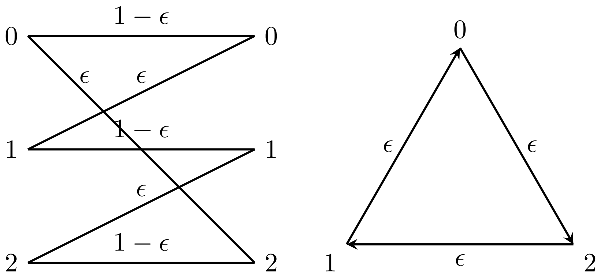
The Zero-Undetected-Error Capacity of the Low-Noise Cyclic Triangle Channel

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



(a) The bipartite channel graph (b) The directed channel graph

Fig. 1. The cyclic triangle channel

***Abstract*—We study the zero-undetected-error capacity of the discrete memoryless channel whose directed channel graph is the cyclic triangle. We show that this capacity is upper-bounded by $\log 2$ and approaches $\log 2$ as the crossover probabilities tend to zero.**

I. Introduction

The zero-undetected-error capacity C_{0u} is the largest rate at which communication is possible over a channel when the decoder must produce either the correct message or declare an erasure (with small probability); it should never produce an incorrect message. This notion of channel capacity was first studied by Forney [1], who noticed that positive rates are achievable on a discrete memoryless channel (DMC) whenever there is an output that is reachable from some but not from all inputs. He also derived a lower bound on C_{0u} using IID random coding, but the bound is not always tight.

More than 40 years after Forney's paper, determining C_{0u} for arbitrary DMCs is still an open problem. In fact, even for seemingly simple (and symmetric) channels, like those considered in this paper, C_{0u} is unknown. The main contributions after Forney's paper include improved lower bounds [2], [3], sufficient conditions for C_{0u} to equal the (ordinary) capacity C [4], [5], a single-letter expression for channels with binary inputs [6], and a single-letter expression when there is a feedback link from the channel output to the encoder [7].

The biggest challenge, it seems, is finding nontrivial upper bounds, i.e., upper bounds better than C. In this paper, we derive such an upper bound for the channel whose bipartite (undirected) channel graph is shown in Figure 1(a).[1] We refer to this channel as the cyclic triangle channel because it can be represented by the directed graph in Figure 1(b).[2] Specifically, we show that for this channel C_{0u} is upper-bounded by $\log 2$ if $0 < \epsilon < 1$. By providing a lower bound, we conclude that C_{0u} approaches $\log 2$ as ϵ tends to zero.

There are a number of reasons for considering the cyclic triangle channel. It is one of the simplest channels for which C_{0u} is not known and apparently hard to compute. A numerical analysis in [3] suggests that the best known single-letter lower bound for C_{0u} (see (35) ahead) is not tight for this channel. Moreover, the channel illustrates nicely how different, in general, the zero-error capacity (zero when $0 < \epsilon < 1$), the capacity (approaches $\log 3$ as $\epsilon \to 0$), and the zero-undetected-error capacity are. Another reason is that the channel belongs to the class of low-noise channels as defined in [3], where it is conjectured that for such channels C_{0u} approaches the Sperner capacity (as defined, e.g., in [8, page 194]) of the directed channel graph as the crossover probabilities tend to zero. The Sperner capacity of the graph in Figure 1(b) is $\log 2$ [9], [10].

It is instructive to compare the cyclic triangle channel to the acyclic triangle channel, whose bipartite and directed channel graphs are depicted in Figure 2. This channel too belongs to the class of low-noise channels. Unlike its cyclic counterpart, its zero-undetected-error capacity approaches $\log 3$ as ϵ tends to zero. Here the converse is trivial because C_{0u} is upper-bounded by C. Achievability can be demonstrated by noting that the Sperner capacity of the directed channel graph is a lower bound to the limiting value of C_{0u} [3, Theorem 2], and the Sperner capacity of the graph in Figure 2(b) is $\log 3$ [9]. An alternative proof is provided in Section IV.

The rest of the paper is organized as follows. In Section II we briefly discuss notation; in Section III we state and

[1]The two disjoint independent vertex sets of the bipartite channel graph of a DMC are the input and output alphabets. There is an edge between an input x and an output y if sending x can produce y at the output. It is customary to draw the inputs on the left and the outputs on the right. The edges are labeled with the transition probabilities.

