the pairs $(R^{(1\to 3)}, R^{(2\to 3)})$ satisfying

$$R^{(2\to 3)} \leq R_2, \tag{91}$$

$$R^{(1\to 3)} \leq R_1 + C_r, \tag{92}$$

$$R^{(2\to 3)} + (M+1)R^{(1\to 3)} \leq (M+1)R_1 + R_2. \tag{93}$$

*Proof (Achievability):* The basic idea of our achievability proof is as follows: node 2 makes use of the list decoding scheme from [18], and then transmits along link $(2,3)$ the entire list of $M+1$ potential messages, in addition to message $W^{2\to 3}$. Along link $(1,3)$, we send part of message $W^{1\to 3}$, in addition to a small hash that allows node 3 to determine which of the $M+1$ messages is the true one. That this is possible with a hash of negligible rate is not quite proved in [18], since in neither scenario is there a list decoding followed by a determination of the true message via side information. Here we use a random linear hash to achieve essentially the same effect as the random choice of channel codes in [11, Lemma 12.8], but in the context of a list code, as we will show in the following.

Fix rates $R^{(1\to 3)}, R^{(2\to 3)}$ satisfying (91)–(93), but with strict inequalities. Fix an integer $q$ and a blocklength $n$. Let $\mathbb{F}_{2^q}$ be the finite field of order $2^q$. We express $W^{(1\to 3)}$ as a vector of elements of $\mathbb{F}_{2^q}$ as follows. Let $\tilde{R}^{(1\to 3)}$ be the largest multiple of $\frac{q}{n}$ no larger than $R^{(1\to 3)}$. Clearly $\tilde{R}^{(1\to 3)} \geq R^{(1\to 3)} - \frac{q}{n}$. Define integers

$$K_1 = \left\lfloor \frac{nR_1}{q} \right\rfloor - 1, \tag{94}$$

$$K_2 = \frac{n\tilde{R}^{(1\to 3)}}{q} - K_1. \tag{95}$$

By the assumption that $R_1 > 0$, for $n$ sufficiently large we have $K_1 \geq 1$. Message $W^{(1\to 3)}$ is chosen from the alphabet $[2^{n\tilde{R}^{(1\to 3)}}]$ and message $W^{(2\to 3)}$ from the alphabet $[2^{nR^{(2\to 3)}}]$, respectively. We may denote $W^{(1\to 3)} = (W_1, \ldots, W_{K_1+K_2})$ where $W_j \in \mathbb{F}_{2^q}$ for all $j \in [K_1+K_2]$, where we for the sake of brevity drop the superscript $(1\to 3)$ for the vector elements. Note that the $W_j$ are independent and each drawn uniformly from $\mathbb{F}_{2^q}$. For convenience, we write $W_{K_1+1}^{K_1+K_2} = (W_{K_1+1}, \ldots, W_{K_1+K_2})$.

At the start of encoding, node 1 generates a hash of the vector $W_{K_1+1}^{K_1+K_2}$. The symbol $W_1$ is used as the random seed for the hash, and the hash itself is given by

$$h = \sum_{j=1}^{K_2} (W_1)^{j-1} W_{K_1+j}, \tag{96}$$

where $(W_1)^{j-1}$ represents exponentiation in the field $\mathbb{F}_{2^q}$. Encoding and decoding proceeds as follows:

1) $(h, W_1, \ldots, W_{K_1})$ is transmitted along link $(1,3)$.
2) $W_{K_1+1}^{K_1+K_2}$ is encoded using an $(M+1)$-list code from [18] and the resulting codeword is transmitted into the AVC $(1,2)$.
3) After receiving the output sequence from the AVC, node 2 decodes the $(M+1)$-length list, denoted $\widehat{W}_{i,K_1+1}^{K_1+K_2} = (\widehat{W}_{i,K_1+1}, \ldots, \widehat{W}_{i,K_1+K_2})$ for $i \in [M+1]$.
4) $(W^{(2\to 3)}, \widehat{W}_{i,K_1+1}^{K_1+K_2} : i \in [M+1])$ is transmitted across link $(2,3)$.
5) Node 3 receives the vectors transmitted on links $(1,3)$ and $(2,3)$ without error. It decodes $W^{(2\to 3)}$ from its received vector on link $(2,3)$. Given $\widehat{W}_{i,K_1+1}^{K_1+K_2}$ for each $i \in [M+1]$ received on link $(2,3)$, node 3 computes

$$\hat{h}_i = \sum_{j=1}^{K_2} (W_1)^{j-1} \widehat{W}_{i,K_1+j}. \tag{97}$$

For the smallest $i$ for which $\hat{h}_i = h$, node 3 declares

$$\widehat{W}^{(2\to 3)} = (W_1, \ldots, W_{K_1}, \widehat{W}_{i,K_1+1}, \ldots, \widehat{W}_{i,K_1+K_2}), \tag{98}$$

where $h$ and $W_1, \ldots, W_{K_1}$ were received on link $(1,3)$.

