Binomial Channel: On the Capacity-Achieving Distribution and Bounds on the Capacity

Luca Barletta†, Ian Zieder*, Antonino Favano†, and Alex Dytso**
† Politecnico di Milano, Milano, 20133, Italy. Email: {antonino.favano, luca.barletta}@polimi.it
* New Jersey Institute of Technology, Newark, NJ 07102, USA. Email: ihz3@njit.edu
** Qualcomm Flarion Technologies, Bridgewater, NJ 08807, USA. Email: odytso2@gmail.com

***Abstract*—This work considers a binomial noise channel. The paper can be roughly divided into two parts. The first part is concerned with the properties of the capacity-achieving distribution. In particular, for the binomial channel, it is not known if the capacity-achieving distribution is unique since the output space is finite (i.e., supported on integers $0, \dots, n$) and the input space is infinite (i.e., supported on the interval $[0, 1]$), and there are multiple distributions that induce the same output distribution. This paper shows that the capacity-achieving input distribution is unique by appealing to the total positivity property of the binomial kernel. In addition, we provide upper and lower bounds on the cardinality of the support of the capacity-achieving distribution. Specifically, an upper bound of order $\frac{n}{2}$ is shown, which improves on the previous upper bound of order n due to Witsenhausen. Moreover, a lower bound of order $\sqrt{n}$ is shown. Finally, additional results about the locations and probability values of the support points are established.**

The second part of the paper focuses on deriving upper and lower bounds on capacity. In particular, firm bounds are established for all n that show that the capacity scales as $\frac{1}{2}\log(n)$.

I. Introduction

We consider a channel for which the relationship between the input $X \in [0, 1]$ and the output $Y \in \{0, \dots, n\}$ is described by the binomial distribution:

$$P_{Y|X}(y|x) = \binom{n}{y} x^y (1-x)^{n-y}. \tag{1}$$

In this work, we are interested in studying the capacity of this channel as a function of the number of trials n, that is

$$C(n) = \max_{P_X : X \in [0,1]} I(X; Y). \tag{2}$$

In addition to studying capacity, we are also interested in studying properties of an optimal capacity-achieving distribution distribution denoted by $P_{X^\star}$.

A. Literature Review

The binomial channel naturally arises in molecular communications and the interested reader is referred to [1]–[4] and references therein. The channel is also useful in the study of the deletion channel [5], [6].

The capacity of the binomial channel was first considered in [7] where the authors used minimax redundancy theorem in [8] to argue that asymptotically the capacity scales as $\frac{1}{2}\log n$. The exact capacity for $n = 1$ case was computed in [1] where binary distribution supported on $\{0, 1\}$ was shown to be capacity achieving. To the best of our knowledge, there are no firm bounds on the capacity of the binomial channel.

Properties of the capacity-achieving distribution have also been looked at. For example, the authors of [1] have designed an algorithm for computing capacity and a capacity-achieving distribution by using a dual representation of the maximization problem. It is also known that, by using the Witsenhausen technique [9], there exists a capacity-achieving distribution with at most $n+1$ mass points. We note, however, that the Witsenhausen technique does not guarantee that the optimal input distribution is unique. In fact, for the binomial channel, uniqueness has not been shown; note that uniqueness is important not just for theoretical purposes, but also for algorithmic purposes. A conventional way to show that the capacity-achieving distribution is unique is by establishing that the mutual information is a *strictly* concave function of the input distribution. However, as will be shown by an example, for the binomial channel, the mutual information is not strictly concave. Other properties, such as location of the support points, are also not well understood. The main goal of this work is to close some of these gaps.

In this work, we also rely on estimation theoretic quantities such as the conditional expectation. For the estimation theoretic treatments of the binomial channel, the interested reader is referred to [10], [11]. Recently, deterministic identification capacity for the binomial channel has been studied in [12].

B. Outline and Contributions

The paper outline and contributions are as follows. The remaining part of Sec. I is dedicated to notation. Sec. II presents the required preliminary and ancillary results. In particular, Sec. II-A presents the Karush-Kuhn-Tucker (KKT) conditions and some important consequences of these conditions, and Sec. II-B establishes properties of estimation theoretic quantities, such as the conditional mean, that will be needed in our analysis. Sec. III constitutes the first part of our main results and focuses on properties of capacity-achieving distributions. In particular, in Sec. III-A it is shown that all capacity-achieving distributions are discrete; in Sec. III-B, the discreetness is used to argue strong concavity of the mutual information, which implies uniqueness of $P_{X^\star}$; in Sec. III-C it is shown that all capacity-achieving distributions are symmetric around $\frac{1}{2}$; Sec. III-D provides additional information about the location of the support points; Sec. III-E provides bounds on the probability values; Sec. III-F improves the upper bound $n+1$ on cardinality of the support, due to Witsenhausen, to the bound of order $\frac{n}{2}$, and provides a lower bound of order $\sqrt{n}$. Sec. IV constitutes the second part of our main results and focuses on bounds on the capacity. In particular, firm lower and upper bounds of order $\frac{1}{2}\log(n)$ are derived.