[2]Every DMC with identical input and output alphabets has a natural representation as a directed graph in which the vertices are the inputs/outputs and there is an edge from x to y if y is not equal to x and if sending x can produce y at the output. The edges are labeled with the transition probabilities. It is implicitly assumed that a given channel input can produce an identical output if the labels of the edges emanating from the vertex do not sum to one.

prove the main result about the cyclic triangle channel; and Section IV contains the analysis of the acyclic triangle channel.

II. Notation

We use $W, \mathcal{X}, \mathcal{Y}$ to denote the channel law and the input and output alphabets (both finite) of a generic DMC. For a PMF Q on $\mathcal{X}$, we write QW for the induced PMF on $\mathcal{Y}$

$$(QW)(y) = \sum_{x \in \mathcal{X}} Q(x) W(y|x), \quad y \in \mathcal{Y}. \tag{1}$$

We write W^n for the n-fold product of W

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{j=1}^{n} W(y_j|x_j), \quad \mathbf{x} \in \mathcal{X}^n, \mathbf{y} \in \mathcal{Y}^n, \tag{2}$$

where x_j and y_j denote the j-th coordinates of $\mathbf{x}$ and $\mathbf{y}$. We often use the shorthand

$$\mathcal{X}^n(\mathbf{y}) = \{\mathbf{x} \in \mathcal{X}^n : W^n(\mathbf{y}|\mathbf{x}) > 0\}, \quad \mathbf{y} \in \mathcal{Y}^n, \tag{3}$$

and, for a PMF Q on $\mathcal{X}^n$,

$$Q\big(\mathcal{X}^n(\mathbf{y})\big) = \sum_{\mathbf{x} \in \mathcal{X}^n(\mathbf{y})} Q(\mathbf{x}). \tag{4}$$

The support $\text{supp}(P)$ of a PMF P on a set $\mathcal{X}$ is defined as

$$\text{supp}(P) = \{x \in \mathcal{X} : P(x) > 0\}. \tag{5}$$

The cardinality of a set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. For the usual information theoretic quantities, we follow the notation in [8]. All logarithms are natural logarithms.

III. The Cyclic Triangle Channel

The following theorem is the main result of this paper.

Theorem III.1. *Let $C_{0u}(\epsilon)$ denote the zero-undetected-error capacity of the cyclic triangle channel with crossover probability ϵ (Figure 1). Then*

$$C_{0u}(\epsilon) \le \log 2, \quad 0 < \epsilon < 1, \tag{6}$$

and

$$\lim_{\epsilon \to 0} C_{0u}(\epsilon) = \log 2. \tag{7}$$

Proof. We begin with the direct part. If we restrict the channel to the inputs 0 and 1, say, then the resulting bipartite channel graph is acyclic. A result by Pinsker and Sheverdyaev [4] asserts that C_{0u} equals C for all DMCs with acyclic bipartite channel graphs. Noting that C of the restricted channel approaches $\log 2$ as ϵ tends to zero completes the direct part.

For the converse, we use the following multi-letter characterization of C_{0u}. For any DMC W,

$$C_{0u} = \lim_{n \to \infty} \frac{1}{n} \max_{Q \in \mathcal{Q}_n} \sum_{\mathbf{y}} (QW^n)(\mathbf{y}) \log \frac{1}{Q(\mathcal{X}^n(\mathbf{y}))}, \tag{8}$$

where $\mathcal{Q}_n$ is the family of PMFs on $\mathcal{X}^n$ that are uniform over a subset of $\mathcal{X}^n$, and where the sum extends over the support of QW^n. The limit on the RHS of (8) exists (and is equal to the supremum) because the sequence without the $1/n$ factor is superadditive.[3] The achievability of the RHS of (8) follows immediately from Forney's lower bound [1] applied to W^n. The converse part of (8) seems to be well-known, but we include it in the appendix for completeness. Using Jensen's inequality, we can upper bound the sum on the RHS of (8) as

$$\sum_{\mathbf{y}} (QW^n)(\mathbf{y}) \log \frac{1}{Q(\mathcal{X}^n(\mathbf{y}))} \le \log \sum_{\mathbf{y}} \frac{(QW^n)(\mathbf{y})}{Q(\mathcal{X}^n(\mathbf{y}))}, \tag{9}$$

