entropy

MDPI

*Article*

# Kolmogorov Capacity with Overlap †

**Anshuka Rangi and Massimo Franceschetti ***

Department of Electrical and Computer Engineering, University of California at San Diego, 9500 Gilman Drive, Mail Code 0407, La Jolla, CA 92093-0407, USA; rangianshukaiitr@gmail.com

* Correspondence: massimo@ece.ucsd.edu

† This article is a revised and expanded version of a paper entitled Insights into [Towards a Non-Stochastic Information Theory], which was presented at [2019 IEEE International Symposium of Information Theory, Paris, France, 7–12 July 2019] and [Channel Coding Theorems in Non-Stochastic Information Theory], which was presented at [2021 IEEE International Symposium of Information Theory, Melbourne, VIC, Australia, 12–20 July 2021].

**Abstract:** The notion of $\delta$-mutual information between non-stochastic uncertain variables is introduced as a generalization of Nair's non-stochastic information functional. Several properties of this new quantity are illustrated and used in a communication setting to show that the largest $\delta$-mutual information between received and transmitted codewords over $\epsilon$-noise channels equals the $(\epsilon, \delta)$-capacity. This notion of capacity generalizes the Kolmogorov $\epsilon$-capacity to packing sets of overlap at most $\delta$ and is a variation of a previous definition proposed by one of the authors. Results are then extended to more general noise models, including non-stochastic, memoryless, and stationary channels. The presented theory admits the possibility of decoding errors, as in classical information theory, while retaining the worst-case, non-stochastic character of Kolmogorov's approach.

**Keywords:** $\epsilon$-capacity; $(\epsilon, \delta)$-capacity; mutual information; non-stochastic uncertainty

## 1. Introduction

Shannon's celebrated channel coding theorem states that the capacity is the supremum of the mutual information between the input and the output of the channel [1]. In this setting, the mutual information is intended as the amount of information obtained regarding the random variable at the input of the channel by observing the random variable at the output of the channel, and the capacity is the largest rate of communication that can be achieved with an arbitrarily small probability of error. In an effort to provide an analogous result for safety-critical control systems where occasional decoding errors can result in catastrophic failures, Nair introduced a non-stochastic mutual information functional and established that this equals the zero-error capacity [2], namely the largest rate of communication that can be achieved with zero probability of error. Nair's approach is based on the calculus of non-stochastic uncertain variables (UVs), and his definition of mutual information in a non-stochastic setting is based on the quantization of the range of uncertainty of a UV induced by the knowledge of the other. While Shannon's theorem leads to a single letter expression, Nair's result is multi-letter, involving the non-stochastic information between codeword blocks of $n$ symbols. The zero-error capacity can also be formulated as a graph-theoretic property, and the absence of a single-letter expression for general graphs is well known [3,4]. Extensions of Nair's non-stochastic approach to characterize the zero-error capacity in the presence of feedback from the receiver to the transmitter using nonstochastic directed mutual information have been considered in [5].

check for updates

Academic Editor: Raúl Alcaraz

Received: 4 March 2025
Revised: 9 April 2025
Accepted: 10 April 2025
Published: 27 April 2025

**Citation:** Rangi, A.; Franceschetti, M. Kolmogorov Capacity with Overlap. *Entropy* **2025**, *27*, 472. https://doi.org/10.3390/e27050472

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Kolmogorov introduced the notion of $\epsilon$-capacity in the context of functional spaces as the logarithm base two of the packing number of the space, namely the logarithm of the maximum number of balls of radius $\epsilon$ that can be placed in the space without overlap [6]. Determining this number is analogous to designing a channel codebook such that the distance between any two codewords is at least $2\epsilon$. In this way, any transmitted codeword that is subject to a perturbation of, at most, $\epsilon$ can be recovered at the receiver without error. It follows that the $\epsilon$-capacity per transmitted symbol (*viz.* per signal dimension) corresponds to the zero-error capacity of an additive channel having arbitrary bounded noise of the radius at most $\epsilon$. Lim and Franceschetti extended this concept introducing the $(\epsilon, \delta)$ capacity [7], defined as the logarithm base two of the largest number of balls of radius $\epsilon$ that can be placed in the space with an average codeword overlap of, at most, $\delta$. In this setting, $\delta$ measures the amount of error that can be tolerated when designing a codebook in a non-stochastic setting, and the $(\epsilon, \delta)$ capacity per transmitted symbol corresponds to the largest rate of communication with error at most $\delta$.

