A Proof of the Ahlswede-Cai-Zhang Conjecture

Christoph Bunte
ETH Zurich
Switzerland
Email: bunte@isi.ee.ethz.ch

Amos Lapidoth
ETH Zurich
Switzerland
Email: lapidoth@isi.ee.ethz.ch

Alex Samorodnitsky
The Hebrew University of Jerusalem
Israel
Email: salex@cs.huji.ac.il

***Abstract*—Ahlswede, Cai, and Zhang proved that, in the noise-free limit, the zero-undetected-error capacity is lower-bounded by the Sperner capacity of the channel graph, and they conjectured equality. Here we derive an upper bound that proves the conjecture.**

I. Introduction

A *zero-undetected-error (z.u.e.) decoder* declares that a message was transmitted only if it is the only message that could have produced the observed output. If the output could have been produced by two or more messages, it declares an erasure. Such a decoder thus never errs: it either produces the correct message or an erasure.

The *z.u.e. capacity* $C_{\text{0-u}}$ of a channel is the supremum of all rates that are achievable with a z.u.e. decoder in the sense that the probability of erasure tends to zero as the blocklength tends to infinity [1], [2]. (It does not matter whether we define $C_{\text{0-u}}$ using an average or a maximal erasure probability criterion.) Clearly, $C_{\text{0-u}}$ never exceeds the Shannon capacity C.

Determining the z.u.e. capacity of general discrete memoryless channels (DMCs) is an open problem. The focus of this paper is the z.u.e. capacity of nearly noise-free channels. More precisely, we focus on *ε-noise channels*, that is, DMCs whose input alphabet $\mathcal{X}$ is a subset of their output alphabet $\mathcal{Y}$ and whose transition law W satisfies

$$W(x|x) \geq 1-\varepsilon \quad \text{for all } x \in \mathcal{X}. \tag{1}$$

Here and throughout we assume that $0 \leq \varepsilon < 1$. For ε-noise channels we derive an upper bound on $C_{\text{0-u}}$. We then apply this result to study the limit of $C_{\text{0-u}}$ as ε tends to zero. Ahlswede, Cai, and Zhang proved that this limit is lower-bounded by the *Sperner capacity* of a certain related graph, and they conjectured equality [2]. Our upper bound proves this conjecture.

The Sperner capacity is defined using graph-theoretic language in Section III. Here we give an alternative characterization in terms of codes (see also [2]). For this we need some standard notation.

A DMC is specified by its transition law $W(y|x)$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite input and output alphabets. Feeding a sequence of input symbols $\mathbf{x} = (x^{(1)}, \ldots, x^{(n)})$ to a DMC of transition law W produces a random sequence of output symbols $\mathbf{Y} = (Y^{(1)}, \ldots, Y^{(n)})$ with distribution

$$W^n(\mathbf{y}|\mathbf{x}) \triangleq \prod_{1 \leq j \leq n} W(y^{(j)}|x^{(j)}), \quad \mathbf{y} \in \mathcal{Y}^n. \tag{2}$$

The support of W is the set of all pairs $(x, y) \in \mathcal{X} \times \mathcal{Y}$ for which $W(y|x)$ is positive; it is denoted by $\mathfrak{S}(W)$. Similarly, if P is a PMF on $\mathcal{X}$, then $\mathfrak{S}(P)$ denotes the set of all $x \in \mathcal{X}$ for which $P(x)$ is positive. We write PW for the PMF on $\mathcal{Y}$ induced by P and the channel W, i.e., $(PW)(y) = \sum_{x \in \mathcal{X}} P(x)W(y|x)$. If $\mathcal{A} \subseteq \mathcal{X}$, then $P(\mathcal{A}) = \sum_{x \in \mathcal{A}} P(x)$. The cardinality of a set $\mathcal{A}$ is denoted by $|\mathcal{A}|$. All logarithms are natural logarithms, and we adopt the convention $0 \log \frac{1}{0} = 0$.

We define a blocklength-n *Sperner code* for a DMC W with $\mathcal{X} \subseteq \mathcal{Y}$ and $W(x|x) > 0$ for all $x \in \mathcal{X}$ as a collection of length-n codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ with the property

$$W^n(\mathbf{x}_m|\mathbf{x}_{m'}) = 0 \quad \text{whenever } m \neq m'. \tag{3}$$

The rate of the code is $n^{-1} \log M$. The largest rate of a Sperner code is a function of the channel law W and the blocklength n. In fact, it depends on W only via its support $\mathfrak{S}(W)$. The supremum over n of the largest rate of blocklength-n Sperner codes is the Sperner capacity C_{Sp} of the channel.

With this notation, we can now state our main result.