where both sums extend over the support of QW^n. For the rest of the proof, let W denote the channel law of the cyclic triangle channel with crossover probability $0 < \epsilon < 1$. Note that $W^n(\mathbf{y}|\mathbf{x}) > 0$ if, and only if, $\mathbf{x}$ can be obtained from $\mathbf{y}$ by adding 1 in mod 3 arithmetic to a subset (perhaps empty) of the coordinates of $\mathbf{y}$. For example, if $n = 2$, then $\mathbf{y} = (0, 2)$ can be reached from $(0, 2), (0, 0), (1, 2), (1, 0)$. For every $\mathbf{y} \in \{0, 1, 2\}^n$ and $\mathcal{U} \subseteq \{1, \dots, n\}$, define $\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U})$ to be the element of $\{0, 1, 2\}^n$ obtained from $\mathbf{y}$ by adding 1 (in mod 3 arithmetic) to the coordinates of $\mathbf{y}$ enumerated in $\mathcal{U}$. With this notation, we can write the sum inside the logarithm on the RHS of (9) as

$$\sum_{\mathbf{y} \in \text{supp}(QW^n)} \sum_{\mathcal{U} \subseteq \{1, \dots, n\}} \frac{W^n(\mathbf{y}|\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U})) Q(\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U}))}{Q(\mathcal{X}^n(\mathbf{y}))}. \tag{10}$$

Since

$$W^n(\mathbf{y}|\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U})) = \epsilon^{|\mathcal{U}|} (1 - \epsilon)^{n - |\mathcal{U}|}, \tag{11}$$

we can rewrite (10) as

$$\sum_{\mathcal{U} \subseteq \{1, \dots, n\}} \epsilon^{|\mathcal{U}|} (1 - \epsilon)^{n - |\mathcal{U}|} \sum_{\mathbf{y} \in \text{supp}(QW^n)} \frac{Q(\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U}))}{Q(\mathcal{X}^n(\mathbf{y}))}. \tag{12}$$

Since $(QW^n)(\mathbf{y}) > 0$ whenever $Q(\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U})) > 0$, we can restrict the inner summation in (12) to all $\mathbf{y}$ such that $Q(\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U})) > 0$. Moreover, for $Q \in \mathcal{Q}_n$ and $\mathbf{y}$ such that $Q(\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U})) > 0$, we have

$$\frac{Q(\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U}))}{Q(\mathcal{X}^n(\mathbf{y}))} = \frac{1}{|\{\mathbf{x} : Q(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) > 0\}|}, \tag{13}$$

so the inner sum in (12) can be written as

$$\sum_{\mathbf{y} : Q(\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U})) > 0} \frac{1}{|\{\mathbf{x} : Q(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) > 0\}|}. \tag{14}$$

For every $\mathbf{x} \in \{0, 1, 2\}^n$ and $\mathcal{U} \subseteq \{1, \dots, n\}$, define $\boldsymbol{\beta}(\mathbf{x}, \mathcal{U})$ to be the element of $\{0, 1, 2\}^n$ obtained from $\mathbf{x}$ by changing 0's into 2's and vice versa in all coordinates of $\mathbf{x}$ enumerated in $\mathcal{U}$. Define the permutation $Q_{\mathcal{U}}$ of Q by

$$Q_{\mathcal{U}}(\mathbf{x}) = Q(\boldsymbol{\beta}(\mathbf{x}, \mathcal{U})), \quad \mathbf{x} \in \{0, 1, 2\}^n. \tag{15}$$

Observe that for every $\mathcal{U} \subseteq \{1, \dots, n\}$,

$$W^n(\mathbf{y}|\mathbf{x}) > 0 \iff W^n\big(\boldsymbol{\beta}(\boldsymbol{\alpha}(\mathbf{y}, \mathcal{U}), \mathcal{U}) \big| \boldsymbol{\beta}(\mathbf{x}, \mathcal{U})\big) > 0.$$

[3] A sequence $\{a_n\}$ of real numbers is superadditive if $a_{m+n} \ge a_m + a_n$ for all positive integers m and n.