*Bit-Pipe Capacity Limits:* We first confirm that in the coding procedure described above, the vectors sent along links $(1,3)$ and $(2,3)$ do not exceed the capacities of these bit-pipes. The number of bits sent along link $(1,3)$ is $(K_1+1)q \leq nR_1$, so its capacity constraint is satisfied.

From (95) we obtain

$$K_2 q = n\tilde{R}^{(1\to 3)} - K_1 q, \tag{99}$$

$$\leq n\tilde{R}^{(1\to 3)} - nR_1 + 2q, \tag{100}$$

$$\leq nR^{(1\to 3)} - nR_1 + 2q, \tag{101}$$

where (100) is due to $qK_1 \leq nR_1 - 2$ from (94). Using (101), the number of bits sent along link $(2,3)$ is now given as

$$(M+1)K_2 q + nR^{(2\to 3)} \leq n(M+1)R^{(1\to 3)} - n(M+1)R_1 + 2q + nR^{(2\to 3)}. \tag{102}$$

Since (93) holds with a strict inequality, this quantity is at most $nR_2$ for sufficiently large $n$.

*Probability of Error:* There are two potential sources of error: (i) the decoded list from the AVC at node 2 does not include the true intended message, and (ii) there exists $i \in [M+1]$ such that $\hat{h}_i = h$ even though $\widehat{W}_{i,K_1+1}^{K_1+K_2} \neq W_{K_1+1}^{K_1+K_2}$. For the first source of error, note that the number of bits in $W_{K_1+1}^{K_1+K_2}$ is $K_2 q$, so the rate of the list code on the AVC can be obtained from (101) as

$$\frac{K_2 q}{n} = \tilde{R}^{(1\to 3)} - \frac{K_1 q}{n} \leq R^{(1\to 3)} - R_1 + \frac{2q}{n} \tag{103}$$

Since (92) holds with a strict inequality, the quantity on the l.h.s. in (103) is less than $C_r$ for sufficiently large $n$. Thus, by the results in [18], the probability that the decoded list does not include the true message vanishes with $n$.

Now consider the second source of error. The content of the decoded list depends only on $W_{K_1+1}^{K_1+K_2}$, the state sequence $S^n$, and the random operation of the AVC. In particular, the list is independent of $W_1$. Thus, for any $w_{K_1+1}^{K_1+K_2}, \hat{w}_{K_1+1}^{K_1+K_2} \in \mathbb{F}_{2^q}^{K_2}$

$$\begin{aligned}\Pr\left(\hat{h}_i = h \middle| W_{K_1+1}^{K_1+K_2} = w_{K_1+1}^{K_1+K_2}, \widehat{W}_{K_1+1}^{K_1+K_2} = \hat{w}_{K_1+1}^{K_1+K_2}\right)\\ = \Pr\left(\sum_{j=1}^{K_2} (W_1)^{j-1}(\hat{w}_j - w_j) = 0\right).\end{aligned} \tag{104}$$

If $w_{K_1+1}^{K_1+K_2} \neq \hat{w}_{K_1+1}^{K_1+K_2}$ then the polynomial in $W_1$ inside the probability is a nonzero polynomial of degree at most $K_2 - 1$, so it has at most $K_2 - 1$ roots. Since $W_{K_2+1}$ is chosen uniformly from $\mathbb{F}_{2^q}$, if $w_{K_1+1}^{K_1+K_2} \neq \hat{w}_{K_1+1}^{K_1+K_2}$

$$\begin{aligned}\Pr(\hat{h}_i = h | W_{K_1+1}^{K_1+K_2} = w_{K_1+1}^{K_1+K_2}, \widehat{W}_{K_1+1}^{K_1+K_2} = \hat{w}_{K_1+1}^{K_1+K_2})\\ \leq \frac{K_2 - 1}{2^q}.\end{aligned} \tag{105}$$

Therefore, the probability that $\hat{h}_i = h$ for any $i$ satisfying $\widehat{W}_{i,K_1+1}^{K_1+K_2} \neq W_{K_1+1}^{K_1+K_2}$ is at most

$$\frac{(K_2 - 1)M}{2^q}. \tag{106}$$

This can be made arbitrarily small for sufficiently large $q$.