Theorem I.1. *For every ε-noise channel,*

$$C_{\text{0-u}} \leq \log\big(e^{C_{\text{Sp}}} + \varepsilon|\mathcal{X}|(|\mathcal{Y}|-1)\big). \tag{4}$$

Combining Theorem I.1 with [2, Theorem 2] proves the following corollary, which was conjectured in [2].

Corollary I.2. *For ε-noise channels,*

$$\lim_{\varepsilon \to 0} C_{\text{0-u}} = C_{\text{Sp}}, \tag{5}$$

where the limit is to be understood in a uniform sense with respect to all ε-noise channels with given $\mathfrak{S}(W)$.

To put this result into perspective, a review of the literature on the zero-undetected-error capacity is provided in the ArXiv-version of this paper [3]. Here we only mention our earlier work [4], where (5) is proved for the "cyclic triangle channel".

A proof of Theorem I.1 is given in Section IV. Before providing an outline of this proof, we try to explain why Corollary I.2 is plausible. If we use a Sperner code in conjunction with a z.u.e. decoder, then an erasure can occur only if the codeword is corrupted, which happens with probability at most $1-(1-\varepsilon)^n$. This suggests that C_{Sp} should be a

The work of A. Samorodnitsky was partially supported by grants from BSF and ISF.

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

lower bound to $C_{0\text{-u}}$ when ε is very small (ignoring the issue that n tends to infinity before ε tends to zero). Conversely, any code whose maximal probability of erasure under z.u.e. decoding is smaller than $(1-\varepsilon)^n$ must be a Sperner code. Since for all rates strictly smaller than $C_{0\text{-u}}$ the probability of erasure can be driven to zero exponentially fast [3], this suggests that C_{Sp} should be an upper bound on $C_{0\text{-u}}$ for small ε (ignoring the issue that the exponent of the erasure probability may become arbitrarily small as ε becomes small and the rate approaches $C_{0\text{-u}}$).

As to the outline of the proof of Theorem I.1, we first show that a multi-letter version of Forney's lower bound on $C_{0\text{-u}}$ is asymptotically tight, even when the input distributions are restricted to be uniform over their support (Section II). We then upper-bound the multi-letter expression using Jensen's inequality followed by algebraic manipulations that yield a still looser bound. Thanks to the input distribution being uniform, this looser bound depends only on ε and the support of W. The final step is to use graph-theoretic techniques, which are introduced in Section III, to obtain the desired upper bound. These techniques include upper-bounding a sum that depends only on the in-degrees of the vertices of a graph G by the maximum size of any induced acyclic subgraph of G. They also include showing that the Sperner capacity of a graph G can be expressed as the limit as n tends to infinity of $1/n$ times the logarithm of the maximum cardinality of any induced acyclic subgraph of the n-fold strong product of G with itself.

II. A Multi-Letter Formula for $C_{0\text{-u}}$

In [5] Forney derived the lower bound

$$C_{0\text{-u}} \geq \max_{P} \sum_{y\in\mathcal{Y}} (PW)(y) \log \frac{1}{P(\mathcal{X}(y))}, \tag{6}$$

where the maximum is over all PMFs on the input alphabet $\mathcal{X}$, and where $\mathcal{X}(y)$ denotes the set of all $x \in \mathcal{X}$ for which $W(y|x)$ is positive.

Since any code for the product channel W^n is also a code for the channel W of n times the blocklength and $1/n$ times the rate, it follows that Forney's bound can be improved by applying it to W^n and normalizing the result by $1/n$. This yields for every n the bound

$$C_{0\text{-u}} \geq n^{-1} \max_{P} \sum_{\mathbf{y}\in\mathcal{Y}^n} (PW^n)(\mathbf{y}) \log \frac{1}{P(\mathcal{X}^n(\mathbf{y}))}, \tag{7}$$

where the maximum is over all PMFs on $\mathcal{X}^n$, and where $\mathcal{X}^n(\mathbf{y})$ denotes the set of all $\mathbf{x} \in \mathcal{X}^n$ for which $W^n(\mathbf{y}|\mathbf{x})$ is positive.

We next show that (7) is asymptotically tight even when the input PMFs are restricted to be uniform over their support.

Theorem II.1. *For any DMC,*

$$C_{0\text{-u}} = \lim_{n\to\infty} n^{-1} \max_{P\in U_n} \sum_{\mathbf{y}\in\mathcal{Y}^n} (PW^n)(\mathbf{y}) \log \frac{1}{P(\mathcal{X}^n(\mathbf{y}))}, \tag{8}$$

where U_n denotes the collection of PMFs on $\mathcal{X}^n$ that are uniform over their support. Moreover, the limit is equal to the supremum.

Proof: It is straightforward to verify that the sequence on the RHS of (8) without the $1/n$ factor is superadditive, which implies that the limit is equal to the supremum.[1] Let us denote this limit by λ. Achievability, i.e., $C_{0\text{-u}} \geq \lambda$, follows because (7) holds for every n.

