# Capacity of Frequency-based Channels: Encoding Information in Molecular Concentrations

Yuval Gerzon[1], Ilan Shomorony[2] and Nir Weinberger[1]

**Abstract**

We consider a molecular channel, in which messages are encoded to the frequency of objects (or concentration of molecules) in a pool, and whose output during reading time is a noisy version of the input frequencies, as obtained by sampling with replacement from the pool. We tightly characterize the capacity of this channel using upper and lower bounds, when the number of objects in the pool of objects is constrained. We apply this result to the DNA storage channel in the short-molecule regime, and show that even though the capacity of this channel is technically zero, it can still achieve a large information density.

## I. Introduction

In molecular communication [1], information is encoded into the presence of objects from various possible types in some restricted physical domain. As a prominent example, in DNA storage systems [2]–[12], information is encoded to $K$ molecules, each is a strand of length $L$ composed from the four possible nucleotides, denoted by $\mathcal{A} := \{\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}\}$. The $K$ molecules are stored in a pool, and the distinctive aspect of this system is that the order of the $K$ molecules in the pool cannot be preserved. So, while the total of $KL$ symbols can be considered as the blocklength of the codeword, unlike standard channel coding, the codeword is, in fact, partitioned to $K$ out-of-order segments of length $L$ each. Given the pool, the message is decoded by randomly sampling molecules from the pool, sequencing each of them to obtain a noisy read of the sequence of nucleotides in the strand, and using the out-of-order output strands to decode the message.

A simple method to resolve the lack of order of the encoded molecules is to devote the first $\log_{|\mathcal{A}|} K$ symbols of the molecule to encode its index. This requires that the molecule length will be large enough to accommodate both the index and the message encoding. Indeed, in [13] it was established that the regime of interest is $L = \beta \log K$ for some $\beta > 0$, and it was shown that indexing achieves the capacity of the DNA

[1] Yuval Gerzon and Nir Weinberger are with the Viterbi Faculty of Electrical and Computer Engineering, Technion-Israel Institute of Technology, Haifa 3200004, Israel (Emails: gerzon.yuval@campus.technion.ac.il, nirwein@technion.ac.il.). [2] Ilan Shomorony is with the Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (Email: ilans@illinois.edu). This paper was accepted in part to the 2024 IEEE International Symposium on Information Theory. The research of N. W. was supported by the Israel Science Foundation (ISF), grant no. 1782/22. The work of I. S. was supported in part by the National Science Foundation under CCF grants 2007597 and 2046991.

storage channel when the sequencing is noiseless. Conversely, if the molecule length $L$ is short in comparison to their number $K$, and so there is not enough symbols to encode the index in the molecule, concretely, if $\beta \leq \frac{1}{\log|\mathcal{A}|}$, then the capacity is *zero*. Consequently, in this regime, the log-cardinality of the optimal codebook (which is the total number of stored bits) scales at most *sub-linearly* with the total number of nucleotides $KL$.

Nonetheless, the DNA storage medium has an extreme information density, and a huge number of nucleotides can be stored in tiny pools. As noted in [14, Sec. 7.3], the amount of stored information can be large even for channels with asymptotically vanishing maximal rate. This observation serves as a strong motivation to study the log-cardinality of the optimal codebook in the short-molecule regime, that is, for $\beta \leq \frac{1}{\log|\mathcal{A}|}$. In this regime, the number of molecules in a codeword $K$ is larger than the number of possible strands of length $L$ from the alphabet $\mathcal{A}$, to wit, $K \geq |\mathcal{A}|^L$. Accordingly, each codeword must contain multiple copies of the same strand (at least for one strand). Hence, in the short-molecule regime, the message is actually encoded by the number of times each of the $|\mathcal{A}|^L$ possible strands of length $L$ appears in the pool. The encoded codeword can thus be represented by the frequency vector that measures the frequency of each type of strands in the pool. We refer to this type of channel as *frequency-based channel*. During reading, molecules are sampled from the pool (with replacement) and so the output is a also a frequency vector. This vector is a noisy version of the input vector for two reasons: First, the frequency of the strands in the codeword is not preserved by the sampling. Second, the synthesis and the sequencing processes are possibly noisy [15].

In this paper, we formulate a general frequency-based channel. We focus on the effect of random sampling, which we model as a multinomial distribution, and thus assume noiseless sequencing. In this channel model, the blocklength $n$ models the number of different types of objects. Each codeword has a total count of $ng_n$ objects from the different types, and which are sampled $nr_n$ times in total. For example, in the DNA storage channel $n = |\mathcal{A}|^L$, as this is the number of different molecules of length $L$ from an alphabet $\mathcal{A}$, and $ng_n = K$, as this is the total number of molecules. In [14, Sec. 7.3] a slightly different Poisson sampling channel was considered, which assumes $g_n = r_n$, and a conjecture was made on the scaling of the log-cardinality of the optimal codebook [14, Conjecture 4]. Concretely, based on the capacity of the average-power-constrained Poisson channel [16] it was conjectured that the capacity scales as $\frac{1}{2}\log r_n + o_n(1)$.

*Main contribution:* In this work, we address that conjecture. Though our multinomial sampling is slightly different, its analysis is, in fact, based on a reduction to a Poisson channel, and so in this sense the multinomial model subsumes the Poisson model. That being said, the reduction itself complicates the analysis of the resulting Poisson channel, and the latter is non-standard from two aspects. First, the total number of counts in the codeword must be common to all codewords in the codebook in order for an input count to accurately model frequencies. Second, since the input symbols also measure the count of a possible objects in the codeword, they must be *integer-valued*, and so the input distribution must be supported on the integers. Our main result (Theorem 2) is an approximate solution of the conjecture of [14, Conjecture 4]: Our converse bound shows that the capacity is less

then $\frac{1}{2}\log[r_n \wedge (eg_n)] + o_n(1)$. That is, increasing $r_n$ beyond $g_n$ may increase capacity, but asymptotically only up to $\frac{1}{2}$[nats]. Our achievable bound requires the condition $n = \omega(g_n)$, and when the ratio $r_n/g_n$ is optimized, it is given by $\frac{1}{2}\log(g_n) - 1.295 + o_n(1)$[nats]. Interestingly, the optimum of the lower bound occurs at $r_n \approx 0.4 g_n$, i.e., when sampling *less* objects than there are in the codeword.

The implication of this result to the DNA storage channel is valid when the molecules are not very short, and specifically, in the regime $\beta \in (\frac{1}{2\log|\mathcal{A}|}, \frac{1}{\log|\mathcal{A}|})$. The result shows that the log-cardinality of the optimal codebook increases as

$$\frac{1 - \beta\log|\mathcal{A}|}{2\beta} \cdot K^{\beta\log|\mathcal{A}|}\log K, \tag{1}$$

up to terms negligible with $K$. A simple numerical example then shows that the resulting information density (in nats per gram) could still be huge, which remarkably reinforces the importance of this, strictly speaking, zero capacity, channel.

*Related work:* The analysis of the DNA-storage channel is an active research area, both from an information-theoretic point of view [13], [14], [17]–[21] and from coding-theoretic point of view [10], [11], [22], [23]. Our channel model is closely-related to the *permutation channel*. Using our formulation, the input to the permutation channel is also a frequency vector, however, it is assumed that each object is sampled exactly once, and the output vector is noisy due to noisy sequencing (in our terms). This channel was considered in [24], [25] with codes termed *multiset codes*. Constructions of such codes were proposed, and combinatorial bounds on the size of optimal codes for a given detection or correction capability were derived. An information-theoretic version of the channel was introduced in [26], with sharp converse bounds obtained in [27]. However, compared to our results, and in our terminology, the blocklength in [26], [27] is considered fixed, whereas in our model it increases without bound (with a certain scaling). A multi-user model of this channel was recently explored in [28]. Another related model for DNA storage is based on composite DNA letters [29], [30], in which many copies of a single molecule are generated, and each letter in the molecule is a *composite letter*, i.e., it is randomly chosen from a subset of $\{\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}\}$, chosen according to the encoded information. In this channel model too, information is stored in the frequency of each DNA letter at the output, though the randomness of each letter is created during synthesis. This leads to a somewhat different mathematical model, of a multinomial channel, and its capacity is discussed, e.g., in [31].

We rely on the analysis of the Poisson channel under an average-power constraint. The capacity of this channel was extensively explored (e.g., [16], [32]–[42]), but for the frequency-based channel we mainly rely on the asymptotic expression of [16]. On its own, the entropy of a Poisson random variable (RV) was also extensively explored [43]–[45], along with ample study of related properties, e.g., [41], [43], [46]–[50].

*Paper outline:* The paper is organized as follows. In Sec. II we shortly describe the DNA storage channel, and then formulate the more general frequency-based channel. In Sec. III we state our main result and outline

its proof. In Sec. IV we provide detailed proofs, and in Sec. V we conclude the paper with a summary and future research directions.

## II. Problem Formulation

*Notation conventions:* For an integer $n \in \mathbb{N}_+$, let $[n] := \{1, 2, \dots n\}$. For $a, b \in \mathbb{R}$, let $a \vee b := \max\{a, b\}$ and $a \wedge b := \min\{a, b\}$. Logarithms and exponents are taken to the natural base. Standard notation for information-theoretic quantities is used [51], e.g., the entropy $H(P_X)$ or $H(X)$ for a discrete RV $X$ with probability mass function (PMF) $P_X$, the mutual information $I(X;Y)$ between two RVs $X$ and $Y$, and $\mathrm{D_{KL}}(P \,||\, Q)$ for the KL divergence between the probability measures $P$ and $Q$. The binary entropy function is denoted by $\mathrm{h_{bin}}(\cdot)$.

Although our results are valid under general molecular storage settings, the DNA storage channel is our main motivation, and so we next briefly review its model, and explain how it translates in the short-molecule regime to a frequency-based channel [14, Sec. 7.3].

### A. The DNA Storage Channel Model

The DNA storage model, also called the *multi-draw noisy shuffling channel* [14], is as follows. Let an alphabet $\mathcal{A}$ be given, e.g., $\mathcal{A} = \{\mathrm{A}, \mathrm{C}, \mathrm{G}, \mathrm{T}\}$ in the case of an actual DNA-based storage system. A codeword is $a^{LK} = (a_1^L, \dots a_K^L)$, where $a_k^L \in \mathcal{A}^L$ for all $k \in [K]$ is called a *molecule* or a *strand*. Thus, there are $K$ molecules in a codeword, each of which is a length-$L$ vector from the alphabet $\mathcal{A}$. A codebook is a set of different codewords, $\mathcal{C} = \{a^{LK}(j)\}_{j \in [M]}$. The codeword is read in a noisy way comprised of two stages. In the first stage, $N$ molecules are uniformly sampled from the $K$ molecules of $a^{LK}$, independently, with replacement. Letting $\{U_i\}_{i \in [N]}$ be independent and identically distributed (IID) such that $U_i \sim \mathrm{Uniform}[K]$, the output of this stage is $\{a_{U_1}^L, \dots, a_{U_N}^L\}$. In the second stage, each of the sampled molecules $a_{U_i}^L$ is sequenced, and the output molecule $b_i^* \in \mathcal{B}^*$ is obtained, where $\mathcal{B}$ is an output alphabet, and $*$ indicates varying length $\mathcal{B}^* = \bigcup_{\ell=1}^{\infty} \mathcal{B}^\ell$. Thus, the length of $b_i^*$ may be different from $L$, and vary from one molecule to the other. The possibly noisy sequencing is modeled as a noisy channel $V_L$ from $\mathcal{A}^L$ to $\mathcal{B}^*$. For example, this channel may include *substitutions* of a letter from $\mathcal{A}$ with a different letter, deletions, and insertions [52]. The channel output is then $(b_1^*, \dots, b_N^*)$.

*Remark* 1. It may seem more natural to model the channel output as obtained via sampling *without* replacement, since each sampled molecule is removed from the DNA pool in order to be sequenced. The reason why we model the channel as performing sampling *with* replacement is because in practical DNA storage systems, *polymerase chain reaction* (PCR) amplification is used to replicate each molecule in the pool a large (but roughly fixed) number of times. Hence, the relative frequency of each DNA molecule in the pool remains roughly fixed, but sampling from this amplified DNA pool is essentially sampling with replacement from the original pool.

We let the length of each molecule scale as $L = L_K$, the number of sampled molecules to scale as $N = N_K$, and denote the maximal cardinality of a codebook with codewords of $K$ molecules and maximal error probability $\epsilon_K$ as $M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)$. As a codeword is composed from a total $KL$ symbols from $\mathcal{A}$, the rate of a codebook of cardinality $M$ is defined as $R := \frac{1}{KL}\log M$. The capacity is the maximal rate with vanishing error probability, that is, $\frac{1}{KL}\log M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)$ with $\epsilon_K \to 0$ as $K \to \infty$.

*The short-molecule regime:* The capacity of the DNA storage channel for discrete memoryless sequencing channels was analyzed in [13], [17], [18], [20]. Specifically, it was shown in [13] that even for *noiseless* sequencing channels, the capacity of DNA storage is strictly positive only if the molecule length scales as $L_K = \beta \log K$ and $\beta > \frac{1}{\log|\mathcal{A}|}$. Intuitively, the lack of order in the codeword can be resolved by using $\log_{|\mathcal{A}|} K = \frac{\log K}{\log|\mathcal{A}|}$ symbols from each molecule to encode its index in the codeword, and using the rest of the symbols to encode the message. However, if $\beta < \frac{1}{\log|\mathcal{A}|}$ then the length of a molecule does not allow for encoding the index, let alone for encoding the message. It can be shown that indexing is optimal for noiseless sequencing channels (though not necessarily for noisy channels), and that no positive rate can be achieved when $\beta \leq \frac{1}{\log|\mathcal{A}|}$. We thus refer to this regime as the *short-molecule regime.* From a different point of view, we may note that the number of distinct molecules of length $L$ is $|\mathcal{A}|^L$. As $K \geq |\mathcal{A}|^L$ in the short-molecule regime, each codeword must contain more than a single copy of the same molecule. Since the molecules of the codeword lack any order, the message is actually encoded in the number of copies of each of the $n = |\mathcal{A}|^L$ possible molecules in $\mathcal{A}^L$, or in their *frequencies*. Let us order the $\mathcal{A}^L$ strings representing the possible molecules in some arbitrary order $[n]$. Then, the input codeword can be equivalently represented by the vector $x^n := (x_1, \ldots, x_n) \in \mathbb{N}^n$ where $x_i$ is the count of the $i$th string in $\mathcal{A}^L$. Thus, it holds that $\sum_{i=1}^n x_i = K$. Similarly, let us denote the number of counts of each of the strings in $\mathcal{A}^L$ at the output codeword by $y^n := (y_1, \ldots, y_n) \in \mathbb{N}^n$. Due to the randomness in the sampling stage, $y^n$ is a noisy version of $x^n$ even for noiseless sequencing channels. We refer to this equivalent channel model as a frequency-based channel, and formally define it in the next subsection, in greater generality. In the rest of the paper we will analyze the capacity of that channel model, which, in turn, leads to bounds on $M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)$ for DNA storage channels. The capacity is zero in the short-molecule regime, and so $\log M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)$ scales *sub-linearly* with $KL$. However, it is still a monotonic non-decreasing function of $K$, for which our goal it to characterize the optimal asymptotic scaling.

### *B. The Frequency-based Channel Model*

Consider a set of $n$ distinguishable types of objects (e.g., molecules). An input message is encoded as a pool of unordered objects from the various types. Thus, the channel input is represented by the count vector $x^n := (x_1, \ldots, x_n) \in \mathbb{N}^n$, where $x_i$ is the number of objects of the $i$th type in the pool of objects. It is assumed that $\sum_{i=1}^n x_i$ is constant for all possible messages. Thus, $\hat{x}_i := x_i/(\sum_{i=1}^n x_i)$ is a *frequency-vector* (or *concentration*) of the $i$th type in the codeword pool. It is further assumed that the total number of objects

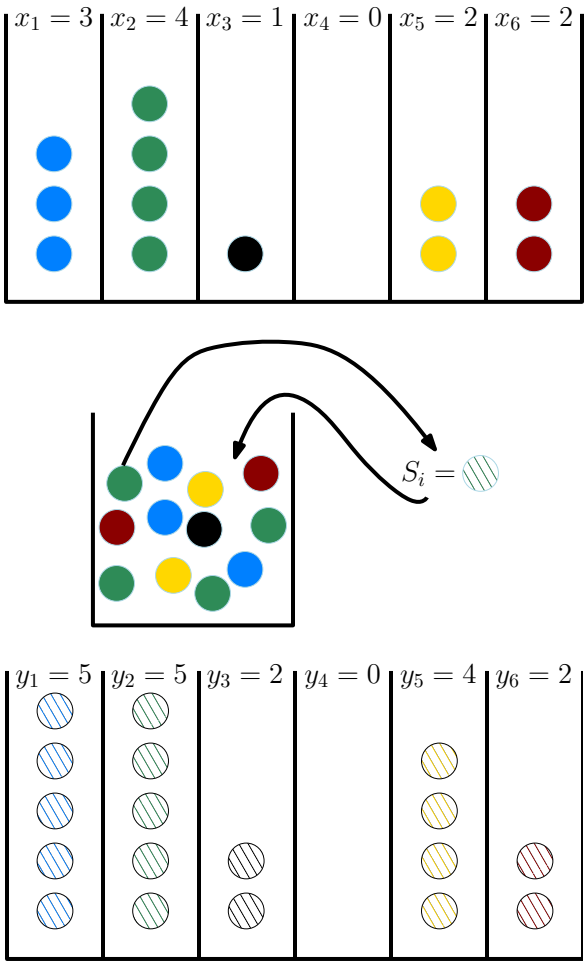


Figure 1. An illustration of the channel model with $n = 6$, $g_n = 2$ and $r_n = 3$. Top: The message is encoded to the codeword $x^6 = (3, 4, 1, 0, 2, 2)$. Middle: The $ng_n = 12$ objects are stored in a pool, and then sampled with replacement $nr_n = 18$ times. At each sample the object type is recorded as $S_i$. Bottom: The output vector is the histogram of $S^{nr_n}$, given by $y^n = (5, 5, 2, 0, 4, 2)$.

is restricted as $\sum_{i=1}^n x_i \le ng_n$, for some given $g_n$. To read the message, $nr_n$ samples are taken, where for each $i \in [nr_n]$, an object is randomly chosen uniformly at random from the set of $\sum_{i=1}^n x_i$ objects in the pool, with replacement. Then, the type of the object is read, by a possibly noisy mechanism. Let $W_n \in \mathbb{R}^{n\times n}$ be a Markov kernel, so that $W_n(j, i)$ represents the probability that an object of type $i$ is determined to be of type $j$ (hence $W_n(j, i) \ge 0$ and $\sum_{j=1}^n W_n(j, i) = 1$). Thus, the $i$th object is $S_i \sim \text{Categorical}(\hat{x}^n W_n)$, where $\hat{x}^n W_n$ represents the standard multiplication of row vector by a matrix, and $S^{nr_n} := (S_1, \cdots, S_{nr_n})$ is a vector of IID RVs. Conditioned on input $x^n$, the output is equivalently a noisy count vector $Y^n := (Y_1, \ldots, Y_n) \in \mathbb{N}^n$ where $Y^n \sim \text{Multinomial}(nr_n, \hat{x}^n W_n)$. A noiseless setting is illustrated in Fig. 1.