*Converse*: Let $(R^{(1\to 3)}, R^{(2\to 3)})$ be an achievable rate pair. Thus there exists a sequence of solutions $\mathsf{S}_n(\mathcal{N})$ of length $n$, rates $R^{(1\to 3)}, R^{(2\to 3)}$ and probability of error going to 0 as $n \to \infty$. In this argument, we use the fact that the capacity region does not change if the state $S^n$ is chosen randomly, as long as this random choice is independent of the message (but it may depend on the code). We consider two specific distributions for $S^n$ under $\mathsf{S}_n(\mathcal{N})$ for some $n$. First, that $S^n$ is chosen randomly from the i.i.d. distribution with marginal $p^\star(s)$ defined in (80) as the saddle-point in the random coding capacity. With this choice, the AVC behaves as a (stateless) stationary memoryless channel with transition probability

$$p^\star(y|x) = \sum_s p^\star(s) p(y|x,s). \tag{107}$$

Note that the channel $p^\star(y|x)$ has capacity $C_r$. Simple applications of the cutset bound yield (91) and (92).

To prove (93), we consider a different distribution on the state. Let $p_{X^n}(x^n)$ be the distribution of the input sequence to the AVC $(1,2)$ under solution $\mathsf{S}_n(\mathcal{N})$. Note that this distribution depends only on the code at node 1, so it is independent of the state $S$ of the AVC. The state sequence $S^n$ is drawn from the distribution

$$\sum_{x_1^n,\ldots,x_M^n} p_{X^n}(x_1^n)\cdots p_{X^n}(x_M^n) \prod_{i=1}^n p(s_i|x_{1i},\ldots,x_{Mi}), \tag{108}$$

where the distribution $p(s|x_1,\ldots,x_m)$ is one for which (90) is symmetric. Let $z_1 \in [2^{nR_1}]$ and $z_2 \in [2^{nR_2}]$ with the corresponding random variables $Z_1$ and $Z_2$ denote the input symbols of links $(1,3)$ and $(2,3)$ respectively. Since these links are bit-pipes, these variables also represent the output symbols of the respective links. We also write $X^n$ and $Y^n$ for the input and output sequences of the AVC $(1,2)$. We may now write the distribution of all relevant random variables, conditioned on state sequence $S^n = s^n$, by

$$\begin{aligned}&p(w^{(1\to 3)}, w^{(2\to 3)}, x^n, y^n, z_1, z_2, \hat{w}^{(1\to 3)}, \hat{w}^{(2\to 3)}|s^n)\\ &= \frac{1}{2^{nR^{(1\to 3)}} 2^{nR^{(2\to 3)}}} p(x^n|w^{(1\to 3)})\, p(z_1|w^{(1\to 3)})\\ &\quad\cdot\left[\prod_{i=1}^n p(y_i|x_i,s_i)\right] p(z_2|y^n, w^{(2\to 3)})\\ &\quad\cdot p(\hat{w}^{(1\to 3)}|z_1,z_2)\, p(\hat{w}^{(2\to 3)}|z_1,z_2),\end{aligned} \tag{109}$$

where the encoding and decoding operations are written as conditional distributions because randomized coding is allowed. Let $V(y|x,x_1,\ldots,x_M)$ be the symmetric distribution in (90). The distribution of $X^n, Y^n$ may be written as

$$\begin{aligned}p_{X^n}(x^n) &\sum_{x_1^n,\ldots,x_M^n} p_{X^n}(x_1^n)\cdots p_{X^n}(x_M^n)\\ &\cdot\prod_{i=1}^n V(y_i|x_i,x_{1i},\ldots,x_{Mi}).\end{aligned} \tag{110}$$

Thus, the distribution of $X^n, Y^n$ is unchanged if we let $X_1^n,\ldots,X_M^n$ be random sequences, each distributed according to $p_{X^n}$, and independent from each other, from $X^n$, and from the messages, and where $Y^n$ is drawn from

$$\prod_{i=1}^n V(y_i|x_i,x_{1i},\ldots,x_{Mi}). \tag{111}$$

This induces a probability law on all variables other than $S^n$ given by

$$\begin{aligned}&p(w^{(1\to 3)}, w^{(2\to 3)}, x^n, x_1^n,\ldots, x_M^n, y^n, z_1, z_2, \hat{w}^{(1\to 3)}, \hat{w}^{(2\to 3)})\\ &= \frac{1}{2^{nR^{(1\to 3)}} 2^{nR^{(2\to 3)}}} p(x^n|w^{(1\to 3)})\, p(z_1|w^{(1\to 3)})\\ &\quad\cdot p_{X^n}(x_1^n)\cdots p_{X^n}(x_M^n)\left[\prod_{i=1}^n V(y_i|x_i,x_{1i},\ldots,x_{Mi})\right]\\ &\quad\cdot p(z_2|y^n, w^{(2\to 3)})\, p(\hat{w}^{(1\to 3)}|z_1,z_2)\, p(\hat{w}^{(2\to 3)}|z_1,z_2).\end{aligned} \tag{112}$$