This equivalence is straightforward to verify for $n = 1$ and easily extends to general n on account of (2). Since $Q_{\mathcal{U}}(\boldsymbol{\beta}(\mathbf{x},\mathcal{U})) = Q(\mathbf{x})$, we thus have

$$\begin{aligned}&\big\{\mathbf{x} : Q(\mathbf{x})W^n(\mathbf{y}|\mathbf{x}) > 0\big\}\\ &= \Big\{\mathbf{x} : Q_{\mathcal{U}}\big(\boldsymbol{\beta}(\mathbf{x},\mathcal{U})\big)W^n\big(\boldsymbol{\beta}(\boldsymbol{\alpha}(\mathbf{y},\mathcal{U}),\mathcal{U})\big|\boldsymbol{\beta}(\mathbf{x},\mathcal{U})\big) > 0\Big\}.\end{aligned}$$

And since $\boldsymbol{\beta}(\cdot,\mathcal{U})$ is bijective, it follows that

$$\begin{aligned}&\big|\big\{\mathbf{x} : Q(\mathbf{x})W^n(\mathbf{y}|\mathbf{x}) > 0\big\}\big|\\ &\quad = \Big|\Big\{\mathbf{x} : Q_{\mathcal{U}}(\mathbf{x})W^n\big(\boldsymbol{\beta}(\boldsymbol{\alpha}(\mathbf{y},\mathcal{U}),\mathcal{U})\big|\mathbf{x}\big) > 0\Big\}\Big|. \qquad (16)\end{aligned}$$

Substituting (16) into (14), the inner sum in (12) becomes

$$\sum_{\mathbf{y}} \frac{1}{\big|\big\{\mathbf{x} : Q_{\mathcal{U}}(\mathbf{x})W^n\big(\boldsymbol{\beta}(\boldsymbol{\alpha}(\mathbf{y},\mathcal{U}),\mathcal{U})\big|\mathbf{x}\big) > 0\big\}\big|}, \qquad (17)$$

where the sum extends over $\mathbf{y}$ such that $Q(\boldsymbol{\alpha}(\mathbf{y},\mathcal{U})) > 0$. Since summing over $\mathbf{y}$ such that $Q(\boldsymbol{\alpha}(\mathbf{y},\mathcal{U})) > 0$ is the same as summing over $\mathbf{y}$ such that $Q_{\mathcal{U}}(\boldsymbol{\beta}(\boldsymbol{\alpha}(\mathbf{y},\mathcal{U}),\mathcal{U})) > 0$, and since $\boldsymbol{\beta}(\cdot,\mathcal{U})$ and $\boldsymbol{\alpha}(\cdot,\mathcal{U})$ are both bijective, (17) is equal to

$$\sum_{\mathbf{y}\in\mathrm{supp}(Q_{\mathcal{U}})} \frac{1}{|\{\mathbf{x} : Q_{\mathcal{U}}(\mathbf{x})W^n(\mathbf{y}|\mathbf{x}) > 0\}|}. \qquad (18)$$

Consequently, if we define

$$\vartheta_n(Q) = \sum_{\mathbf{y}\in\mathrm{supp}(Q)} \frac{1}{|\{\mathbf{x} : Q(\mathbf{x})W^n(\mathbf{y}|\mathbf{x}) > 0\}|}, \qquad (19)$$

then it follows from (10), (12), (17), and (18) that the sum inside the logarithm on the RHS of (9) is equal to

$$\sum_{\mathcal{U}\subseteq\{1,\ldots,n\}} \epsilon^{|\mathcal{U}|}(1-\epsilon)^{n-|\mathcal{U}|}\,\vartheta_n(Q_{\mathcal{U}}). \qquad (20)$$

The maximum of (20) taken over all $Q \in \mathcal{Q}_n$ can be upper bounded as

$$\begin{aligned}&\max_{Q\in\mathcal{Q}_n} \sum_{\mathcal{U}\subseteq\{1,\ldots,n\}} \epsilon^{|\mathcal{U}|}(1-\epsilon)^{n-|\mathcal{U}|}\,\vartheta_n(Q_{\mathcal{U}})\\ &\quad\le \sum_{\mathcal{U}\subseteq\{1,\ldots,n\}} \epsilon^{|\mathcal{U}|}(1-\epsilon)^{n-|\mathcal{U}|} \max_{Q\in\mathcal{Q}_n} \vartheta_n(Q_{\mathcal{U}})\\ &\quad= \max_{Q\in\mathcal{Q}_n} \vartheta_n(Q) \sum_{\mathcal{U}\subseteq\{1,\ldots,n\}} \epsilon^{|\mathcal{U}|}(1-\epsilon)^{n-|\mathcal{U}|}\\ &\quad= \max_{Q\in\mathcal{Q}_n} \vartheta_n(Q), \qquad (21)\end{aligned}$$