As to the converse, let $\mathbf{x}_1, \ldots, \mathbf{x}_M$ be a codebook of blocklength n and rate R with maximal probability of erasure under z.u.e. decoding less than some $\delta \in (0,1)$:

$$\max_{1\leq m\leq M} \sum_{\mathbf{y}\in\mathcal{Y}^n : M(\mathbf{y})>1} W^n(\mathbf{y}|\mathbf{x}_m) < \delta, \tag{9}$$

where $M(\mathbf{y})$ denotes the number of messages that cannot be ruled out when $\mathbf{y}$ is observed at the output

$$M(\mathbf{y}) = \big|\big\{1 \leq m \leq M : W^n(\mathbf{y}|\mathbf{x}_m) > 0\big\}\big|. \tag{10}$$

Condition (9) implies that $\mathbf{x}_m \neq \mathbf{x}_{m'}$ when $m \neq m'$ because otherwise, as we next argue, the conditional probability of erasure given that the m-th message was sent would be one. Indeed, if $\mathbf{x}_m = \mathbf{x}_{m'}$ for some $m \neq m'$, then $M(\mathbf{y}) \geq 2$ whenever $W^n(\mathbf{y}|\mathbf{x}_m) > 0$ because then also $W^n(\mathbf{y}|\mathbf{x}_{m'}) > 0$, and hence

$$\sum_{\mathbf{y}\in\mathcal{Y}^n : M(\mathbf{y})>1} W^n(\mathbf{y}|\mathbf{x}_m) = 1. \tag{11}$$

Having established that the codewords are distinct, we choose P to be the uniform PMF on the codebook. Then $P \in U_n$ and

$$P\big(\mathcal{X}^n(\mathbf{y})\big) = \frac{M(\mathbf{y})}{M}, \quad \text{for all } \mathbf{y} \in \mathcal{Y}^n. \tag{12}$$

We further observe that

$$\begin{aligned}
\lambda &\geq n^{-1} \sum_{\mathbf{y}\in\mathcal{Y}^n} (PW^n)(\mathbf{y}) \log \frac{1}{P(\mathcal{X}^n(\mathbf{y}))} && (13)\\
&= R - n^{-1} \sum_{\mathbf{y}\in\mathcal{Y}^n : M(\mathbf{y})>1} (PW^n)(\mathbf{y}) \log M(\mathbf{y}) && (14)\\
&\geq R\bigg(1 - \sum_{\mathbf{y}\in\mathcal{Y}^n : M(\mathbf{y})>1} (PW^n)(\mathbf{y})\bigg) && (15)\\
&= R\bigg(1 - M^{-1} \sum_{m=1}^{M} \sum_{\mathbf{y}\in\mathcal{Y}^n : M(\mathbf{y})>1} W^n(\mathbf{y}|\mathbf{x}_m)\bigg) && (16)\\
&> R(1-\delta), && (17)
\end{aligned}$$

where (13) follows because λ is the supremum of a sequence whose n-th term is no smaller than the RHS of (13); where (14) follows from (12) and the fact that $\log 1 = 0$; where (15) follows because $M(\mathbf{y}) \leq M$; where (16) follows from the choice of P; and where (17) follows from (9). Thus, for any sequence of blocklength-n rate-R codebooks with maximal probability of erasure approaching zero, we must have $R \leq \lambda$. A standard expurgation argument shows that this is also true when we replace the maximal probability of erasure with the average (over the messages) probability of erasure. ∎

[1] A sequence $a_1, a_2, \ldots$ of real numbers is superadditive if $a_{n+m} \geq a_n + a_m$ for every n and m. For superadditive sequences a_n/n tends to $\sup_n a_n/n$ [6, Problem 98].

III. Graph-Theoretic Preliminaries

A *directed graph* (or simply a *graph*) G is described by its finite *vertex set* $V(G)$ and its *edge set* $E(G) \subset V(G) \times V(G)$. We say that there is an edge from x to y in G if $(x, y) \in E(G)$. We always assume that G does not contain self-loops, i.e., that $(x, x) \notin E(G)$ for all $x \in V(G)$.

The *strong product* of two graphs G and H is denoted by $G \times H$; its vertex set is $V(G) \times V(H)$, and there is an edge from (x, y) to (x', y') in $G \times H$ if either $(x, x') \in E(G)$ and $(y, y') \in E(H)$, or if $(x, x') \in E(G)$ and $y = y'$, or if $x = x'$ and $(y, y') \in E(H)$. The n-fold strong product of G with itself is denoted by G^n.

The *subgraph of G induced by $A \subseteq V(G)$* is the graph whose vertex set is A and whose edge set is $E(G) \cap (A \times A)$.