The first contribution of this paper is to consider a generalization of Nair's mutual information based on a quantization of the range of uncertainty of a UV given the knowledge of another, which reduces the uncertainty to, at most, $\delta$, and to show that this new notion corresponds to the $(\epsilon, \delta)$ capacity. Our definition of $(\epsilon, \delta)$ capacity is a variation of the one in [7], as it is required to bound the overlap between any pair of balls, rather than the average overlap. For $\delta = 0$, we recover Nair's result for the Kolmogorov $\epsilon$-capacity or, equivalently, for the zero-error capacity of an additive, bounded noise channel. We then extend the results to more general channels where the noise can be different across codewords and is not necessarily contained within a ball of radius $\epsilon$. Finally, we consider the class of non-stochastic, memoryless, stationary uncertain channels, where the noise experienced by a codeword of $n$ symbols factorizes into $n$ identical terms describing the noise experienced by each codeword symbol. This is the non-stochastic analog of a discrete memoryless channel (DMC), where the current output symbol depends only on the current input symbol, not on any of the previous input symbols, and where the noise distribution is constant across symbol transmissions and differs from Kolmogorov's $\epsilon$-noise channel, where the noise experienced by one symbol affects the noise experienced by other symbols (in Kolmogorov's setting, the noise occurs within a ball of radius $\epsilon$. It follows that for any realization where the noise along one dimension (*viz.* symbol) is close to $\epsilon$, the noise experienced by all other symbols lying in the remaining dimensions must be close to zero.). Letting $(1 - \delta_n)$ be the confidence of correct decoding after transmitting $n$ symbols, we introduce several notions of capacity and establish coding theorems in terms of mutual information for all of them, including a generalization of the zero-error capacity that requires the error sequence $\{\delta_n\}$ to remain constant and a non-stochastic analog of Shannon's capacity that requires the error sequence to vanish, as in $n \to \infty$.

Finally, since in Nair's case, all of our results are multi-letter, in the Supplementary Materials, we provide some sufficient conditions for the factorization of the mutual information leading to a single-letter expression for the non-stochastic capacity of stationary, memoryless, uncertain channels, provide some examples in which these conditions are satisfied, and compute the corresponding capacity.

The rest of the paper is organized as follows: Section 2 introduces the mathematical framework of non-stochastic uncertain variables that are used throughout the paper. Section 3 introduces the concept of non-stochastic mutual information. Section 4 gives an operational definition of the capacity of a communication channel and relates it to the mutual information. Section 5 extends the results to more general channel models, and Section 6 concentrates on the special case of stationary, memoryless, uncertain channels. Section 7 draws conclusions and discusses future directions.

## 2. Uncertain Variables

We start by reviewing the mathematical framework used in [2] to describe UVs. A UV $X$ is a mapping from a sample space $\Omega$ to a set $\mathscr{X}$, i.e., for all $\omega \in \Omega$, we have $x = X(\omega) \in \mathscr{X}$, namely

$$\begin{aligned} X: \quad & \Omega \to \mathscr{X} \\ & \omega \mapsto X(\omega) = x. \end{aligned} \tag{1}$$

Given a UV $X$, the marginal range of $X$ is

$$[\![X]\!] = \{X(\omega) : \omega \in \Omega\}. \tag{2}$$

The joint range of the two UVs $X$ and $Y$ is

$$[\![X, Y]\!] = \{(X(\omega), Y(\omega)) : \omega \in \Omega\}. \tag{3}$$

Given a UV $Y$, the conditional range of $X$ given $Y = y$ is

$$[\![X|y]\!] = \{X(\omega) : Y(\omega) = y, \omega \in \Omega\}, \tag{4}$$

and the conditional range of $X$ given $Y$ is

$$[\![X|Y]\!] = \{[\![X|y]\!] : y \in [\![Y]\!]\}. \tag{5}$$

Thus, $[\![X|Y]\!]$ denotes the uncertainty in $X$, given the realization of $Y$ and that $[\![X, Y]\!]$ represents the total joint uncertainty of $X$ and $Y$, namely

$$[\![X, Y]\!] = \cup_{y \in [\![Y]\!]} [\![X|y]\!] \times \{y\}. \tag{6}$$

Finally, two UVs, $X$ and $Y$, are independent if for all $x \in [\![X]\!]$,

$$[\![Y|x]\!] = [\![Y]\!], \tag{7}$$

which also implies that for all $y \in [\![Y]\!]$,

$$[\![X|y]\!] = [\![X]\!]. \tag{8}$$

## 3. $\delta$-Mutual Information

### *3.1. Uncertainty Function*

We now introduce a class of functions that are used to express the amount of uncertainty in determining one UV given another. In our setting, an uncertainty function associates a positive number with a given set, which expresses the "massiveness" or "size" of that set.