Finally, some of the proofs and additional results are delegated to the extended version of the paper [13].

C. Notation

All logarithms are to the base e. Deterministic scalar quantities are denoted by lower-case letters and random variables are denoted by uppercase letters. For a random variable X and every measurable subset $\mathcal{A} \subseteq \mathbb{R}$ the probability distribution is written as $P_X(\mathcal{A}) = \mathbb{P}[X \in \mathcal{A}]$. The support set of P_X is

$$\mathsf{supp}(P_X) = \{x : \ P_X(\mathcal{D}) > 0 \text{ for every open set } \mathcal{D} \ni x \ \}. \tag{3}$$

When X is discrete, we write $P_X(x)$ for $P_X(\{x\})$, i.e., P_X is a probability mass function (pmf). The relative entropy of the distributions P and Q is $D\left(P \,\|\, Q\right)$.

Given a function $f : \mathbb{R} \to \mathbb{R}$ and a set $\mathcal{B} \subseteq \mathbb{R}$, the number of zeros of f in $\mathcal{B}$ is given by

$$\mathsf{N}\left(\mathcal{B}; f\right) = \left|\{x : f(x) = 0\} \cap \mathcal{B}\right|, \tag{4}$$

where $|\cdot|$ denotes the cardinality.

The set of the first n positive integers is denoted by $[n]$. The entry in position (i, j) of matrix A is denoted by $[\mathsf{A}]_{i,j}$.

II. Preliminaries

We now presents some of the tools needed in our analysis.

A. KKT Conditions

The next result provides KKT conditions for the optimization problem in (2), allowing the study of the support properties of an optimal input distribution (see for example [14]).

Lemma 1. *$P_{X^\star}$ is a capacity-achieving input distribution if and only if the following conditions hold:*

$$i(x; P_{Y^\star}) \le C(n), \qquad x \in [0, 1], \tag{5}$$

$$i(x; P_{Y^\star}) = C(n), \qquad x \in \mathsf{supp}(P_{X^\star}) \tag{6}$$

where $P_{X^\star} \to P_{Y|X} \to P_{Y^\star}$ (i.e., the optimal output distribution) and

$$i(x; P_{Y^\star}) = D\left(P_{Y|X}(\cdot|x) \,\middle\|\, P_{Y^\star}\right). \tag{7}$$

We also define the following set, which will be useful in our study of the uniqueness of $P_{X^\star}$:

$$\mathcal{A}_n = \{x \in [0, 1] : \ i(x; P_{Y^\star}) - C(n) = 0\}. \tag{8}$$

The importance of $\mathcal{A}_n$ is demonstrated in the following lemma.

Lemma 2. *For a given n*

- *$\mathcal{A}_n$ is unique; and*
- *$\mathsf{supp}(P_{X^\star}) \subseteq \mathcal{A}_n$ for every $P_{X^\star}$.*

Proof. Note that, for a given n, both $P_{Y^\star}$ and $C(n)$ are unique (even if $P_{X^\star}$ is not unique) [15], and since $\mathcal{A}_n$ only depends on these quantities, the uniqueness follows.

The second part follows from the KKT conditions in Lemma 1, because $x \in \mathsf{supp}(P_{X^\star})$ implies $x \in \mathcal{A}_n$. □

B. Estimation Theoretic Preliminaries

Estimation theoretic quantities will play an important role in our analysis. In what follows, the quantity $\mathbb{E}^{n-1}\left[f(Y) \mid X = x\right]$ denotes expectation with respect to a binomial distribution with $n-1$ trials and success probability x per trial, and

$$\ell_b(x, \hat{x}) = x \log\left(\frac{x(1-\hat{x})}{(1-x)\hat{x}}\right) - \frac{x - \hat{x}}{1 - \hat{x}}, \quad (x, \hat{x}) \in (0, 1)^2 \tag{9}$$

represents the Bregman divergence for the binomial channel.

We now summarize some of these preliminary results.