A subset $A \subseteq V(G)$ is an *independent set* in G if the subgraph of G it induces has no edges, i.e., if $E(G) \cap (A \times A) = \emptyset$. The maximum cardinality of an independent set in G is denoted by $\omega(G)$. We define the *Sperner capacity* of G as[2]

$$\Sigma(G) = \lim_{n\to\infty} n^{-1} \log \omega(G^n), \tag{18}$$

where the limit on the RHS is equal to the supremum because the sequence $\omega(G^1), \omega(G^2), \ldots$ is supermultiplicative.[3]

A *path* in G is a sequence of $n \geq 2$ distinct vertices $x_1, \ldots, x_n$ such that $(x_j, x_{j+1}) \in E(G)$ for all $j \in \{1, \ldots, n-1\}$. The first vertex in this path is x_1, and the last vertex is x_n. We say that there is a path from x to y in G if there is a path in G whose first vertex is x and whose last vertex is y.

A *cycle* is a path $x_1, \ldots, x_n$ with $(x_n, x_1) \in E(G)$. We say that G is *acyclic* if it does not contain a cycle. The maximum cardinality of a subset $A \subseteq V(G)$ that induces an acyclic subgraph of G is denoted by $\rho(G)$.

The following two results will be key in the proof of Theorem I.1. The first is that ω can be replaced with ρ in (18).

Theorem III.1. *For every graph G,*

$$\Sigma(G) = \lim_{n\to\infty} n^{-1} \log \rho(G^n), \tag{19}$$

and the limit is equal to the supremum.

In particular, Theorem III.1 asserts that

$$\rho(G^n) \leq e^{n\Sigma(G)}, \quad \text{for all } n. \tag{20}$$

A proof of Theorem III.1 is provided in the appendix.

The number of edges of G ending in a vertex x is called the *in-degree* of x in G and is denoted by $\mathrm{d_{in}}(x, G)$, i.e.,

$$\mathrm{d_{in}}(x, G) = \big|\big\{x' \in V(G) : (x', x) \in E(G)\big\}\big|. \tag{21}$$

The next result is a slight generalization of [8, p. 95, Thm 1].

[2]Some authors prefer to define Sperner capacity in terms of cliques instead of independent sets (see, e.g., [7]).

[3]A sequence $a_1, a_2, \ldots$ of real numbers is supermultiplicative if $a_{n+m} \geq a_n a_m$ for all m and n.

Theorem III.2. *For every graph G,*

$$\sum_{x\in V(G)} \frac{1}{1 + \mathrm{d_{in}}(x, G)} \leq \rho(G). \tag{22}$$

A proof of Theorem III.2 is provided in the appendix.

For DMCs W with $\mathcal{X} \subseteq \mathcal{Y}$ and $W(x|x) > 0$ for every $x \in \mathcal{X}$, we define the *associated graph* $G(W)$ to have vertex set $\mathcal{X}$ and edge set comprising all ordered pairs (x, y) of distinct elements of $\mathcal{X}$ for which $W(y|x) > 0$. Thus, for such channels

$$C_{\mathrm{Sp}}(W) = \Sigma\big(G(W)\big). \tag{23}$$

Indeed, every Sperner code for W of blocklength n is an independent set in $G(W)^n$ and vice versa.

IV. Proof of Theorem I.1

Applying Jensen's Inequality to the RHS of (8) yields

$$C_{\text{0-u}} \leq \sup_{n\geq 1} n^{-1} \max_{P\in U_n} \log \sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(\mathcal{X}^n(\mathbf{y}))}. \tag{24}$$

It thus suffices to show that for all $P \in U_n$,

$$\sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(\mathcal{X}^n(\mathbf{y}))} \leq \big(e^{C_{\mathrm{Sp}}} + \varepsilon|\mathcal{X}|(|\mathcal{Y}| - 1)\big)^n. \tag{25}$$

Fix then some $P \in U_n$. Since the labels do not matter, we may assume for simplicity of notation that $\mathcal{X} = \{0, \ldots, |\mathcal{X}| - 1\}$ and $\mathcal{Y} = \{0, \ldots, |\mathcal{Y}| - 1\}$, where $|\mathcal{Y}| \geq |\mathcal{X}|$. The distribution on $\mathcal{Y}^n$ induced by P and W^n can be written as

$$(PW^n)(\mathbf{y}) = \sum_{\substack{\mathbf{z}\in\mathcal{Y}^n:\\ \mathbf{y}+\mathbf{z}\in\mathcal{X}^n}} P(\mathbf{y}+\mathbf{z}) W^n(\mathbf{y}|\mathbf{y}+\mathbf{z}), \tag{26}$$

where addition is to be understood component-wise modulo $|\mathcal{Y}|$. The ε-noise property (1) implies