**Definition 1.** *Given the set of non-negative real numbers $\mathbb{R}_0^+$ and any set of $\mathscr{X}$, $m_{\mathscr{X}} : 2^{\mathscr{X}} \to \mathbb{R}_0^+$ is an uncertainty function if it is finite and strongly transitive:*

*We have $m_{\mathscr{X}}(\emptyset) = 0$, and for all $\mathscr{S} \subseteq \mathscr{X}$, $\mathscr{S} \neq \emptyset$,*

$$0 < m_{\mathscr{X}}(\mathscr{S}) < \infty. \tag{9}$$

*For all $\mathscr{S}_1, \mathscr{S}_2 \subseteq \mathscr{X}$, we have*

$$\max\{m_{\mathscr{X}}(\mathscr{S}_1), m_{\mathscr{X}}(\mathscr{S}_2)\} \leq m_{\mathscr{X}}(\mathscr{S}_1 \cup \mathscr{S}_2). \tag{10}$$

In the case where $\mathscr{X}$ is measurable, an uncertainty function can easily be constructed using a measure. In the case where $\mathscr{X}$ is a bounded (not necessarily measurable) metric space and the input set $\mathscr{S}$ contains at least two points, an example of an uncertainty function is the diameter.

### *3.2. Association and Dissociation Between UVs*

We now introduce notions of association and dissociation between UVs. In the following definitions, we let $m_{\mathscr{X}}(.)$ and $m_{\mathscr{Y}}(.)$ be uncertainity functions defined over sets $\mathscr{X}$ and $\mathscr{Y}$ corresponding to UVs $X$ and $Y$. We use the notation $\mathscr{A} \succ \delta$ to indicate that for all $a \in \mathscr{A}$, we have $a > \delta$. Similarly, we use $\mathscr{A} \preceq \delta$ to indicate that for all $a \in \mathscr{A}$, we have $a \leq \delta$, where $\mathscr{A} \subseteq \mathbb{R}$ and $\delta \in \mathbb{R}$. For $\mathscr{A} = \emptyset$, we assume that $\mathscr{A} \preceq \delta$ is always satisfied, while $\mathscr{A} \succ \delta$ is not. Whenever we consider $i \neq j$, we also assume that $y_i \neq y_j$ and $x_i \neq x_j$, where $x_i, x_j \in [\![X]\!]$ and $y_i, y_j \in [\![Y]\!]$.

**Definition 2.** *The sets of association for UVs $X$ and $Y$ are*

$$\mathscr{A}(X;Y) = \left\{ \frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!])}{m_{\mathscr{X}}([\![X]\!])} : y_1, y_2 \in [\![Y]\!] \right\} \setminus \{0\}, \tag{11}$$

$$\mathscr{A}(Y;X) = \left\{ \frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!])}{m_{\mathscr{Y}}([\![Y]\!])} : x_1, x_2 \in [\![X]\!] \right\} \setminus \{0\}. \tag{12}$$

The sets of association are used to describe the correlation between two uncertain variables. Since $m_{\mathscr{X}}(\emptyset) = 0$, the exclusion of the zero value in (11) and (12) occurs when there is no overlap between the two conditional ranges.

**Definition 3.** *For any $\delta_1, \delta_2 \in [0,1)$, UVs $X$ and $Y$ are disassociated at levels $(\delta_1, \delta_2)$ if the following inequalities hold:*

$$\mathscr{A}(X;Y) \succ \delta_1, \tag{13}$$

$$\mathscr{A}(Y;X) \succ \delta_2, \tag{14}$$

*and, in this case, we write $(X,Y) \overset{d}{\leftrightarrow} (\delta_1, \delta_2)$.*

Having UVs $X$ and $Y$ be disassociated at levels $(\delta_1, \delta_2)$ indicates that at least two conditional ranges $[\![X|y_1]\!]$ and $[\![X|y_2]\!]$ have non-zero overlap and that, given any two conditional ranges, either they do not overlap or the uncertainty associated with their overlap is greater than a $\delta_1$ fraction of the total uncertainty associated with $[\![X]\!]$; the same holds for conditional ranges $[\![Y|x_1]\!]$ and $[\![Y|x_2]\!]$ and for level $\delta_2$. The levels of disassociation can be viewed as lower bounds in the amount of residual uncertainty in each variable when the other is known. If $X$ and $Y$ are independent, then all the conditional ranges completely overlap, $\mathscr{A}(X;Y)$ and $\mathscr{A}(Y;X)$ contain only the element one, and the variables are maximally disassociated (see Figure 1a).

