The Zero-Undetected-Error Capacity Approaches the Sperner Capacity

Christoph Bunte, Amos Lapidoth, Alex Samorodnitsky

September 16, 2013

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.

1 Introduction

A *zero-undetected-error decoder* (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 *zero-undetected-error capacity* (z.u.e. capacity) $C_{0\text{-u}}$ 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].[1] Restricting the decoding rule cannot help, so $C_{0\text{-u}}$ never exceeds the Shannon capacity C.

Although partial results exist (see below), the z.u.e. capacity of general discrete memoryless channels (DMCs) is still unknown. 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 X is a subset of their output alphabet Y and whose transition law W satisfies

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

Here and throughout we assume that $0 \leq \varepsilon < 1$. For ε-noise channels we shall derive an upper bound on $C_{0\text{-u}}$. We shall then apply this result to study the limit of $C_{0\text{-u}}$ 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 3. 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 $W(y|x)$, $x \in X$, $y \in Y$, where X and Y are finite input and output alphabets. Feeding a sequence of input symbols $\mathbf{x} = (x^{(1)}, \ldots, x^{(n)})$ to a

[1] It does not matter whether we define $C_{0\text{-u}}$ using an average or a maximal erasure probability criterion.

DMC of transition law W produces a random sequence of output symbols $\mathbf{Y} = (Y^{(1)}, \ldots, Y^{(n)})$ whose distribution (PMF) is

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

The *support of* W is the set of all pairs $(x, y) \in X \times Y$ for which $W(y|x)$ is positive; it is denoted by $\mathfrak{S}(W)$. Similarly, if P is a PMF on X, then $\mathfrak{S}(P)$ denotes the set of all $x \in X$ for which $P(x)$ is positive. We write PW for the PMF on Y induced by P and the channel W

$$(PW)(y) = \sum_{x \in X} P(x)W(y|x), \quad y \in Y. \tag{3}$$

If $A \subseteq X$, then we write $P(A)$ in lieu of $\sum_{x \in A} P(x)$. The Cartesian product of two set A and B is denoted by $A \times B$. The n-fold Cartesian product of A with itself is denoted by A^n, and the cardinality of A is denoted by $|A|$. 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 $X \subseteq Y$ and $W(x|x) > 0$ for all $x \in X$ as a collection of codewords $\mathbf{x}_1, \ldots, \mathbf{x}_M$ with the property

$$W^n(\mathbf{x}_m|\mathbf{x}_{m'}) = 0 \quad \text{whenever } m \neq m'. \tag{4}$$

The rate of the code is $n^{-1} \log M$, and the largest such rate 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 1.1. *For every ε-noise channel,*

$$C_{\text{0-u}} \le \log\bigl(e^{C_{\text{Sp}}} + \varepsilon|X|(|Y| - 1)\bigr). \tag{5}$$

Combining Theorem 1.1 with [2, Theorem 2] proves the following corollary, which was conjectured in [2].

Corollary 1.2. *For ε-noise channels,*

$$\lim_{\varepsilon \to 0} C_{\text{0-u}} = C_{\text{Sp}}, \tag{6}$$

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

The proof of Theorem 1.1 is in Section 4. Before providing an outline of this proof, we try to explain why Corollary 1.2 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 lower bound to $C_{\text{0-u}}$ 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 below $C_{\text{0-u}}$ the probability of erasure can be driven to zero exponentially fast (see below), this suggests that C_{Sp} should

be an upper bound on $C_{0\text{-u}}$ for small ε (ignoring the issue that the exponential decay of the erasure probability may become arbitrarily slow as ε becomes small).

As to the outline of the proof of Theorem 1.1, we will 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 2). We then upper-bound Forney's 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 3, 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.

To put our result into perspective, we briefly review some of the literature on the z.u.e. capacity and related concepts. Forney derived the lower bound [3]

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

where the maximum is over all PMFs on the input alphabet X, and where $X(y)$ denotes the set of all $x \in X$ for which $W(y|x)$ is positive. It can be proved using standard random coding where each component of each codeword is drawn IID from a PMF P. Forney's bound is not always tight.[2] A tighter lower bound was derived in [4, 2, 1] using random coding over constant composition codes

$$C_{0\text{-u}} \geq \max_{P} \min_{\substack{V \ll W: \\ PV = PW}} I(P, V). \tag{8}$$

The minimum in (8) is over all auxiliary DMCs $V(y|x), x \in X, y \in Y$ such that $V(y|x) = 0$ whenever $W(y|x) = 0$ (in short $V \ll W$) and such that V induces the same output distribution under P as the true channel W.