$$W^n(\mathbf{y}|\mathbf{y}+\mathbf{z}) \leq \varepsilon^{\|\mathbf{z}\|_0}, \quad \text{if } \mathbf{y}+\mathbf{z} \in \mathcal{X}^n, \tag{27}$$

where $\|\mathbf{z}\|_0$ denotes the number of nonzero components of $\mathbf{z}$. Thus, starting with the LHS of (25),

$$\sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(\mathcal{X}^n(\mathbf{y}))} \tag{28}$$

$$= \sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \sum_{\substack{\mathbf{z}\in\mathcal{Y}^n:\\ \mathbf{y}+\mathbf{z}\in\mathcal{X}^n}} \frac{P(\mathbf{y}+\mathbf{z}) W^n(\mathbf{y}|\mathbf{y}+\mathbf{z})}{P(\mathcal{X}^n(\mathbf{y}))} \tag{29}$$

$$= \sum_{\mathbf{z}\in\mathcal{Y}^n} \sum_{\substack{\mathbf{y}\in\mathcal{Y}^n:\\ \mathbf{y}+\mathbf{z}\in\mathcal{X}^n\\ P(\mathbf{y}+\mathbf{z})>0\\ W^n(\mathbf{y}|\mathbf{y}+\mathbf{z})>0}} \frac{P(\mathbf{y}+\mathbf{z}) W^n(\mathbf{y}|\mathbf{y}+\mathbf{z})}{P(\mathcal{X}^n(\mathbf{y}))} \tag{30}$$

$$\leq \sum_{\mathbf{z}\in\mathcal{Y}^n} \varepsilon^{\|\mathbf{z}\|_0} \sum_{\substack{\mathbf{y}\in\mathcal{Y}^n:\\ \mathbf{y}+\mathbf{z}\in\mathcal{X}^n\\ P(\mathbf{y}+\mathbf{z})>0\\ W^n(\mathbf{y}|\mathbf{y}+\mathbf{z})>0}} \frac{P(\mathbf{y}+\mathbf{z})}{P(\mathcal{X}^n(\mathbf{y}))} \tag{31}$$

$$= \sum_{\mathbf{z}\in\mathcal{Y}^n} \varepsilon^{\|\mathbf{z}\|_0} \sum_{\substack{\mathbf{y}\in\mathcal{X}^n:\\ P(\mathbf{y})>0\\ W^n(\mathbf{y}-\mathbf{z}|\mathbf{y})>0}} \frac{1}{|\{\mathbf{x}\in\mathfrak{S}(P) : W^n(\mathbf{y}-\mathbf{z}|\mathbf{x}) > 0\}|}, \tag{32}$$

where (29) follows from (26); where (30) follows by changing the order of summation and dropping terms that are zero; where (31) follows from (27); and where (32) follows by substituting $\mathbf{y}$ for $\mathbf{y}+\mathbf{z}$ and because P is uniform over its support. For every $\mathbf{z} \in \mathcal{Y}^n$, let $P_{\mathbf{z}}$ be any PMF on $\mathcal{X}^n$ of support

$$\mathfrak{S}(P_{\mathbf{z}}) = \{\mathbf{x} \in \mathcal{X}^n : P(\mathbf{x})W^n(\mathbf{x}-\mathbf{z}|\mathbf{x}) > 0\}. \tag{33}$$

(In fact, $P_{\mathbf{z}}$ could be any nonnegative function with the above support.) Also define for every $\mathbf{z} \in \mathcal{Y}^n$ the channel

$$W_{\mathbf{z}}(\mathbf{y}|\mathbf{x}) = W^n(\mathbf{y}-\mathbf{z}|\mathbf{x}), \tag{34}$$

with input alphabet $\mathfrak{S}(P_{\mathbf{z}})$ and output alphabet $\mathcal{Y}^n$. Since $\mathfrak{S}(P_{\mathbf{z}}) \subseteq \mathfrak{S}(P)$,

$$\begin{aligned}|\{\mathbf{x} \in \mathfrak{S}(P) : W^n(\mathbf{y}-\mathbf{z}|\mathbf{x}) > 0\}| \\ \geq |\{\mathbf{x} \in \mathfrak{S}(P_{\mathbf{z}}) : W_{\mathbf{z}}(\mathbf{y}|\mathbf{x}) > 0\}|.\end{aligned} \tag{35}$$

Using (35) we can upper-bound the inner sum on the RHS of (32) by

$$\sum_{\mathbf{y} \in \mathfrak{S}(P_{\mathbf{z}})} \frac{1}{|\{\mathbf{x} \in \mathfrak{S}(P_{\mathbf{z}}) : W_{\mathbf{z}}(\mathbf{y}|\mathbf{x}) > 0\}|}. \tag{36}$$