A code is a set of $M$ input count vectors $\mathcal{C}_M := \{x^n(1), \ldots, x^n(M)\}$ for which $\sum_{i=1}^n x_i(m)$ is constant for all $m \in [M]$. The size of the largest code for $n$ object types, normalized total count of input objects $g_n$, normalized number of sampled objects $r_n$, a reading kernel $W_n$, under a given error probability $\epsilon_n \in (0, 1)$ is denoted by $M^*(n \mid \epsilon_n, g_n, r_n, W_n)$. Our goal is to accurately determine the growth rate of $M^*(n \mid \epsilon_n, g_n, r_n, W_n)$,

or the rate of the codebook, given by $\frac{1}{n}\log M^*(\cdot)$. We assume that both $g_n, r_n$ are monotonic non-decreasing functions of $n$, and aim to accurately characterize the dependency of $M^*(n \mid \epsilon_n, g_n, r_n, W_n)$ on these sequences. For reasons that will be clear in what follows, we focus on the regime $r_n = \Theta(g_n)$. In this paper, we focus on the randomness stemming from the sampling channel, and thus assume that $W_n = I_n$, the noiseless kernel for all $n$ (we discuss possible extensions to noisy channels in Sec. V). For the error probability, we just assume that $\epsilon_n \to 0$ as $n \to \infty$, though in possibly arbitrarily slow rate. Our main theorem provides upper and lower bounds on the rate $\frac{1}{n}\log M^*(\cdot)$ in this regime.

In [14, Sec. 7.3], a closely-related channel model was considered, in which $r_n = g_n$, the sequencing is noiseless $W_n = I_n$, and the output is $Z^n = (Z_1, \ldots, Z_n)$ is such that conditioned on $X^n = x^n$, it holds that $Z_i \sim \text{Poisson}(x_i)$, and the $Z_i$'s are independent. Based on the known asymptotic capacity expression of the Poisson channel under an average-power constraint [16, Thm. 7], it was conjectured that the capacity of the frequency-based channel whose output is $Z^n$ is given by $\frac{1}{2}\log r_n + o_n(1)$. However, the capacity of the Poisson channel is asymptotically achieved by a gamma distributed input $X_n \sim \text{Gamma}(\frac{1}{2}, 2g_n)$ [16], [36], which is a continuous distribution, whereas the frequency-based channel only allows for integer inputs. Our bounds will be proved by modifying the multinomial output to a Poisson output. However, the analysis of the resulting Poisson channel, along with integer-input constraints will require additional technical steps.

## III. Main Result

Assume that reading operation is noiseless, and so $W = I_n$. The channel is given by

$$Y^n \sim \text{Multinomial}\left(nr_n, \frac{1}{\sum_{i=1}^n x_i} x^n\right). \tag{2}$$

For $\mu \in \mathbb{R}_+$, let

$$\Psi(\mu) := (\mu+1)\cdot \text{h}_{\text{bin}}\left(\frac{1}{\mu+1}\right), \tag{3}$$

which, as is well known, is the maximum entropy for non-negative integer-supported RVs with mean bounded by $\mu$ (see Lemma 14).

**Theorem 2.** *Assume $W_n = I_n$, that $g_n \to \infty$, and that $\underline{c} g_n \le r_n \le e g_n$ for some $\underline{c} \in (0, e)$.*

- *A (weak) converse bound: For any $\epsilon_n \to 0$*

$$\frac{1}{n}\log M^*(n \mid \epsilon_n, g_n, r_n, W_n) \le \frac{1}{2}\log\left[r_n \wedge (eg_n)\right] + o_n(1). \tag{4}$$

- *An achievability bound: Further assume that $n = \Omega(g_n^{1+\zeta})$ for some $\zeta > 0$. Then,*

$$\frac{1}{n}\log M^*(n \mid \epsilon_n, g_n, r_n, W_n) \ge \frac{1}{2}\log(r_n) - \Psi\left(\frac{r_n}{g_n}\right) + o_n(1). \tag{5}$$

*Comparison to the standard Poisson channel:* For a Poisson channel with an average-power input constraint $\mathbb{E}[X] \leq g_n$,[1] and gain $\frac{r_n}{g_n}$, that is, $Z \mid X = x \sim \text{Poisson}(\frac{r_n}{g_n}x)$, the capacity is asymptotically given by $\frac{1}{2}\log r_n + o_n(1)$, [16, Thm. 7].[2] Nonetheless, this rate is achieved with input distribution $X \sim \text{Gamma}(\frac{1}{2}, 2g_n)$, which is a continuous distribution, and is unsuitable for the frequency-based channel, which accepts non-negative *integer* inputs. This restriction on the input affects both the converse and the achievability bound. For the converse part, it leads to an upper bound $\frac{1}{2}\log(eg_n)$ on the maximal rate, which, as we discuss in what follows, is a result of the log-cardinality of the set of possible inputs. Thus, unlike the standard, continuous-input, Poisson channel, there is no motivation to increase $r_n$ beyond $eg_n$, at least in terms of rate. For the achievability bound, the restriction of the input to be integer valued leads to the loss additive term $\Psi(\frac{r_n}{g_n})$. This leads to a delicate issue: For the decoder, increasing $r_n$, the number of samples from the pool of objects, only leads to higher mutual information and rate, since due to the data-processing theorem, the decoder can always ignore output observations. However, in our model increasing $r_n$ also put a more restrictive constraint on the input integer constraint, and this has the opposite effect on the mutual information, as it does not allow to achieve the output entropy obtained as in the standard case. Thus, it is not obvious that increasing $r_n$ also increases the mutual information. To further inspect this, let us write the lower bound, without the asymptotically vanishing terms, as

$$\frac{1}{2}\log(g_n) + \frac{1}{2}\log\left(\frac{r_n}{g_n}\right) - \Psi\left(\frac{r_n}{g_n}\right). \tag{6}$$

We may then optimize it over $r_n \leq eg_n$. Interestingly, the function $\mu \to \frac{1}{2}\log(\mu) - \Psi(\mu)$ has a unique global maximum at $\mu \approx 0.398$ and equals $-1.295$[nats], which is better than its value at $\mu = 1$, given by $-1.386$[nats]. Thus, to optimize the lower bound of Theorem 2, the optimal choice is $r_n \approx 0.4g_n$, that is, *surely* not sampling some of the objects in the pool optimizes this bound. The optimized lower bound is then (in nats)

$$\frac{1}{n}\log M^*(n \mid \epsilon_n, g_n, r_n, W_n) \geq \frac{1}{2}\log(g_n) - 1.295 + o_n(1). \tag{7}$$

Naturally, an interesting open question is whether the rate can go beyond $\frac{1}{2}\log(g_n)$, and if it can match the upper bound, and what is the optimal value of $r_n$.

*Implication on the data stored in DNA storage systems:* As discussed, strictly speaking, the capacity of the DNA storage channel in the regime of interest is zero. Following [14, Sec. 7.1], let the pseudo-rate of a DNA storage be defined as

$$\tilde{R}_{\text{DNA}} := \frac{\log M}{LK^{\beta \log|\mathcal{A}|}}, \tag{8}$$

[1]For the Poisson channel, the average input power is modeled as $\mathbb{E}[X]$, as opposed to the more common Gaussian channel, in which the average input power is $\mathbb{E}[X^2]$.

[2]In [16, Thm. 7], the Poisson channel is assumed to have unity gain, i.e., $\frac{r_n}{g_n} = 1$. The capacity expression can be easily generalized to non-unity gain by scaling of input codewords.

and the pseudo-capacity be defined as the maximal achievable pseudo-rate such that $\epsilon_K \to 0$ as $K \to \infty$. We thus obtain the following corollary:

**Corollary 3.** *Assume that* $\beta > \frac{1}{2\log|\mathcal{A}|}$. *Then,*

$$\tilde{R}_{\text{DNA}} = \frac{1-\beta\log|\mathcal{A}|}{2\beta}. \tag{9}$$

Specifically, this settles [14, Conjecture 4], under the more demanding multinomial channel, yet under the restrictive constraint on $\beta \in (\frac{1}{2\log|\mathcal{A}|}, \frac{1}{\log|\mathcal{A}|})$, which excludes very short molecules.

*Proof:* For the DNA storage channel, the number of unique objects is the number of unique molecules of length $L_K = \beta \log K$, given by $n \equiv |\mathcal{A}|^{L_K} = K^{\beta\log|\mathcal{A}|}$, and the total number of objects is $ng_n \equiv K$, that is, $g_n = K^{1-\beta\log|\mathcal{A}|}$, whereas $N_K = r_n$. The converse bound of Theorem 2 then implies that

$$\frac{\log M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)}{K^{\beta\log|\mathcal{A}|}} \le \frac{1}{2}\log\left[N_K \wedge (eK^{1-\beta\log|\mathcal{A}|})\right] + o_K(1). \tag{10}$$

If, e.g., $N_K = K^{1-\beta\log|\mathcal{A}|}$ (that is, $r_n = g_n$) we get

$$\frac{\log M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)}{LK^{\beta\log|\mathcal{A}|}} \le \frac{1-\beta\log|\mathcal{A}|}{2\beta} + o\left(\frac{1}{\log K}\right). \tag{11}$$

The achievability bound of Theorem 2 requires the condition $n = \Omega(g_n^{1+\zeta})$, which translates into $\beta > \frac{1}{2\log|\mathcal{A}|}$, and then implies that

$$\frac{\log M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)}{LK^{\beta\log|\mathcal{A}|}} \ge \frac{1}{2\beta}\frac{\log(N_K)}{\log K} - \Psi\left(\frac{N_K}{K^{1-\beta\log|\mathcal{A}|}}\right) + o\left(\frac{1}{\log K}\right). \tag{12}$$

Using $N_K = K^{1-\beta\log|\mathcal{A}|}$ results

$$\frac{\log M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)}{LK^{\beta\log|\mathcal{A}|}} \ge \frac{1}{2\beta}\frac{\log(K^{1-\beta\log|\mathcal{A}|})}{\log K} - \frac{\Psi(1)}{\beta\log K} + o\left(\frac{1}{\log K}\right) \tag{13}$$

$$= \frac{1-\beta\log|\mathcal{A}|}{2\beta} - \frac{2.773}{2\beta}\cdot\frac{1}{\log K} + o\left(\frac{1}{\log K}\right). \tag{14}$$

Thus, in this regime for $\beta$ the converse and achievable bounds only differ in $O(\log^{-1} K)$ term. It should be mentioned that using the approximately optimized value of $N_K = 0.4K^{1-\beta\log|\mathcal{A}|}$ slightly improves the factor of $1/\log K$ from $\frac{2.773}{2\beta}$ to $\frac{2.59}{2\beta}$. ■

**Example 4.** Consider a DNA storage system with $|\mathcal{A}| = 4$. We compare the number of achievable bits, that is, the asymptotic lower bound on $\log M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)$ of (14) (without the $o(\log^{-1} K)$ term) to the total number of nucleotides, in the short-molecule regime. As mentioned in [14, Sec. 1], just 5[grams] of DNA contain about $KL = 4\cdot 10^{21}$ nucleotides. Thus, e.g., if $\beta = \frac{0.76}{\log(4)}$ then Fig. 2 shows that these 5[grams] store over $1.253\cdot 10^{16}$[nats] $= 1.8\cdot 10^{16}$[bits], while $L = \beta\,\mathrm{W}(\frac{KL}{\beta}) \approx 26$, where here W is the Lambert W function. This is a huge amount of stored data, while the molecule length is rather short and thus amenable for efficient

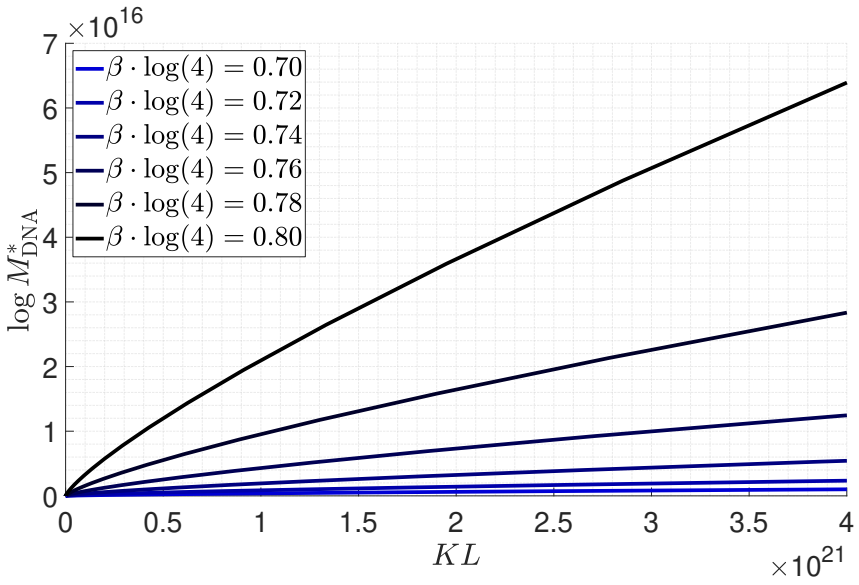


Figure 2. The lower bound on $\log M^*_{\text{DNA}}(L_K, V_{L_K}, N_K, \epsilon_K)$ of (14) vs. $KL$. A darker line corresponds to larger $\beta$.

synthesis and sequencing.

*The assumptions of the theorem:*

1) A simple application of the data-processing theorem implies that the converse bound is valid for any Markov kernel $W_n$, not just noiseless.
2) The condition $n = \Omega(g_n^{1+\zeta})$ can be relaxed to just $n = \omega(g_n)$. We have used the polynomial factor $\frac{n}{g_n} = \Omega(g_n^{\zeta})$ for simplicity of exposition.

### A. *Proof Outline of the Converse Bound of Theorem 2*

The proof follows the standard Fano's inequality, which requires bounding $I(X^n; Y^n)$. The frequency-based channel from $X^n$ to $Y^n$ is a multinomial channel $Y^n \sim \text{Multinomial}(nr_n, \hat{x}^n)$, which is not a memoryless channel, and so a direct analysis of the mutual information is difficult. Nonetheless, as is well known, the multinomial distribution can be converted into a Poisson distribution (see Appendix B). The Poisson distribution is memoryless, and so the resulting mutual information is amenable for evaluation. Specifically, we consider $Z^n$ to be the output of a memoryless Poisson channel, with the same input $X^n$, and then relate $I(X^n; Y^n)$ to $I(X^n; Z^n)$. An optimal input distribution for the Poisson channel is memoryless $P_{X^n} = P_X^{\otimes n}$, which, in turn, requires bounding a single-letter $I(X; Z)$. This is bounded using the known bounds on the average-power-constrained Poisson channel [16, Thm. 7], in the asymptotic regime of high power. This results in the term $\frac{1}{2}\log(r_n) + o_n(1)$ in the upper bound. Next, we also note that $I(X^n; Y^n) \le H(X^n)$, and that since $X^n$ is non-negative integer-valued and $\sum_{i=1}^n X_i \le ng_n$, this puts an immediate constraint on the cardinality of the alphabet of $X^n$, and thus on its entropy. Stirling's bound then shows that it is bounded as $\frac{n}{2}\log(eg_n)$, which results the term $\frac{1}{2}\log(eg_n) + o_n(1)$ in the upper bound.

### B. Steps of the Proof of the Achievability Bound of Theorem 2

The proof of achievability Theorem 2 is based on the three propositions that will be described next. Here we will state these propositions, and briefly outline their proofs. The detailed proof will appear in Sec. IV-B. The proof is based on Feinstein's maximal coding bound [53], which bounds the maximal error probability of the optimal codebook of a given cardinality via the cumulative distribution function (CDF) of the *information density* of the channel (i.e., the *information spectrum*). Concretely, we use the extended version stated in [54, Thm. 20.7], which also takes into account input constraints. Let $P_{Y^n|X^n}$ denote the Markov kernel from the input $X^n$ to the output $Y^n$, for which $Y^n \mid X^n = x^n \sim \text{Multinomial}(nr_n, \hat{x}^n)$. Let $P_{X^n}$ denote the input distribution, and $P_{Y^n}$ be the corresponding output distribution. Let the information density be

$$i(x^n; y^n) := \log \frac{P_{Y^n|X^n}(y^n \mid x^n)}{P_{Y^n}(y^n)}. \tag{15}$$

The extended Feinstein bound assures the following: For any $\gamma > 0$ and $M \in \mathbb{N}_+$ there exists a code $\mathcal{C}_M = \{x^n(1), \ldots, x^n(M)\}$ such that $x^n(j) \in F_n$ for all $j \in [M]$, and whose maximal error probability is $\epsilon_n$, where

$$\epsilon_n P_{X^n}(F_n) \le \mathbb{P}\left[i(X^n; Y^n) \le \log \gamma\right] + \frac{M}{\gamma}. \tag{16}$$

Here too, the fact that the channel from $X^n$ to $Y^n$ is not memoryless makes a direct analysis of the information spectrum challenging, and similarly, it is altered to a memoryless Poisson distribution. To obtain a useful bound, however, it is required to restrict the input distribution to a finite support. The result is that we show that there exists a codebook whose number of codewords is roughly $M \approx e^{nI(X;Z)}$, where $I(X;Z)$ is the mutual information of the Poisson channel, and the error probability upper bounded, by a bound which can be made vanishing. This first step is summarized in the following proposition.

**Proposition 5.** *Let $P_X$ be a distribution such that $\text{supp}(P_X) \subseteq [s_n] = \{1, 2, \ldots, s_n\}$ for some $s_n \in \mathbb{N}_+$. Let $F_n := \{x^n \in \mathbb{N}^n \colon \frac{1}{n}\sum_{i=1}^n x_i = g_n\}$ be a set of input vectors. Also let $Z \mid X = x \sim \text{Poisson}(\frac{r_n}{g_n} x)$, and let $\delta_n \in (0, \frac{r_n}{g_n} s_n)$, where $\frac{r_n}{g_n} s_n \ge 12\pi e^2$. Then, there exists a code $\mathcal{C}_M \subset F_n$ of $M$ codewords with*

$$\log M = nI(X;Z) - 3n\delta_n - \frac{1}{2}\log(6\pi n r_n), \tag{17}$$

*whose maximal error probability $\epsilon_n$ on the multinomial channel from $X^n$ to $Y^n$ is bounded as*

$$\epsilon_n \le \frac{11}{P_X^{\otimes n}(F_n)} \left[\sqrt{nr_n} \exp\left[-n\delta_n^2 \cdot \left(\frac{2}{\log^2 \frac{r_n s_n}{g_n}} \wedge \frac{g_n}{19 r_n s_n \log^2 s_n}\right)\right] + e^{-n\delta_n}\right]. \tag{18}$$

*Proof outline of Prop. 5:* Feinstein's bound is based on the information spectrum $\mathbb{P}[i(X^n; Y^n) \le \log \gamma]$ for $\gamma > 0$. In order to relate this to the information spectrum of a memoryless Poisson channel $(X^n, Z^n)$ we

first relate the information density $i(x^n;y^n)$ to that of Poisson, i.e., to

$$\tilde{i}(x^n;z^n) := \log \frac{P_{Z^n|X^n}(z^n \mid x^n)}{P_{Z^n}(z^n)}. \tag{19}$$

We show that the modification of the information density leads to an additive loss term in $\log M$ given by $\frac{1}{2}\log(6\pi n r_n)$, which will be negligible after normalizing by $n$. Second, we replace the randomness over $(X^n, Y^n)$ in the information spectrum with that of $(X^n, Z^n)$, using the Poissonization of the multinomial effect (see Fact 20). As a result, the analysis of the information spectrum of the channel from $X^n$ to $Y^n$ is altered to the analysis of the information spectrum $\mathbb{P}[\tilde{i}(X^n;Z^n) \le \log\gamma]$. Since the Poisson channel is memoryless, if we further restrict $P_{X^n}$ to a product distribution $P_X^{\otimes n}$, then $\tilde{i}(X^n;Z^n)$ is a sum of IID RVs, for which tail bounds can be readily derived. Before discussing the derivation of this bound, we highlight that the required bound should decay faster than its decay for standard analysis of memoryless channels. In the standard analysis, $\gamma$ is chosen so that $\mathbb{P}[\tilde{i}(X^n;Z^n) \le \log\gamma] \to 0$ as $n \to \infty$, albeit with an arbitrary slow rate (e.g., in the proof of [54, Thm. 19.8]). Here, this probability is multiplied by a term that scales as $\Theta(\frac{\sqrt{nr_n}}{P_X^{\otimes n}(F_n)})$, and so obtaining a vanishing upper bound on $\epsilon_n$ requires a bound which is $o(\frac{P_X^{\otimes n}(F_n)}{\sqrt{nr_n}})$. To obtain the desired upper bound, we separate the analysis of the randomness of $Z^n$ conditioned on $X^n = x^n$ from the randomness of $X^n$. To analyze the randomness over $Z^n$ conditioned on $X^n = x^n$, we note that $\tilde{i}(x^n;Z^n)$ is a function of $n$ independent Poisson RVs. We show that under the restricted support assumption, that is, $\mathrm{supp}(P_X) \subseteq [s_n]$, it holds that $\tilde{i}(x^n;Z^n)$ is a Lipschitz function with semi-norm $\log s_n$. In turn, this allows us to use the concentration bound of Lipschitz functions of Poisson RVs due to Bobkov and Ledoux [55, Prop. 11] (see Appendix D for a brief overview). Thus we show that $\tilde{i}(x^n;Z^n)$ is close to its expected value, denoted as $J(x^n) := \sum_{i=1}^n \mathbb{E}[\log P_{Z|X}(Z_i \mid x_i) \mid X_i = x_i] = \sum_{i=1}^n J(x_i)$.[3] Hence, under the choice of memoryless input distribution, $J(x^n)$ is also a sum of independent RVs. We then prove that $J(x) \in [-\log\frac{r_n s_n}{g_n}, 0]$, that is, $J(X^n)$ is a sum of *bounded* RVs. An application of Hoeffding's inequality then shows that $J(X^n)$ concentrates to its expected value $-H(Z \mid X)$. Combining the concentration results of both $J(X^n)$ and $\tilde{i}(x^n;Z^n)$ leads to an upper bound on $\mathbb{P}[\tilde{i}(X^n;Z^n) \le \log\gamma]$, then to an upper bound on $\mathbb{P}[i(X^n;Y^n) \le \log\gamma]$, and finally, to the claimed upper bound on the error probability $\epsilon_n$, via Feinstein's bound. ■

Further evaluation of the Feinstein-based bound (18) in Prop. 5 requires two tasks: First, evaluating the mutual information $I(X;Z)$ over the Poisson channel, and second, evaluating the probability that a randomly chosen codeword meets the constraint, that is $P_X^{\otimes n}(F_n)$. This is the content of the next two propositions, beginning with the former.