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 the bounds (7) and (8) can be improved by applying them to W^n and normalizing the result by $1/n$. For example, the n-letter version of (7) is

$$C_{0\text{-u}} \geq \frac{1}{n} \max_{P} \sum_{\mathbf{y} \in Y^n} (PW^n)(\mathbf{y}) \log \frac{1}{P(X^n(\mathbf{y}))}, \tag{9}$$

where the maximum is over all PMFs P on X^n, and where $X^n(\mathbf{y})$ denotes the set of all $\mathbf{x} \in X^n$ for which $W^n(\mathbf{y}|\mathbf{x}) > 0$. A numerical evaluation in [2] of the single-letter and two-letter versions of (8) for a particular channel suggests that a strict improvement is possible and hence that (8) is not always tight.

The n-letter version of (8) becomes tight as n tends to infinity [1, 2]. Since the proof of (8) shows that for all rates less than the RHS of (8) the probability of erasure can be driven to zero exponentially fast, it follows that this is also true for all rates less than the n-letter version of the bound, and hence for all rates less than $C_{0\text{-u}}$.

[2] An example where it is not tight is the Z-channel [4].

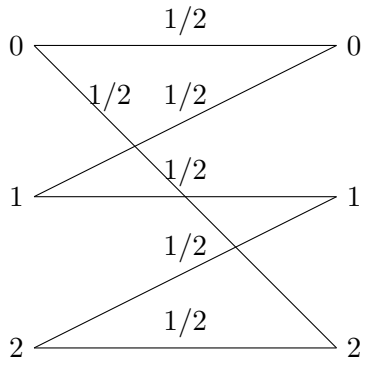

Figure 1: The graph contains a cycle but the channel law factorizes.

In Section 2 we prove that also the n-letter version of the weaker bound (7) is asymptotically tight, and that this is true even when the maximization is restricted to PMFs that are uniform over their support. This result will be crucial in the proof of Theorem 1.1.

We have already pointed out that the Shannon capacity C is an upper bound to $C_{0\text{-u}}$. Since computing C is easy and computing $C_{0\text{-u}}$ seems hard, it is desirable to know when equality holds. Pinsker and Sheverdyaev proved that $C_{0\text{-u}} = C$ whenever the *bipartite channel graph* is acyclic [5]. The bipartite channel graph is the undirected bipartite graph whose two independent sets of vertices are the input and output alphabets of the channel, and where there is an edge between an input x and an output y if $W(y|x) > 0$ (it is customary to draw the inputs on the left and the outputs on the right). Acyclic means that we cannot find an integer $\ell \geq 2$, distinct inputs $x_1, \ldots, x_\ell$ and distinct outputs $y_1, \ldots, y_\ell$ such that $W(y_j|x_j) > 0$ and $W(y_j|x_{j+1}) > 0$ for all $j \in \{1, \ldots, \ell\}$ where $x_{\ell+1} = x_1$. Important examples of channels with acyclic bipartite channel graphs are the binary erasure channel and the Z-channel.

Csiszár and Narayan proved that $C_{0\text{-u}} = C$ whenever there exist positive functions A and B such that

$$W(y|x) = A(x)B(y), \quad \text{whenever } W(y|x) > 0. \tag{10}$$

They also proved that in this case (8) is tight [1]. Their result is stronger than Pinsker-Sheverdyaev because every DMC with an acyclic bipartite channel graph possesses a factorization of its channel law in the sense of (10), but the converse is not true. For example, the graph shown in Figure 1 contains a cycle, yet the channel law factorizes with the choice $A(x) = 1$ for all x and $B(y) = 1/2$ for all y.

It is conjectured in [1] that a necessary condition for $C_{0\text{-u}}$ to equal C is for (10) to hold on some capacity-achieving subset of inputs (which is clearly also sufficient).

The fact that $C_{0\text{-u}} = C$ for the Z-channel can be used to characterize channels with positive z.u.e. capacity. Indeed, suppose there exist two input symbols x and x' and an output symbol y such that $W(y|x) > 0$ and $W(y|x') = 0$. By combining all output symbols other than y into a single distinct output symbol, and by using only the inputs x and x', we can reduce the channel to a Z-channel with crossover probability $1 - W(y|x)$. For this channel $C > 0$, and hence also $C_{0\text{-u}} > 0$. Conversely, if $W(y|x) > 0$ implies that $W(y|x') > 0$ for all $x' \in X$, then any received sequence of output symbols can be produced by every codeword and the decoder must always erase, so $C_{0\text{-u}}$ must be zero.