where the first equality follows because if Q is in $\mathcal{Q}_n$, then so is every permutation of Q. Combing (21), (20), (9), and (8) shows that

$$C_{0\mathrm{u}}(\epsilon) \le \varliminf_{n\to\infty} \frac{1}{n}\log\Big(\max_{Q\in\mathcal{Q}_n} \vartheta_n(Q)\Big), \quad 0 < \epsilon < 1. \qquad (22)$$

Note that the RHS of (22) does not depend on ϵ. The proof will be completed by showing that $\vartheta_n(Q) \le 2^n$ for all n and $Q \in \mathcal{Q}_n$. To this end, we will formulate the problem in the language of graph theory. For positive integers n, consider the directed graph $G_n = (V(G_n), E(G_n))$ with vertex set $V(G_n) = \{0,1,2\}^n$ and edge set

$$E(G_n) = \big\{(\mathbf{x},\mathbf{y}) : \mathbf{x} \ne \mathbf{y},\, W^n(\mathbf{y}|\mathbf{x}) > 0\big\}. \qquad (23)$$

In other words, there is an edge from $\mathbf{x}$ to $\mathbf{y}$ if, and only if, $\mathbf{x}$ is not equal to $\mathbf{y}$ and $W^n(\mathbf{y}|\mathbf{x}) > 0$. For every subset $F \subseteq \{0,1,2\}^n$ and $\mathbf{x} \in F$ let $d(F,\mathbf{x})$ denote the in-degree (the number of incoming edges) of the vertex $\mathbf{x}$ in the subgraph of G_n induced by F, i.e., the graph with vertex set F and edge set $E(G_n) \cap (F \times F)$. Using this notation, we see from (19) that

$$\max_{Q\in\mathcal{Q}_n} \vartheta_n(Q) = \max_{F\subseteq\{0,1,2\}^n} \sum_{\mathbf{x}\in F} \frac{1}{1 + d(F,\mathbf{x})}. \qquad (24)$$

To upper bound the RHS of (24), we use a technique similar to that in [11, page 95]. Fix a positive integer n, and fix a nonempty $F \subseteq \{0,1,2\}^n$. A bijective map O_F of the form

$$O_F \colon F \to \{1,\ldots,|F|\}, \qquad (25)$$

is called an ordering of F. Let O_F be an ordering of F and consider the set

$$I(O_F) = \big\{\mathbf{x} : O_F(\mathbf{x}') < O_F(\mathbf{x}) \text{ if } (\mathbf{x}',\mathbf{x}) \in E(G_n)\big\}. \qquad (26)$$

In other words, $I(O_F) \subseteq F$, and $I(O_F)$ contains the vertex $\mathbf{x} \in F$ if, and only if, for every $\mathbf{x}' \in F$ such that there is an edge from $\mathbf{x}'$ to $\mathbf{x}$, the vertex $\mathbf{x}'$ is lower in the ordering O_F than $\mathbf{x}$. If O_F is drawn uniformly at random among all possible orderings of F, then the probability that a particular $\mathbf{x} \in F$ is in $I(O_F)$ is $1/(1 + d(F,\mathbf{x}))$. Thus, the expected cardinality of $I(O_F)$ can be computed as

$$\begin{aligned}\mathrm{E}\big[|I(O_F)|\big] &= \sum_{\mathbf{x}\in F} \Pr\big(\mathbf{x} \in I(O_F)\big)\\ &= \sum_{\mathbf{x}\in F} \frac{1}{1 + d(F,\mathbf{x})}. \qquad (27)\end{aligned}$$

Since the average is upper-bounded by the maximum,

$$\sum_{\mathbf{x}\in F} \frac{1}{1 + d(F,\mathbf{x})} \le \max_{O_F} |I(O_F)|. \qquad (28)$$