Note in particular that $X^n, X_1^n, \ldots, X_M^n, Y^n$ are distributed according to

$$p_{X^n}(x^n)p_{X^n}(x_1^n)\cdots p_{X^n}(x_M^n)\prod_{i=1}^{n} V(y_i|x_i, x_{1i}, \ldots, x_{Mi}). \tag{113}$$

By Fano's inequality and the data processing inequality,

$$nR^{(2\to 3)} \le I(W^{(2\to 3)}; Z_2) + n\epsilon_n \tag{114}$$

where $\epsilon_n \to 0$ as $n \to \infty$. Applying Fano's inequality again, we have

$$\begin{aligned} nR^{(1\to 3)} &= H(W^{(1\to 3)}) && (115)\\ &\le I(W^{(1\to 3)}; Z_1, Z_2) + n\epsilon_n && (116)\\ &= I(W^{(1\to 3)}; Z_2) + I(W^{(1\to 3)}; Z_1|Z_2) + n\epsilon_n && (117)\\ &\le I(W^{(1\to 3)}; Z_2) + nR_1 + n\epsilon_n && (118)\\ &\le I(X^n; Z_2) + nR_1 + n\epsilon_n && (119)\end{aligned}$$

where in (119) we have used the fact that $W^{(1\to 3)} \to X^n \to Z_2$ is a Markov chain. By symmetry of $(X^n, X_1^n, \ldots, X_M^n)$, we have $I(X_k^n; Z_2) = I(X^n; Z_2)$ for all $k \in [M]$. Thus, defining $\epsilon_n' = (M+2)\epsilon_n$ and $X_0^n = X^n$,

$$\begin{aligned} &nR^{(2\to 3)} + (M+1)nR^{(1\to 3)} && (120)\\ &\le I(W^{(2\to 3)}; Z_2) + \sum_{k=0}^{M} I(X_k^n; Z_2) + (M+1)nR_1 + n\epsilon_n' && (121)\\ &\le I(W^{(2\to 3)}; Z_2) + \sum_{k=0}^{M} I(X_k^n; Z_2|W^{(2\to 3)}, X_0^n, \ldots, X_{k-1}^n) \\ &\quad + (M+1)nR_1 + n\epsilon_n' && (122)\\ &= I(W^{(2\to 3)}, X_0^n, \ldots, X_M^n; Z_2) + (M+1)nR_1 + n\epsilon_n' && (123)\\ &\le nR_2 + (M+1)nR_1 + n\epsilon_n' && (124)\end{aligned}$$

where (122) follows because $(W^{(2\to 3)}, X_0^n, \ldots, X_M^n)$ are mutually independent. Dividing by $n$ and taking the limit as $n \to \infty$ yields (93). ∎

Suppose that in the example network the AVC were replaced by a bit-pipe of capacity $\tilde{R}$. It is easy to see that the resulting set of achievable $(R^{(1\to 3)}, R^{(2\to 3)})$ pairs is given by

$$\begin{aligned} R^{(2\to 3)} &\le R_2, && (125)\\ R^{(1\to 3)} &\le R_1 + \tilde{R}, && (126)\\ R^{(1\to 3)} + R^{(2\to 3)} &\le R_1 + R_2. && (127)\end{aligned}$$

This region does not correspond to (91)–(92) for any value of $\tilde{R}$, as long as $M \ge 1$ (i.e., the AVC is symmetrizable). Therefore, the AVC in Fig. 5 is not equivalent to any fixed capacity bit-pipe.

## X. Relation to the "Edge Removal" Problem

Consider two networks $\mathcal{N}$ and $\mathcal{N}'$ with identical topologies except for a single edge, which has capacity $C_e$ in network $\mathcal{N}$, but capacity $C_e' = C_e - \delta$ in network $\mathcal{N}'$. Herein, $\delta > 0$ is a small constant. Particular attention has been devoted recently to the so called *edge removal* problem which describes the special case of this scenario for $C_e = \delta$. It has been shown in [19] and [20] that for a variety of demand types for which the network coding capacity can be described by the cut-set bound, the capacity of every cut is reduced by at most $\delta$ for each dimension. This means that if a rate vector $\mathcal{R}$ is achievable in network $\mathcal{N}$, a rate vector $\mathcal{R} - \delta\mathcal{I}$ is achievable in $\mathcal{N}'$, where $\mathcal{I}$ denotes the unit rate vector. Examples include single and multisource multicast and single source cases with non-overlapping demands, but also scenarios for which the cut-set bound is not tight, for example a specific class of multiple unicast networks [20]. Further, in [21] the edge removal problem has also been connected to the problem whether a network coding instance allows a reconstruction with $\epsilon$ and zero error, respectively. However, so far only various special cases have been considered, and it is not clear how to formulate the edge removal problem for general demands and topologies.