We also mention that z.u.e. capacity is a special case of d-capacity [1], and the lower bound (8) is a special case of a lower bound on d-capacity [6, 7, 1]. It was proved in [8] that

this lower bound on d-capacity is tight for binary-input channels. The proof is complicated and is, in fact, not needed when one is interested only in z.u.e. capacity. Indeed, the binary-input case is easily solved using the Pinsker-Sheverdyaev result and the following proposition, which can also be used to improve on (5) when $|Y|$ is larger than $|X|+2^{|X|}-1$ (see Section 5).

Proposition 1.3. *Suppose that the output symbols $y, y' \in Y$ are such that for every $x \in X$, $W(y|x) > 0$ if, and only if, $W(y'|x) > 0$. Then the z.u.e. capacity is unaltered when we combine y and y' into a single output symbol distinct from all other output symbols.*

Proof. The set of messages that cannot be ruled out when $\mathbf{y}$ is observed at the output is unchanged when any occurrence of y in $\mathbf{y}$ is replaced with y' or vice versa. □

Using Proposition 1.3 we can reduce the output alphabet of a given DMC W with input alphabet X to at most $2^{|X|}-1$ symbols without changing the z.u.e. capacity (there are $2^{|X|}-1$ possible combinations of inputs that can connect to a given output). In particular, when the input alphabet is binary, we can reduce the channel to an asymmetric binary erasure channel (possibly with some transition probabilities equal to zero). Since this channel has an acyclic bipartite channel graph, computing the z.u.e. capacity of a binary-input channel can thus be reduced to computing the Shannon capacity of an asymmetric binary erasure channel.

Unlike the Shannon capacity, the z.u.e. capacity can be increased by feedback [9]. In fact, with feedback the z.u.e. capacity $C_{\text{0-u,fb}}$ is [10, 9]

$$C_{\text{0-u,fb}} = \begin{cases} C & \text{if } C_{\text{0-u}} > 0, \\ 0 & \text{if } C_{\text{0-u}} = 0. \end{cases} \tag{11}$$

A concept closely related to the z.u.e. capacity is the *listsize capacity* [2, 11]: Suppose that instead of erasing when more than one message could have produced the observed output, the decoder instead produces a list of all messages that it cannot rule out. The listsize capacity is the supremum of all rates that are achievable in the sense that the ρ-th moment of the size of the list produced by the decoder tends to one as the blocklength tends to infinity; it is denoted by $C_{\text{a-l}}(\rho)$. Here ρ can be any positive number; the case $\rho = 1$ has been called *average-listsize capacity* [2]. It follows from the definition that $C_{\text{a-l}}(\rho)$ never exceeds $C_{\text{0-u}}$, and it is not difficult to show that $C_{\text{a-l}} > 0$ if, and only if, $C_{\text{0-u}} > 0$ [11]. Lower bounds on $C_{\text{a-l}}(\rho)$ analogous to (7) and (8) were derived in [3, 2, 11]. Results for channels with feedback can be found in [12], where, in particular, it is shown that feedback can increase the listsize capacity.

It was proved in [2] that Corollary 1.2 is true when $C_{\text{0-u}}$ is replaced with $C_{\text{a-l}}(1)$

$$\lim_{\epsilon \to 0} C_{\text{a-l}}(\rho)\Big|_{\rho=1} = C_{\text{Sp}}. \tag{12}$$

This, in fact, holds for all $\rho > 0$. Indeed, Theorem 1.1 and the fact that $C_{\text{a-l}}(\rho)$ is upper-bounded by $C_{\text{0-u}}$ for all $\rho > 0$ imply that

$$\lim_{\epsilon \to 0} C_{\text{a-l}}(\rho) \leq C_{\text{Sp}}. \tag{13}$$

The reverse inequality for $0 < \rho < 1$ follows from (12) and the fact that $C_{\text{a-l}}(\rho)$ is nonincreasing in ρ. The case $\rho \geq 1$ is not difficult to obtain from a generalization to all $\rho > 0$ of Forney's lower bound on $C_{\text{a-l}}(1)$ [13].