Let $\tilde{X}$ be a continuous input RV. As mentioned, under the average-power input constraint $\mathbb{E}[\tilde{X}] \le g_n$, the optimal input distribution for a Poisson channel is $\tilde{X} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$ [16], [36]. This is a continuous

[3] With a slight abuse of notation, we use $J(\cdot)$ for both scalar and vector inputs, with the common convention that the value of vector inputs is the sum of the value of the scalar function of the coordinates.

distribution supported on $\mathbb{R}_+$, and thus unsuitable to the frequency-based channel, which accepts non-negative *integer* inputs. Furthermore, the bound of Prop. 5 is based on the assumption that $\mathrm{supp}(P_X) \subseteq [s_n]$, where $s_n \in \mathbb{N}_+$ is finite. In order to obtain a valid lower bound on the mutual information, we modify the gamma distribution of $\tilde{X}$ by first truncating (or restricting) it to a judicious choice of interval $\mathcal{S}_n$, and then rounding it to be integer valued so that the resulting RV is supported on $[s_n]$. We will use the following definition for truncation:

**Definition 6.** Let $A$ be a real RV, and let $\mathcal{S} \subset \mathbb{R}$ be such that $\mathbb{P}[A \in \mathcal{S}] > 0$. The truncation of $A$ to a support $\mathcal{S}$ is the RV $A_{|\mathcal{S}}$ which satisfies that for any Borel set $\mathcal{A} \in \mathfrak{B}(\mathbb{R})$,

$$\mathbb{P}\left[A_{|\mathcal{S}} \in \mathcal{A}\right] = \frac{\mathbb{P}\left[A_{|\mathcal{S}} \in \mathcal{A} \cap \mathcal{S}\right]}{\mathbb{P}\left[A \in \mathcal{S}\right]}. \tag{20}$$

**Proposition 7.** *Assume that $g_n \to \infty$, and that $\underline{c} g_n \le r_n \le e g_n$ for some $\underline{c} \in (0, e)$. Let $\rho \in (0,1)$ be given, and consider the interval*

$$\mathcal{S}_n = \left[\frac{1}{g_n^{1+3\rho}}, g_n^{1+\rho}\right] \in \mathbb{R}_+. \tag{21}$$

*Let $\tilde{X} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$, and let $X = \lceil \tilde{X}_{|\mathcal{S}_n} \rceil$, i.e., $\tilde{X}$ is first truncated to $\mathcal{S}_n$ and then rounded upward to the nearest integer. Further let $Z \mid X = x \sim \mathrm{Poisson}(\frac{r_n}{g_n} x)$ for $x \in \mathbb{R}_+$. Then, there exists $n_0$ (which depends on $(\underline{c}, \rho)$ and $\{g_n\}$), such that for all $n \ge n_0$*

$$I(X;Z) \ge \frac{1}{2}\log r_n - \Psi\left(\frac{r_n}{g_n}\right) - o_n(1). \tag{22}$$

*Proof outline of Prop. 7:* The proof bounds the loss in the achievable mutual information $I(X;Z)$ when the asymptotically ideal $\tilde{X} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$ is truncated to $\mathcal{S}_n$ and then upward rounded to an integer. We begin by analyzing $\overline{X} := \tilde{X}_{|\mathcal{S}_n}$. A direct analysis of the reduction in mutual information when modifying $\tilde{X}$ to $\overline{X}$ appears to be cumbersome, and we thus take an indirect route, which exploits the relation between mutual information and optimal estimation over the Poisson channel [37], [41], [42], [49], [50], [56], [57], [58, Ch. 8], which we next briefly review. Let $\ell(u,v) \equiv \ell_{\mathrm{Poi}}(u,v) := v - u + u\log\frac{u}{v}$ be the Poisson error function. For a positive random variable $U$, we let $V \mid U = u \sim \mathrm{Poisson}(u)$. Let $\hat{U}$ be an estimator of $U$ based on $V$. Then, since $\ell(u,v)$ is the Bregman divergence [59] associated with the Poisson distribution, it holds that the minimal estimation error is obtained by the expected mean $\mathbb{E}[U \mid V]$ and the *minimum mean Poisson error* (MMPE) is

$$\mathrm{mmpe}(U) = \min_{\hat{U}} \mathbb{E}\left[\ell(U, \hat{U})\right] \tag{23}$$

$$= \mathbb{E}\left[\ell\left(U, \mathbb{E}[U \mid V]\right)\right] \tag{24}$$

$$= \mathbb{E}\left[U \log \frac{U}{\mathbb{E}[U \mid V]}\right]. \tag{25}$$

The following relation between the MMPE and the mutual information was established in [56, Corollary 1]:

**Theorem 8** (I-MMPE relation [56, Corollary 1]). *Assume that* $\mathbb{E}[U \log U] < \infty$ *and let* $V_a \mid U = u \sim \mathrm{Poisson}(au)$ *for* $a > 0$. *Then,*

$$I(U; V_\gamma) = \int_0^\gamma \mathrm{mmpe}(aU) \cdot \frac{\mathrm{d}a}{a}. \tag{26}$$

Following a similar analysis for the Gaussian channel [60, Lemma 2], we analyze the difference between $\mathrm{mmpe}(aU)$ and $\mathrm{mmpe}(a\overline{U})$, where $\overline{U}$ is a truncated version of $U$ (Lemma 12). We show that this difference is controlled by three terms

$$\gamma \cdot s_{\max} \cdot \mathbb{P}\left[U \in [0, s_{\min})\right] + \gamma \mathbb{E}\left[U \log U \cdot \mathbb{1}\left\{U \in (s_{\max}, \infty)\right\}\right] + \gamma \mathbb{E}\left[U \log \frac{1}{s_{\min}} \cdot \mathbb{1}\left\{U \in (s_{\max}, \infty)\right\}\right] \tag{27}$$

where $\mathcal{S} := [s_{\min}, s_{\max}]$ is the set used for truncation. Accordingly, we show that this difference is small for the specific gamma distribution of $\tilde{X}$ and the truncation set of interest (Lemma 13) when $U$ follows the distribution of $\tilde{X} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$. The difference between the MMPE of $\tilde{X}$ and $\overline{X}$ is then translated, via the I-MMPE relation in Theorem 8, to a difference between their mutual information, which is eventually shown to be a negligible $o_n(1)$ term in the regime of interest. Thus, there is no essential loss in mutual information due to the truncation of the gamma distribution.

Next, we consider the influence of rounding $\overline{X}$ to the integer $X = \lceil \overline{X} \rceil$. In contrast to truncation, it appears that the rounding operation leads to a loss in the mutual information, and we upper bound this loss as $\Psi(\frac{r_n}{g_n})$ (Lemma 15). This is the source of the additive loss term $\Psi(\frac{r_n}{g_n})$ that appears in the statement of the proposition. Specifically, from $I(X;Z) = H(Z) - H(Z \mid X)$ and $I(\overline{X}; \overline{Z}) = H(\overline{Z}) - H(\overline{Z} \mid \overline{X})$ with $\overline{Z} \mid \overline{X} = \overline{x} \sim \mathrm{Poisson}(\frac{r_n}{g_n}\overline{x})$, we may compare the mutual information values by separately comparing the conditional entropy values and the output entropy values. First, we use properties of the entropy of the Poisson PMF and the input gamma distribution to show that $H(\overline{Z} \mid \overline{X}) \leq H(Z \mid X) + o_n(1)$. Second, we show that $H(\overline{Z})$ is only larger than $H(Z)$ by at most $\Psi(\frac{r_n}{g_n})$. The proof of this result relies on the infinite divisibility property of the Poisson distribution. By writing $X = \overline{X} + D$, where $D \in [0, 1]$, we may also write $Z = \overline{Z} + \acute{Z}$ where $\acute{Z} \mid D = d \sim \mathrm{Poisson}(\frac{r_n}{g_n} d)$ (note, however, that $Z$ are $\acute{Z}$ are not independent). We then use the bounding method used, e.g., in [61, Prop. 8]. We relate $H(\overline{Z}) - H(Z)$ to the *maximum entropy* that is possible for a non-negative integer-valued RV whose expectation is less than $\mathbb{E}[Z - \overline{Z}]$. This maximum entropy is well-known to be the entropy of a proper geometric RV (Lemma 14), given by the function $\Psi(\mu)$, assuming that the allowed mean is $\mu$. Combining the comparison between the values of the conditional entropy and the values of the output entropy leads to the required comparison between the mutual information values. The final bound is obtained by combining the effects of both the truncation and the rounding. ■

Following the analysis of $I(X;Z)$ for a proper choice of $X$, it remains to bound $P_{X^n}(F_n)$. Since the term multiplying $P_{X^n}(F_n)$ in (18) decays super-polynomially with $n$, the error probability $\epsilon_n$ may decay to zero, even if $P_{X^n}(F_n)$ decays to zero, although polynomially.

**Proposition 9.** *Let $\zeta > 0$ and $\rho \in (0, \frac{1}{4} \vee \frac{2\zeta}{3})$ be given. For $i \in [n]$, let $\{\tilde{X}_i\}_{i\in[n]}$ be IID with $\tilde{X}_i \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$, and let $X_i = \lceil \tilde{X}_{|\mathcal{S}_n} \rceil$ where $\mathcal{S}_n = [g_n^{-(1+3\rho)}, g_n^{1+\rho}]$. Let $F_n(\tau) := \{x^n \in \mathbb{N}^n \colon \frac{1}{n}\sum_{i=1}^n x_i = \tau\}$. Assume that $n = \Omega(g_n^{1+\zeta})$. Then, there exists a sequence $\varsigma_n = o_n(1)$ and $\tau_n \in [g_n(1+\varsigma_n)]$ such that*

$$P_{X^n}\left[F_n(\tau_n)\right] \geq \frac{1}{3ng_n} \tag{28}$$

*for all $n$ sufficiently large.*

*Proof outline of Prop. 9:* The main technical aspect of the proof is to find a right-tail bound on $\sum_{i=1}^n \overline{X}_i$, where $\{\overline{X}_i\}_{i\in[n]}$ are IID, and each is distributed according to the $\mathrm{Gamma}(\frac{1}{2}, 2g_n)$ distribution *truncated* to $\mathcal{S}_n = [g_n^{-(1+3\rho)}, g_n^{1+\rho}]$. The gist of the proof is to define a proper generative model for $\{\overline{X}_i\}_{i\in[n]}$. To this end, we define $\{\tilde{X}_{i,j}\}_{i\in[n],j\in\mathbb{N}_+}$ to be a double-index array of IID RV, distributed according to $\tilde{X}_{i,j} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$. We then define $\overline{X}_i = \tilde{X}_{J^*(i)}$ where $J^*(i) \in \mathbb{N}_+$ is the minimal index such that $\tilde{X}_{i,j} \in \mathcal{S}_n$. Evidently, $\{\overline{X}_i\}_{i\in[n]}$ are IID, and have the required truncated gamma distribution. Now, using the fact that $\mathcal{S}_n$ is a high probability set, we show that, with high probability, $\tilde{X}_{i,1}$ is in $\mathcal{S}_n$ for most indices, and so $\overline{X}_i = \tilde{X}_1$ for most indices, which we denote by $n - \ell$. For the remaining $\ell$ indices, it holds that $\overline{X}_i \leq g_n^{1+\rho}$, and so their effect on $\frac{1}{n}\sum_{i=1}^n \overline{X}_i$ is controlled. Consequently, loosely speaking, $\sum_{i=1}^n \overline{X}_i$ is dominated by $\sum_{i=1}^n \tilde{X}_{i,1}$, up to terms which are eventually negligible. Taking into account the number of possibilities for $\{J^*(i)\}_{i\in[n]}$ (under the high probability event), and using standard tail bounds on the gamma distribution, we show that $\sum_{i=1}^n \overline{X}_i \leq ng_n(1+\bar{\varsigma}_n)$ with probability larger than, say, $1/2$. In turn, this is also true for the upward rounded $X_i = \lceil \overline{X}_i \rceil$ with $\bar{\varsigma}_n$ replaced by $\varsigma_n$ which is essentially the same. Since $\sum_{i=1}^n X_i$ is an integer, roughly upper bounded by $ng_n$, there must exists $n\tau_n \in \mathbb{N}_+$, with $\tau_n \leq g_n(1+\varsigma_n)$ such that the probability of $F_n(\tau_n)$, as defined in the proposition, is $\Omega(\frac{1}{ng_n})$, as claimed. ∎

The proof of the achievability bound of Theorem 2 then directly combines the above propositions: First, Prop. 5 leads to a Feinstein-based bound on the number of codewords $M$ and error probability $\epsilon_n$, which depends on $I(X;Z)$ of the Poisson channel with integer inputs, and the probability of an input vector with a fixed sum $\tau_n$ below $ng_n$. Prop. 7 lower bounds $I(X;Z)$, and Prop. 9 bounds $P_{X^n}[F_n(\tau_n)]$. Analyzing the leading terms in the resulting mutual information and error probability leads to the claimed result.

## IV. Proof of Theorem 2

### A. Proof of the converse bound of Theorem 2

*Proof of the converse bound of Theorem 2:* Let $\mathcal{C}_M$ be a code of cardinality $M$, whose maximal error probability is $\epsilon_n$, and for which each codeword $x^n(j) \in \mathcal{C}_M$ satisfies $\sum_{i=1}^n x_i(j) = n\underline{g}_n$ for some $\underline{g}_n \leq g_n$. By

Fano's inequality (e.g. [54, Thm. 20.6]) it holds that

$$\log M \le \frac{1}{1-\epsilon_n}\left(\mathrm{h_{bin}}(\epsilon_n) + \sup_{P_{X^n}:\mathrm{supp}(P_{X^n})\subset\mathbb{N}^n,\, P_{X^n}\left(\frac{1}{n}\sum_{i=1}^n X_i=\underline{g}_n\right)=1} I(X^n;Y^n)\right) \tag{29}$$

where $Y^n \sim \mathrm{Multinomial}(nr_n, \frac{1}{n\underline{g}_n}X^n)$. We evaluate this bound by further analyzing $I(X^n;Y^n)$. Set $\eta\in(0,1)$, and let $Z^n=(Z_1,\ldots,Z_n)$ be a vector of independent components, such that $Z_i \mid X_i=x_i \sim \mathrm{Poisson}(\frac{1}{1-\eta}\cdot\frac{r_n}{\underline{g}_n}x_i)$. We next relate $I(X^n;Z^n)$ to $I(X^n;Y^n)$. Let $Q:=\sum_{i=1}^n Z_i \sim \mathrm{Poisson}(\frac{1}{1-\eta}nr_n)$ be the random number of output objects in the Poisson model. Let $\{S_i\}_{i=1}^\infty$ be drawn as in the problem formulation (Sec. II-B), and with a slight abuse of notation, let $Y^n(q)\sim\mathrm{Multinomial}(q,\frac{1}{\underline{g}_n}X^n)$ be a sequence of RVs, with the coupling that $Y^n(q)$ is the histogram of $S^q$. Thus, $Y^n \overset{d}{=} Y^n(nr_n)$. Furthermore, since $\mathbb{P}[S_1,S_2,\ldots,S_q \mid X^n=x^n] = \mathbb{P}[S_{\pi(1)},S_{\pi(2)},\ldots,S_{\pi(q)} \mid X^n=x^n]$ for any permutation $\pi$ in the symmetric group of degree $q$, it holds that $I(X^n;S^q)=I(X^n;Y^n(q))$. The data-processing inequality then implies that for any $q_1\le q_2$

$$I(X^n;Y^n(q_1)) = I(X^n;S^{q_1}) \le I(X^n;S^{q_2}) = I(X^n;Y^n(q_2)). \tag{30}$$