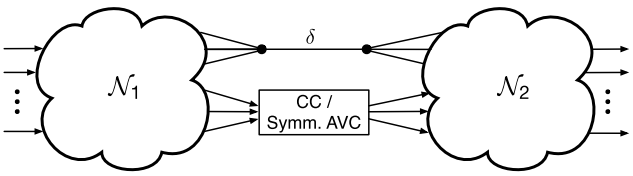


Fig. 6. The network $\mathcal{N}$ consists of two arbitrary networks $\mathcal{N}_1$ and $\mathcal{N}_2$ connected by an edge with capacity $\delta > 0$ and a CC or alternatively, a symmetrizable AVC.

In the following, based on the discussion in Sections VII and VIII, we extend the edge removal problem to networks with state. We formulate our result for both the CC and the AVC case in the following theorem.

*Theorem 22:* Given a network $\mathcal{N}$ with state according to (1) and assume that a non-zero rate vector $\mathcal{R}(\mathcal{N})$ is achievable. Further, assume that there exists a single edge with capacity $\delta$ in the network. Let the network $\mathcal{N}'$ be defined as the network $\mathcal{N}$ with the $\delta$-capacitated edge removed. Then, there exists a network $\mathcal{N}$ such that for the corresponding edge-removed network $\mathcal{N}'$, $\mathcal{R}(\mathcal{N}') < \mathcal{R}(\mathcal{N}) - \delta\mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix.

*Proof:* We show this by considering the example in Fig. 6, where two networks $\mathcal{N}_1$ and $\mathcal{N}_2$ are connected via a CC or a symmetrizable point-to-point AVC, resp., and an edge of capacity $\delta$. Suppose that this connection also represents the min-cut of the overall network $\mathcal{N}$. For the AVC case, as the capacity of the symmetrizable AVC is either 0 or $C_r$, removing the $\delta$-capacitated edge leads to a network capacity of $\mathcal{R}_{\text{AVC}}(\mathcal{N}') = 0$ according to Theorem 19. For the CC case the capacity of the CC is either $\underline{C}$ or $\bar{C}$ (see (65) and (66)). By removing the $\delta$-capacitated feedback edge the network capacity is reduced from $\mathcal{R}_{\text{CC}}(\mathcal{N}) = \bar{C}$ to $\mathcal{R}_{\text{CC}}(\mathcal{N}') = \underline{C}$, where $\bar{C} - \underline{C}$ can be larger than $\delta$. ∎

## XI. Conclusion

We have considered reliable communication over noisy network in the presence of active adversaries. This is modeled by

a subset of independent point-to-point channels consisting of AVCs or CCs. For these cases we have identified scenarios for which the capacity of the corresponding noisy state-dependent network equals the capacity of another state-less network in which the AVCs or CCs are replaced by noiseless bit-pipes. Our results indicate that, in the network setting, the equivalent capacity of these channels is not necessarily equal to their capacity in an isolated point-to-point scenario. For example, the point-to-point AVC represents a pessimistic model for the action of an adversary, leading to zero capacity in some cases. We have shown that in a network setting such a pessimistic model becomes much more optimistic and leads to a positive rate if additional network connectivity exits between the head and the tail node of the AVC or CC under consideration. As most modern communication is performed in an underlying networking framework, this suggests that existing results may be insufficient for characterizing networks in the presence of active adversaries.

## Appendix
## Proof of Lemma 18

We make use of the method of types, adopting notation from [22]. Specifically, given a sequence $x^n$, define its type as

$$P_{x^n}(x) = \frac{|\{i : x_i = x\}|}{n}. \tag{128}$$

Similarly define the joint type of a pair of sequences $(x^n, y^n)$ as $P_{x^n,y^n}$. Given a type $P_X$, define the type class $T(P_X)$ as the set of sequences $x^n$ with $P_{x^n} = P_X$.

Fix $\epsilon > 0$, and define $\tilde{\mathcal{P}}_n$ to be the set of $n$-length types $P_{XY}$ such that

$$|P_{XY}(x, y) - P_X(x)q(y|x)| \leq \epsilon P_X(x)q(y|x) \quad \text{for all } x \in \mathcal{X}, \quad y \in \mathcal{Y}, \tag{129}$$

where $q(y|x)$ is the channel to be simulated. Observe that $P_{x^n,y^n} \in \tilde{\mathcal{P}}_n$ if and only if $(x^n, y^n)$ is robustly typical [23] with respect to the distribution $P_{x^n}(x)q(y|x)$. By the Conditional Typicality Lemma from [24, Ch. 2], since $x^n$ is trivially robustly typical with respect to $P_{x^n}$ (indeed, with parameter $\epsilon = 0$), if $Y^n \sim \prod_{i=1}^n q(y_i|x_i)$, then with probability approaching 1, $P_{x^n,Y^n} \in \tilde{\mathcal{P}}_n$.