Proposition 1. *For $n \ge 2$ and $x \in (0, 1)$, we have*

$$i'(x; P_Y) = \frac{n}{x}\mathbb{E}^{n-1}\left[\ell_b(x, \mathbb{E}^{n-1}\left[X \mid Y\right]) \;\middle|\; X = x\right] + \frac{n}{x}\mathbb{E}^{n-1}\left[\frac{x - \mathbb{E}^{n-1}\left[X \mid Y\right]}{1 - \mathbb{E}^{n-1}\left[X \mid Y\right]} \;\middle|\; X = x\right] \tag{10}$$

and

$$i''(x; P_Y) = \frac{n}{x(1-x)} + \frac{1}{(1-x)^2} G(x) \tag{11}$$

where $G(x)$ is defined in (12) *(at the top of next page).*

The Bregman divergence in (10) appeared previously in a different but related result. Specifically, in [10, Prop. 2] it was shown that for $a \in (0, 1)$

$$\frac{\partial}{\partial a} I(X; \mathcal{B}_n(aX)) = \frac{n}{a}\mathbb{E}\left[\ell_b\left(aX, \mathbb{E}[aX|\mathcal{B}_{n-1}(aX')]\right)\right] \tag{13}$$

where $Y = \mathcal{B}_n(aX)$ denotes the transformation of input aX through a binomial channel with n trials.

Finally, we also need to show the monotonicity of the conditional mean.

Lemma 3. *The function $y \mapsto \mathbb{E}\left[X \mid Y = y\right]$ is non-decreasing.*

III. Properties of the Capacity-Achieving Distributions

In this section we study properties of capacity-achieving distributions.

A. Discreteness

As already mentioned in Sec. I-A, from the Witsenhausen approach we only know that there exists a discrete distribution with at most $n+1$ mass points. This, however, does not rule out the existence of other capacity-achieving distributions (*e.g.*, continuous capacity-achieving distributions).

We now show that all capacity-achieving distributions are discrete and provide a preliminary bound on the support.

Proposition 2. $|\mathsf{supp}(P_{X^\star})| \le |\mathcal{A}_n| \le n + 1$.

This bound will be improved in Section III-F.

$$G(x) = \mathbb{E}\left[(n-Y)(n-Y-1)\log\frac{\mathbb{E}[X \mid Y=Y]}{\mathbb{E}[1-X \mid Y=Y+1]}\frac{\mathbb{E}[1-X \mid Y=Y+2]}{\mathbb{E}[X \mid Y=Y+1]}\,\middle|\, X=x\right] \tag{12}$$

B. Uniqueness of the Input Distribution

In this section, we show and discuss uniqueness of the capacity-achieving input distribution. To aid our discussion, it is useful to parameterize the mutual information in terms of distributions instead of random variables, that is

$$I(P_X; P_{Y|X}) = I(X;Y). \tag{14}$$

We also let $\mathcal{P}_{\mathcal{X}}$ be the set of all distributions over the set $\mathcal{X}$. In particular, the optimization in (2) can be written as

$$\max_{P_X \in \mathcal{P}_{[0,1]}} I(P_X; P_{Y|X}). \tag{15}$$

A typical way to show that there is a unique maximizer is to show that the mapping $P_X \mapsto I(P_X; P_{Y|X})$ over the set $\mathcal{P}_{[0,1]}$ is *strictly* concave [16]. However, since the output space of the binomial channel is finite and the input space is uncountable, the mutual information is not strictly concave over $\mathcal{P}_{[0,1]}$. For example, when $n = 1$ any distribution symmetric around $x = \frac{1}{2}$ will induce

$$P_Y(0) = P_Y(1) = \frac{1}{2} \tag{16}$$

which is the capacity-achieving output distribution for $n = 1$. Therefore, to show uniqueness of the capacity-achieving input distribution a new or slightly different argument is needed.

We begin by showing the following result.

Proposition 3. *Consider an arbitrary sequence $0 \le x_1 < \ldots < x_{n+1} \le 1$ and define the matrix $\mathsf{A} \in \mathbb{R}^{(n+1)\times(n+1)}$ as*

$$[\mathsf{A}]_{i,k} = P_{Y|X}(i-1|x_k), \quad i \in [n+1],\, k \in [n+1]. \tag{17}$$

Then, A *is full rank.*

Proof. First of all, we argue that considering $x_1 = 0$ and $x_{n+1} = 1$ comes without loosing generality. In fact, in this case the first and last columns of A are $\mathbf{e}_1$ and $\mathbf{e}_{n+1}$, respectively, where $\mathbf{e}_i$ is a zero vector with a 1 in the i-th position. As a consequence, we have $\det(\mathsf{A}) = \det(\tilde{\mathsf{A}})$, where

$$[\tilde{\mathsf{A}}]_{i,k} = [\mathsf{A}]_{i+1,k+1}, \quad i \in [n-1],\, k \in [n-1]. \tag{18}$$