In this case, knowledge of $Y$ does not reduce the uncertainty of $X$, and vice versa. On the other hand, when the uncertainty associated with any of the non-zero intersections of the conditional ranges decreases but remains positive, $X$ and $Y$ become less disassociated in the sense that knowledge of $Y$ can reduce the residual uncertainty of $X$, and vice versa (see Figure 1b). When the intersection between every pair of conditional ranges becomes empty, the variables cease to be disassociated (see Figure 1c). Note that excluding the value of 1 in the definition of disassociation allows us to distinguish the case of disassociation from the case of full independence.

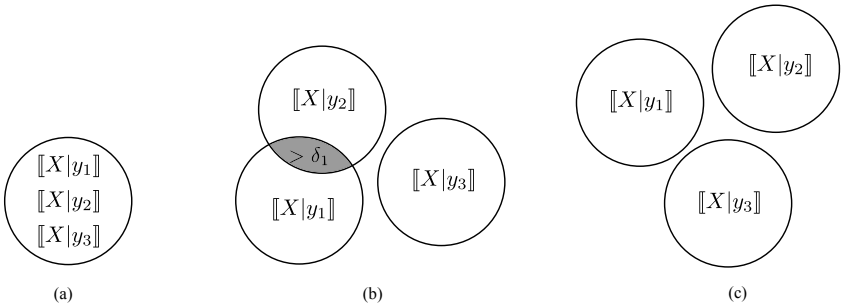



**Figure 1.** Illustration of disassociation between UVs. Case (**a**): variables are maximally disassociated, and all conditional ranges completely overlap, in that all conditional ranges are equal to $[\![X]\!]$ (or $[\![Y]\!]$). Case (**b**): variables are disassociated at some levels $(\delta_1, \delta_2)$, and there is some overlap between at least two conditional ranges. Case (**c**): variables are not disassociated at any levels, and there is no overlap between the conditional ranges.

An analogous definition of association is given to provide upper bounds on the residual uncertainty of one uncertain variable when the other is known.

**Definition 4.** *For any $\delta_1, \delta_2 \in [0,1]$, we say that UVs $X$ and $Y$ are associated at levels $(\delta_1, \delta_2)$ if the following inequalities hold:*

$$\mathscr{A}(X;Y) \preceq \delta_1, \tag{15}$$

$$\mathscr{A}(Y;X) \preceq \delta_2, \tag{16}$$

*and in this case, we write $(X,Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$.*

Note that $\delta_1, \delta_2 = 1$ is included in Definition 4 and not in Definition 3. This is because in Definition 3, we have a strict lower bound on the uncertainty in the association sets.

The following lemma provides the necessary and sufficient conditions for association to hold at given levels. These conditions are stated for all points in the marginal ranges $[\![Y]\!]$ and $[\![X]\!]$. They show that in the case of association, one can also include in the definition the conditional ranges that have zero intersection. This is not the case for disassociation.

**Lemma 1.** *For any $\delta_1, \delta_2 \in [0,1]$, $(X,Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$ if and only if for all $y_1, y_2 \in [\![Y]\!]$, we have*

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} \leq \delta_1, \tag{17}$$

*and for all $x_1, x_2 \in [\![X]\!]$, we have*

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} \leq \delta_2. \tag{18}$$

**Proof.** The proof is given in Appendix A. □

An immediate yet important consequence of our definitions is that both association and disassociation at given levels $(\delta_1, \delta_2)$ cannot hold simultaneously. We also understand that, given any two UVs, one can always select $\delta_1$ and $\delta_2$ to be large enough such that they are associated at levels $(\delta_1, \delta_2)$. In contrast, as the smallest value in the sets $\mathcal{A}(X;Y)$ and $\mathcal{A}(Y;X)$ tends towards zero, the variables eventually cease to be disassociated. Finally, it is possible that two uncertain variables are neither associated nor disassociated at given levels $(\delta_1, \delta_2)$. Also, any two uncertain variables are associated at level $(1,1)$ trivially by definition.