Let $I(P_X, P_{Y|X})$ be the mutual information between $X$ and $Y$ where $(X, Y) \sim P_X P_{Y|X}$. By continuity of mutual information, for any $\gamma > 0$, there exists $\epsilon$ small enough so that for all $P_{XY} \in \tilde{\mathcal{P}}_n$,

$$I(P_X, P_{Y|X}) \leq I(P_X, q(y|x)) + \gamma \leq C + \gamma. \tag{130}$$

In particular, if we choose $\gamma = (R-C)/2$, then for sufficiently small $\epsilon$,

$$I(P_X, P_{Y|X}) \leq R - \gamma. \tag{131}$$

We construct a noiseless channel simulation code out of a number of codebooks, one for each type $P_{XY} \in \tilde{\mathcal{P}}_n$. A codebook of joint type $P_{XY}$, denoted $\mathcal{C}(P_{XY})$, is a subset of $T(P_Y)$. We say a codebook with joint type $P_{XY}$ is *feasible* if, for all $x^n \in T(P_X)$, there exists a sequence $y^n \in \mathcal{C}(P_{XY})$ where $P_{x^n,y^n} = P_{XY}$. Define

$$M = 2^{n(R-\delta)} \tag{132}$$

where $0 < \delta < \gamma$. We claim that for sufficiently large $n$, for all $P_{XY} \in \tilde{\mathcal{P}}_n$ there exists a feasible codebook of size at most $M$. To prove this, consider a random choice of codebook $\mathcal{C}(P_{XY})$ consisting of $M$ sequences chosen uniformly and independently from $T(P_Y)$. Note that the codebook will contain fewer than $M$ unique sequences if the same sequence is chosen more than once. We show that with positive probability this codebook is feasible. For each $x^n \in T(P_X)$ define the event

$$\mathcal{E}(x^n) := \{P_{x^n,y^n} \neq P_{XY} \text{ for all } y^n \in \mathcal{C}(P_{XY})\}. \tag{133}$$

Note that the only random variable in this event is the codebook itself. Define the conditional type class

$$T_{P_{XY}}(x^n) := \{y^n : P_{x^n,y^n} = P_{XY}\}. \tag{134}$$

Note that

$$\mathcal{E}(x^n) = \{T_{P_{XY}}(x^n) \cap \mathcal{C}(P_{XY}) = \emptyset\}. \tag{135}$$

By using standard bounds on the size of type classes, for any $x^n \in T(P_X)$

$$\frac{|T_{P_{XY}}(x^n)|}{|T(P_Y)|} \geq \frac{1}{(n+1)^{|\mathcal{Y}|-1}} 2^{-nI(P_X,P_{Y|X})}. \tag{136}$$

For any $x^n \in T(P_X)$, we may bound the probability of event $\mathcal{E}(x^n)$ by

$$\mathbb{P}(\mathcal{E}(x^n)) = \left(1 - \frac{|T_{P_{XY}}(x^n)|}{|T(P_Y)|}\right)^M \tag{137}$$

$$\leq \left(1 - \frac{1}{(n+1)^{|\mathcal{Y}|-1}} 2^{-nI(P_X,P_{Y|X})}\right)^M \tag{138}$$

$$\leq \exp\left\{-\frac{1}{(n+1)^{|\mathcal{Y}|-1}} M 2^{-nI(P_X,P_{Y|X})}\right\}. \tag{139}$$

Thus, by the union bound

$$\mathbb{P}\left(\bigcup_{x^n \in T(P_X)} \mathcal{E}(x^n)\right) \leq |\mathcal{X}|^n \exp\left\{-\frac{1}{(n+1)^{|\mathcal{Y}|-1}} M 2^{-nI(P_X,P_{Y|X})}\right\} \tag{140}$$

$$= |\mathcal{X}|^n \exp\left\{-\frac{1}{(n+1)^{|\mathcal{Y}|-1}} 2^{n(R-I(P_X,P_{Y|X})-\delta)}\right\} \tag{141}$$

$$\leq |\mathcal{X}|^n \exp\left\{-\frac{1}{(n+1)^{|\mathcal{Y}|-1}} 2^{n(\gamma-\delta)}\right\}. \tag{142}$$

This quantity is vanishing in $n$ since $\delta < \gamma$, so for sufficiently large $n$ there exists at least one feasible codebook $\mathcal{C}(P_{XY})$ of size at most $M$.