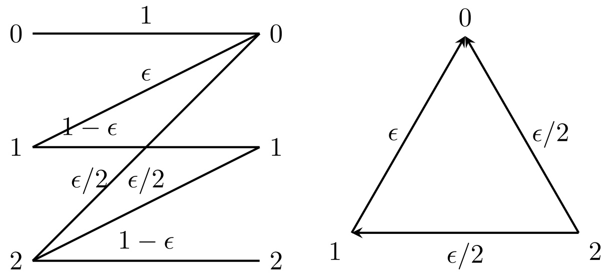
For every ordering O_F of F, the subgraph of G_n induced by the vertices in $I(O_F)$ is acyclic. Consequently, if Γ_n denotes the maximum cardinality of a subset of $\{0,1,2\}^n$ that induces an acyclic subgraph of G_n, then (28) implies

$$\max_{F\subseteq\{0,1,2\}^n} \sum_{\mathbf{x}\in F} \frac{1}{1 + d(F,\mathbf{x})} \le \Gamma_n, \quad n \ge 1. \qquad (29)$$

It thus suffices to show that $\Gamma_n \le 2^n$. We use an argument similar to that in [10]. Let $J \subseteq \{0,1,2\}^n$ be a subset of cardinality Γ_n that induces an acyclic subgraph of G_n, and consider the set of multivariate polynomials $\{p_{\mathbf{y}}\}_{\mathbf{y}\in J}$ over GF(3) defined as

$$p_{\mathbf{y}}(\mathbf{x}) = \prod_{j=1}^{n} (y_j - x_j - 1), \quad \mathbf{x} \in \{0,1,2\}^n. \qquad (30)$$

Then $p_{\mathbf{y}}(\mathbf{y}) \ne 0$ for every $\mathbf{y} \in J$ and $p_{\mathbf{y}}(\mathbf{x}) = 0$ if $W^n(\mathbf{y}|\mathbf{x}) = 0$. We show that the polynomials $\{p_{\mathbf{y}}\}_{\mathbf{y}\in J}$ are



(a) The bipartite channel graph (b) The directed channel graph

Fig. 2. The acyclic triangle channel

linearly independent. Assume that for some $m \geq 2$, the m elements $\mathbf{y}_1, \ldots, \mathbf{y}_m$ in J are distinct and that $a_1, \ldots, a_m$ are nonzero coefficients such that for all $\mathbf{x} \in \{0,1,2\}^n$,

$$a_1 p_{\mathbf{y}_1}(\mathbf{x}) + a_2 p_{\mathbf{y}_2}(\mathbf{x}) + \ldots + a_m p_{\mathbf{y}_m}(\mathbf{x}) = 0. \tag{31}$$

Since the subgraph of G_n induced by J is acyclic, so is the subgraph induced by $\mathbf{y}_1, \ldots, \mathbf{y}_m$. Consequently, there must be some $k \in \{1, \ldots, m\}$ such that there is no edge from $\mathbf{y}_k$ to $\mathbf{y}_j$ for all $j \neq k$, and hence $p_{\mathbf{y}_j}(\mathbf{y}_k) = 0$ for all $j \neq k$. In view of (31) this implies

$$a_k p_{\mathbf{y}_k}(\mathbf{y}_k) = 0, \tag{32}$$

which is a contradiction because a_k is assumed to be nonzero and $p_{\mathbf{y}_k}(\mathbf{y}_k) \neq 0$. We conclude that the set $\{p_{\mathbf{y}}\}_{\mathbf{y} \in J}$ is linearly independent. Since these polynomials are contained in the span of the polynomials

$$\Big\{\mathbf{x} \mapsto \prod_{\ell \in \mathcal{L}} x_\ell : \mathcal{L} \subseteq \{1, \ldots, n\}\Big\}, \tag{33}$$

and since this set has 2^n elements, it follows that $\Gamma_n \leq 2^n$. Combining this with (29), (24) and (22) completes the proof. □

IV. The Acyclic Triangle Channel

As pointed out in Section I, the following proposition is an immediate consequence of [3, Theorem 2] and the fact that the Sperner capacity of the acyclic triangle (Figure 2(b)) is $\log 3$ [9].

Proposition IV.1. *Let $C_{0\mathrm{u}}(\epsilon)$ denote the zero-undetected-error capacity of the acyclic triangle channel (Figure 2). Then*

$$\lim_{\epsilon \to 0} C_{0\mathrm{u}}(\epsilon) = \log 3. \tag{34}$$

We give an alternative proof based on the following lower bound [2], [3]. For any DMC W,

$$C_{0\mathrm{u}} \geq \max_{Q} \min_{\substack{V \ll W \\ QV = QW}} I(Q, V), \tag{35}$$

where the maximum is over all PMFs Q on $\mathcal{X}$, and where the minimum is over all auxiliary channels V such that $V(y|x) = 0$ whenever $W(y|x) = 0$ (in short $V \ll W$) and such that $QV = QW$. We use the following lemma; the proof is in the appendix.