Next, note that we can rewrite the binomial law as

$$P_{Y|X}(y|x) = \binom{n}{y}(1+\mathrm{e}^t)^{-n}\mathrm{e}^{ty} \tag{19}$$

where $x = \frac{\mathrm{e}^t}{1+\mathrm{e}^t}$. The matrix B with $[\mathsf{B}]_{y,k} = \mathrm{e}^{t_k y}$ and $y \in [n-1]$ is a Vandermonde matrix, which is full rank since the t_k's are all distinct [17]. Thanks to the multilinear property of the determinant, we can write that

$$\det(\tilde{\mathsf{A}}) = \det(\mathsf{B})\prod_{y=1}^{n-1}\binom{n}{y}\prod_{k=2}^{n}(1+\mathrm{e}^{t_k})^{-n} > 0 \tag{20}$$

where the last step is due to $\det(\mathsf{B}) > 0$ and to the positivity of the products. As a consequence, A is a full rank matrix. □

With the aid of Proposition 3, we show the following result.

Theorem 1. *Let $\mathcal{X} \subset [0,1]$ be a discrete set of cardinality $n+1$. Then, $P_X \mapsto I(P_X; P_{Y|X})$ is* strictly *concave over $\mathcal{P}_{\mathcal{X}}$.*

Proof. Let $P_X, Q_X \in \mathcal{P}_{\mathcal{X}}$, and let $P_X^\epsilon = (1-\epsilon)P_X + \epsilon Q_X$ for $\epsilon \in (0,1)$, which is also in $\mathcal{P}_{\mathcal{X}}$. Moreover, let $P_X \to P_{Y|X} \to P_Y$, $Q_X \to P_{Y|X} \to Q_Y$ and $P_X^\epsilon \to P_{Y|X} \to P_Y^\epsilon$. Then, first note that

$$\begin{aligned} &I(P_X^\epsilon; P_{Y|X}) \\ &- (1-\epsilon)I(P_X; P_{Y|X}) - \epsilon I(Q_X; P_{Y|X}) && (21)\\ &= D(P_{Y|X} \| P_Y^\epsilon | P_X^\epsilon) \\ &- (1-\epsilon)D(P_{Y|X} \| P_Y | P_X) - \epsilon D(P_{Y|X} \| Q_Y | Q_X) && (22)\\ &= (1-\epsilon)D(P_Y \| P_Y^\epsilon) + \epsilon D(Q_Y \| P_Y^\epsilon). && (23) \end{aligned}$$

We now show that every $P_X \in \mathcal{P}_{\mathcal{X}}$ induces a distinct output distribution (i.e., $P_X \to P_{Y|X} \to P_Y$ is an injective mapping), which implies that (23) is strictly positive and, therefore, the mutual information is strictly concave. Define the following:

$$\mathbf{p}_X = [P_X(x_1), \ldots, P_X(x_{n+1})], \quad x_k \in \mathcal{X}, \tag{24}$$

$$\mathbf{p}_Y = [P_Y(0), \ldots, P_Y(n)]. \tag{25}$$

Then, the mapping $P_X \to P_{Y|X} \to P_Y$ can be written as the following system of linear equations:

$$\mathsf{A}\mathbf{p}_X = \mathbf{p}_Y \tag{26}$$

where the matrix $\mathsf{A} \in \mathbb{R}^{(n+1)\times(n+1)}$ is such that

$$[\mathsf{A}]_{i,k} = P_{Y|X}(i-1|x_k), \quad i \in [n+1],\, x_k \in \mathcal{X}. \tag{27}$$

From Proposition 3, we know that A is full rank for any $\mathcal{X}$ of cardinality $n+1$. Therefore, from standard linear algebra result, it follows that the mapping in (26) is injective (i.e., every $\mathbf{p}_X$ induces a distinct $\mathbf{p}_Y$). Therefore, we conclude that (23) is positive and mutual information is strictly concave. □

Note that since by Proposition 2, $\mathcal{A}_n$ has cardinality of at most $n+1$, from Theorem 1 we have the following corollary.

Corollary 1. *$P_X \mapsto I(P_X; P_{Y|X})$ is strictly concave over $\mathcal{P}_{\mathcal{A}_n}$. Consequently, P_{X^*} is unique.*

C. Symmetry

The binomial channel exhibits the following symmetry

$$P_{Y|X}(y|x) = P_{Y|X}(n-y|1-x), \quad x \in [0,1],\, y \in \{0\} \cup [n], \tag{28}$$

which immediately leads to the following result.