Now, from the Poissonization of the multinomial distribution effect (Fact 20) it holds that $Z^n \mid X^n, Q=q \overset{d}{=} Y^n(q)\mid X^n$ for any $q\in\mathbb{N}$. Hence,

$$\begin{aligned}
I(X^n;Z^n) &= I(X^n;Z^n,Q) && (31)\\
&= I(X^n;Q) + I(X^n;Z^n\mid Q) && (32)\\
&\ge I(X^n;Z^n\mid Q) && (33)\\
&= \sum_{q=0}^\infty \mathbb{P}[Q=q]\cdot I(X^n;Y^n(q)) && (34)\\
&\ge \sum_{q=nr_n}^\infty \mathbb{P}[Q=q]\cdot I(X^n;Y^n(q)) && (35)\\
&\overset{(a)}{\ge} \mathbb{P}[Q\ge nr_n]\cdot I(X^n;Y^n) && (36)\\
&\overset{(b)}{\ge} \left(1-e^{-\frac{\eta^2}{2(1-\eta)}nr_n}\right)\cdot I(X^n;Y^n), && (37)
\end{aligned}$$

where $(a)$ follows from the monotonicity property in (30), and $(b)$ from Chernoff's bound for Poisson RVs (Lemma 22), which implies that $\mathbb{P}[Q\ge nr_n]\ge 1-\exp[-\frac{\eta^2}{2(1-\eta)}nr_n]$ for any $\eta\in(0,1)$. Therefore,

$$\begin{aligned}
&\sup_{P_{X^n}:\mathrm{supp}(P_{X^n})\subset\mathbb{N}^n,\, P_{X^n}\left(\frac{1}{n}\sum_{i=1}^n X_i=\underline{g}_n\right)=1} I(X^n;Y^n)\\
&\le \sup_{P_{X^n}:\mathrm{supp}(P_{X^n})\subset\mathbb{N}^n,\, \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i]=\underline{g}_n} I(X^n;Y^n) && (38)\\
&\le \sup_{P_{X^n}:\mathrm{supp}(P_{X^n})\subset\mathbb{N}^n,\, \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i]\le\underline{g}_n} I(X^n;Y^n) && (39)
\end{aligned}$$

$$\overset{(a)}{\leq} \frac{1}{1-e^{-\frac{\eta^2}{2(1-\eta)}nr_n}} \sup_{P_{X^n}:\mathrm{supp}(P_{X^n})\subset\mathbb{R}^n,\, \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i]\leq \underline{g}_n} I(X^n;Z^n) \tag{40}$$

$$\overset{(b)}{=} \frac{1}{1-e^{-\frac{\eta^2}{2(1-\eta)}nr_n}} \sup_{P_{X^n}:\mathrm{supp}(P_{X^n})\subset\mathbb{R}^n,\, \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i]\leq \frac{r_n}{1-\eta}} I(X^n;\hat{Z}^n) \tag{41}$$

$$= \frac{1}{1-e^{-\frac{\eta^2}{2(1-\eta)}nr_n}} n \cdot \sup_{P_X:\mathrm{supp}(P_X)\subset\mathbb{R}^n,\, \mathbb{E}[X]\leq \frac{r_n}{1-\eta}} I(X;\hat{Z}) \tag{42}$$

$$\overset{(c)}{=} n\left(\frac{1}{1-e^{-\frac{\eta^2}{2(1-\eta)}nr_n}}\right)\cdot\left[\frac{1}{2}\log\left(\frac{r_n}{1-\eta}\right)+o_{r_n}(1)\right] \tag{43}$$

$$= n\left(\frac{1}{1-e^{-\frac{\eta^2}{2(1-\eta)}nr_n}}\right)\cdot\left[\frac{1}{2}\log(r_n)-\frac{1}{2}\log(1-\eta)+o_{r_n}(1)\right] \tag{44}$$

$$= n\cdot\left[\frac{1}{2}\log(r_n)+\frac{e^{-\frac{\eta^2}{2(1-\eta)}nr_n}}{1-e^{-\frac{\eta^2}{2(1-\eta)}nr_n}}\frac{1}{2}\log r_n-\left(\frac{1}{1-e^{-\frac{\eta^2}{2(1-\eta)}nr_n}}\right)\frac{1}{2}\log(1-\eta)+o_{r_n}(1)\right], \tag{45}$$

where $(a)$ follows from (37), in $(b)$ we have defined $\hat{Z}_n$ as a Poisson channel with a unity gain, that is $\hat{Z}_i \mid X_i = x_i \sim \mathrm{Poisson}(x_i)$, $(c)$ follows from the asymptotic expression of the mean-constrained Poisson channel capacity in [16, Thm. 7, Eq. (23)]. Choosing $\eta \equiv \eta_n = (nr_n)^{1/2-\rho}$ for some $\rho \in (0, 1/2)$ shows that if $\epsilon_n \to 0$ as as $n \to \infty$ then it must hold that

$$\log M \leq \frac{1}{2}n\left[\log(r_n)+o_n(1)\right]. \tag{46}$$

Next, it also holds that

$$\sup_{P_{X^n}:\mathrm{supp}(P_{X^n})\subset\mathbb{N}^n,\, P_{X^n}\left(\frac{1}{n}\sum_{i=1}^n X_i=\underline{g}_n\right)=1} I(X^n;Y^n)$$

$$\leq \sup_{P_{X^n}:\mathrm{supp}(P_{X^n})\subset\mathbb{N}^n,\, P_{X^n}\left(\frac{1}{n}\sum_{i=1}^n X_i=\underline{g}_n\right)=1} H(X^n) \tag{47}$$

$$\leq \log\left|\left\{x^n\subset\mathbb{N}^n\colon\ \frac{1}{n}\sum_{i=1}^n x_i=\underline{g}_n\right\}\right| \tag{48}$$

$$\overset{(a)}{=} \log\binom{n\underline{g}_n+n-1}{n-1} \tag{49}$$

$$\overset{(b)}{\leq} \log\binom{ng_n+n-1}{n-1} \tag{50}$$

$$\overset{(c)}{=} \log\left[\left(\frac{(ng_n+n-1)e}{n-1}\right)^{n-1}\frac{1}{\sqrt{2\pi(n-1)}}\exp\left(-\frac{(n-1)^2(1+o(1))}{2(ng_n+n-1)}\right)\right]+o(1) \tag{51}$$

$$\leq n\left[\log eg_n+o(1)\right], \tag{52}$$

where $(a)$ follows from the stars and bars model, $(b)$ follows since $\underline{g}_n \leq g_n$ and the monotonicity of the binomial coefficient, $(c)$ follows from Stirling's approximation of the binomial coefficient (see (A.2) in Fact 17, Appendix A). Combining both (45) and (52) in Fano's inequality (29) results the claimed bound. ■

### B. Proof of the Achievability Bound of Theorem 2

In this section, we prove the achievability bound of Theorem 2. To this end, we prove Props. 5, 7 and 9 one after the other, and then combine them in order to complete the proof of the bound.

*Proof of Prop. 5:* Our goal is to analyze the probability on the right-hand side of (16), which is typically simple whenever $i(X^n;Y^n)$ is a sum of IID RVs. To approach this, we further choose a scalar distribution $P_X$, and, as common, restrict $P_{X^n}$ to be the product distribution $P_{X^n} = P_X^{\otimes n}$. However, even under this choice, $i(X^n;Y^n)$ is not a sum of IID RVs, since $P_{Y^n|X^n}$ is not a memoryless channel. We will transform the analysis of this probability to the analysis of sum of IID RVs in two steps, first by relating $i(x^n;y^n)$ to the information density of a memoryless Poisson channel, and second, by relating the probability of events under the original channel to the probability of events under this Poisson channel. Concretely, let $Z^n$ be the output of a channel, such that conditioned on $X^n = x^n$ it holds that $Z_i \sim \mathrm{Poisson}(\frac{r_n}{g_n} x_i)$, and where the components of $Z^n$ are independent. Let $P_{Z^n|X^n}$ denote the Markov kernel from the input $X^n$ to the output $Z^n$. Recall that $Q = \sum_{i=1}^n Z_i$, and $\sum_{i=1}^n Y_i = nr_n$ with probability 1. Then, for any $x^n \in \mathbb{N}^n$ and $y^n$ with $\sum_{i=1}^n y_i = nr_n$, it holds that

$$\frac{P_{Y^n|X^n}(y^n \mid x^n)}{P_{Z^n|X^n}(y^n \mid x^n)} \overset{(a)}{=} \frac{P_{Z^n|X^n,Q}(y^n \mid x^n, nr_n)}{P_{Z^n|X^n}(y^n \mid x^n)} \tag{53}$$

$$= \frac{P_{Z^n|X^n,Q}(y^n \mid x^n, nr_n)}{\sum_{q=0}^{\infty} P_{Q|X^n}(q \mid x^n) \cdot P_{Z^n|X^n,Q}(y^n \mid x^n, q)} \tag{54}$$

$$\overset{(b)}{=} \frac{1}{P_{Q|X^n}(nr_n \mid x^n)} \tag{55}$$

$$\overset{(c)}{=} \frac{(nr_n)!}{(nr_n)^{nr_n} e^{-nr_n}}, \tag{56}$$

where $(a)$ follows from the Poissonization of the multinomial (Fact 20), $(b)$ follows since

$$P_{Z^n|X^n,Q}(y^n \mid x^n, q) = 0 \tag{57}$$

if $q \neq \sum_{i=1}^n y_i = nr_n$, and $(c)$ follows since $Q \mid X^n = x^n \sim \mathrm{Poisson}(nr_n)$ (i.e., $Q$ is independent of $X^n$). Now, Stirling's bound (Fact 16 in Appendix A) implies that, with probability 1

$$\sqrt{2\pi nr_n} \leq \frac{P_{Y^n|X^n}(y^n \mid x^n)}{P_{Z^n|X^n}(y^n \mid x^n)} \leq \sqrt{6\pi nr_n}. \tag{58}$$

Then using $\sum a_i / \sum b_i \geq \min_i(a_i/b_i)$ for reals $\{(a_i, b_i)\}$, we also have

$$\frac{P_{Z^n}(y^n)}{P_{Y^n}(y^n)} = \frac{\sum_{x^n} P_{X^n}(x^n) P_{Z^n|X^n}(y^n \mid x^n)}{\sum_{x^n} P_{X^n}(x^n) P_{Y^n|X^n}(y^n \mid x^n)} \geq \frac{1}{\sqrt{6\pi nr_n}}. \tag{59}$$

So, for any $nr_n \geq 2$,

$$\mathbb{P}\left[i(X^n;Y^n) \leq \log \gamma\right]$$

$$= \mathbb{P}\left[\log \frac{P_{Y^n|X^n}(Y^n \mid X^n)}{P_{Y^n}(Y^n)} \leq \log \gamma\right] \tag{60}$$

$$= \mathbb{P}\left[\log \frac{P_{Z^n|X^n}(Y^n \mid X^n)}{P_{Z^n}(Y^n)} + \log \frac{P_{Y^n|X^n}(Y^n \mid X^n)}{P_{Z^n|X^n}(Y^n \mid X^n)} + \log \frac{P_{Z^n}(Y^n)}{P_{Y^n}(Y^n)} \leq \log \gamma\right] \tag{61}$$

$$\overset{(a)}{\leq} \mathbb{P}\left[\log \frac{P_{Z^n|X^n}(Y^n \mid X^n)}{P_{Z^n}(Y^n)} - \frac{1}{2}\log(6\pi n r_n) \leq \log \gamma\right] \tag{62}$$

$$\overset{(b)}{\leq} e\sqrt{n r_n}\mathbb{P}\left[\log \frac{P_{Z^n|X^n}(Z^n \mid X^n)}{P_{Z^n}(Z^n)} - \frac{1}{2}\log(6\pi n r_n) \leq \log \gamma\right] \tag{63}$$

$$\overset{(c)}{=} e\sqrt{n r_n}\mathbb{P}\left[\sum_{i=1}^{n} \log \frac{P_{Z|X}(Z_i \mid X_i)}{P_Z(Z_i)} \leq \log \gamma + \frac{1}{2}\log(6\pi n r_n)\right], \tag{64}$$

where $(a)$ follows since $\log\sqrt{2\pi n r_n} \geq 0$, $(b)$ follows since the probability of any event of a multinomial is upper bounded, with a proper factor, by the probability of that event under its Poissonized version [62, Thm. 5.7 and Corollary 5.9] (see Lemma 21 in Appendix B), and $(c)$ holds since $P_{Z^n|X^n}(Z^n \mid X^n)$ is a product Markov kernel, which combined with the restriction $P_{X^n} = P_X^{\otimes n}$ results that $\log \frac{P_{Z^n|X^n}(Z^n|X^n)}{P_{Z^n}(Z^n)}$ is decomposed to a sum of IID RVs.

We continue to upper bound the probability in (64) over $(X^n, Z^n)$. Due to the pre-factor $\Theta(\sqrt{n r_n})$, we will need to show that this probability decays sufficiently fast in order to obtain a sufficiently strong bound on the probability in the extended Feinstein bound (16). We again bound in two steps. First, we condition on $X^n = x^n$, and analyze the probability with respect to (w.r.t.) the randomness of $Z^n$ and second, we analyze the resulting upper bound w.r.t. the randomness of $X^n$.

We begin with the first step, for which we recall that the $\mathrm{supp}(P_X) \subseteq [s_n] = \{1, 2, \ldots, s_n\}$, and specifically, that $P_X(0) = 0$. We use this assumption to establish that for any $x^n \in \mathrm{supp}(P_X^{\otimes n})$, the RV

$$f_{x^n}(Z^n) := \sum_{i=1}^{n} \log \frac{P_{Z|X}(Z_i \mid x_i)}{P_Z(Z_i)} \tag{65}$$

concentrates fast around its expected value

$$I(Z^n; X^n = x^n) := \sum_{i=1}^{n} \mathbb{E}\left[\log \frac{P_{Z|X}(Z_i \mid x_i)}{P_Z(Z_i)} \,\middle|\, X_i = x_i\right]. \tag{66}$$

We achieve this using a concentration bound of Lipschitz functions of Poisson RVs due to Bobkov and Ledoux [55, Prop. 11] stated in Lemma 27 (Appendix D). The result is as follows:

**Lemma 10.** *Assume that* $\mathrm{supp}(P_X) \subseteq [s_n]$ *for some* $s_n \in \mathbb{N}_+$. *Let* $x^n \in ([s_n])^{\otimes n}$. *Then, for any* $\delta \in (0, \frac{r_n}{g_n} s_n)$

$$\mathbb{P}\left[f_{x^n}(Z^n) < I(Z^n; X^n = x^n) - n\delta \mid X^n = x^n\right] \leq \exp\left[-n\frac{g_n \delta^2}{19 r_n s_n \log^2 s_n}\right]. \tag{67}$$

*Proof:* To establish this, we begin by showing that if $x^n \in ([s_n])^{\otimes n}$ then $f_{x^n}(z^n)$ is Lipschitz with semi-norm $\beta = \log s_n$, as follows. We denote by $e^n(i) = (0, 0, \ldots, 1, 0..)$ the $i$th standard basis vector in $\mathbb{R}^n$. Let

$Z \mid X = x \sim \text{Poisson}(\frac{r_n}{g_n}x)$. Then, it holds for any $x \in \mathbb{R}_+$ and $z \in \mathbb{N}$ that

$$\frac{P_{Z|X}(z+1 \mid x)}{P_{Z|X}(z \mid x)} = \frac{e^{-r_n x/g_n}\left(\frac{r_n x}{g_n}\right)^{z+1}}{(z+1)!} \cdot \frac{z!}{e^{-r_n x/g_n}\left(\frac{r_n x}{g_n}\right)^{z}} \tag{68}$$

$$= \frac{r_n}{g_n}\frac{x}{z+1}. \tag{69}$$

Let $P_Z$ be the marginal resulting from $P_X \otimes P_{Z|X}$. Then, similarly,

$$\frac{P_Z(z)}{P_Z(z+1)} = \frac{\sum_{\tilde{x}\in\text{supp}(P_X)} P_X(\tilde{x})P_{Z|X}(z \mid \tilde{x})}{\sum_{\tilde{x}\in\text{supp}(P_X)} P_X(\tilde{x})P_{Z|X}(z+1 \mid \tilde{x})} \tag{70}$$

$$\leq \max_{\tilde{x}\in\text{supp}(P_X)} \frac{P_{Z|X}(z \mid \tilde{x})}{P_{Z|X}(z+1 \mid \tilde{x})} \tag{71}$$

$$= \max_{\tilde{x}\in\text{supp}(P_X)} \frac{g_n}{r_n}\frac{z+1}{\tilde{x}}. \tag{72}$$

Hence,

$$\frac{P_{Z|X}(z+1 \mid x)}{P_{Z|X}(z \mid x)}\frac{P_Z(z)}{P_Z(z+1)} \leq \max_{\tilde{x}\in\text{supp}(P_X)} \frac{x}{\tilde{x}} \leq s_n. \tag{73}$$

Analogously, we can prove that

$$\frac{P_{Z|X}(z+1 \mid x)}{P_{Z|X}(z \mid x)}\frac{P_Z(z)}{P_Z(z+1)} \geq \min_{\tilde{x}\in\text{supp}(P_X)} \frac{x}{\tilde{x}} \geq 1. \tag{74}$$

Thus, for any $x \in \text{supp}(P_X) \subseteq [s_n]$ and $z \in \mathbb{N}$

$$\left|\log \frac{P_{Z|X}(z+1 \mid x)}{P_Z(z+1)} - \log \frac{P_{Z|X}(z \mid x)}{P_Z(z)}\right| \leq \log s_n. \tag{75}$$

The additive form of $f_{x^n}(z^n)$ then implies that

$$\max_{z^n\in\mathbb{N}^n} |f_{x^n}(z^n + e^n(i)) - f_{x^n}(z^n)| \leq \log s_n. \tag{76}$$

This Lipschitz property results in a left-tail concentration of $f_{x^n}(Z^n)$, by invoking a variant of the Bobkov–Ledoux concentration inequality [55, Prop. 11] (see Lemma 27) on the function $-f_{x^n}(Z^n)$ of the Poisson RVs $Z_i \mid X_i = x \sim \text{Poisson}(\frac{r_n}{g_n}x_i)$. Specifically, since (76) implies that $f_{x^n}(z^n)$ is Lipschitz with semi-norm $\beta = \log s_n$, Lemma 27 results

$$\mathbb{P}\left[f_{x^n}(Z^n) - I(Z^n; X^n = x^n) < -n\delta \mid X^n = x^n\right] \leq \exp\left[-n \cdot \frac{\delta^2}{16\beta^2\overline{\lambda} + 3\beta\delta}\right], \tag{77}$$

where $\overline{\lambda} \leq \max_{i\in[n]} \frac{r_n}{g_n}x_i \leq \frac{r_n}{g_n}s_n$. The concentration result stated in the lemma then follows by utilizing the assumption that $x_i \in [s_n]$ for all $i \in [n]$, and by slightly loosening the bound, using the assumption $\delta \leq \frac{r_n}{g_n}s_n$. ■

We continue to the second step in analyzing the probability in (64), which is the analysis of the randomness

of $X^n$. To this end, we denote

$$J(x^n) := \sum_{i=1}^{n} \mathbb{E}\left[\log P_{Z|X}(Z_i \mid x_i) \,\middle|\, X_i = x_i\right] \tag{78}$$

$$= \sum_{i=1}^{n} I(Z_i; X_i = x_i) - H(Z_i). \tag{79}$$

**Lemma 11.** *Assume that* $\mathrm{supp}(P_X) \subseteq [s_n]$, *and* $r_n s_n \geq 12\pi e^2 g_n$. *Then,*

$$\mathbb{P}\left[J(X^n) + H(Z^n) < I(X^n; Z^n) - n\delta\right] \leq \exp\left[-n \cdot \frac{2\delta^2}{\log^2 \frac{r_n s_n}{g_n}}\right]. \tag{80}$$

*Proof:* We note that $\mathbb{E}[J(X^n)] = -H(Z^n \mid X^n)$, and show that $J(X^n)$ concentrates to its expected value using Hoeffding's inequality (Fact 18, Appendix A). We begin by noting that it holds that

$$\mathbb{E}\left[\log P_{Z|X}(Z_i \mid x_i) \mid X_i = x_i\right] \leq 0. \tag{81}$$

Also,

$$-\log P_{Z|X}(z \mid x) = \log \frac{z!}{e^{-r_n x/g_n}\left(\frac{r_n x}{g_n}\right)^z} \tag{82}$$

$$\leq \frac{r_n x}{g_n} + \log z! + z \cdot \log\left(\frac{g_n}{r_n x}\right). \tag{83}$$

For $z \geq 1$, using Stirling's bound (Fact 16 in Appendix A)

$$\log z! \leq z \log z - z + \frac{1}{2}\log(6\pi z), \tag{84}$$

and for $z = 0$ it holds that $\log z! = 0$. Hence, for $Z \mid X = x \sim \mathrm{Poisson}(\frac{r_n}{g_n} x)$

$$\mathbb{E}\left[\log Z! \mid X = x\right]$$

$$\leq \mathbb{E}\left[\left(Z \log Z - Z + \frac{1}{2}\log(6\pi Z)\right) \cdot \mathbb{1}\{Z > 0\} \,\middle|\, X = x\right] \tag{85}$$

$$\overset{(a)}{\leq} \mathbb{E}\left[Z \log Z - Z + \frac{1}{2}\log(6\pi(Z+1)) \,\middle|\, X = x\right] \tag{86}$$

$$\overset{(b)}{\leq} \frac{r_n x}{g_n}\log\left(1 + \frac{r_n x}{g_n}\right) - \frac{r_n x}{g_n} + \mathbb{E}\left[\frac{1}{2}\log(6\pi(Z+1)) \,\middle|\, X = x\right] \tag{87}$$

$$\overset{(c)}{\leq} \frac{r_n x}{g_n}\log\left(1 + \frac{r_n x}{g_n}\right) - \frac{r_n x}{g_n} + \frac{1}{2}\log\left(6\pi\left(\frac{r_n x}{g_n} + 1\right)\right), \tag{88}$$

where $(a)$ follows by analytically completing $Z \log Z = 0$ for $Z = 0$, and upper bounding $\log Z \leq \log(Z+1)$, $(b)$ follows since if $V \sim \mathrm{Poisson}(\lambda)$ then, $\mathbb{E}[V \log V] \leq \lambda \log(1+\lambda)$ (see Lemma 24 in Appendix B for a

proof), $(c)$ follows from Jensen's inequality for the concave logarithm function. So,

$$\mathbb{E}\left[-\log P_{Z|X}(Z \mid x_i) \mid X = x_i\right] \tag{89}$$

$$= \frac{r_n x_i}{g_n} + \mathbb{E}\left[\log Z! \mid X = x_i\right] + \mathbb{E}\left[Z \mid x = x_i\right] \cdot \log\left(\frac{g_n}{r_n x_i}\right) \tag{90}$$

$$= \frac{r_n x_i}{g_n} + \mathbb{E}\left[\log Z! \mid X = x_i\right] + \frac{r_n x_i}{g_n} \log\left(\frac{g_n}{r_n x_i}\right) \tag{91}$$

$$\overset{(a)}{\leq} \frac{r_n x_i}{g_n}\left[1 + \log\left(\frac{g_n}{r_n x_i}\right)\right] + \frac{r_n x_i}{g_n} \log\left(1 + \frac{r_n x_i}{g_n}\right) - \frac{r_n x_i}{g_n} + \frac{1}{2}\log\left(6\pi\left(\frac{r_n x_i}{g_n} + 1\right)\right) \tag{92}$$

$$= \frac{r_n x_i}{g_n} \log\left(1 + \frac{g_n}{r_n x_i}\right) + \frac{1}{2}\log\left(6\pi\left(\frac{r_n x_i}{g_n} + 1\right)\right) \tag{93}$$

$$\overset{(b)}{\leq} 1 + \frac{1}{2}\log\left(6\pi\left(\frac{r_n x_i}{g_n} + 1\right)\right) \tag{94}$$

$$\overset{(c)}{\leq} \frac{1}{2}\log\left(6\pi e^2\left(\frac{r_n s_n}{g_n} + 1\right)\right) \tag{95}$$

$$\overset{(d)}{\leq} \frac{1}{2}\log\left(12\pi e^2 \frac{r_n s_n}{g_n}\right), \tag{96}$$

$$\overset{(e)}{\leq} \log \frac{r_n s_n}{g_n}, \tag{97}$$

where $(a)$ follows from (88), $(b)$ follows from $\log(1+t) \leq t$, $(c)$ follows since under the assumption of the lemma $x_i \leq s_n$, and both $(d)$ and $(e)$ follow by the assumption $\frac{r_n s_n}{g_n} \geq 12\pi e^2 \geq 1$.