Lemma IV.2. *Let $\mathcal{V}(Q, W)$ denote the set of auxiliary channels $V \ll W$ satisfying $QV = QW$. If $V^\star \in \mathcal{V}(Q, W)$ and there exist functions $A\colon \mathcal{X} \to (0, \infty)$ and $B\colon \mathcal{Y} \to (0, \infty)$ such that*

$$V^\star(y|x) = A(x)B(y), \quad \textit{whenever } Q(x)W(y|x) > 0, \tag{36}$$

then

$$I(Q, V^\star) \leq I(Q, V), \quad \textit{for all } V \in \mathcal{V}(Q, W), \tag{37}$$

with equality if, and only if, $V(y|x) = V^\star(y|x)$ for all $y \in \mathcal{Y}$ and $x \in \mathrm{supp}(Q)$.

It should be noted that if $V^\star = W$ and Q is capacity achieving in Lemma IV.2, then (35) implies that $C_{0\mathrm{u}} = C$. This was observed in [5].

Proof of Proposition IV.1. Let W denote the channel law of the acyclic triangle channel with $0 < \epsilon < 1$, and let U be the uniform PMF on $\{0, 1, 2\}$. Choose

$$\big(A(0), A(1), A(2)\big) = \left(\frac{2(1+\epsilon)}{3\epsilon^2}, \frac{1}{\epsilon}, 1\right) \tag{38}$$

and

$$\big(B(0), B(1), B(2)\big) = \left(\frac{3\epsilon^2}{2(1+\epsilon)}, \frac{\epsilon(2-\epsilon)}{2(1+\epsilon)}, 1-\epsilon\right). \tag{39}$$

It is straightforward to verify that

$$V^\star(y|x) = \begin{cases} A(x)B(y) & W(y|x) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{40}$$

defines a channel $V^\star \ll W$ that satisfies $UV^\star = UW$. From (35) and Lemma IV.2, it thus follows that

$$C_{0\mathrm{u}}(\epsilon) \geq I(U, V^\star), \quad 0 < \epsilon < 1. \tag{41}$$

But $V^\star$ approaches the identity matrix as ϵ tends to zero, so by the continuity of mutual information

$$\lim_{\epsilon \to 0} I(U, V^\star) = \log 3. \tag{42}$$

This completes the proof because clearly $C_{0\mathrm{u}}(\epsilon)$ is upper-bounded by $\log 3$ for every $0 < \epsilon < 1$. □

Appendix

Converse Proof of (8). We first recall the precise definition of $C_{0\mathrm{u}}$. Let R be a positive number and n a positive integer. An injective mapping

$$\mathbf{f}_n\colon \big\{1, \ldots, \lceil e^{nR} \rceil\big\} \to \mathcal{X}^n, \tag{43}$$

is called a rate-R blocklength-n encoder. The domain of $\mathbf{f}_n$ is called the message set. The number of messages compatible with a received sequence $\mathbf{y} \in \mathcal{Y}^n$ is

$$L_n(\mathbf{y}) = \big|\big\{m : W^n\big(\mathbf{y}|\mathbf{f}_n(m)\big) > 0\big\}\big|. \tag{44}$$

A zero-undetected-error decoder declares an erasure whenever $L_n(\mathbf{y}) > 1$; otherwise it produces the only compatible message. A rate R is said to be achievable if there exists a sequence $\{\mathbf{f}_n\}$ of rate-R blocklength-n encoders with maximum probability of erasure tending to zero:

$$\eta_n = \max_m \sum_{\mathbf{y}:L_n(\mathbf{y})>1} W^n\big(\mathbf{y}|\mathbf{f}_n(m)\big) \to 0, \quad n \to \infty. \tag{45}$$

We define $C_{0\mathrm{u}}$ as the supremum of all achievable rates. (The definition does not depend on whether we use an average or maximal probability of erasure criterion.)