Proposition 4. *If $X^\star$ is capacity-achieving, then $X^\star \stackrel{d}{=} 1 - X^\star$.*[1]

[1]Here $\stackrel{d}{=}$ denotes equality in distribution.

D. On the Location of Support Points

Following the same lines of [18, Sec. V] we have that.

Proposition 5. *Let $P_{X^\star}$ be a capacity-achieving input distribution. Then,* $\{0,1\} \in \mathsf{supp}(P_{X^\star})$.

An important consequence of Proposition 5 is given next.

Corollary 2. *The channel capacity is equal to*

$$C(n) = \log \frac{1}{P_{Y^\star}(0)} = \log \frac{1}{P_{Y^\star}(n)}. \tag{29}$$

Proof. Thanks to Proposition 5, we know that $0 \in \mathsf{supp}(P_{X^\star})$. By using the KKT condition (6), we can write

$$C(n) = i(0; P_{Y^\star}) = \sum_{y=0}^{n} \binom{n}{y} 0^y \log \frac{\binom{n}{y} 0^y}{P_{Y^\star}(y)} = \log \frac{1}{P_{Y^\star}(0)}.$$

By symmetry, we can argue that $P_{Y^\star}(0) = P_{Y^\star}(n)$. □

We next show that there is at most one support point in the interval $\left(0, \frac{1}{n}\right]$ and, by symmetry, at most one point in $\left[1-\frac{1}{n}, 1\right)$. The proof technique we use was developed in [19] in the context of Poisson noise channel.

Proposition 6. *For all $n \geq 1$, we have*

$$\left| \mathsf{supp}(P_{X^\star}) \cap \left(0, \frac{1}{n}\right] \right| \leq 1, \tag{30}$$

$$\left| \mathsf{supp}(P_{X^\star}) \cap \left[1-\frac{1}{n}, 1\right) \right| \leq 1. \tag{31}$$

E. Bounds on the Probabilities

We begin by recalling that for $P_X \to P_{Y|X} \to P_Y$ and $Q_X \to P_{Y|X} \to Q_Y$, we have that

$$D\left(P_X \,\|\, Q_X\right) = D\left(P_Y \,\|\, Q_Y\right) + D(P_{X|Y} \| Q_{X|Y} | P_Y), \tag{32}$$

where the conditional relative entropy is defined as

$$D(P_{X|Y} \| Q_{X|Y} | P_Y) = \sum_{y=0}^{n} P_Y(y) D\left(P_{X|Y}(\cdot|y) \,\|\, Q_{X|Y}(\cdot|y)\right). \tag{33}$$

The key to finding bounds on the probabilities is the following lemma.

Lemma 4. *For $x^\star \in \mathsf{supp}(P_{X^\star})$*

$$P_{X^\star}(x^\star) = \mathrm{e}^{-C(n) - \mathcal{D}(x^\star)}, \tag{34}$$

where

$$\mathcal{D}(x^\star) = D\left(\delta_{x^\star} \| P_{X^\star|Y} | P_{Y|X}(\cdot|x^\star)\right). \tag{35}$$

Proof. Using the equality condition in the KKT (6), we have that for $x^\star \in \mathsf{supp}(P_{X^\star})$

$$\begin{aligned} C(n) &= D\left(P_{Y|X}(\cdot|x^\star) \,\|\, P_{Y^\star}\right) && (36)\\ &= D\left(P_{Y_{x^\star}} \,\|\, P_{Y^\star}\right) && (37)\\ &= D\left(\delta_{x^\star} \,\|\, P_{X^\star}\right) - D(\delta_{x^\star} \| P_{X^\star|Y} | P_{Y_{x^\star}}) && (38)\\ &= \log \frac{1}{P_{X^\star}(x^\star)} - D(\delta_{x^\star} \| P_{X^\star|Y} | P_{Y_{x^\star}}), && (39) \end{aligned}$$

where (37) follows by defining $\delta_{x^\star} \to P_{Y|X} \to P_{Y_{x^*}}$; and (38) follows by using (32).