We now describe a channel simulation code. Assume $n$ is large enough such there exists at least one feasible codebook for each $P_{XY} \in \tilde{\mathcal{P}}_n$.

*Encoder:* Given input sequence $x^n$, randomly choose a sequence

$$\tilde{Y}^n \sim \prod_{i=1}^n q(y_i|x_i). \tag{143}$$

Let $P_{XY} = P_{x^n,\tilde{Y}^n}$. If $P_{XY} \in \tilde{\mathcal{P}}_n$, randomly choose a codebook $\mathcal{C}(P_{XY})$ uniformly from among all feasible codebooks of size at most $M$ for this type. If $P_{XY} \notin \tilde{\mathcal{P}}_n$, declare an error. Of the sequences $y^n \in \mathcal{C}(P_{XY}) \cap T_{P_{XY}}(x^n)$ (there must be at least one, since the codebook is feasible), choose one uniformly at random, which we denote $Y^n$. The encoder outputs two bit-strings:

1) A string of length $\lceil \log |\tilde{\mathcal{P}}_n| \rceil$ denoting the type $P_{XY}$.
2) A string of length $\log M$ denoting the index of $Y^n$ in $\mathcal{C}(P_{XY})$.

Note that for sufficiently large $n$, the total number of bits is at most $nR$, since $|\tilde{\mathcal{P}}_n| \leq 2^{n\gamma}$ for sufficiently large $n$ and $\gamma > 0$.

*Decoder:* Upon learning $P_{XY}$, the decoder can determine the chosen feasible codebook $\mathcal{C}(P_{XY})$, since it has access to the same randomness as the encoder, and thus it can recover $Y^n$.

To bound the variational distance, we first note that, for a given joint type $P_{XY}$, if there is at least one feasible codebook of size at most $M$, then each sequence $y^n \in T(P_Y)$ appears in exactly the same number of such codebooks. Indeed, consider two sequences $y_1^n, y_2^n \in T(P_Y)$. There exists a permutation that takes $y_1^n$ to $y_2^n$. Applying this permutation to the codebook preserves feasibility, because both the input type class $T(P_X)$ and the output type class $T(P_Y)$ are unchanged by permutation. Thus, the permutation constitutes a bijection between feasible codebooks containing $y_1^n$ and feasible codebooks containing $y_2^n$. This implies that they are equal in number. Thus, if $P_{XY} = P_{x^n,\tilde{Y}^n} \in \tilde{\mathcal{P}}_n$, then the randomly chosen codebook $\mathcal{C}(P_{XY})$ is equally likely to contain any sequence $y^n \in T_{P_{XY}}(x^n)$, and hence $Y^n$ is uniformly distributed among $T_{P_{XY}}(x^n)$. Hence, for any pair of sequences $x^n, y^n$ where $P_{x^n,y^n} \in \tilde{\mathcal{P}}_n$, the induced distribution from the simulation code is given by

$$p(y^n|x^n) = \mathbb{P}(P_{x^n,\tilde{Y}^n} = P_{x^n,y^n}) \frac{1}{|T_{P_{x^n,y^n}}(x^n)|}. \quad (144)$$

Now, for the discrete memoryless channel $q(y|x)$, the probability $q(y^n|x^n)$ depends only on the joint type of $(x^n, y^n)$. Thus, conditioning on a particular joint type, the output sequence is uniformly distributed among the conditional type class. In other words, the right-hand side of (144) is precisely equal to $q(y^n|x^n)$ for all $x^n, y^n$. Since (144) only holds if $P_{x^n,y^n} \in \tilde{\mathcal{P}}_n$, the total variational distance between $p(y^n|x^n)$ and $q(y^n|x^n)$ is at most the probability that $P_{x^n,\tilde{Y}^n} \notin \tilde{\mathcal{P}}_n$, which, as argued above, vanishes as $n \to \infty$.

## REFERENCES


[1] S. Jaggi *et al.*, "Resilient network coding in the presence of Byzantine adversaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2596–2603, Jun. 2008.

[2] S. Kim, T. Ho, M. Effros, and S. Avestimehr, "Network error correction with unequal link capacities," in *Proc. 47th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2009, pp. 1387–1394.

[3] O. Kosut, L. Tong, and D. Tse, "Nonlinear network coding is necessary to combat general Byzantine attacks," in *Proc. 47th Annu. Allerton Conf. Commun. Control, Comput.*, Monticello, IL, USA, Oct. 2009, pp. 593–599.