We deduce from (81) and (97) that $J(X^n)$ is a sum of $n$ independent RVs $\mathbb{E}_{Z_i|X_i}[\log P_{Z|X}(Z_i \mid X_i)]$, each of which is bounded, with probability 1, in $[-\log \frac{r_n s_n}{g_n}, 0]$. Consequently, Hoeffding's inequality (Fact 18 in Appendix A) implies that

$$\mathbb{P}\left[J(X^n) + H(Z^n \mid X^n) < -n\delta\right] \leq \exp\left[-n \cdot \frac{2\delta^2}{\log^2 \frac{r_n s_n}{g_n}}\right], \tag{98}$$

which then implies the claim of the lemma, by adding $I(X^n; Z^n)$ to both sides in the inequality defining the event of interest. ■

Setting $\delta_n \in (0, \frac{r_n}{g_n} s_n)$, and then $\log \gamma = nI(X;Z) - 2n\delta_n - \frac{1}{2}\log(6\pi n r_n)$. Let us define the event

$$\mathcal{E}_n(x^n) := \left\{f_{x^n}(Z^n) < I(Z^n; X^n = x^n) - n\delta_n\right\}. \tag{99}$$

Then,

$$\mathbb{P}\left[\sum_{i=1}^n \log \frac{P_{Z|X}(Z_i \mid X_i)}{P_Z(Z_i)} \leq \log\gamma + \frac{1}{2}\log(6\pi n r_n)\right]$$

$$= \mathbb{P}\left[\sum_{i=1}^n \log \frac{P_{Z|X}(Z_i \mid X_i)}{P_Z(Z_i)} \leq nI(X^n; Z^n) - 2n\delta_n\right] \tag{100}$$

$$= \sum_{x^n \in [s_n]^n} \mathbb{P}[X^n = x^n] \cdot \mathbb{P}\left[f_{x^n}(Z^n) \leq nI(X^n; Z^n) - 2n\delta_n \,|\, X^n = x^n\right] \tag{101}$$

$$\leq \sum_{x^n\in[s_n]^n} \mathbb{P}[X^n=x^n]\cdot\mathbb{P}\left[\{f_{x^n}(Z^n)\leq nI(X^n;Z^n)-2n\delta_n\}\cap\{\mathcal{E}_n^c(x^n)\}\,|\,X^n=x^n\right]$$
$$+\sum_{x^n\in[s_n]^n}\mathbb{P}[X^n=x^n]\cdot\mathbb{P}\left[\mathcal{E}_n(x^n)\,|\,X^n=x^n\right] \tag{102}$$
$$\overset{(a)}{\leq} \sum_{x^n\in[s_n]^n}\mathbb{P}[X^n=x^n]\cdot\mathbb{1}\left[\{I(Z^n;X^n=x^n)-n\delta_n\leq nI(X^n;Z^n)-2n\delta_n\}\right]$$
$$+\sum_{x^n\in[s_n]^n}\mathbb{P}[X^n=x^n]\cdot\exp\left[-n\cdot\frac{2\delta_n^2}{\log^2\frac{r_ns_n}{g_n}}\right] \tag{103}$$
$$=\mathbb{P}\left[J(X^n)+H(Z^n)<I(X^n;Z^n)-n\delta\right]+\exp\left[-n\cdot\frac{2\delta_n^2}{\log^2\frac{r_ns_n}{g_n}}\right] \tag{104}$$
$$\overset{(b)}{\leq}\exp\left[-n\cdot\frac{2\delta_n^2}{\log^2\frac{r_ns_n}{g_n}}\right]+\exp\left[-n\cdot\frac{g_n\delta_n^2}{19r_ns_n\log^2 s_n}\right] \tag{105}$$
$$\leq 2\exp\left[-n\delta_n^2\cdot\left(\frac{2}{\log^2\frac{r_ns_n}{g_n}}\wedge\frac{g_n}{19r_ns_n\log^2 s_n}\right)\right], \tag{106}$$

where $(a)$ follows from the concentration bound in Lemma 10, and $(b)$ follows from Lemma 11. We substitute this back into (64), and then in the extended Feinstein's bound (16) to obtain

$$\epsilon_n P_X^{\otimes n}(F_n)\leq 4e\sqrt{nr_n}\exp\left[-n\delta_n^2\cdot\left(\frac{2}{\log^2\frac{r_ns_n}{g_n}}\wedge\frac{g_n}{19r_ns_n\log^2 s_n}\right)\right]+\frac{M}{e^{nI(X^n;Z^n)-2n\delta_n-\frac{1}{2}\log(6\pi nr_n)}}. \tag{107}$$

The claim of the proposition is then proved by choosing $M=\exp\left[nI(X;Z)-3n\delta_n-\frac{1}{2}\log(6\pi nr_n)\right]$, and performing minor algebraic simplifications. ■

We now turn to prove Prop. 7.

*Proof of Prop. 7:* We analyze the reduction in mutual information over the Poisson channel, resulting from modifying the ideal gamma distribution of $\tilde{X}$ to the truncated $\overline{X}:=\tilde{X}_{|\mathcal{S}_n}$, and then to upward rounded $X:=\lceil\overline{X}\rceil$. We begin by analyzing the reduction in mutual information due to the truncation operation, using the I-MMPE relation (Theorem 8). We begin with the following general result.

**Lemma 12.** *Let $U$ be a non-negative RV, which satisfies $\mathbb{E}[U^2\log^2 U]<\infty$, and let $\overline{U}\equiv U_{|\mathcal{S}}$ be distributed as $U$ truncated to an interval $\mathcal{S}:=[s_{\min},s_{\max}]\subset\mathbb{R}_+$, as in Definition 6 where $s_{\min}<1<s_{\max}$. Let $a>0$ be given, assume that $V_a\,|\,U=u\sim\text{Poisson}(au)$, and let $\overline{V}_a\,|\,\overline{U}=u\sim\text{Poisson}(au)$. Then, for any $\gamma>0$*

$$I(U;V_\gamma)\leq I(\overline{U};\overline{V}_\gamma)+\gamma\cdot s_{\max}\cdot\mathbb{P}\left[U\in[0,s_{\min})\right]$$
$$+\gamma\mathbb{E}\left[U\log U\cdot\mathbb{1}\left\{U\in(s_{\max},\infty)\right\}\right]$$
$$+\gamma\mathbb{E}\left[U\log\frac{1}{s_{\min}}\cdot\mathbb{1}\left\{U\in(s_{\max},\infty)\right\}\right] \tag{108}$$

*Proof:* Let $a>0$ be given. Let $\widehat{aU}(v)=\mathbb{E}[aU\mid V_a=v]$ be the MMPE estimator of $aU$ based on the measurement $V_a$. Similarly, let $\widehat{a\overline{U}}(v)=\mathbb{E}[a\overline{U}\mid\overline{V}_a=v]$ be the MMPE estimator of $a\overline{U}$ based on the

measurement $\overline{V}_a$. Recall that $\ell(u,v) = v - u + u\log\frac{u}{v}$ is Poisson error function. Then,

$$
\begin{aligned}
&\mathrm{mmpe}(aU) \\
&= \mathbb{E}\left[\ell\left(aU, \widehat{a\overline{U}}(V_a)\right)\right] && (109)\\
&\overset{(a)}{\leq} \mathbb{E}\left[\ell\left(aU, \widehat{a\overline{U}}(V_a)\right)\right] && (110)\\
&\overset{(b)}{=} a\mathbb{E}\left[\ell\left(U, \widehat{\overline{U}}(V_a)\right)\right] && (111)\\
&= a\mathbb{E}\left[\ell\left(U, \widehat{\overline{U}}(V_a)\right)\Big| U \in \mathcal{S}\right]\cdot\mathbb{P}[U\in\mathcal{S}] + a\mathbb{E}\left[\ell\left(U, \widehat{\overline{U}}(V_a)\right)\cdot\mathbb{1}\{U\notin\mathcal{S}\}\right] && (112)\\
&\overset{(c)}{=} a\mathbb{E}\left[\ell\left(\overline{U}, \widehat{\overline{U}}(\overline{V}_a)\right)\right]\cdot\mathbb{P}[U\in\mathcal{S}] + a\mathbb{E}\left[\ell\left(U, \widehat{\overline{U}}(V_a)\right)\cdot\mathbb{1}\{U\notin\mathcal{S}\}\right] && (113)\\
&\overset{(d)}{\leq} \mathrm{mmpe}(a\overline{U}) + a\mathbb{E}\left[\ell\left(U, \widehat{\overline{U}}(V_a)\right)\cdot\mathbb{1}\{U\notin\mathcal{S}\}\right] && (114)\\
&= \mathrm{mmpe}(a\overline{U}) + a\mathbb{E}\left[\ell\left(U, \widehat{\overline{U}}(V_a)\right)\cdot\mathbb{1}\left\{U\in[0,s_{\min})\right\}\right] + a\mathbb{E}\left[\ell\left(U, \widehat{\overline{U}}(V_a)\right)\cdot\mathbb{1}\left\{U\in(s_{\max},\infty)\right\}\right] && (115)\\
&\overset{(e)}{\leq} \mathrm{mmpe}(a\overline{U}) + a\mathbb{E}\left[\ell\left(U, s_{\max}\right)\cdot\mathbb{1}\left\{U\in[0,s_{\min})\right\}\right] + a\mathbb{E}\left[\ell\left(U, s_{\min}\right)\cdot\mathbb{1}\left\{U\in(s_{\max},\infty)\right\}\right] && (116)\\
&\overset{(f)}{\leq} \mathrm{mmpe}(a\overline{U}) + as_{\max}\cdot\mathbb{P}\left[U\in[0,s_{\min})\right] + a\cdot\mathbb{E}\left[U\log\frac{U}{s_{\min}}\cdot\mathbb{1}\left\{U\in(s_{\max},\infty)\right\}\right], && (117)
\end{aligned}
$$

where $(a)$ follows from the sub-optimality of $\widehat{a\overline{U}}$ for estimating $aU$, $(b)$ follows from the homogeneity property of the loss function $\ell(au,av) = a\ell(u,v)$ for $a\geq 0$, $(c)$ follows since conditioned on $U\in\mathcal{S}$, the distribution of $U$ equals that of $\overline{U}$ and the distribution of $V_a$ equals that of $\overline{V}_a$, $(d)$ follows again from the homogeneity property and $\mathbb{P}[U\in\mathcal{S}]\leq 1$ and the term multiplying $\mathbb{P}[U\in\mathcal{S}]$ is non-negative (the expected value of the Poisson loss function), $(e)$ follows since $\widehat{\overline{U}}(V_a)\in\mathcal{S} = [s_{\min}, s_{\max}]$ and since the loss function $v\to\ell(u,v)$ is monotonic increasing (resp. decreasing) for $v\geq u$ (resp. $v\leq u$), $(f)$ follows since for $u\leq s_{\min} < 1$ it holds that

$$
\ell\left(u, s_{\max}\right) = s_{\max} - u + u\log\frac{u}{s_{\max}} \leq s_{\max} + u\log\frac{u}{s_{\max}} \leq s_{\max}, \tag{118}
$$

and for $u\geq s_{\max} > 1 > s_{\min}$ it holds that

$$
\ell\left(u, s_{\min}\right) = s_{\min} - u + u\log\frac{u}{s_{\min}} \leq s_{\min} - u + u\log\frac{u}{s_{\min}} \leq u\log\frac{u}{s_{\min}}. \tag{119}
$$

Using twice the I-MMPE relation (Theorem 8), and the bound (117) directly leads to the stated claim of the lemma, as

$$
\begin{aligned}
I(U;V_\gamma) &= \int_0^\gamma \mathrm{mmpe}(aU)\frac{\mathrm{d}a}{a} && (120)\\
&\leq \int_0^\gamma \mathrm{mmpe}(a\overline{U})\frac{\mathrm{d}a}{a} + \\
&\qquad \int_0^\gamma \left\{as_{\max}\cdot\mathbb{P}\left[U\in[0,s_{\min})\right] + a\cdot\mathbb{E}\left[U\log\frac{U}{s_{\min}}\cdot\mathbb{1}\left\{U\in(s_{\max},\infty)\right\}\right]\right\}\frac{\mathrm{d}a}{a} && (121)
\end{aligned}
$$

$$\leq I(\overline{U};\overline{V}_\gamma) + \gamma \cdot s_{\max} \cdot \mathbb{P}\left[U \in [0, s_{\min})\right] + \gamma \mathbb{E}\left[U \log \frac{U}{s_{\min}} \cdot \mathbb{1}\left\{U \in (s_{\max}, \infty)\right\}\right]. \tag{122}$$

■

Lemma 12 shows that the difference in mutual information between the input $U$ and its truncated version consists of three terms. Our next goal is to specifically evaluate these terms for the distribution and support of interest, and show that they are negligible $o_n(1)$.

**Lemma 13.** *Let* $\tilde{X} \sim \text{Gamma}(\frac{1}{2}, 2g_n)$*, and let* $s_{\min} = \frac{1}{g_n^{1+3\rho}}$ *and* $s_{\max} = g_n^{1+\rho}$ *for some* $\rho \in (0,1)$*. Then,*

$$s_{\max} \cdot \mathbb{P}\left[\tilde{X} \in [0, s_{\min})\right] \leq \frac{1}{g_n^{\rho/2}}. \tag{123}$$

*Also, there exists* $n_0(\rho)$ *such that for all* $n \geq n_0(\rho)$

$$\mathbb{E}\left[\tilde{X} \log \tilde{X} \cdot \mathbb{1}\left\{\tilde{X} \in (s_{\max}, \infty)\right\}\right] \leq \exp\left[-\frac{g_n^\rho}{4}\right], \tag{124}$$

*and*

$$\mathbb{E}\left[\tilde{X} \log \frac{1}{s_{\min}} \cdot \mathbb{1}\left\{\tilde{X} \in (s_{\max}, \infty)\right\}\right] \leq \exp\left[-\frac{g_n^\rho}{4}\right]. \tag{125}$$

*Proof:* We begin with the first term in (123). From the properties of the gamma probability distribution function (PDF) of consideration (Lemma 25 in Appendix C), it holds that

$$s_{\max} \cdot \mathbb{P}\left[\tilde{X} \in [0, s_{\min})\right] = g_n^{1+\rho} \cdot \mathbb{P}\left[\tilde{X} \leq \frac{1}{g_n^{1+3\rho}}\right] \tag{126}$$

$$\leq \frac{g_n^{1+\rho}}{g_n^{(2+3\rho)/2}} \leq \frac{1}{g_n^{\rho/2}}. \tag{127}$$

We now move on to the second term in (124). For $t \in [g_n^{1+\rho}, \infty)$, using the expression for the gamma PDF (see Appendix C)

$$\mathbb{E}\left[\tilde{X} \log \tilde{X} \cdot \mathbb{1}\left\{\tilde{X} \geq g_n^{1+\rho}\right\}\right] = \int_{g_n^{1+\rho}}^{\infty} \frac{1}{\sqrt{2\pi \tilde{x} g_n}} e^{-\tilde{x}/(2g_n)} \cdot \tilde{x} \log \tilde{x} \cdot \mathrm{d}\tilde{x} \tag{128}$$

$$\overset{(a)}{\leq} \int_{g_n^{1+\rho}}^{\infty} \frac{\tilde{x}}{\sqrt{2\pi g_n}} e^{-\tilde{x}/(2g_n)} \cdot \mathrm{d}\tilde{x} \tag{129}$$

$$= \sqrt{\frac{2g_n}{\pi}} \int_{g_n^{1+\rho}}^{\infty} \frac{\tilde{x}}{2g_n} e^{-\tilde{x}/(2g_n)} \cdot \mathrm{d}\tilde{x} \tag{130}$$

$$\overset{(b)}{=} \frac{(2g_n)^{3/2}}{\sqrt{\pi}} \int_{\frac{1}{2}g_n^\rho}^{\infty} s e^{-s} \cdot \mathrm{d}s \tag{131}$$

$$\overset{(c)}{=} \frac{(2g_n)^{3/2}}{\sqrt{\pi}} \left(\frac{1}{2} g_n^\rho + 1\right) e^{-\frac{1}{2}g_n^\rho} \tag{132}$$

$$\overset{(d)}{\leq} 4 \exp\left[-\frac{1}{2} g_n^\rho + \left(\frac{3}{2} + \rho\right) \log g_n\right], \tag{133}$$

where $(a)$ follows since $\log \tilde{x} \leq \sqrt{\tilde{x}}$ for $\tilde{x} \in \mathbb{R}_+$, $(b)$ is using the change of variables $s = \frac{\tilde{x}}{2g_n}$, $(c)$ by solving

the integral $\int se^{-s} \cdot \mathrm{d}s = -(s+1)e^{-s}$, and $(d)$ follows since $g_n \geq 1$. We finally move to the third term in (124). It holds that

$$\mathbb{E}\left[\tilde{X} \cdot \mathbb{1}\left\{\tilde{X} \geq g_n^{1+\rho}\right\}\right] \overset{(a)}{\leq} \mathbb{E}\left[\tilde{X}^{3/2} \cdot \mathbb{1}\left\{\tilde{X} \in (s_{\max}, \infty)\right\}\right] \tag{134}$$

$$\overset{(b)}{\leq} \exp\left[-\frac{g_n^{\rho}}{3}\right], \tag{135}$$

where $(a)$ holds since $s_{\max} = g_n^{1+\rho} \geq 1$, and $(b)$ holds as for the second term. The third term is bounded as (124) since

$$\log \frac{1}{s_{\min}} = (1+3\rho)\log g_n \leq 4 \log g_n \leq \exp\left[-\frac{g_n^{\rho}}{12}\right]. \tag{136}$$

■

Up until now we have considered the effect on the mutual information of the Poisson channel when truncating the asymptotically optimal input $\tilde{X} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$ to $\overline{X} = \tilde{X}_{|\mathcal{S}}$. We next consider the effect on the mutual information of upward rounding a continuous input $\overline{X}$ to an integer input $X = \lceil \overline{X} \rceil$. To this end, we will decompose the mutual information $I(X;Z) = H(Z) - H(Z \mid X)$ and analyze how each of these terms changes due to the rounding operation.