By rearranging (39), and recognizing that $P_{Y_{x^*}}(\cdot) = P_{Y|X}(\cdot|x^\star)$, we arrive at: for $x^\star \in \mathsf{supp}(P_{X^\star})$

$$P_{X^\star}(x^\star) = \mathrm{e}^{-C(n) - D\left(\delta_{x^\star} \| P_{X^\star|Y} | P_{Y|X}(\cdot|x^\star)\right)}. \tag{40}$$

□

The term $\mathcal{D}(x^\star)$ measures how on average the $P_{X^\star|Y}$ is close to a point measure. We refer to $\mathcal{D}(x^\star)$ as the *crest-factor*.[2]

From Lemma 4, by using $\mathcal{D}(x^\star) \geq 0$, which follows from the non-negativity of the relative-entropy, we immediately arrive at the following bound:

$$P_{X^\star}(x^\star) \leq \mathrm{e}^{-C(n)}, \qquad x^\star \in \mathsf{supp}(P_{X^\star}). \tag{41}$$

The bound in (41) might appear ineffective due to the fact that the capacity is unknown. However, note that for any $\tilde{X}$, from the definition of the capacity we have that

$$P_{X^\star}(x^\star) \leq \mathrm{e}^{-I(\tilde{X};\tilde{Y})}, \qquad x^\star \in \mathsf{supp}(P_{X^\star}), \tag{42}$$

which implies that any good guess results in an upper bound.

Tighter lower bounds on $\mathcal{D}(x)$ than $\mathcal{D}(x) \geq 0$ are shown in the extended version in [13]. The upper bounds on $\mathcal{D}(x)$ so far have been elusive.

F. Bounds on the Cardinality

We now provide upper and lower bounds on the cardinality of the support of $P_{X^\star}$. We start with the following exact formula for the number of support points.

Proposition 7. *For $n \geq 1$*

$$|\mathsf{supp}(P_{X^\star})| = \frac{\mathrm{e}^{C(n)}}{\mathbb{E}\left[\mathrm{e}^{-\mathcal{D}(U^*)}\right]}, \tag{43}$$

where U^* is uniformly distributed on $\mathsf{supp}(P_{X^\star})$.

Proof. Starting with Lemma 4 and summing over $x^\star \in \mathsf{supp}(P_{X^\star})$, we arrive at

$$1 = \mathrm{e}^{-C(n)} \sum_{x^\star \in \mathsf{supp}(P_{X^\star})} \mathrm{e}^{-\mathcal{D}(x^\star)}. \tag{44}$$

Dividing both sides of (44) by $|\mathsf{supp}(P_{X^\star})|$ and rearranging, we arrive at the desired result. □

From Proposition 7 and non-negativity of $\mathcal{D}$, we arrive at

$$|\mathsf{supp}(P_{X^\star})| \geq \mathrm{e}^{C(n)} = \Theta(\sqrt{n}) \tag{45}$$

where the order of the lower bound follows from the fact that $C(n)$ scales as $\frac{1}{2}\log(n)$ as will be shown in Theorem 3 (Section IV-B below).

We now move on to showing upper bounds. We already have demonstrated a bound of order $n+1$ in Proposition 2. We now improve this upper bound by a factor of two.

[2]In signal processing, the crest factor measures how peaky the waveform is. Specifically, it compares the peak amplitude of a waveform relative to its average value.

n	$C(n)$	$\mathcal{X} \equiv \mathsf{supp}(P_{X^\star})$	$\{P_{X^\star}(x),\, x \in \mathcal{X}\}$	$\{P_{Y^\star}(y),\, y \in \{0\} \cup [n]\}$
1	$\log(2)$	$\{0,1\}$	$\left\{\frac{1}{2}, \frac{1}{2}\right\}$	$\left\{\frac{1}{2}, \frac{1}{2}\right\}$
2	$\log\left(\frac{17}{8}\right)$	$\left\{0, \frac{1}{2}, 1\right\}$	$\left\{\frac{15}{34}, \frac{2}{17}, \frac{15}{34}\right\}$	$\left\{\frac{8}{17}, \frac{1}{17}, \frac{8}{17}\right\}$
3	$\log\left(\frac{19}{8}\right)$	$\left\{0, \frac{1}{2}, 1\right\}$	$\left\{\frac{15}{38}, \frac{4}{19}, \frac{15}{38}\right\}$	$\left\{\frac{8}{19}, \frac{3}{38}, \frac{3}{38}, \frac{8}{19}\right\}$

TABLE I: Capacity and capacity-achieving distributions.

Theorem 2. *For $n \geq 1$*

$$|\mathsf{supp}(P_{X^\star})| \leq 2 + \left\lfloor \frac{1}{2} \mathsf{N}\left((0,1);\, i''(x; P_{Y^\star})\right) \right\rfloor \tag{46}$$

$$\leq 2 + \left\lfloor \frac{n}{2} \right\rfloor. \tag{47}$$

Proof. First of all, note that by Proposition 5 and the KKT conditions we know that the function $i(\cdot; P_{Y^\star})$ starts with a local maximum at $x^\star = 0$, and from Proposition 1 we know that $i''(x; P_{Y^\star}) > 0$ for $x \to 0$. Now, by continuity of $i(\cdot; P_{Y^\star})$, if $i(\cdot; P_{Y^\star})$ changes concavity k times, then it has at most $2 + \left\lfloor \frac{k}{2} \right\rfloor$ local maxima. Moreover, from the KKT conditions we know that all the zeros of $i(\cdot; P_{Y^\star}) - C(n)$ are local maxima.