**Example 1.** *Consider three individuals, a, b, and c, are going for a walk along a path. Assume they take, at most, 15, 20, and 10 min to finish their walk, respectively. Assume a starts walking at time 5:00, b starts walking at 5:10, and c starts walking at 5:20. Figure 2 shows the possible time intervals for the walkers on the path. Let an uncertain variable W represent the set of walkers that are present on the path at any time and an uncertain variable T represent the time at which any walker on the path finishes their walk. Then, we have the following marginal ranges:*

$$[\![W]\!] = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{b, c\}\}, \tag{19}$$

$$[\![T]\!] = [5:00, 5:30]. \tag{20}$$

*We also have the following conditional ranges:*

$$[\![T|\{a\}]\!] = [5:00, 5:15], \tag{21}$$

$$[\![T|\{b\}]\!] = [5:10, 5:30], \tag{22}$$

$$[\![T|\{c\}]\!] = [5:20, 5:30], \tag{23}$$

$$[\![T|\{a, b\}]\!] = [5:10, 5:15], \tag{24}$$

$$[\![T|\{b, c\}]\!] = [5:20, 5:30]. \tag{25}$$

*For all* $t \in [5{:}00, 5{:}10)$*, we have*

$$[\![W|t]\!] = \{\{a\}\}; \tag{26}$$

*for all* $t \in [5{:}10, 5{:}15]$*, we have*

$$[\![W|t]\!] = \{\{a, b\}, \{a\}, \{b\}\}; \tag{27}$$

*for all* $t \in (5{:}15, 5{:}20)$*, we have*

$$[\![W|t]\!] = \{\{b\}\}; \tag{28}$$

*and for all* $t \in [5{:}20, 5{:}30]$*, we have*

$$[\![W|t]\!] = \{\{b, c\}, \{b\}, \{c\}\}. \tag{29}$$

*Now, let the uncertainty function of a time set* $\mathscr{S}$ *be*

$$m_{\mathscr{T}}(\mathscr{S}) = \begin{cases} \mathcal{L}(\mathscr{S}) + 10 \text{ if } \mathscr{S} \neq \emptyset, \\ 0 \text{ otherwise }, \end{cases} \tag{30}$$

*where* $\mathcal{L}(\cdot)$ *is the Lebesgue measure. Let the uncertainty function* $m_{\mathscr{W}}(.)$ *associated with a set of individuals be the cardinality of the set. Then, the sets of association are*

$$\mathscr{A}(W; T) = \{1/5, 3/5\}, \tag{31}$$

$$\mathscr{A}(T; W) = \{3/8, 1/2\}. \tag{32}$$

*It follows that for all* $\delta_1 < 1/5$ *and* $\delta_2 < 3/8$*, we have*

$$(W, T) \overset{d}{\leftrightarrow} (\delta_1, \delta_2), \tag{33}$$

*and the residual uncertainty in W given T is at least a* $\delta_1$ *fraction of the total uncertainty in W, while the residual uncertainty in T given W is at least a* $\delta_2$ *fraction of the total uncertainty in T. On the other hand, for all* $\delta_1 \geq 3/5$ *and* $\delta_2 \geq 1/2$*, we have*

$$(W, T) \overset{a}{\leftrightarrow} (\delta_1, \delta_2), \tag{34}$$

*and the residual uncertainty in $W$ given $T$ is at most a $\delta_1$ fraction of the total uncertainty in $W$, while the residual uncertainty in $T$ given $W$ is at most a $\delta_2$ fraction of the total uncertainty in $T$.*

*Finally, if $1/5 \leq \delta_1 < 3/5$ or $3/8 \leq \delta_2 < 1/2$, then $W$ and $T$ are neither associated nor disassociated.*

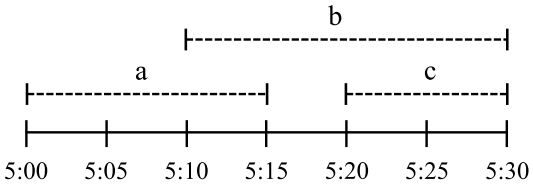



**Figure 2.** Illustration of the possible time intervals for the walkers on the path.

### *3.3. δ-Mutual Information*

We now introduce the mutual information between uncertain variables in terms of some structural properties of covering sets. Intuitively, for any $\delta \in [0, 1]$ the $\delta$-mutual information, expressed in bits, represents the most refined knowledge that one uncertain variable provides about the other at a given level of confidence $(1 - \delta)$. We express this idea by considering the quantization of the range of uncertainty of one variable, induced by the knowledge of the other. Such quantization ensures that the variable can be identified with uncertainty at most $\delta$. The notions of association and disassociation introduced above are used to ensure that the mutual information is well defined, in that it can be positive and exhibits a certain symmetric property.