This sum can also be written as

$$\sum_{\mathbf{y} \in V(G(W_{\mathbf{z}}))} \frac{1}{1+\mathrm{d}_{\mathrm{in}}(\mathbf{y}, G(W_{\mathbf{z}}))}, \tag{37}$$

where $G(W_{\mathbf{z}})$ is the graph associated with the channel $W_{\mathbf{z}}$ (see Section III). Since (37) is upper-bounded by $\rho(G(W_{\mathbf{z}}))$ (Theorem III.2), we thus have

$$\sum_{\mathbf{y} \in \mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(\mathcal{X}^n(\mathbf{y}))} \leq \sum_{\mathbf{z} \in \mathcal{Y}^n} \varepsilon^{\|\mathbf{z}\|_0} \rho\big(G(W_{\mathbf{z}})\big). \tag{38}$$

We next argue that

$$\rho\big(G(W_{\mathbf{z}})\big) \leq |\mathcal{X}|^{\|\mathbf{z}\|_0} \rho\big(G(W)^{n-\|\mathbf{z}\|_0}\big), \tag{39}$$

where we define $\rho(G(W)^0) = 1$. When $\|\mathbf{z}\|_0 = n$, then (39) is trivial, so we assume that $0 \leq \|\mathbf{z}\|_0 < n$. Let $\mathbf{x}(\mathbf{z})$ denote the restriction of $\mathbf{x} \in \mathcal{X}^n$ to the nonzero components of $\mathbf{z}$, and let $\mathbf{x}(\mathbf{z}^{\mathrm{c}})$ denote the restriction of $\mathbf{x}$ to the zero components of $\mathbf{z}$. We will prove (39) by contradiction. In order to reach a contradiction, assume that for some integer η strictly larger than the RHS of (39) there exist distinct vertices $\mathbf{x}_1, \ldots, \mathbf{x}_\eta$ in $\mathfrak{S}(P_{\mathbf{z}})$ that induce an acyclic subgraph of $G(W_{\mathbf{z}})$. Partition this collection of vertices by placing into the same class all $\mathbf{x}_j$'s that have the same restriction $\mathbf{x}_j(\mathbf{z})$. Since there are $|\mathcal{X}|^{\|\mathbf{z}\|_0}$ such classes, one of them must contain $\kappa > \rho(G(W)^{n-\|\mathbf{z}\|_0})$ vertices; call them $\mathbf{x}'_1, \ldots, \mathbf{x}'_\kappa$. Since $\mathbf{x}'_1, \ldots, \mathbf{x}'_\kappa$ are distinct, and since their restrictions to the nonzero components of $\mathbf{z}$ are identical, their restrictions to the zero components of $\mathbf{z}$, i.e., $\mathbf{x}'_1(\mathbf{z}^{\mathrm{c}}), \ldots, \mathbf{x}'_\kappa(\mathbf{z}^{\mathrm{c}})$ must all be distinct. Also, if $\mathbf{x}, \mathbf{y} \in \mathfrak{S}(P_{\mathbf{z}})$ and $\mathbf{x}(\mathbf{z}) = \mathbf{y}(\mathbf{z})$, then

$$W_{\mathbf{z}}(\mathbf{y}|\mathbf{x}) > 0 \iff W^{n-\|\mathbf{z}\|_0}\big(\mathbf{y}(\mathbf{z}^{\mathrm{c}})\big|\mathbf{x}(\mathbf{z}^{\mathrm{c}})\big) > 0. \tag{40}$$

It follows that the subgraph of $G(W_{\mathbf{z}})$ induced by $\mathbf{x}'_1, \ldots, \mathbf{x}'_\kappa$ is isomorphic to the subgraph of $G(W^{n-\|\mathbf{z}\|_0})$ induced by $\mathbf{x}'_1(\mathbf{z}^{\mathrm{c}}), \ldots, \mathbf{x}'_\kappa(\mathbf{z}^{\mathrm{c}})$.[4] And since the former is acyclic, so must the latter be, which is a contradiction because $G(W^{n-\|\mathbf{z}\|_0}) = G(W)^{n-\|\mathbf{z}\|_0}$ and $\kappa > \rho(G(W)^{n-\|\mathbf{z}\|_0})$.