Then, we can write

$$|\mathsf{supp}(P_{X^\star})| \leq \mathsf{N}\left([0,1];\, i(\cdot; P_{Y^\star}) - C(n)\right) \tag{48}$$

$$\leq 2 + \left\lfloor \frac{1}{2} \mathsf{N}\left((0,1);\, i''(\cdot; P_{Y^\star})\right) \right\rfloor \tag{49}$$

$$\leq 2 + \left\lfloor \frac{n}{2} \right\rfloor, \tag{50}$$

where (50) follows from the fact that $x \mapsto x(x-1) i''(x; P_{Y^\star})$ is a polynomial of degree n (see Proposition 1). □

A few remarks are now in order:

- The proof of Theorem 2 does not rely on the uniqueness of P_{X^*}. Therefore, it improves on the Witsenhausen bound by a factor of two. Furthermore, the key part of the proof leading to (46) is independent of the fact that the channel is binomial: Indeed, this fact is only used in (47). Consequently, we posit that this bound may prove more beneficial for channels where it is feasible to establish bounds on the number of zeros in $i''(x; P_{X^\star})$.
- The lower bound in (45) and the upper bound in (47) do not match in their respective orders. This lack of alignment is perhaps unsurprising, considering the inherent difficulty in establishing tight bounds on the cardinality of the support. For further exploration of this challenging problem, the interested reader is directed to [20]–[22]. We suspect that neither the upper nor the lower bounds are tight.

IV. Capacity and Bounds on the Capacity

In this section, we provide exact values of the capacity for $n \leq 3$. For the remaining regime we provide upper and lower bounds on capacity.

A. Exact Capacity for $n \leq 3$

The exact capacity can be computed by first making a guess of the capacity-achieving distribution according to the properties outlined in Section III. Then, this guess can be checked against the sufficient and necessary KKT conditions in Lemma 1. These somewhat tedious computations are performed in [13], and the results are displayed in Table I.

B. Bounds on the Capacity

We now provide bounds on the capacity. The upper bound relies on the dual representation of the capacity as:

$$C(n) = \inf_{q} \max_{x \in [0,1]} D\left(P_{Y|X}(\cdot|x) \,\middle\|\, q\right), \tag{51}$$

which, by properly choosing an auxiliary output distribution q, often leads to order-tight bounds. The reader is referred to [23]–[25] for applications to other channels. It will also be convenient to work with continuous output, and we will use the following channel output: $\tilde{Y} = Y + U$, where $U \sim \mathcal{U}(0,1)$. Note that, because the distance between original Y points is one, the additive noise U can be completely filtered out, and we have $I(X;Y) = I(X; Y+U)$ for all X. This trick has been used before in the context of the Poisson channel in [23].

The lower bound on the capacity will follow from choosing a convenient input distribution. The exact computation, however, will not be possible, and some further bounds on the entropy of the binomial distribution will be needed. Therefore, in [13], we also provide a new upper bound on the entropy of a binomial distribution. Bounds on the entropy of a binomial distribution have been considered before in [26], [27].

Theorem 3. *For $n \geq 1$, the channel capacity is bounded below by*

$$C(n) \geq \max\left\{\log(2), \log(\pi n) - \frac{1}{2}\log\left(2\pi e\left(\frac{n}{8} + \frac{1}{12}\right)\right) + \frac{1}{\sqrt{\pi\left(n + \frac{1}{4}\right)}} \log\left(\frac{1}{16n^2}\right) - \log(4) - 1\right\} \tag{52}$$

and bounded above by

$$C(n) \leq \min\left\{\log\left(3 + \left\lfloor \frac{(n-1)}{2} \right\rfloor\right), \log(\pi(n+1)) - \frac{1}{2}\log(n) + \frac{3}{2} + \frac{1}{2^{n+1}}\log(n) + \frac{1}{2}\log\left(\frac{3}{2}\left(1 + \frac{1}{n}\right)\right)\right\}. \tag{53}$$

REFERENCES

[1] N. Farsad, W. Chuang, A. Goldsmith, C. Komninakis, M. Médard, C. Rose, L. Vandenberghe, E. E. Wesel, and R. D. Wesel, "Capacities and optimal input distributions for particle-intensity channels," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 6, no. 3, pp. 220–232, 2020.