To prove the converse, suppose $\{\mathbf{f}_n\}$ is a sequence of rate-R blocklength-n encoders with $\eta_n \to 0$ as $n \to \infty$. Let Q_n be the uniform PMF on the range of $\mathbf{f}_n$. Then $Q_n \in \mathcal{Q}_n$, and

$$Q_n\big(\mathcal{X}^n(\mathbf{y})\big) = \frac{L_n(\mathbf{y})}{\lceil e^{nR} \rceil}. \tag{46}$$

Consequently,

$$\begin{aligned}
&\sum_{\mathbf{y}\in\mathrm{supp}(Q_nW^n)} (Q_nW^n)(\mathbf{y}) \log \frac{1}{Q_n(\mathcal{X}^n(\mathbf{y}))} \\
&\quad = \log\lceil e^{nR} \rceil - \sum_{\mathbf{y}:L_n(\mathbf{y})>1} (Q_nW^n)(\mathbf{y}) \log L_n(\mathbf{y}) \\
&\quad \geq \log\lceil e^{nR} \rceil (1-\eta_n) \\
&\quad \geq nR(1-\eta_n),
\end{aligned} \tag{47}$$

where the first inequality follows because $L_n(\mathbf{y}) \leq \lceil e^{nR} \rceil$. Dividing by $n(1-\eta_n)$, taking the maximum over $\mathcal{Q}_n$, and letting $n \to \infty$ shows that R cannot exceed the RHS of (8). □

Proof of Lemma IV.2. Suppose $V^\star$ satisfies the hypothesis of the lemma and $V \in \mathcal{V}(Q,W)$. Since $QV^\star = QV = QW$, we have $I(Q,V^\star) = H(QW) - H(V^\star|Q)$ and $I(Q,V) = H(QW) - H(V|Q)$. Thus, it suffices to show that $H(V^\star|Q) \geq H(V|Q)$. To this end, observe that

$$\begin{aligned}
H(V^\star|Q) = &\sum_{x,y:Q(x)W(y|x)>0} Q(x)V^\star(y|x) \log \frac{1}{A(x)} \\
&+ \sum_{x,y:Q(x)W(y|x)>0} Q(x)V^\star(y|x) \log \frac{1}{B(y)}.
\end{aligned}$$

The first sum on the RHS is the expectation of $-\log A(X)$ when X has PMF Q, and the second sum is the expectation of $-\log B(Y)$ when Y has PMF $QV^\star$. Since $QV^\star = QV$ and $V \ll W$, we may replace $V^\star$ with V in both sums to obtain

$$\begin{aligned}
H(V^\star|Q) &= \sum_{x,y:Q(x)W(y|x)>0} Q(x)V(y|x) \log \frac{1}{V^\star(y|x)} \\
&= \sum_{x,y:Q(x)V(y|x)>0} Q(x)V(y|x) \log \frac{V(y|x)}{V^\star(y|x)} \\
&\quad + H(V|Q) \\
&\geq H(V|Q),
\end{aligned} \tag{48}$$

where we used Jensen's inequality in the last line. Equality holds in (48) if, and only if, $V^\star(y|x) = V(y|x)$ for all $y \in \mathcal{Y}$ and $x \in \mathrm{supp}(Q)$. □

References

[1] 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.

[2] I. Telatar and R. Gallager, "New exponential upper bounds to error and erasure probabilities," in *Information Theory, 1994. Proceedings., 1994 IEEE International Symposium on*. IEEE, 1994, p. 379.

[3] 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.

[4] M. Pinsker and A. Sheverdyaev, "Transmission capacity with zero error and erasure," *Problemy Peredachi Informatsii*, vol. 6, no. 1, pp. 20–24, 1970.

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

[6] V. Balakirsky, "A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1889–1902, 1995.

[7] C. Bunte and A. Lapidoth, "The zero-undetected-error capacity of discrete memoryless channels with feedback," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*. IEEE, 2012, pp. 1838–1842.

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

[9] A. Calderbank, P. Frankl, R. Graham, W. Li, and L. Shepp, "The Sperner capacity of linear and nonlinear codes for the cyclic triangle," *Journal of Algebraic Combinatorics*, vol. 2, no. 1, pp. 31–48, 1993.

[10] A. Blokhuis, "On the Sperner capacity of the cyclic triangle," *Journal of Algebraic Combinatorics*, vol. 2, no. 2, pp. 123–124, 1993.

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