Having established (39), we further note that by (20) and (23),

$$\rho\big(G(W)^{n-\|\mathbf{z}\|_0}\big) \leq e^{(n-\|\mathbf{z}\|_0)C_{\mathrm{Sp}}}. \tag{41}$$

By combining (38), (39), and (41), we obtain

$$\sum_{\mathbf{y} \in \mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(\mathcal{X}^n(\mathbf{y}))} \tag{42}$$

$$\leq \sum_{\mathbf{z} \in \mathcal{Y}^n} \varepsilon^{\|\mathbf{z}\|_0} |\mathcal{X}|^{\|\mathbf{z}\|_0} e^{(n-\|\mathbf{z}\|_0)C_{\mathrm{Sp}}} \tag{43}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (|\mathcal{Y}|-1)^k \varepsilon^k |\mathcal{X}|^k e^{(n-k)C_{\mathrm{Sp}}}, \tag{44}$$

where the equality follows because the summand on the RHS of (43) depends on $\mathbf{z}$ only via $\|\mathbf{z}\|_0$ and because there are $\binom{n}{k}(|\mathcal{Y}|-1)^k$ elements in $\mathcal{Y}^n$ with exactly k nonzero components. This completes the proof because the RHS of (44) is equal to the RHS of (25). ∎

V. Remarks

1) In Theorem I.1 we may replace $|\mathcal{Y}|$ with $|\mathcal{X}| + 2^{|\mathcal{X}|} - 1$. See [3] for a proof.
2) For some channels the bound in Theorem I.1 can be sharpened. See [4] for an interesting example.

Appendix

Proof of Theorem III.1: We shall need the elementary fact that the vertices of any acyclic graph G can be labeled with the numbers $1, \ldots, |V(G)|$ such that $(x, y) \in E(G)$ only if $x < y$ (see, e.g., [9, Section 5.7]).[5]

Using this fact, we first show that the sequence $\rho(G^1), \rho(G^2), \ldots$ is supermultiplicative, which will imply that the limit on the RHS of (19) equals the supremum. Choose for each n some $A_n \subseteq V(G)^n$ that achieves $\rho(G^n)$, i.e., A_n induces an acyclic subgraph of G^n and $|A_n| = \rho(G^n)$. We show that $A_n \times A_m$ induces an acyclic subgraph of G^{n+m} and hence that

$$\rho\big(G^{n+m}\big) \geq |A_n \times A_m| \tag{45}$$

$$= \rho(G^n)\rho(G^m). \tag{46}$$

Label the vertices in A_n with the numbers $1, \ldots, |A_n|$ so that $(x, x') \in E(G^n) \cap (A_n \times A_n)$ implies $x < x'$. Similarly label the vertices in A_m. To reach a contradiction, assume that $(x_1, y_1), \ldots, (x_\eta, y_\eta)$ is a cycle in the subgraph of G^{n+m} induced by $A_n \times A_m$. From the definition of strong product and the labeling of the vertices it follows that $x_1 < x_\eta$ or $y_1 < y_\eta$. Consequently, there cannot be an edge from (x_η, y_η) to (x_1, y_1) in this subgraph, which contradicts the assumption that $(x_1, y_1), \ldots, (x_\eta, y_\eta)$ is a cycle.

[4] The isomorphism is $\mathbf{x} \mapsto \mathbf{x}(\mathbf{z}^{\mathrm{c}})$.

[5] A different way to state this is that any partial order on a finite set can be extended to a total order on this set.

As to (19), we first show that

$$\Sigma(G) = \log|V(G)|, \quad \text{for all acyclic } G. \tag{47}$$

Note that this will prove Theorem III.1 in the special case where G is acyclic. Indeed, in this case $\rho(G) = |V(G)|$, so (46) implies $\rho(G^n) \geq |V(G)|^n$. And since clearly $\rho(G^n) \leq |V(G)|^n$, we thus have

$$\rho(G^n) = |V(G)|^n, \quad \text{for all acyclic } G. \tag{48}$$

To prove (47), note that $\omega(G^n) \leq |V(G)|^n$ and hence $\Sigma(G) \leq \log|V(G)|$ (this is true for any G, not just acyclic), so it only remains to prove the reverse inequality. Since G is acyclic, we may label its vertices with the numbers $1, \ldots, |V(G)|$ so that there is an edge from x to y in G only if $x < y$. We then define the weight of a vertex $\mathbf{x}$ in G^n as the sum of the labels of its n components. Thus, the weight is a number between n and $n|V(G)|$.

As we next show, if A is a subset of $V(G)^n$ all of whose members have the same weight, then A is an independent set in G^n. Indeed, if $\mathbf{x}$ and $\mathbf{y}$ are distinct vertices in A, then $x^{(j)} > y^{(j)}$, say, for some $j \in \{1, \ldots, n\}$. Since $\mathbf{x}$ and $\mathbf{y}$ have equal weight, there must also be some $k \neq j$ for which $x^{(k)} < y^{(k)}$. Thus, $(x^{(j)}, y^{(j)}) \notin E(G)$ and $(y^{(k)}, x^{(k)}) \notin E(G)$, so there is no edge from $\mathbf{x}$ to $\mathbf{y}$ and no edge from $\mathbf{y}$ to $\mathbf{x}$ in G^n.