It is shown in [12] that if the channel law factorizes in the sense of (10), then

$$C_{\text{a-l}}(\rho) = \max_P \frac{E_0(\rho, P)}{\rho}, \tag{14}$$

where $E_0(\rho, P)$ is Gallager's function (see [14]). The RHS of (14) is also known as the cutoff rate [15]. In fact, the listsize capacity relates to the cutoff rate in much the same way that the z.u.e. capacity relates to the Shannon capacity; see [12, 13] for details.

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

In this section we show that (9) is asymptotically tight even when the input PMFs are restricted to be uniform over their support.

Theorem 2.1. *For any DMC,*

$$C_{0\text{-u}} = \lim_{n\to\infty} \frac{1}{n} \max_{P\in U_n} \sum_{\mathbf{y}\in Y^n} (PW^n)(\mathbf{y}) \log \frac{1}{P(X^n(\mathbf{y}))}, \tag{15}$$

where U_n denotes the collection of PMFs on 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 (15) without the $1/n$ factor is superadditive, which implies that the limit is equal to the supremum.[3] Let us denote this limit by λ. Achievability, i.e., $C_{0\text{-u}} \geq \lambda$, follows because (9) 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 $\delta \in (0, 1)$

$$\max_{1\leq m\leq M} \sum_{\substack{\mathbf{y}\in Y^n:\\ M(\mathbf{y})>1}} W^n(\mathbf{y}|\mathbf{x}_m) < \delta, \tag{16}$$

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

$$M(\mathbf{y}) = \big|\big\{1 \leq m \leq M : W^n(\mathbf{y}|\mathbf{x}_m) > 0\big\}\big|. \tag{17}$$

Condition (16) 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

$$\sum_{\substack{\mathbf{y}\in Y^n:\\ M(\mathbf{y})>1}} W^n(\mathbf{y}|\mathbf{x}_m) = 1. \tag{18}$$

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

$$P\big(X^n(\mathbf{y})\big) = \frac{M(\mathbf{y})}{M}, \quad \text{for all } \mathbf{y} \in Y^n. \tag{19}$$

[3] 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 $\sup_n a_n/n$ [16, Problem 98].

We further observe that

$$\lambda \geq \frac{1}{n} \sum_{\mathbf{y} \in Y^n} (PW^n)(\mathbf{y}) \log \frac{1}{P(X^n(\mathbf{y}))} \tag{20}$$

$$= R - \frac{1}{n} \sum_{\substack{\mathbf{y} \in Y^n: \\ M(\mathbf{y})>1}} (PW^n)(\mathbf{y}) \log M(\mathbf{y}) \tag{21}$$

$$\geq R\left(1 - \sum_{\substack{\mathbf{y} \in Y^n: \\ M(\mathbf{y})>1}} (PW^n)(\mathbf{y})\right) \tag{22}$$

$$= R\left(1 - \frac{1}{M} \sum_{m=1}^{M} \sum_{\substack{\mathbf{y} \in Y^n: \\ M(\mathbf{y})>1}} W^n(\mathbf{y}|\mathbf{x}_m)\right) \tag{23}$$

$$\geq R(1-\delta), \tag{24}$$

where (20) follows because λ is the supremum of a sequence whose n-th term is no smaller than the RHS of (20); where (21) follows from (19) and the fact that $\log 1 = 0$; where (22) follows because $M(\mathbf{y}) \leq M$; where (23) follows from the choice of P; and where (24) follows from (16). Thus, for any sequence of blocklength-n rate-R codebooks with maximal probability of erasure approaching zero,

$$R \leq \lambda. \tag{25}$$

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. □

3 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[4]

$$\Sigma(G) = \lim_{n \to \infty} \frac{1}{n} \log \omega(G^n), \tag{26}$$

[4]Many authors prefer to define Sperner capacity in terms of *cliques* instead of independent sets (see, e.g., [17]). Which of the two definition we choose does not matter as long as we keep in mind that an independent set in G is a clique in the "complement of G" and vice versa, where "complement" means that the vertex set is the same and the edge set is the set complement of $E(G)$ in $V(G) \times V(G)$ minus the self-loops.

where the limit on the RHS is equal to the supremum because the sequence $\omega(G^1), \omega(G^2), \ldots$ is supermultiplicative.[5]

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 $j \in \{1, \ldots, n-1\}$. 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 1.1; their proofs are in the Appendix. The first is that ω can be replaced with ρ in (26).

Theorem 3.1. *For every graph* G*,*

$$\Sigma(G) = \lim_{n\to\infty} \frac{1}{n} \log \rho(G^n), \tag{27}$$

and the limit is equal to the supremum.