[2] A. Einolghozati, M. Sardari, and F. Fekri, "Design and analysis of wireless communication systems using diffusion-based molecular communication among bacteria," *IEEE Transactions on Wireless Communications*, vol. 12, no. 12, pp. 6096–6105, 2013.

[3] N. Farsad, C. Rose, M. Médard, and A. Goldsmith, "Capacity of molecular channels with imperfect particle-intensity modulation and detection," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 2468–2472.

[4] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel modeling for diffusive molecular communication—A tutorial review," *Proceedings of the IEEE*, vol. 107, no. 7, pp. 1256–1301, 2019.

[5] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," in *Soviet Physics Doklady*, vol. 10, no. 8. Soviet Union, 1966, pp. 707–710.

[6] M. Cheraghchi, "Capacity upper bounds for deletion-type channels," *Journal of the ACM (JACM)*, vol. 66, no. 2, pp. 1–79, 2019.

[7] C. Komninakis, L. Vandenberghe, and R. D. Wesel, "Capacity of the binomial channel, or minimax redundancy for memoryless sources," in *IEEE International Symposium on Information Theory*, 2001, pp. 127–127.

[8] Q. Xie and A. R. Barron, "Minimax redundancy for the class of memoryless sources," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 646–657, 1997.

[9] H. Witsenhausen, "Some aspects of convexity useful in information theory," *IEEE Trans. Inf. Theory*, vol. 26, no. 3, pp. 265–271, 1980.

[10] C. G. Taborda, D. Guo, and F. Perez-Cruz, "Information-estimation relationships over binomial and negative binomial models," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2630–2646, 2014.

[11] A. Dytso, M. Fauß, and H. V. Poor, "Bayesian risk with Bregman loss: A Cramér–Rao type bound and linear estimation," *IEEE Trans. Inf. Theory*, vol. 68, no. 3, pp. 1985–2000, 2022.

[12] M. J. Salariseddigh, V. Jamali, H. Boche, C. Deppe, and R. Schober, "Deterministic identification for MC binomial channel," in *IEEE Intern. Symp. Inf. Theory (ISIT)*. IEEE, 2023, pp. 448–453.

[13] I. Zieder, A. Favano, L. Barletta, and A. Dytso, "Binomial channel: On the capacity-achieving distribution and bounds on the capacity," *arXiv preprint arXiv:2401.12818*, 2024.

[14] A. Dytso, M. Goldenbaum, H. V. Poor, and S. S. Shitz, "When are discrete channel inputs optimal? — optimization techniques and some new results," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, 2018, pp. 1–6.

[15] J. Kemperman, "On the Shannon capacity of an arbitrary channel," in *Indagationes Mathematicae (Proceedings)*, vol. 77, no. 2. North-Holland, 1974, pp. 101–115.

[16] J. G. Smith, "The information capacity of amplitude-and variance-constrained scalar Gaussian channels," *Information and Control*, vol. 18, no. 3, pp. 203–219, 1971.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computations*. JHU press, 2013.

[18] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, 2001.

[19] R. McEliece, E. Rodemich, and A. Rubin, "The practical limits of photon communication," *Jet Propulsion Laboratory Deep Space Network Progress Reports*, vol. 42, pp. 63–67, 1979.

[20] A. Dytso, S. Yagli, H. V. Poor, and S. S. Shitz, "The capacity achieving distribution for the amplitude constrained additive Gaussian channel: An upper bound on the number of mass points," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, pp. 2006–2022, 2019.

[21] A. Dytso, L. Barletta, and S. Shamai Shitz, "Properties of the support of the capacity-achieving distribution of the amplitude-constrained Poisson noise channel," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 7050–7066, 2021.

[22] N. Kashyap and M. Krishnapur, "How many modes can a constrained Gaussian mixture have," *arXiv preprint arXiv:2005.01580*, 2020.

[23] A. Lapidoth and S. M. Moser, "On the capacity of the discrete-time Poisson channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 303–322, 2008.

[24] A. L. McKellips, "Simple tight bounds on capacity for the peak-limited discrete-time channel," in *Proceedings International Symposium on Information Theory (ISIT)*. IEEE, 2004, pp. 348–348.

[25] A. Thangaraj, G. Kramer, and G. Böcherer, "Capacity bounds for discrete-time, amplitude-constrained, additive white Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4172–4182, 2017.

[26] C. Knessl, "Integral representations and asymptotic expansions for Shannon and Renyi entropies," *Applied Mathematics Letters*, vol. 11, no. 2, pp. 69–74, 1998.

[27] M. Cheraghchi, "Expressions for the entropy of binomial-type distributions," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 2520–2524.