If we partition $V(G)^n$ by putting in the same class all vertices of the same weight, then one of the classes must have at least

$$\frac{|V(G)|^n}{n|V(G)| - n + 1}$$

members. Thus,

$$n^{-1} \log \omega(G^n) \geq \log|V(G)| - n^{-1} \log\big(n|V(G)| - n + 1\big),$$

and letting n tend to infinity establishes $\Sigma(G) \geq \log|V(G)|$ and hence proves (47).

To complete the proof of (19), let G be any graph (not necessarily acyclic) and let λ denote the limit of $n^{-1} \log \rho(G^n)$ as n tends to infinity (i.e., the supremum). For a given $\delta > 0$ select ν so that

$$\nu^{-1} \log \rho(G^\nu) \geq \lambda - \delta. \tag{49}$$

Choose $A \subseteq V(G)^\nu$ that achieves $\rho(G^\nu)$ and let H denote the acyclic subgraph of G^ν it induces. Since H^m is the subgraph of $G^{\nu m}$ induced by A^m,

$$(\nu m)^{-1} \log \omega(G^{\nu m}) \geq (\nu m)^{-1} \log \omega(H^m). \tag{50}$$

Letting m tend to infinity, we obtain

$$\Sigma(G) \geq \nu^{-1} \Sigma(H). \tag{51}$$

Since H is acyclic, we can substitute it for G in (47) to obtain

$$\nu^{-1} \Sigma(H) = \nu^{-1} \log|A| \tag{52}$$

$$= \nu^{-1} \log \rho\big(G^\nu\big), \tag{53}$$

where (53) follows because A achieves $\rho(G^\nu)$. Combining (51), (53), and (49) shows that $\Sigma(G) \geq \lambda - \delta$. Since this is true for every $\delta > 0$, we must in fact have $\Sigma(G) \geq \lambda$. On the other hand, a graph with no edges is trivially acyclic, so $\omega(G^n) \leq \rho(G^n)$ and hence $\Sigma(G) \leq \lambda$. ∎

Proof of Theorem III.2: Let $<$ be a total ordering of the vertices of G and consider the subset $A \subseteq V(G)$ comprising all $x \in V(G)$ such that if $(x', x) \in E(G)$ for some $x' \in V(G)$, then $x' < x$. The subgraph of G induced by A is acyclic. Indeed, if $x_1, \ldots, x_\eta$ is a path in this subgraph, then $x_1 < x_\eta$, so we cannot have $(x_\eta, x_1) \in E(G)$. Thus,

$$|A| \leq \rho(G). \tag{54}$$

Suppose now that $<$ is drawn uniformly at random among all total orderings of $V(G)$. Then

$$\Pr(x \in A) = \frac{1}{1 + \mathrm{d}_{\text{in}}(x, G)}, \quad \text{for all } x \in V(G). \tag{55}$$

Indeed, x is in A if, and only if, it is the greatest vertex in the set

$$B = \{x\} \cup \{x' : (x', x) \in E(G)\}. \tag{56}$$

Since $<$ is drawn uniformly at random, every vertex in B has the same probability of being the greatest element in B, so (55) follows by noting that $|B| = 1 + \mathrm{d}_{\text{in}}(x, G)$.

Summing both sides of (55) over all vertices of G yields

$$\sum_{x \in V(G)} \frac{1}{1 + \mathrm{d}_{\text{in}}(x, G)} = \sum_{x \in V(G)} \Pr(x \in A). \tag{57}$$

By writing $\Pr(x \in A)$ as the expectation of the indicator function of the event $\{x \in A\}$ and by swapping summation and expectation, we see that the RHS of (57) is the expected cardinality of A. This expected cardinality cannot exceed $\rho(G)$ because (54) holds for every realization of $<$. ∎

REFERENCES

[1] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, 1995.

[2] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 55–62, 1996.

[3] C. Bunte, A. Lapidoth, and A. Samorodnitsky, "The zero-undetected-error capacity approaches the Sperner capacity," Sep. 2013, arXiv:1309.4930 [cs.IT]. [Online]. Available: http://arxiv.org/abs/1309.4930

[4] C. Bunte, A. Lapidoth, and A. Samorodnitsky, "The zero-undetected-error capacity of the low-noise cyclic triangle channel," in *Information Theory Proceedings (ISIT), 2013 IEEE International Symposium on*, 2013, pp. 91–95.

[5] G. Forney Jr, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, pp. 206–220, 1968.

[6] G. Pólya and G. Szegő, *Problems and Theorems in Analysis I*. Berlin Heidelberg: Springer-Verlag, 1978.

[7] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York: Cambridge University Press, 2011.

[8] N. Alon and J. H. Spencer, *The Probabilistic Method*, 3rd ed. Hoboken, NJ: Wiley, 2008.

[9] K. Thulasiraman and M. N. S. Swamy, *Graphs: Theory and Algorithms*. New York: John Wiley & Sons, 1992.