In particular, Theorem 3.1 asserts that

$$\rho(G^n) \leq e^{n\Sigma(G)}, \quad \text{for all } n. \tag{28}$$

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

$$\mathrm{d}_{\mathrm{in}}(x, G) = \big|\big\{x' \in V(G) : (x', x) \in E(G)\big\}\big|. \tag{29}$$

The next result is a slight generalization of [18, p. 95, Theorem 1].

Theorem 3.2. *For every graph* G*,*

$$\sum_{x\in V(G)} \frac{1}{1 + \mathrm{d}_{\mathrm{in}}(x, G)} \leq \rho(G). \tag{30}$$

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

$$C_{\mathrm{Sp}}(W) = \Sigma\big(G(W)\big). \tag{31}$$

We note the identity

$$G(W^n) = G(W)^n. \tag{32}$$

4 Proof of Theorem 1.1

Applying Jensen's Inequality to the RHS of (15) yields

$$C_{\text{0-u}} \leq \sup_{n\geq 1} \frac{1}{n} \max_{P\in U_n} \log \sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(X^n(\mathbf{y}))}. \tag{33}$$

[5] 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.

Thus, it suffices to show that

$$\sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(X^n(\mathbf{y}))} \le \left(e^{C_{\mathrm{Sp}}} + \varepsilon|X|(|Y|-1)\right)^n, \quad \text{for all } P \in U_n. \tag{34}$$

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

$$(PW^n)(\mathbf{y}) = \sum_{\substack{\mathbf{z}\in Y^n:\\ \mathbf{y}+\mathbf{z}\in X^n}} P(\mathbf{y}+\mathbf{z})W^n(\mathbf{y}|\mathbf{y}+\mathbf{z}), \quad \text{for all } \mathbf{y} \in Y^n, \tag{35}$$

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

$$W^n(\mathbf{y}|\mathbf{y}+\mathbf{z}) \le \varepsilon^{|\mathbf{z}|}, \quad \text{if } \mathbf{y}+\mathbf{z} \in X^n, \tag{36}$$

where $|\mathbf{z}|$ denotes the number of nonzero components of $\mathbf{z}$. Thus, starting with the LHS of (34),

$$\sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(X^n(\mathbf{y}))} = \sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \sum_{\substack{\mathbf{z}\in Y^n:\\ \mathbf{y}+\mathbf{z}\in X^n}} \frac{P(\mathbf{y}+\mathbf{z})W^n(\mathbf{y}|\mathbf{y}+\mathbf{z})}{P(X^n(\mathbf{y}))} \tag{37}$$

$$= \sum_{\mathbf{z}\in Y^n} \sum_{\substack{\mathbf{y}\in Y^n:\\ \mathbf{y}+\mathbf{z}\in X^n\\ P(\mathbf{y}+\mathbf{z})>0\\ W^n(\mathbf{y}|\mathbf{y}+\mathbf{z})>0}} \frac{P(\mathbf{y}+\mathbf{z})W^n(\mathbf{y}|\mathbf{y}+\mathbf{z})}{P(X^n(\mathbf{y}))} \tag{38}$$

$$\le \sum_{\mathbf{z}\in Y^n} \varepsilon^{|\mathbf{z}|} \sum_{\substack{\mathbf{y}\in Y^n:\\ \mathbf{y}+\mathbf{z}\in X^n\\ P(\mathbf{y}+\mathbf{z})>0\\ W^n(\mathbf{y}|\mathbf{y}+\mathbf{z})>0}} \frac{P(\mathbf{y}+\mathbf{z})}{P(X^n(\mathbf{y}))} \tag{39}$$

$$= \sum_{\mathbf{z}\in Y^n} \varepsilon^{|\mathbf{z}|} \sum_{\substack{\mathbf{y}\in X^n:\\ P(\mathbf{y})>0\\ W^n(\mathbf{y}-\mathbf{z}|\mathbf{y})>0}} \frac{1}{|\{\mathbf{x}\in X^n : P(\mathbf{x})W^n(\mathbf{y}-\mathbf{z}|\mathbf{x})>0\}|}, \tag{40}$$

where (37) follows from (35); where (38) follows by changing the order of summation and dropping terms that are zero; where (39) follows from (36); and where (40) follows by substituting $\mathbf{y}$ for $\mathbf{y}+\mathbf{z}$ and because P is uniform over its support. For every $\mathbf{z} \in Y^n$, let $P_{\mathbf{z}}$ be any PMF on X^n of support

$$\mathfrak{S}(P_{\mathbf{z}}) = \{\mathbf{x}\in X^n : P(\mathbf{x})W^n(\mathbf{x}-\mathbf{z}|\mathbf{x})>0\}. \tag{41}$$