Let $\Psi(\mu) : \mathbb{R}_+ \to \mathbb{R}_+$ be the maximum entropy for non-negative integer distributions, under a mean constraint $\mu$. With a slight abuse of the notation in (3) we denote

$$\Psi(\mu) := \max_{P_A} \{H(A) \colon \mathrm{supp}(P_A) \subseteq \mathbb{N}_+,\ \mathbb{E}[A] \leq \mu\}. \tag{137}$$

**Lemma 14.** $\Psi(\mu) = (\mu+1) \cdot \mathrm{h_{bin}}\left(\frac{1}{\mu+1}\right)$. *The function $\mu \to \Psi(\mu)$ is monotonic non-decreasing and concave in $\mu$.*

*Proof:* It is well known that the maximum entropy distribution among distributions over the non-negative integers with a mean constraint is geometric. For completeness, a standard proof is as follows. Let $A \sim P_A$ where $p_i := P_A(i)$ for $i \in \mathbb{N}$. Assume that $P_A$ satisfies the mean constraint. Let $Q_A^{(\lambda)}$ be a distribution defined by

$$q_i(\lambda) = \frac{e^{\lambda(\mu - i)}}{\sum_j e^{\lambda(\mu-j)}} \tag{138}$$

for $i \in \mathbb{N}$. Then, for any $\lambda \geq 0$

$$H(P_A) \leq \sum_{i=0}^{\infty} -p_i \log p_i + \lambda\left(\mu - \sum_{i=0}^{\infty} i p_i\right) \tag{139}$$

$$= \sum_{i=0}^{\infty} p_i \log \frac{e^{\lambda(\mu-i)}}{p_i} \tag{140}$$

$$= -\sum_{i=0}^{\infty} p_i \log \frac{p_i}{q_i(\lambda)} + \log\left(\sum_j e^{\lambda(\mu-j)}\right) \tag{141}$$

$$= -\mathrm{D}_{\mathrm{KL}}(P_A \,||\, Q_A^{(\lambda)}) + \log\left(\sum_j e^{\lambda(\mu-j)}\right) \tag{142}$$

$$\leq \log\left(\sum_j e^{\lambda(\mu-j)}\right), \tag{143}$$

where equality holds if both $\mathbb{E}_P[A] = \sum_{i=0}^{\infty} ip_i = \mu$ and $P_A \equiv Q_A^{(\lambda)}$ holds. Now, $Q_A^{(\lambda)}$ is readily identified as a geometric distribution over $\mathbb{N} = \{0, 1, 2, ..\}$. Using the standard parametrization of the geometric distribution, if $A \sim \mathrm{Geo}(\theta)$ then $\mathbb{E}[A] = \frac{1-\theta}{\theta}$ and $H(A) = \frac{\mathrm{h_{bin}}(\theta)}{\theta}$. Thus, $H(P_A) = (\mu+1) \cdot \mathrm{h_{bin}}(\frac{1}{\mu+1})$, and this is the maximum entropy. Monotonicity is trivial, and concavity is assured by the concavity of the entropy, or directly from the closed-form expression of $\Psi(\mu)$. ■

Lemma 14 will next be used to compare the output entropy of the Poisson channel when the input is a continuous $\overline{X}$, to that entropy when the input is an integer rounded version of $\overline{X}$. We will then also compare the conditional entropy $H(Z \mid X)$ under the different input distributions can be easily compared, and combining these two results we obtain a relation between the mutual information values. Concretely:

**Lemma 15.** *Assume that $\underline{c} g_n \leq r_n \leq e g_n$ for some $\underline{c} \in (0, e)$ and let $\rho \in (0, 1)$ be given, and assume that $g_n \to \infty$ as $n \to \infty$. Let $\tilde{X} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$, $\overline{X} = \tilde{X}_{|\mathcal{S}_n}$ with $\mathcal{S}_n = [g_n^{-(1+3\rho)}, g_n^{1+\rho}]$ and let $X = \lceil \overline{X} \rceil$ be its upward rounding to the nearest integer. Let $Z \sim \mathrm{Poisson}(\frac{r_n}{g_n} X)$ and let $\overline{Z} \sim \mathrm{Poisson}(\frac{r_n}{g_n} \overline{X})$. Then,*

$$I(X;Z) \geq I(\overline{X};\overline{Z}) - \Psi\left(\frac{r_n}{g_n}\right) - o_n(1). \tag{144}$$

*Proof:* First, we compare the conditional entropy values $H(\overline{Z} \mid \overline{X})$ and $H(Z \mid X)$. Let $f_{\overline{X}}$ denote the density of $\overline{X}$ w.r.t. the Lebesgue measure $\lambda$, and let $P_X$ denote the PMF of $X$. We have that

$$H(\overline{Z} \mid \overline{X}) = \int_0^{\infty} f_{\overline{X}}(\overline{x}) \cdot H(\overline{Z} \mid \overline{X} = \overline{x}) \cdot \lambda(\mathrm{d}\overline{x}) \tag{145}$$

where $H(\overline{Z} \mid \overline{X} = \overline{x})$ is the entropy of a Poisson RV, and, similarly,

$$H(Z \mid X) = \sum_{i=0}^{\infty} P_X(i) \cdot H(Z \mid X = i). \tag{146}$$

We show that the contribution to this sum by "small" indices $i \in [\lfloor g_n^{1-\rho} \rfloor]$ is negligible. Indeed, by Lemma 23, it holds that for any $i \in [\lfloor g_n^{1-\rho} \rfloor]$

$$H(Z \mid X = i) \leq \frac{1}{2} \log\left[2\pi e \left(i \frac{r_n}{g_n} + \frac{1}{12}\right)\right] \tag{147}$$

$$\leq \frac{1}{2} \log\left[2\pi e \left(2 \frac{r_n}{g_n^{\rho}} + \frac{1}{12}\right)\right] \tag{148}$$

$$\leq \frac{1}{2} \log(35 r_n) \tag{149}$$

assuming that $\rho$ is sufficiently small so that $2\frac{r_n}{g_n^\rho} \geq \frac{1}{12}$. Now, Lemma 25 (Appendix C) implies that it holds that

$$\mathbb{P}[\tilde{X} \notin \mathcal{S}_n] = \mathbb{P}[\tilde{X} \leq g_n^{-(1+3\rho)}] + \mathbb{P}[\tilde{X} \geq g_n^{1+\rho}] \tag{150}$$

$$\leq \frac{1}{g_n^{1+3\rho/2}} + 2e^{-g_n^\rho/2} \tag{151}$$

$$\leq \frac{1}{2} \tag{152}$$

for all $n \geq n_0(\rho)$. Hence, using again (C.5) Lemma 25, it holds that

$$\mathbb{P}\left[X \leq \lfloor g_n^{1-\rho} \rfloor\right] = \mathbb{P}\left[\overline{X} \leq \lfloor g_n^{1-\rho} \rfloor\right] \tag{153}$$

$$\leq \frac{\mathbb{P}[\tilde{X} \leq g_n^{1-\rho}]}{\mathbb{P}[\tilde{X} \in \mathcal{S}_n]} \tag{154}$$

$$\leq \frac{1/g_n^{\rho/2}}{\mathbb{P}[\tilde{X} \in \mathcal{S}_n]} \tag{155}$$

$$\leq \frac{2}{g_n^{\rho/2}}. \tag{156}$$

Furthermore, assume that $i \geq g_n^{1-\rho}$. Then, for any $\overline{x} \in [i-1, i]$ it holds from the asymptotic expression for the Poisson entropy in Lemma 23 (Appendix B) that

$$H(Z \mid X = i) - H(Z \mid X = \overline{x}) \tag{157}$$

$$= \frac{1}{2}\log\left[2\pi e i \frac{r_n}{g_n}\right] + O\left(\frac{1}{i\frac{r_n}{g_n}}\right) - \frac{1}{2}\log\left[2\pi e \overline{x}\frac{r_n}{g_n}\right] + O\left(\frac{1}{\overline{x}\frac{r_n}{g_n}}\right) \tag{158}$$

$$= \frac{1}{2}\log\left[\frac{i}{\overline{x}}\right] + O\left(\frac{1}{\overline{x}\frac{r_n}{g_n}}\right) \tag{159}$$

$$\leq \frac{1}{2}\log\left[\frac{\overline{x}+1}{\overline{x}}\right] + O\left(\frac{1}{\overline{x}\frac{r_n}{g_n}}\right) \tag{160}$$

$$= O\left(\frac{1}{c\overline{x}}\right) \tag{161}$$

$$= O\left(\frac{1}{c g_n^{1-\rho}}\right). \tag{162}$$

Thus,

$$H(Z \mid X) = \sum_{i=0}^{\lfloor g_n^{1-\rho} \rfloor} P_X(i) \cdot H(Z \mid X = i) + \sum_{i=\lceil g_n^{1-\rho} \rceil}^{\infty} P_X(i) \cdot H(Z \mid X = i) \tag{163}$$

$$\overset{(a)}{\leq} \mathbb{P}[X \leq \lfloor g_n^{1-\rho} \rfloor] \cdot H(Z \mid X = \lfloor g_n^{1-\rho} \rfloor) + \sum_{i=\lceil g_n^{1-\rho} \rceil}^{\infty} P_X(i) \cdot H(Z \mid X = i) \tag{164}$$

$$\overset{(b)}{\leq} \frac{2}{g_n^{\rho/2}} \cdot \frac{1}{2}\log(35 r_n) + \sum_{i=\lceil g_n^{1-\rho} \rceil}^{\infty} P_X(i) \cdot H(Z \mid X = i) \tag{165}$$

$$= \frac{2}{g_n^{\rho/2}} \cdot \frac{1}{2} \log(35 r_n) + \sum_{i=\lceil g_n^{1-\rho} \rceil}^{\infty} \left[ \int_{i-1}^{i} f_{\overline{X}}(\overline{x}) \lambda(\mathrm{d}\overline{x}) \right] \cdot H(Z \mid X = i) \tag{166}$$

$$\overset{(c)}{\leq} \frac{2}{g_n^{\rho/2}} \cdot \frac{1}{2} \log(35 r_n) + \int_{\lceil g_n^{1-\rho} \rceil}^{\infty} f_{\overline{X}}(\overline{x}) \cdot H(\overline{Z} \mid \overline{X} = \overline{x}) \cdot \lambda(\mathrm{d}\overline{x}) + O\left( \frac{1}{\underline{c} g_n^{1-\rho}} \right) \tag{167}$$

$$\leq o_n(1) + \int_{0}^{\infty} f_{\overline{X}}(\overline{x}) \cdot H(\overline{Z} \mid \overline{X} = \overline{x}) \cdot \lambda(\mathrm{d}\overline{x}) \tag{168}$$

$$= o_n(1) + H(\overline{Z} \mid \overline{X}), \tag{169}$$

where $(a)$ follows from the monotonicity of the Poisson entropy as a function of its parameter, $(b)$ follows from (156) and (149), $(c)$ follows from (162).

Second, we compare the output entropy $H(\overline{Z})$ and $H(Z)$. To this end, we decompose $X = \overline{X} + D$ where $D \in [0,1]$ (note, however, that $X$ and $D$ are statistically dependent). Conditioned on $X = x$, or equivalently, on $(\overline{X}, D) = (\overline{x}, d)$, it holds that $Z \sim \mathrm{Poisson}(\frac{r_n}{g_n}(\overline{x} + d))$. By the infinite divisibility of the Poisson distribution, we may write $Z \overset{d}{=} \overline{Z} + \acute{Z}$, where $\overline{Z} \sim \mathrm{Poisson}(\frac{r_n}{g_n}\overline{x})$ and $\acute{Z} \sim \mathrm{Poisson}(\frac{r_n}{g_n} d)$ are statistically independent. We thus let $Z = \overline{Z} + \acute{Z}$, and then note that that $Z \geq \overline{Z}$, with probability 1, and that both are integer valued discrete RVs. We continue similarly to the bound in [61, Prop. 8]. It holds that

$$H(\overline{Z}) - H(Z) \leq H(Z, \overline{Z}) - H(Z) \tag{170}$$

$$= H(\overline{Z} \mid Z) \tag{171}$$

$$= H(Z - \overline{Z} \mid Z) \tag{172}$$

$$\overset{(a)}{\leq} \mathbb{E}\left[ \Psi\left( \mathbb{E}\left[ Z - \overline{Z} \mid Z \right] \right) \right] \tag{173}$$

$$\overset{(b)}{\leq} \Psi\left( \mathbb{E}\left[ Z - \overline{Z} \right] \right) \tag{174}$$

$$= \Psi\left( \frac{r_n}{g_n} \mathbb{E}\left[ X - \overline{X} \right] \right) \tag{175}$$

$$\overset{(c)}{\leq} \Psi\left( \frac{r_n}{g_n} \right), \tag{176}$$

where $(a)$ follows from the operational definition of $\Psi(\mu)$, $(b)$ follows since $\mu \to \Psi(\mu)$ is concave (Lemma 14) along with Jensen's inequality, and $(c)$ follows since $\mu \to \Psi(\mu)$ is monotonic non-decreasing in $\mu$ (Lemma 14) and as $0 \leq X - \overline{X} \leq 1$.

Concluding, utilizing both (169) and (176) we obtain the claimed bound. ■

We may now conclude the proof of Prop. 7. Let $\rho \in (0,1)$ be given, and let $\mathcal{S}_n = [g_n^{-(1+3\rho)}, g_n^{1+\rho}]$. Let $\tilde{X} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$, let $\overline{X} = \tilde{X}_{|\mathcal{S}_n}$, and let $X = \lceil \overline{X} \rceil$. Let $\tilde{Z} \mid \tilde{X} = \tilde{x} \sim \mathrm{Poisson}(\frac{r_n}{g_n}\tilde{x})$, $\overline{Z} \mid \overline{X} = \overline{x} \sim \mathrm{Poisson}(\frac{r_n}{g_n}\overline{x})$, and $Z \mid X = x \sim \mathrm{Poisson}(\frac{r_n}{g_n} x)$. It then holds for all $n \geq n_0$ (which depends on $\underline{c}$, $\rho$, $\{g_n\}$),

$$I(X;Z) \overset{(a)}{\geq} I(\overline{X}; \overline{Z}) - \Psi\left( \frac{r_n}{g_n} \right) - o_n(1) \tag{177}$$

$$\overset{(b)}{\geq} I(\tilde{X};\tilde{Z}) - \frac{r_n}{g_n}\cdot\left(\frac{1}{g_n^{\rho/2}} + 2e^{-\frac{1}{4}g_n^{\rho}}\right) - \Psi\left(\frac{r_n}{g_n}\right) - o_n(1) \tag{178}$$

$$= I(\tilde{X};\tilde{Z}) - \Psi\left(\frac{r_n}{g_n}\right) - o_n(1) \tag{179}$$

$$\overset{(c)}{\geq} \frac{1}{2}\log r_n - \Psi\left(\frac{r_n}{g_n}\right) - o_n(1), \tag{180}$$

where $(a)$ follows from Lemma 15, $(b)$ follows from from Lemmas 12 and 13, and $(c)$ follows from the known lower bound [16, Thm. 7] on the average-power-constrained Poisson channel. ■

We continue with the proof of Prop. 9:

*Proof of Prop. 9:* Essentially, our goal is to analyze the probability that $\frac{1}{n}\sum_{i=1}^{n} X_i$ is significantly larger than its non-truncated mean (that is, the mean of $\mathrm{Gamma}(\frac{1}{2}, 2g_n)$ distribution). Recall that $\{X_i\}$ are drawn from a *rounded and truncated* Gamma distribution. We next mainly discuss the truncation operation, as this has larger effect on the analysis of that probability than the rounding operation. Let us describe a generative model for RVs whose distribution is the truncated gamma. Let $\{\tilde{X}_{i,j}\}_{i\in[n],j\in\mathbb{N}_+}$ be a double-index array of IID RVs, where $\tilde{X}_{i,j} \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$. Let

$$J^*(i) := \min\left\{j \in \mathbb{N}_+ \colon \tilde{X}_{i,j} \in \mathcal{S}_n\right\} \tag{181}$$

where $\mathcal{S}_n = [g_n^{-(1+3\rho)}, g_n^{1+\rho}]$, and let $\overline{X}_i = \tilde{X}_{J^*(i)}$. Then, $\overline{X}_i$ is distributed according to the $\mathrm{Gamma}(\frac{1}{2}, 2g_n)$ distribution, truncated to $\mathcal{S}_n$, as required. Now, using Lemma 25 (Appendix C), it holds for any fixed $(i,j)$ that

$$\mathbb{P}\left[\tilde{X}_{i,j} \notin \mathcal{S}_n\right] = \mathbb{P}\left[\tilde{X}_{i,j} \leq g_n^{-(1+3\rho)}\right] + \mathbb{P}\left[\tilde{X}_{i,j} \geq g_n^{1+\rho}\right] \leq \frac{1}{g_n^{1+3\rho/2}} + 2e^{-g_n^{\rho}/2} \leq \frac{2}{g_n^{1+3\rho/2}}, \tag{182}$$

where the last inequality holds for all $n$ sufficiently large. Let $L := \sum_{i=1}^{n} \mathbb{1}\{J^*(i) > 1\}$ be the number of indices for which $\tilde{X}_{i,1} \notin \mathcal{S}_n$, and so also $\overline{X}_i \neq \tilde{X}_{i,1}$. Hence, $\mathbb{E}[L] \leq \frac{2n}{g_n^{1+3\rho/2}}$, that is, $J^*(i) = 1$ for almost all $i \in [n]$. More sharply, the event $\mathcal{G} := \{L \geq \frac{3n}{g_n^{1+3\rho/2}}\}$ has low probability, and indeed, the relative Chernoff inequality (setting $\xi = \frac{1}{2}$ in Fact 19 in Appendix A) implies that

$$\mathbb{P}[\mathcal{G}] = \mathbb{P}\left[L \geq \frac{3n}{g_n^{1+3\rho/2}}\right] \leq \exp\left[-\frac{1}{5}\frac{n}{g_n^{1+3\rho/2}}\right]. \tag{183}$$