[4] O. Kosut, L. Tong, and D. Tse, "Polytope codes against adversaries in networks," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, pp. 2423–2427.

[5] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.

[6] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Statist.*, vol. 30, no. 4, pp. 1229–1241, Dec. 1959.

[7] J. Wolfowitz, *Coding Theorems of Information Theory*. Berlin, Germany: Springer-Verlag, 1978.

[8] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacities of certain channel classes under random coding," *Ann. Math. Statist.*, vol. 31, no. 3, pp. 558–567, 1960.

[9] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Probab. Theory Rel. Fields*, vol. 44, no. 2, pp. 159–175, 1978.

[10] I. Csiszár and P. Narayan, "The capacity of the arbitrarily varying channel revisited: Positivity, constraints," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 181–193, Mar. 1988.

[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011

[12] M. Bakshi, M. Effros, and T. Ho, "On equivalence for networks of noisy channels under Byzantine attacks," in *Proc. IEEE Int. Symp. Inf. Theory*, Saint Petersburg, Russia, Jul. 2011, pp. 973–977.

[13] R. Koetter, M. Effros, and M. Médard, "A theory of network equivalence—Part I: Point-to-point channels," *IEEE Trans. Inf. Theory*, vol. 57, no. 2, pp. 972–995, Feb. 2011.

[14] R. Duan. "Super-activation of zero-error capacity of noisy quantum channels." [Online]. Available: https://arxiv.org/abs/0906.2527

[15] H. Boche, R. F. Schaefer, and H. V. Poor, "On the continuity of the secrecy capacity of compound and arbitrarily varying wiretap channels," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2531–2546, Dec. 2015.

[16] J. Nötzel, M. Wiese, and H. Boche, "The arbitrarily varying wiretap channel—Secret randomness, stability, and super-activation," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3504–3531, Jun. 2016.

[17] Y. Xiang and Y-H. Kim, "A few meta-theorems in network information theory," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2014, pp. 77–81.

[18] B. L. Hughes, "The smallest list for the arbitrarily varying channel," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 803–815, May 1997.

[19] T. Ho, M. Effros, and S. Jalali, "On equivalence between network topologies," in *Proc. 48th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2010, pp. 391–398.

[20] S. Jalali, M. Effros, and T. Ho, "On the impact of a single edge on the network coding capacity," in *Proc. Inf. Theory Appl. Workshop*, San Diego, CA, USA, Jan. 2011, pp. 1–5.

[21] M. Langberg and M. Effros, "Network coding: Is zero error always possible?" in *Proc. 49th Annu. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Oct. 2011, pp. 1478–1485.

[22] T. M. Cover and J. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1991.

[23] A. Orlitsky and J. R. Roche, "Coding for computing," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 903–917, Mar. 2001.

[24] A. El Gamal and Y. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.



**Oliver Kosut** (S'06–M'10) received B.S. degrees in electrical engineering and mathematics from the Massachusetts Institute of Technology, Cambridge, MA in 2004 and the Ph.D. degree in electrical and computer engineering from Cornell University, Ithaca, NY in 2010.

Since 2012, he has been an Assistant Professor in the school of electrical, computer and energy engineering at Arizona State University, Tempe, AZ. Previously, he was a Postdoctoral Research Associate in the Laboratory for Information and Decision Systems at MIT from 2010 to 2012. His research interests include information theory, cyber-security, and power systems. Prof. Kosut received the NSF Faculty Early Career Development (CAREER) Award in 2015.

**Jörg Kliewer** (S'97–M'99–SM'04) received the Dipl.-Ing. (M.Sc.) degree in electrical engineering from Hamburg University of Technology, Hamburg, Germany, in 1993 and the Dr.-Ing. degree (Ph.D.) in electrical engineering from the University of Kiel, Germany, in 1999, respectively.

From 1993 to 1998, he was a research assistant at the University of Kiel, and from 1999 to 2004, he was a senior researcher and lecturer with the same institution. In 2004, he visited the University of Southampton, U.K., for one year, and from 2005 until 2007, he was with the University of Notre Dame, IN, as a visiting assistant professor. From 2007 until 2013 he was with New Mexico State University, Las Cruces, NM, most recently as an associate professor. He is now with the New Jersey Institute of Technology, Newark, NJ, as an associate professor. His research interests span information and coding theory, graphical models, and statistical algorithms, which includes applications to networked communication and security, data storage, and biology.

Dr. Kliewer was the recipient of a Leverhulme Trust Award and a German Research Foundation Fellowship Award in 2003 and 2004, respectively. He was an associate editor of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 until 2014, and since 2015 serves as an area editor for the same journal. He is also member of the editorial board of the IEEE Information Theory Society Newsletter since 2012.