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

$$W_{\mathbf{z}}(\mathbf{y}|\mathbf{x}) = W^n(\mathbf{y}-\mathbf{z}|\mathbf{x}), \tag{42}$$

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

$$|\{\mathbf{x}\in X^n : P(\mathbf{x})W^n(\mathbf{y}-\mathbf{z}|\mathbf{x})>0\}| \ge |\{\mathbf{x}\in\mathfrak{S}(P_{\mathbf{z}}) : W_{\mathbf{z}}(\mathbf{y}|\mathbf{x})>0\}|. \tag{43}$$

Using (43) we can upper-bound the inner sum on the RHS of (40) by

$$\sum_{\mathbf{y}\in\mathfrak{S}(P_{\mathbf{z}})} \frac{1}{|\{\mathbf{x}\in\mathfrak{S}(P_{\mathbf{z}}) : W_{\mathbf{z}}(\mathbf{y}|\mathbf{x})>0\}|}. \tag{44}$$

This sum can also be written as

$$\sum_{\mathbf{y}\in V(G(W_{\mathbf{z}}))} \frac{1}{1+\mathrm{d_{in}}(\mathbf{y},G(W_{\mathbf{z}}))}, \tag{45}$$

where $G(W_{\mathbf{z}})$ is the graph associated with the channel $W_{\mathbf{z}}$ (see Section 3). Since (45) is upper-bounded by $\rho(G(W_{\mathbf{z}}))$ on account of Theorem 3.2, we thus have

$$\sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(X^n(\mathbf{y}))} \le \sum_{\mathbf{z}\in Y^n} \varepsilon^{|\mathbf{z}|}\rho(G(W_{\mathbf{z}})). \tag{46}$$

We next argue that

$$\rho\big(G(W_{\mathbf{z}})\big) \le |X|^{|\mathbf{z}|}\rho\big(G(W)^{n-|\mathbf{z}|}\big). \tag{47}$$

To this end, let $\mathbf{x}(\mathbf{z})$ denote the restriction of $\mathbf{x}\in X^n$ 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 (47) by contradiction. In order to reach a contradiction, assume that for some integer η strictly larger than the RHS of (47) 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 $|X|^{|\mathbf{z}|}$ such classes, one of them must contain $\kappa > \rho(G(W)^{n-|\mathbf{z}|})$ 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

$$W_{\mathbf{z}}(\mathbf{y}|\mathbf{x})>0 \iff W^{n-|\mathbf{z}|}\big(\mathbf{y}(\mathbf{z}^{\mathrm{c}})\big|\mathbf{x}(\mathbf{z}^{\mathrm{c}})\big)>0. \tag{48}$$

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 $G(W^{n-|\mathbf{z}|})$ induced by $\mathbf{x}'_1(\mathbf{z}^{\mathrm{c}}),\ldots,\mathbf{x}'_\kappa(\mathbf{z}^{\mathrm{c}})$.[6] And since the former is acyclic, so must be latter, which is a contradiction because $G(W^{n-|\mathbf{z}|})=G(W)^{n-|\mathbf{z}|}$ and $\kappa>\rho(G(W)^{n-|\mathbf{z}|})$.

Having established (47), we further note that by (28) and (31),

$$\rho\big(G(W)^{n-|\mathbf{z}|}\big) \le e^{(n-|\mathbf{z}|)C_{\mathrm{Sp}}}. \tag{49}$$

Combining (46), (47), and (49), and noting that any dependence on $\mathbf{z}$ is only via $|\mathbf{z}|$ and that there are $\binom{n}{k}(|Y|-1)^k$ elements in Y^n with k of the n components equal to zero, we obtain

$$\sum_{\mathbf{y}\in\mathfrak{S}(PW^n)} \frac{(PW^n)(\mathbf{y})}{P(X^n(\mathbf{y}))} \le \sum_{k=0}^{n}\binom{n}{k}(|Y|-1)^k\varepsilon^k|X|^k e^{(n-k)C_{\mathrm{Sp}}}. \tag{50}$$

This completes the proof because the RHS is equal to the RHS of (34). □

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

5 Concluding Remarks

1. One may take issue with the fact that the upper bound in Theorem 1.1 is an unbounded function of $|Y|$. This "flaw" may be fixed by replacing $|Y|$ with $|X| + 2^{|X|} - 1$. Indeed, using Proposition 1.3 and noting that the ε-noise property and C_{Sp} are preserved if we combine only output symbols in $Y \setminus X$, we can reduce the output alphabet to at most $|X| + 2^{|X|} - 1$ symbols.