So, letting $t > 0$, we may decompose the probability of interest as

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}\overline{X}_i - g_n \geq t\right] \leq \mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\overline{X}_i - g_n \geq t\right\}\bigcap \mathcal{G}^c\right] + \mathbb{P}[\mathcal{G}] \tag{184}$$

$$\overset{(a)}{\leq} \sum_{\ell=0}^{\lceil\frac{3n}{g_n^{1+3\rho/2}}\rceil} \mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\overline{X}_i - g_n \geq t\right\}\bigcap\{L=\ell\}\right] + \exp\left[-\frac{1}{5}\frac{n}{g_n^{1+3\rho/2}}\right] \tag{185}$$

$$\leq \sum_{\ell=0}^{\lceil\frac{3n}{g_n^{1+3\rho/2}}\rceil} \mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\overline{X}_i - g_n \geq t\right\}\bigcap\{L=\ell\}\right] + o_n(1), \tag{186}$$

where $(a)$ follows from the assumption $n=\Omega(g_n^{1+\zeta})$ and $\rho\leq\frac{2\zeta}{3}$. We focus on a single term in the summation above. Given that $L=\ell$ there are $\ell$ indices for which $\overline{X}_i\neq\tilde{X}_{i,1}$. There are $\binom{n}{\ell}$ possible ways to choose those indices, and further conditioning on one specific choice, all the conditional probabilities are the same. Hence,

$$\begin{aligned}
&\mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\overline{X}_i-g_n\geq t\right\}\bigcap\{L=\ell\}\right]\\
&\overset{(a)}{\leq}\binom{n}{\ell}\cdot\mathbb{P}\left[\left\{\frac{1}{n}\left(\sum_{i=1}^{\ell}\overline{X}_i+\sum_{i=\ell+1}^{n}\tilde{X}_{i,1}\right)-g_n\geq t\right\}\cap\bigcap_{i=1}^{\ell}\{\overline{X}_i\neq\tilde{X}_{i,1}\}\cap\bigcap_{i=\ell+1}^{n}\{\overline{X}_i=\tilde{X}_{i,1}\}\right] &(187)\\
&\leq\binom{n}{\ell}\cdot\mathbb{P}\left[\frac{1}{n}\left(\sum_{i=1}^{\ell}\overline{X}_i+\sum_{i=\ell+1}^{n}\tilde{X}_{i,1}\right)-g_n\geq t\right] &(188)\\
&\overset{(b)}{\leq}\binom{n}{\ell}\cdot\mathbb{P}\left[\frac{1}{n}\left(\sum_{i=\ell+1}^{n}\tilde{X}_{i,1}\right)+\ell\frac{g_n^{1+\rho}}{n}-g_n\geq t\right] &(189)\\
&=\binom{n}{\ell}\cdot\mathbb{P}\left[\frac{1}{n-\ell}\sum_{i=\ell+1}^{n}\left(\tilde{X}_{i,1}-g_n\right)+\frac{\ell(g_n^{1+\rho}-g_n)}{n-\ell}\geq\frac{n}{n-\ell}t\right] &(190)\\
&\overset{(c)}{\leq}\binom{n}{\ell}\cdot\mathbb{P}\left[\frac{1}{n-\ell}\sum_{i=\ell+1}^{n}\left(\tilde{X}_{i,1}-g_n\right)\geq t-1\right] &(191)\\
&\overset{(d)}{\leq}\binom{n}{\ell}\cdot\left\{\exp\left[-\frac{(n-\ell)(t-1)}{4g_n}\right]+\exp\left[-\frac{(n-\ell)(t-1)^2}{8g_n^2}\right]\right\} &(192)\\
&\overset{(e)}{\leq}\exp\left[\frac{15n}{g_n^{1+3\rho/2}}\log(g_n)\right]\cdot\left\{\exp\left[-\frac{(n-\ell)(t-1)}{4g_n}\right]+\exp\left[-\frac{(n-\ell)(t-1)^2}{8g_n^2}\right]\right\}, &(193)
\end{aligned}$$

where $(a)$ follows from the union bound, $(b)$ follows since $\overline{X}_i\in\mathcal{S}_n=[g_n^{-(1+3\rho)},g_n^{1+\rho}]$ with probability 1, $(c)$ follows since $n=\Omega(g_n)$ and so

$$\frac{\ell(g_n^{1+\rho}-g_n)}{n-\ell}\leq\frac{\frac{3n}{g_n^{1+3\rho/2}}g_n^{1+\rho}}{n-\frac{3n}{g_n^{1+3\rho/2}}}=\frac{3}{g_n^{\rho/2}-\frac{3}{g_n^{1+\rho}}}\leq 1 \tag{194}$$

for all $n$ large enough (which depends on $\{g_n\}$) as $g_n\to\infty$, $(d)$ holds since $\frac{1}{n-\ell}\sum_{i=\ell+1}^{n}\tilde{X}_{i,1}\sim\text{Gamma}(\frac{n-\ell}{2},\frac{2g_n}{n-\ell})$ so that $\mathbb{E}[\frac{1}{n-\ell}\sum_{i=\ell+1}^{n}\tilde{X}_{i,1}]=g_n$, and using the tail inequality of sub-gamma RVs in Lemma 26 (Appendix C), $(e)$ follows since $\ell<\frac{n}{2}$, and as $\binom{n}{\ell}$ is monotonic non-increasing for $\ell<\frac{n}{2}$; Then, by Stirling's bound (Fact 17 in Appendix A)

$$\binom{n}{\ell}\leq\exp\left[n\cdot\text{h}_{\text{bin}}\left(\frac{3}{g_n^{1+3\rho/2}}\right)\right]\leq\exp\left[\frac{15n}{g_n^{1+3\rho/2}}\log(g_n)\right], \tag{195}$$

where the last inequality assumes that $\rho\in(0,1)$, and uses $\text{h}_{\text{bin}}(t)\leq-2t\log t$ for $t\in[0,\frac{1}{2}]$, which is valid since $\frac{3}{g_n^{1+3\rho/2}}\leq\frac{1}{2}$ for all $n\geq n_0(\rho)$ as $g_n\to\infty$. So, choosing $t=g_n^{3/4+\rho}+1:=\bar{t}$ in (193) assures that

$$\mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\overline{X}_i-g_n\geq\bar{t}\right\}\bigcap\{L=\ell\}\right]$$

$$\leq \exp\left[\frac{15n}{g_n^{1+3\rho/2}}\log(g_n)\right]\cdot\left\{\exp\left[-\frac{n(1-\frac{\ell}{n})g_n^{3/4+\rho}}{4g_n}\right]+\exp\left[-\frac{n(1-\frac{\ell}{n})g_n^{3/2+2\rho}}{8g_n^2}\right]\right\} \tag{196}$$

$$\leq \exp\left[-c\frac{n}{g_n^{1/4-\rho}}\right], \tag{197}$$

where the last inequality holds since $(1-\frac{\ell}{n})\to 1$ as $n\to\infty$ assuming $\ell\leq\lceil\frac{3n}{g_n^{1+3\rho/2}}\rceil=o(n)$. Hence, from (186)

$$\mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^n\overline{X}_i-g_n\geq\bar{t}\right\}\bigcap\mathcal{G}^c\right]\leq\lceil\frac{3n}{g_n^{1+3\rho/2}}\rceil\cdot\exp\left[-c\frac{n}{g_n^{1/4-\rho}}\right]=o_n(1), \tag{198}$$

which, by substituting back to 186 then implies that

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n\overline{X}_i-g_n\geq\bar{t}\right]=o_n(1). \tag{199}$$

Consequently, and as $\rho\in(0,1/4)$ was assumed, it holds for all $n$ sufficiently large that

$$\sum_{i=1}^n\overline{X}_i\leq n\left(g_n+g_n^{3/4+\rho}+1\right)=:ng_n(1+\bar{\varsigma}_n) \tag{200}$$

with probability larger than $1/2$, where $\bar{\varsigma}_n:=g_n^{-1/4+\rho}+g_n^{-1}=o_n(1)$. Now, the upward integer rounding implies that $X_i=\lceil\overline{X}_i\rceil$ for all $i\in[n]$. Letting $\varsigma_n:=g_n^{-1/4+\rho}+2g_n^{-1}=o_n(1)$, it also holds that

$$\sum_{i=1}^n X_i\leq\sum_{i=1}^n(\overline{X}_i+1)\leq ng_n(1+\varsigma_n) \tag{201}$$

with probability larger than $1/2$. Since $\sum_{i=1}^n X_i$ is integer, there must exists $k\in\mathbb{N}$ such that $0\leq k\leq ng_n(1+\varsigma_n)$ such that

$$\mathbb{P}\left[\sum_{i=1}^n X_i=k\right]\geq\frac{1}{2ng_n(1+\varsigma_n)}\geq\frac{1}{3ng_n}, \tag{202}$$

for all $n$ large enough. ■

We are now ready to prove the achievability bound of Theorem 2.

*Proof of the achievability bound of Theorem 2:* Recall the assumption $n=\Omega(g_n^{1+\zeta})$. Choose $\rho\in(0,\frac{1}{4}\wedge\frac{\zeta}{4})$, and $\chi\in(0,1)$ and set $\underline{g}_n=\frac{g_n}{1+\chi}$. Then, using Prop. 9 for $\underline{g}_n$ instead of $g_n$ implies that there exists $\tau_n\in[g_n]$ such that

$$P_{X^n}\left[F_n(\tau_n)\right]\geq\frac{1}{3n\underline{g}_n} \tag{203}$$

for all $n$ sufficiently large. We also note that the input distribution of $X$ used by Prop. 9 is supported on $[1,s_n]$ with $s_n=\lceil g_n^{1+\rho}\rceil$. Let us choose $\delta_n=g_n^{-\zeta/8}=o_n(1)$. Then, under the theorem assumptions $\delta_n\in(0,\frac{r_n}{g_n}s_n)$ as and so the condition of Prop. 5 is fulfilled. It then implies that there is a codebook of cardinality $M$ which

satisfies

$$\frac{1}{n}\log M \geq I(X;Z) - 3\delta_n - \frac{1}{2n}\log(6\pi n r_n) \tag{204}$$

$$\overset{(a)}{\geq} \frac{1}{2}\log r_n - o_n(1) - \Psi\left(\frac{r_n}{\underline{g}_n}\right) \tag{205}$$

$$\overset{(b)}{\geq} \frac{1}{2}\log r_n - o_n(1) - \Psi\left(\frac{r_n}{\underline{g}_n}\right), \tag{206}$$

where the last inequality follows from Prop. 7. At the same time, the maximal error probability of the codebook satisfies

$$\epsilon_n \overset{(a)}{\leq} 33 n \underline{g}_n \left[\sqrt{n r_n}\exp\left[-n\delta_n^2 \cdot \left(\frac{2}{\log^2\left(r_n \underline{g}_n^{\rho}\right)} \wedge \frac{1}{19 r_n \underline{g}_n^{\rho}(1+\rho)^2 \log^2 \underline{g}_n}\right)\right] + e^{-n\delta_n}\right] \tag{207}$$

$$\overset{(b)}{\leq} n^4\left[\exp\left[-c\frac{n\delta_n^2}{g_n^{1+2\rho}}\cdot\right] + e^{-n\delta_n}\right] \tag{208}$$

$$\overset{(c)}{\leq} n^4\left[\exp\left[-c g_n^{\zeta-2\rho}\delta_n^2\cdot\right] + e^{-n\delta_n}\right] \tag{209}$$

$$\overset{(d)}{\leq} n^4\left[\exp\left[-c g_n^{\zeta/2}\delta_n^2\cdot\right] + e^{-n\delta_n}\right] \tag{210}$$

$$\overset{(e)}{=} o_n(1), \tag{211}$$

where $(a)$ is obtained from (203) and setting $s_n = \underline{g}_n^{1+\rho}$, $(b)$ follows by simplifying with $n \geq \underline{g}_n \geq \underline{g}_n^{\rho}$,and $n \geq \sqrt{r_n}$ as well as $n \geq 33$ and $r_n \leq e g_n \leq e(1+\chi)g_n$, which all hold for sufficiently large $n$, and some numerical constant $c > 0$, $(c)$ holds since $n = \Omega(g_n^{1+\zeta})$ for some $\zeta > 0$, $(d)$ holds due to the choice $\rho \leq \frac{\zeta}{4}$, and $(e)$ holds by the choice $\delta_n = g_n^{-\zeta/8} = o_n(1)$. The result then follows by taking $n \to \infty$, and then $\chi \to 0$. ■

## V. Conclusion and Future Research

In this paper, we have considered the capacity of frequency-based channel with multinomial sampling, provided upper and lower bounds on its capacity, and applied it to the log-cardinality scaling of optimal DNA-storage codebooks in the short-molecule regime. There are multiple avenues for future research. First, while our bounds are rather tight, there is still a gap between the upper and lower bound, and specifically, it is interesting to settle the optimal choice of the normalized number of samples $r_n$. Second, the achievable bound of Theorem 2 is only applicable under the condition $n = \Omega(g_n^{1+\zeta})$. This condition is limiting, and specifically, it limits the application of the achievable bound to the DNA storage channel to $\beta > \frac{1}{2\log|\mathcal{A}|}$, that is, very short molecules are excluded. Inspecting the proof, this condition stems from the concentration inequality for the information spectrum in the Poisson channel in Prop. 5, which results an upper bound on the concentration probability, for which one of the terms is $\exp[-n\delta_n^2 \cdot \frac{g_n}{19 r_n s_n \log^2 s_n}]$, under the assumption that the input $X$ is supported on $[s_n]$. However, in order for the truncation of the optimal input of the Poisson channel to $[s_n]$ to have a

negligible effect on the mutual information, Prop. 7 requires that $s_n = \lceil g_n^{1+\rho} \rceil$ for some $\rho > 0$. In turn, roughly speaking, the above probability only decays when $n = \omega(g_n)$, and this is the source of the condition in the theorem. Consequently, possible removal of this condition requires finding tighter bounds on the concentration of the information spectrum of the Poisson channel, or a completely different approach. Third, it is of interest to analyze noisy sequencing channels. Inspecting the proof of Theorem 2, it appears that this would require analyzing the capacity of a channel with input $X^n$ and output $Z^n \mid X^n \sim \text{Poisson}(\hat{X}^n W_n)$ (where $\hat{X}^n$ is the normalized version of $X^n$). This is a Poisson channel with non-standard memory between the symbols, that is, inter-symbol interference [63], or a multiple-input multiple-output (MIMO) Poisson channel. Informally speaking, even if $W_n$ is a invertible matrix, there are two differences compared to the noiseless case. First, the input $X^n$ is still restricted as $\sum \hat{X}_i \leq 1$, and the channel $\hat{X}^n W_n$ may reduce this sum at the input of the Poisson channel. Second, the achievable lower bound on the capacity of the Poisson channel is obtained by lower bounding the output entropy $H(Z^n)$ with the differential entropy of the input $H(Z^n) \geq h(X^n)$ [16, Prop. 11]. Here, $h(\hat{X}^n W_n)$ has a reduced differential entropy by $\log\det W_n$, which will further reduce the capacity. A rigorous analysis appears challenging, and thus is left for future work. Finally, as common, a more accurate analysis of the decay of the error probability $\epsilon_n$ and establishing a strong converse are also of interest.

## APPENDIX A
## USEFUL MATHEMATICAL RESULTS

**Fact 16** (Stirling's bound)**.** *For* $n \in \mathbb{N}$

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq \sqrt{2\pi e n}\left(\frac{n}{e}\right)^n. \tag{A.1}$$

**Fact 17.** *For* $k_n = o(n)$ *as* $n \to \infty$*, it holds that*

$$\binom{n}{k} \sim \left(\frac{ne}{k}\right)^k \frac{1}{\sqrt{2\pi k}} \exp\left(-\frac{k^2(1+o(1))}{2n}\right) \tag{A.2}$$

*where* $a_n \sim b_n$ *means that* $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$*. Also,*

$$\binom{n}{k} \leq 2^{n \mathrm{h}_{\mathrm{bin}}(k/n)}. \tag{A.3}$$

**Fact 18** (Hoeffding's inequality [64])**.** *If* $\{X_i\}_{i\in[n]}$ *are independent RVs and* $a_i \leq X_i \leq b_i$ *with probability* 1 *then*

$$\mathbb{P}\left[\sum_{i=1}^n X_i - \mathbb{E}[X_i] \geq t\right] \leq \exp\left[-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right]. \tag{A.4}$$

**Fact 19** (The relative (multiplicative) Chernoff bound)**.** *For* $B_i \sim \text{Bernoulli}(p)$ *IID for* $i \in [n]$

$$\mathbb{P}\left[\frac{1}{n}\sum_{i=1}^n B_i - p \geq \xi p\right] \leq \exp\left[-\frac{\xi^2 p}{2+\xi}\right] \tag{A.5}$$

*for any* $\xi > 0$.

## APPENDIX B
## PROPERTIES OF THE POISSON DISTRIBUTION

**Fact 20** (Poissonization of the multinomial distribution)**.** *Let* $\tilde{M} \sim \mathrm{Poisson}(M)$*, and let* $\tilde{G}$ *be a random vector such that* $\tilde{G} \sim \mathrm{Multinomial}(\tilde{M}, (p_1, p_2, \dots p_J\})$ *conditioned on* $\tilde{M}$*, where* $\sum_{j\in[J]} p_j = 1$ *and* $p_j > 0$*. Then,* $\{\tilde{G}(j)\}_{j\in[J]}$ *are statistically independent and* $\tilde{G}(j) \sim \mathrm{Poisson}(Mp_j)$ *(unconditioned on* $\tilde{M}$*).*

Fact 20 can be verified by spelling out the conditional PMF of $\tilde{G}$ conditioned on $\tilde{M}$ [62, Thm. 5.6] in case $\{p_j\}$ are all equal, and can be easily extended to the non-uniform case (as in, e.g., [65, Lecture 11, Thm. 3.2]). The following then follows from [62, Corollary 5.9]:

**Lemma 21.** *Let* $G \sim \mathrm{Multinomial}(M, (p_1, p_2, \dots p_J\})$*, and let* $\tilde{G}$ *be an independent Poisson vector of the same dimension so that* $\mathbb{E}[\tilde{G}(j)] = \mathbb{E}[G(j)] = Mp_j$*. Then, for any event* $\mathcal{E}$

$$\mathbb{P}\left[G \in \mathcal{E}\right] \le \sqrt{eM} \cdot \mathbb{P}\left[\tilde{G} \in \mathcal{E}\right]. \tag{B.1}$$

**Lemma 22** (Chernoff's bound for Poisson RVs [62, Thm. 5.4])**.** *Let* $Z \sim \mathrm{Poisson}(\lambda)$*. Then, for* $\alpha \le 1$

$$\mathbb{P}\left[Z \le \alpha\lambda\right] \le e^{-\lambda}\left(\frac{e}{\alpha}\right)^{\alpha\lambda} = e^{-\lambda(1-\alpha\log(e/\alpha))} \le e^{-\frac{\lambda}{2}(1-\alpha)^2}. \tag{B.2}$$

**Lemma 23** (Poisson entropy)**.** *Let* $Z_\lambda \sim \mathrm{Poisson}(\lambda)$*. Then,*

$$H(Z_\lambda) = \frac{1}{2}\log\left[2\pi e\lambda\right] + O\left(\frac{1}{\lambda}\right). \tag{B.3}$$

*Also,*

$$H(Z_\lambda) \le \frac{1}{2}\log\left[2\pi e\left(\lambda + \frac{1}{12}\right)\right]. \tag{B.4}$$

*Finally,* $H(Z_\lambda)$ *is monotonic non-decreasing in* $\lambda$*.*

*Proof:* For the first properties, see [16, Lemma 10, Lemma 17b, Lemma 19]. For the monotonicity property, note that by the infinite divisibility of the Poisson distribution, if $\lambda_2 > \lambda_1$ then $Z_{\lambda_2} \stackrel{d}{=} Z_{\lambda_1} + \breve{Z}$ where $Z_{\lambda_1}$ and $\breve{Z} \sim \mathrm{Poisson}(\lambda_2 - \lambda_1)$ are independent. As conditioning reduces entropy

$$H(Z_{\lambda_2}) = H(Z_{\lambda_1} + \breve{Z}) \ge H(Z_{\lambda_1} + \breve{Z} \mid \breve{Z}) = H(Z_{\lambda_1} \mid \breve{Z}) = H(Z_{\lambda_1}). \tag{B.5}$$

■

**Lemma 24.** *Let* $V \sim \mathrm{Poisson}(\lambda)$*. Then,*

$$\mathbb{E}\left[V\log V\right] \le \lambda\log(1+\lambda). \tag{B.6}$$

*Proof:* We follow the idea in [66]. For any $v > 0$ and $u > 0$ it holds that $\log \frac{v}{u} \leq \frac{v}{u} - 1$ and so

$$v \log v = v \log \frac{v}{u} + v \log u \leq \frac{v^2}{u} + v \log \frac{u}{e}. \tag{B.7}$$

Hence,

$$\mathbb{E}\left[V \log V\right] \leq \mathbb{E}\left[\frac{V^2}{u} + V \log \frac{u}{e}\right] \tag{B.8}$$

$$= \frac{\lambda + \lambda^2}{u} + \lambda \log \frac{u}{e} \tag{B.9}$$

$$= \lambda \log(1 + \lambda), \tag{B.10}$$

when choosing $u = 1 + \lambda$. ■

## APPENDIX C
## PROPERTIES OF THE GAMMA DISTRIBUTION

Let $X \sim \mathrm{Gamma}(k, \theta)$ where $k > 0$ and $\theta > 0$. Then, the PDF is

$$f_{\mathrm{Gamma}}\left(x \mid k, \theta\right) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta} \tag{C.1}$$

and the CDF is

$$F_{\mathrm{Gamma}}\left(x \mid k, \theta\right) = \frac{1}{\Gamma(k)} \gamma(k, \frac{x}{\theta}) \tag{C.2}$$

where

$$\gamma\left(k, x\right) := \int_0^x t^{k-1} e^{-t} \mathrm{d}t \tag{C.3}$$

is the *incomplete gamma function*. For the special case of $k = \frac{1}{2}$ it holds that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$, and

$$\gamma\left(\frac{1}{2}, x\right) = \sqrt{\pi}\, \mathrm{erf}(\sqrt{x}) = 2 \int_0^{\sqrt{x}} e^{-t^2} \mathrm{d}t, \tag{C.4}$$

where $\mathrm{erf}(x) = 1 - 2\,\mathrm{Q}(\sqrt{2}x)$ and $\mathrm{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-t^2/2} \mathrm{d}t$ is the Q-function (the tail distribution function of the standard normal distribution). Also recall that for $X \sim \mathrm{Gamma}(k, \theta)$ it holds that $\mathbb{E}[X] = k\theta$ and $\mathrm{Var}[X] = k\theta^2$.

**Lemma 25.** *Let* $X \sim \mathrm{Gamma}(\frac{1}{2}, 2g_n)$. *Then, for* $\eta \in (-\infty, 1)$

$$\mathbb{P}\left[X \leq g_n^\eta\right] \leq \frac{1}{g_n^{(1-\eta)/2}} \tag{C.5}$$

*and for any* $\rho \in (0, \infty)$

$$\mathbb{P}\left[X \geq g_n^{1+\rho}\right] \leq 2e^{-g_n^\rho/2}. \tag{C.6}$$

*Thus,*

$$\mathbb{P}\left[X \notin [g_n^{\rho}, g_n^{1+\rho}]\right] \leq [1+o_n(1)] \cdot \frac{1}{g_n^{(1-\rho)/2}}. \tag{C.7}$$

*Proof:* It holds that

$$\mathbb{P}\left[X \leq g_n^{\eta}\right] = F_{\text{Gamma}}\left(g_n^{\eta} \mid \frac{1}{2}, 2g_n\right) \tag{C.8}$$

$$= \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{\frac{g_n^{\eta}}{2g_n}}} e^{-t^2} \mathrm{d}t \tag{C.9}$$

$$\leq \frac{2}{\sqrt{\pi}} \cdot \sqrt{\frac{g_n^{\eta}}{2g_n}} \tag{C.10}$$

$$\leq \frac{1}{g_n^{(1-\eta)/2}}. \tag{C.11}$$

Next,

$$\mathbb{P}\left[X \geq g_n^{1+\rho}\right] \leq 1 - F_{\text{Gamma}}\left(g_n^{1+\rho} \mid \frac{1}{2}, 2g_n\right) \tag{C.12}$$

$$= 1 - \operatorname{erf}\left(\sqrt{\frac{g_n^{\rho}}{2}}\right) \tag{C.13}$$

$$= 2\,\mathrm{Q}(g_n^{\rho/2}) \tag{C.14}$$

$$\leq 2e^{-g_n^{\rho}/2}, \tag{C.15}$$

using Chernoff's bound on the Q-function. ■

**Lemma 26.** *Let $X \sim \text{Gamma}(k, \theta)$ where $k > 0$ and $\theta > 0$. Then, for $t > 0$*

$$\mathbb{P}\left[X \geq \mathbb{E}[X] + t\right] = \mathbb{P}\left[X \geq k\theta + t\right] \tag{C.16}$$

$$\leq e^{-\frac{t}{2\theta}} + e^{-\frac{t^2}{4k\theta^2}}. \tag{C.17}$$

*Proof:* It holds that $\text{Var}[X] = k\theta^2$. Now, [67, Sec. 2.4] states that $X - \mathbb{E}[X]$ is sub-gamma RV on the right tail, with parameters $(v, c) = (k\theta^2, \theta)$. Hence, for any $s \geq 0$

$$\mathbb{P}\left[X - \mathbb{E}[X] \geq \sqrt{2k\theta^2 s} + \theta s\right] \leq e^{-s}. \tag{C.18}$$

Taking $t = \sqrt{2k\theta^2 s} + \theta s$ we have that $t \leq 2(\sqrt{2k\theta^2 s} \vee \theta s)$ (a sum is less than twice the maximum), and so $s \geq \frac{t}{2\theta} \wedge \frac{t^2}{4k\theta^2}$. Hence,

$$e^{-s} \leq \exp\left[-\left(\frac{t}{2\theta} \wedge \frac{t^2}{4k\theta^2}\right)\right] \leq e^{-\frac{t}{2\theta}} + e^{-\frac{t^2}{4k\theta^2}}. \tag{C.19}$$

■

## APPENDIX D
## POISSON CONCENTRATION OF LIPSCHITZ FUNCTIONS

Assume that $V \sim \text{Poisson}(\lambda)$. Then, Bobkov and Ledoux have shown the following logarithmic Sobolev inequality [55, Corollary 4]: It holds for any strictly positive function $f\colon \mathbb{N} \to \mathbb{R}_+$ that

$$\text{Ent}\left[f(V)\right] := \mathbb{E}\left[f(V)\log(f(V)\right] - \mathbb{E}\left[f(V)\right]\mathbb{E}\left[\log(f(V)\right] \tag{D.1}$$

$$\leq \lambda \mathbb{E}\left[\frac{1}{f(V)} \cdot |Df(V)|^2\right], \tag{D.2}$$

where $Df(v) := f(v+1) - f(v)$ for $v \in \mathbb{N}$ is the discrete derivative. Consequently, they have shown that for any function $g\colon \mathbb{N} \to \mathbb{R}$ with $\max_{v\in\mathbb{N}}|Dg(v)| \leq \tau$ it holds that [55, Eq. (24)]

$$\text{Ent}\left[e^{g(V)}\right] \leq \lambda e^{2\tau} \cdot \mathbb{E}\left[|Dg(V)|^2 \cdot e^{g(V)}\right]. \tag{D.3}$$

In turn, this implies the following concentration result:

**Lemma 27** (Poisson concentration of Lipschitz functions, a variant of [55, Prop. 11])**.** *Let $V_i \sim \text{Poisson}(\lambda_i)$ for $i \in [n]$ be independent, and let $\overline{\lambda} \geq \max_{i\in[n]} \lambda_i$ be given. Also let $f\colon \mathbb{N}^n \to \mathbb{R}$ be such that*

$$\max_{v^n \in \mathbb{N}^n} |f(v^n + e^n(i)) - f(v^n)| \leq \beta \tag{D.4}$$

*where $e^n(i)$ is the $i$th standard basis vector in $\mathbb{R}^n$. Then, for any $t > 0$*

$$\mathbb{P}\left[f(V^n) - \mathbb{E}[f(V^n)] > n\delta\right] \leq \exp\left[-n \cdot \frac{\delta^2}{16\beta^2\overline{\lambda} + 3\beta\delta}\right] \tag{D.5}$$

*Proof:* The condition in the lemma trivially implies that choosing $\alpha^2 = n\beta^2$ results

$$\sum_{i=1}^{n} |f(v^n + e^n(i)) - f(v^n)|^2 \leq \alpha^2 \tag{D.6}$$

in the notation of [55, Prop. 11]. The proof therein then relies on the tensorization property (subadditivity) of the entropy functional, which is stated for IID $\{V_i\}_{i\in[n]}$, but holds more generally when they are just independent [67, Thm. 4.22]. Then, since

$$\text{Ent}\left[e^{g(V_i)}\right] \leq \overline{\lambda} e^{2\tau} \cdot \mathbb{E}\left[|Dg(V_i)|^2 \cdot e^{g(V_i)}\right], \tag{D.7}$$

Herbst argument and the entropy method can be used in the exact same manner to show that

$$\mathbb{P}\left[f(V^n) - \mathbb{E}[f(V^n)] > n\delta\right] \leq \exp\left[-\frac{n\delta}{4\beta}\log\left(1 + \frac{\delta}{2\beta\overline{\lambda}}\right)\right]. \tag{D.8}$$

(we set $\alpha^2 = n\beta^2$, $c_1 = \overline{\lambda}$ and $c_2 = 2$ in the bound therein). Finally, we note that

$$u\log(1+u) = (1+u)\log(1+u) - u + u - \log(1+u) \tag{D.9}$$

$$\overset{(*)}{\geq} \frac{u^2}{2(1+u/3)} + u - \log(1+u) \tag{D.10}$$

$$\overset{(**)}{\geq} \frac{u^2}{2(1+u/3)}, \tag{D.11}$$

where $(*)$ was stated in [67, Exercise 2.8], and $(**)$ follows from $u \geq \log(1+u)$ for $u \geq 0$. Using this bound in (D.8) with $u = \frac{\delta}{2\beta\overline{\lambda}}$ establishes the claim of the lemma. ■

## REFERENCES


[1] A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Information theory of molecular communication: Directions and challenges," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 2, pp. 120–142, 2016.

[2] G. M. Church, Y. Gao, and S. Kosuri, "Next-generation digital information storage in DNA," *Science*, vol. 337, no. 6102, pp. 1628–1628, 2012.

[3] N. Goldman, S. Bertone, P.and Chen, C. Dessimoz, E. M. LeProust, B. Sipos, and E. Birney, "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA," *Nature*, vol. 494, no. 7435, pp. 77–80, 2013.

[4] R. N. Grass, R. Heckel, M. Puddu, D. Paunescu, and W. J. S., "Robust chemical preservation of digital information on DNA in silica with error-correcting codes," *Angewandte Chemie International Edition*, vol. 54, no. 8, pp. 2552–2555, 2015.

[5] S. M. H. T. Yazdi, Y. Yuan, J. Ma, H. Zhao, and O. Milenkovic, "A rewritable, random-access DNA-based storage system," *Scientific reports*, vol. 5, no. 1, pp. 1–10, 2015.

[6] H. M. Kiah, G. J. Puleo, and O. Milenkovic, "Codes for DNA sequence profiles," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3125–3146, 2016.

[7] Y. Erlich and D. Zielinski, "DNA fountain enables a robust and efficient storage architecture," *Science*, vol. 355, no. 6328, pp. 950–954, 2017.

[8] F. Sala, R. Gabrys, C. Schoeny, and L. Dolecek, "Exact reconstruction from insertions in synchronization codes," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2428–2445, 2017.

[9] L. Organick, S. D. Ang, Y. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M. Z. Racz, G. Kamath, P. Gopalan, and B. Nguyen, "Random access in large-scale DNA data storage," *Nature biotechnology*, vol. 36, no. 3, pp. 242–248, 2018.

[10] A. Lenz, P. H. Siegel, A. Wachter-Zeh, and E. Yaakobi, "Anchor-based correction of substitutions in indexed sets," in *IEEE International Symposium on Information Theory*, pp. 757–761, IEEE, 2019.

[11] J. Sima, N. Raviv, and J. Bruck, "On coding over sliced information," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2793–2807, 2021.

[12] Y. Tang and F. Farnoud, "Error-correcting codes for noisy duplication channels," *IEEE Transactions on Information Theory*, vol. 67, no. 6, pp. 3452–3463, 2021.

[13] I. Shomorony and R. Heckel, "DNA-based storage: Models and fundamental limits," *IEEE Transactions on Information Theory*, vol. 67, no. 6, pp. 3675–3689, 2021.

[14] I. Shomorony and R. Heckel, "Information-theoretic foundations of DNA data storage," *Foundations and Trends® in Communications and Information Theory*, vol. 19, no. 1, pp. 1–106, 2022.

[15] O. Sabary, Y. Orlev, R. Shafir, L. Anavy, E. Yaakobi, and Z. Yakhini, "Solqc: Synthetic oligo library quality control tool," *Bioinformatics*, vol. 37, no. 5, pp. 720–722, 2021.

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

[17] A. Lenz, P. H. Siegel, A. Wachter-Zeh, and E. Yaakobi, "An upper bound on the capacity of the DNA storage channel," in *IEEE Information Theory Workshop*, pp. 1–5, IEEE, 2019.

[18] A. Lenz, P. H. Siegel, A. Wachter-Zeh, and E. Yaakohi, "Achieving the capacity of the DNA storage channel," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8846–8850, IEEE, 2020.

[19] N. Weinberger, "Error probability bounds for coded-index DNA storage channels," *IEEE Transactions on Information Theory*, vol. 68, no. 11, pp. 7005–7022, 2022.

[20] N. Weinberger and N. Merhav, "The DNA storage channel: Capacity and error probability bounds," *IEEE Transactions on Information Theory*, vol. 68, no. 9, pp. 5657–5700, 2022.

[21] P. K. Vippathalla and N. Kashyap, "The secure storage capacity of a DNA wiretap channel model," *IEEE Transactions on Information Theory*, 2023.

[22] A. Lenz, P. H. Siegel, A. Wachter-Zeh, and E. Yaakobi, "Coding over sets for DNA storage," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2331–2351, 2019.

[23] A. Lenz, L. Welter, and S. Puchinger, "Achievable rates of concatenated codes in DNA storage under substitution errors," in *International Symposium on Information Theory and Its Applications*, pp. 269–273, IEEE, 2020.

[24] M. Kovačević and V. Y. Tan, "Coding for the permutation channel with insertions, deletions, substitutions, and erasures," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 1933–1937, IEEE, 2017.

[25] M. Kovačević and V. Y. Tan, "Codes in the space of multisets – coding for permutation channels with impairments," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 5156–5169, 2018.

[26] A. Makur, "Coding theorems for noisy permutation channels," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6723–6748, 2020.

[27] J. Tang and Y. Polyanskiy, "Capacity of noisy permutation channels," *IEEE Transactions on Information Theory*, 2023.

[28] W. Lu and A. Makur, "Permutation capacity region of adder multiple-access channels," *IEEE Transactions on Information Theory*, 2024.

[29] Y. Choi, T. Ryu, A. C. Lee, H. Choi, H. Lee, J. Park, S.-H. Song, S. Kim, H. Kim, W. Park, *et al.*, "High information capacity dna-based data storage with augmented encoding characters using degenerate bases," *Scientific reports*, vol. 9, no. 1, p. 6582, 2019.

[30] L. Anavy, I. Vaknin, O. Atar, R. Amit, and Z. Yakhini, "Data storage in dna with fewer synthesis cycles using composite dna letters," *Nature biotechnology*, vol. 37, no. 10, pp. 1229–1236, 2019.

[31] A. Kobovich, E. Yaakobi, and N. Weinberger, "M-dab: An input-distribution optimization algorithm for composite dna storage by the multinomial channel," *arXiv preprint arXiv:2309.17193*, 2023.

[32] M. R. Frey, "Information capacity of the Poisson channel," *IEEE Transactions on Information Theory*, vol. 37, no. 2, pp. 244–256, 1991.

[33] D. Brady and S. Verdú, "The asymptotic capacity of the direct detection photon channel with a bandwidth constraint," in *Proc. 28th Allerton Conf. Communication, Control and Computing*, pp. 691–700, 1990.

[34] S. Shamai and A. Lapidoth, "Bounds on the capacity of a spectrally constrained Poisson channel," *IEEE Transactions on Information Theory*, vol. 39, no. 1, pp. 19–29, 1993.

[35] A. Lapidoth and S. Shamai, "The Poisson multiple-access channel," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 488–501, 1998.

[36] A. Martinez, "Spectral efficiency of optical direct detection," *JOSA B*, vol. 24, no. 4, pp. 739–749, 2007.

[37] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," *IEEE Transactions on Information theory*, vol. 58, no. 3, pp. 1302–1318, 2012.

[38] J. Cao, S. Hranilovic, and J. Chen, "Capacity-achieving distributions for the discrete-time Poisson channel – part I: General properties and numerical techniques," *IEEE transactions on communications*, vol. 62, no. 1, pp. 194–202, 2013.

[39] J. Cao, S. Hranilovic, and J. Chen, "Capacity-achieving distributions for the discrete-time Poisson channel – part II: Binary inputs," *IEEE transactions on communications*, vol. 62, no. 1, pp. 203–213, 2013.

[40] M. Cheraghchi and J. Ribeiro, "Improved upper bounds and structural results on the capacity of the discrete-time Poisson channel," *IEEE Transactions on Information Theory*, vol. 65, no. 7, pp. 4052–4068, 2019.

[41] A. Dytso, M. Fauß, and H. V. Poor, "The vector Poisson channel: On the linearity of the conditional mean estimator," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5894–5903, 2020.

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

[43] P. Harremoës, "Binomial and poisson distributions as maximum entropy distributions," *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 2039–2041, 2001.

[44] J. A. Adell, A. Lekuona, and Y. Yu, "Sharp bounds on the entropy of the Poisson law and related quantities," *IEEE transactions on information theory*, vol. 56, no. 5, pp. 2299–2306, 2010.

[45] I. Sason, "Entropy bounds for discrete random variables via maximal coupling," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7118–7131, 2013.

[46] I. Kontoyiannis, P. Harremoës, and O. Johnson, "Entropy and the law of small numbers," *IEEE Transactions on information theory*, vol. 51, no. 2, pp. 466–472, 2005.

[47] O. Johnson, "Log-concavity and the maximum entropy property of the poisson distribution," *Stochastic Processes and their Applications*, vol. 117, no. 6, pp. 791–802, 2007.

[48] Y. Yu, "Monotonic convergence in an information-theoretic law of small numbers," *IEEE transactions on information theory*, vol. 55, no. 12, pp. 5412–5422, 2009.

[49] L. Wang, D. E. Carlson, M. R. Rodrigues, R. Calderbank, and L. Carin, "A Bregman matrix and the gradient of mutual information for vector Poisson and Gaussian channels," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2611–2629, 2014.

[50] A. Dytso and H. V. Poor, "Estimation in Poisson noise: Properties of the conditional mean estimator," *IEEE Transactions on Information Theory*, vol. 66, no. 7, pp. 4304–4323, 2020.

[51] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley-Interscience, 2006.

[52] R. Heckel, G. Mikutis, and R. N. Grass, "A characterization of the DNA data storage channel," *Scientific reports*, vol. 9, no. 1, p. 9663, 2019.

[53] A. Feinstein, "A new basic theorem of information theory," 1954.

[54] Y. Polyanskiy and Y. Wu, *Information Theory: From Coding to Learning*. Cambridge University Press, 2023+.

[55] S. G. Bobkov and M. Ledoux, "On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures," *Journal of functional analysis*, vol. 156, no. 2, pp. 347–365, 1998.

[56] D. Guo, S. Shamai, and S. Verdú, "Mutual information and conditional mean estimation in Poisson channels," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 1837–1849, 2008.

[57] J. Jiao, K. Venkat, and T. Weissman, "Pointwise relations between information and estimation in the Poisson channel," in *2013 IEEE International Symposium on Information Theory*, pp. 449–453, IEEE, 2013.

[58] D. Guo, S. Shamai, S. Verdú, *et al.*, "The interplay between information and estimation measures," *Foundations and Trends® in Signal Processing*, vol. 6, no. 4, pp. 243–429, 2013.

[59] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR computational mathematics and mathematical physics*, vol. 7, no. 3, pp. 200–217, 1967.

[60] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1289–1301, 2011.

[61] Y. Polyanskiy and Y. Wu, "Wasserstein continuity of entropy and outer bounds for interference channels," *IEEE Transactions on Information Theory*, vol. 62, no. 7, pp. 3992–4002, 2016.

[62] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge University Press, 2017.

[63] A. Martinez, "Capacity bounds for MIMO Poisson channels with inter-symbol interference," *Optical Communication Theory and Techniques*, pp. 37–44, 2005.

[64] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *The collected works of Wassily Hoeffding*, pp. 409–426, 1994.

[65] C. Seshadhri, "Lecture notes for cse 290a: Randomized algorithms," 2020. Available at https://users.soe.ucsc.edu/~sesh/Teaching/2020/CSE290A/Slides/Lecture11.pdf.

[66] I. P. (https://mathoverflow.net/users/36721/iosif pinelis), "What is (approximately) the expected value of $X \log X$ where $X$ is binomial (or Poisson)?." MathOverflow. URL:https://mathoverflow.net/q/321933 (version: 2019-01-29).

[67] S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.