2. The Sperner capacity of a graph and the Sperner capacity of a channel are of course just different formulations of the same problem. Indeed, in Section 3 we noted that $C_{\text{Sp}}(W) = \Sigma(G(W))$, where $G(W)$ is the graph associated with the ε-noise channel W. Conversely, we may associate with each directed graph G a canonical ϵ-noise channel $W_\epsilon(G)$ by choosing X and Y equal to $V(G)$ and

$$W_\epsilon(G)(y|x) = \begin{cases} 1-\epsilon & y = x, \\ \frac{\epsilon}{\mathrm{d}_{\text{out}}(x,G)} & (x,y) \in E(G), \quad \text{if } \mathrm{d}_{\text{out}}(x,G) \geq 1, \\ 0, & \text{otherwise}, \end{cases} \tag{51}$$

   and

$$W_\epsilon(G)(y|x) = \begin{cases} 1, & \text{if } y = x \\ 0, & \text{if } y \neq x \end{cases} \quad \text{if } \mathrm{d}_{\text{out}}(x,G) = 0, \tag{52}$$

   where $\mathrm{d}_{\text{out}}(x,G)$ denotes the *out-degree* of x in G, i.e., the number of edges of G emanating from the vertex x. Then

$$\Sigma(G) = C_{\text{Sp}}(W_\epsilon(G)). \tag{53}$$

3. We mentioned in the introduction that a proof of $\underline{\lim}_{\epsilon\to 0} C_{\text{0-u}} \geq C_{\text{Sp}}$ was given in [2]. Here, we offer a simple proof based on the n-letter version of Forney's lower bound (9). Given $\delta > 0$ choose n so that there exists a Sperner code of size $e^{n(C_{\text{Sp}}-\delta)}$. Let P be the uniform PMF on the codebook and note that

$$(PW^n)(\mathbf{x}) \geq P(\mathbf{x})(1-\epsilon)^n \quad \text{for all } \mathbf{x} \in X^n. \tag{54}$$

   Then, for this P,

$$\begin{aligned} C_{\text{0-u}} &\geq \frac{1}{n} \sum_{\mathbf{y}\in Y^n} (PW^n)(\mathbf{y}) \log \frac{1}{P(X^n(\mathbf{y}))} && (55)\\ &\geq \frac{1}{n} \sum_{m=1}^{e^{n(C_{\text{Sp}}-\delta)}} (PW^n)(\mathbf{x}_m) \log \frac{1}{P(X^n(\mathbf{x}_m))} && (56)\\ &= \frac{1}{n} \sum_{m=1}^{e^{n(C_{\text{Sp}}-\delta)}} (PW^n)(\mathbf{x}_m) \log \frac{1}{P(\mathbf{x}_m)} && (57)\\ &\geq (1-\epsilon)^n \frac{1}{n} \sum_{m=1}^{e^{n(C_{\text{Sp}}-\delta)}} P(\mathbf{x}_m) \log \frac{1}{P(\mathbf{x}_m)} && (58)\\ &= (1-\epsilon)^n (C_{\text{Sp}} - \delta), && (59) \end{aligned}$$

where (55) follows from (9); where (56) follows because the summand is nonnegative; where (57) follows from the property of Sperner codes (4); and where (58) follows from (54). Letting first $\epsilon \to 0$ and then $\delta \to 0$ completes the proof.

4. For some channels the bound in Theorem 1.1 can be sharpened. For example, for the canonical ε-noise channel associated with the cyclic orientation of a triangle $C_{\text{Sp}} = \log 2$ [19, 20], and it is shown in [21] that $C_{\text{0-u}} \leq \log 2$ for every $\epsilon \in (0, 1)$.

A Proof of Theorem 3.1

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., [22, Section 5.7]).[7]

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 (27) 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

$$\begin{aligned} \rho\big(G^{n+m}\big) &\geq |A_n \times A_m| \\ &= \rho(G^n)\rho(G^m). \end{aligned} \tag{60}$$

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.

As to (27), we first show that

$$\Sigma(G) = \log|V(G)|, \quad \text{for all acyclic } G. \tag{61}$$

Note that this will prove Theorem 3.1 in the special case where G is acyclic. Indeed, in this case $\rho(G) = |V(G)|$, so (60) implies $\rho(G^n) \geq |V(G)|^n$. And since clearly $\rho(G^n) \leq |V(G)|^n$, we thus have

$$\rho(G^n) = |V(G)|^n, \quad \text{for all acyclic } G. \tag{62}$$

To prove (61), 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 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 $(y^{(k)}, x^{(k)}) \notin E(G)$, which implies that there is no edge from $\mathbf{x}$ to $\mathbf{y}$ and no edge from $\mathbf{y}$ to $\mathbf{x}$ in G^n.

[7] 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.

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

$$\frac{|V(G)|^n}{n|V(G)| - n + 1} \tag{63}$$

members. Thus,

$$\frac{1}{n}\log\omega(G^n) \geq \log|V(G)| - \frac{1}{n}\log\big(n|V(G)| - n + 1\big), \tag{64}$$

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

To complete the proof of (27), let G be any graph (not necessarily acyclic) and let λ denote the limit of $\frac{1}{n}\log\rho(G^n)$ as n tends to infinity (i.e., the supremum). For a given $\delta > 0$ select ν so that

$$\frac{1}{\nu}\log\rho(G^\nu) \geq \lambda - \delta. \tag{65}$$

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,

$$\frac{1}{\nu m}\log\omega(G^{\nu m}) \geq \frac{1}{\nu m}\log\omega(H^m). \tag{66}$$

Letting m tend to infinity, we obtain

$$\Sigma(G) \geq \frac{1}{\nu}\Sigma(H). \tag{67}$$

Since H is acyclic, we can substitute it for G in (61) to obtain

$$\frac{1}{\nu}\Sigma(H) = \frac{1}{\nu}\log|A| \tag{68}$$

$$= \frac{1}{\nu}\log\rho\big(G^\nu\big), \tag{69}$$

where (69) follows because A achieves $\rho(G^\nu)$. Combining (67), (69), and (65) shows that $\Sigma(G) \geq \lambda - \delta$. Since this is true for every $\delta > 0$,

$$\Sigma(G) \geq \lambda. \tag{70}$$

On the other hand, a graph with no edges is trivially acyclic, so $\omega(G^n) \leq \rho(G^n)$ and hence $\Sigma(G) \leq \lambda$. □

B Proof of Theorem 3.2

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 because 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,

$$|A| \leq \rho(G). \tag{71}$$

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

$$\Pr(x \in A) = \frac{1}{1 + \mathrm{d}_{\mathrm{in}}(x, G)}, \quad \text{for all } x \in V(G). \tag{72}$$

Indeed, x is in A if, and only if, it is greater than all the vertices x' for which $(x', x) \in E(G)$. Since there are $\mathrm{d}_{\mathrm{in}}(x, G)$ such vertices, x must therefore be the greatest among a total of $1 + \mathrm{d}_{\mathrm{in}}(x, G)$ vertices to be in A. And since each of these vertices has the same probability of being the greatest (because $<$ is drawn uniformly at random), (72) follows.

Summing both sides of (72) over all vertices of G yields

$$\sum_{x \in V(G)} \frac{1}{1 + \mathrm{d}_{\mathrm{in}}(x, G)} = \sum_{x \in V(G)} \Pr(x \in A). \tag{73}$$

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 is the expected cardinality of A. This expected cardinality cannot exceed $\rho(G)$ because (71) holds for every outcome 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] 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.

[4] I. E. Telatar, "Multi-access communications with decision feedback decoding," Ph.D. dissertation, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, May 1992.

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

[6] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, Massachusetts Institute of Technology, Dept. of Electrical Engineering and Computer Science, November 1983.

[7] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, 1981.

[8] 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.

[9] 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.

[10] B. Nakiboğlu and L. Zheng, "Errors-and-erasures decoding for block codes with feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 24–49, 2012.

[11] I. E. Telatar, "Zero-error list capacities of discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1977–1982, 1997.

[12] C. Bunte and A. Lapidoth, "On the cutoff rate and the average-listsize capacity of discrete memoryless channels with feedback," in *Information Theory Workshop (ITW), 2013 IEEE*, 2013, pp. ?–?

[13] ——, manuscript in preparation.

[14] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, 1968.

[15] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 99–105, 1996.

[16] G. Pólya and G. Szegő, *Problems and Theorems in Analysis I*, ser. Classics in Mathematics. Berlin Heidelberg: Springer Berlin Heidelberg, 1978.

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

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

[19] 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.

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

[21] 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. ?–?

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