**Definition 5.** *$\delta$-Connectedness and $\delta$-isolation.*

- *For any $\delta \in [0, 1]$, points $x_1, x_2 \in [\![X]\!]$ are $\delta$-connected via $[\![X|Y]\!]$ and are denoted by $x_1 \overset{\delta}{\rightsquigarrow} x_2$ if there exists a finite sequence $\{[\![X|y_i]\!]\}_{i=1}^{N}$ of conditional sets such that $x_1 \in [\![X|y_1]\!]$, $x_2 \in [\![X|y_N]\!]$ and for all $1 < i \leq N$, we have*

$$\frac{m_{\mathscr{X}}([\![X|y_i]\!] \cap [\![X|y_{i-1}]\!])}{m_{\mathscr{X}}([\![X]\!])} > \delta. \tag{35}$$

  *If $x_1 \overset{\delta}{\rightsquigarrow} x_2$ and $N = 1$, then we say that $x_1$ and $x_2$ are singly $\delta$-connected via $[\![X|Y]\!]$, i.e., there exists a $y$ such that $x_1, x_2 \in [\![X|y]\!]$.*
- *A set $\mathscr{S} \subseteq [\![X]\!]$ is $\delta$-connected via $[\![X|Y]\!]$ if every pair of points in the set is $\delta$-connected via $[\![X|Y]\!]$.*
- *A set $\mathscr{S} \subseteq [\![X]\!]$ is singly $\delta$-connected via $[\![X|Y]\!]$ if there exists a $y \in [\![Y]\!]$ such that every point in the set is contained in $[\![X|y]\!]$, namely $\mathscr{S} \subseteq [\![X|y]\!]$.*
- *Two sets $\mathscr{S}_1, \mathscr{S}_2 \subseteq [\![X]\!]$ are $\delta$-isolated via $[\![X|Y]\!]$ if no point in $\mathscr{S}_1$ is $\delta$-connected to any point in $\mathscr{S}_2$.*

**Example 2.** *Consider the same setting discussed in Example 1. For $\delta = 2/8$, two points at times 5:05 and 5:25 $\in [\![T]\!]$ are $\delta$-connected. The sequence of conditional sets connecting the two points is $\{[\![T|\{a\}]\!], [\![T|\{b\}]\!]\}$, where the sets are defined in (21) and (22). This is because 5:05 $\in [\![T|\{a\}]\!]$, 5 : 25 $\in [\![T|\{b\}]\!]$ and*

$$\frac{m_{\mathscr{T}}([\![T|\{a\}]\!] \cap [\![T|\{b\}]\!])}{m_{\mathscr{T}}([\![T]\!])} = \frac{3}{8} > \delta. \tag{36}$$

*For all $\delta \geq 0$, two points at times 5:00 and 5:05 $\in [\![T]\!]$ are singly $\delta$-connected since 5:00, 5:05 $\in [\![T|\{a\}]\!]$.*

*Likewise, for all $\delta \geq 0$, the set $\mathscr{S} = \{[\![T|\{a\}]\!]\}$ is singly $\delta$-connected by definition.*

*For $\delta = 2/8$, the set $\mathscr{S} = \{[\![T|\{a\}]\!], [\![T|\{b\}]\!]\}$ is $\delta$-connected. This is because for all $x_1, x_2 \in \mathscr{S}$, one of the following scenarios holds:*

- $x_1, x_2 \in [\![T|\{a\}]\!]$.
- $x_1, x_2 \in [\![T|\{b\}]\!]$.
- $x_1 \in [\![T|\{a\}]\!]$ *and* $x_2 \in [\![T|\{b\}]\!]$, *or vice-versa.*

*In the first two scenarios, the points $x_1$ and $x_2$ are singly $\delta$-connected. In the third scenario, the points are $\delta$-connected since*

$$\frac{m_{\mathscr{T}}([\![T|\{a\}]\!] \cap [\![T|\{b\}]\!])}{m_{\mathscr{T}}([\![T]\!])} = \frac{3}{8} > \delta. \tag{37}$$

*For all $\delta > 0$, the two sets $[\![T|\{a\}]\!]$ and $[\![T|\{c\}]\!]$ are $\delta$-isolated since there is no overlap between the two sets.*

**Definition 6.** *$\delta$-overlap family.*
*For any $\delta \in [0, 1]$, a $[\![X|Y]\!]$ $\delta$-overlap family of $[\![X]\!]$, denoted by $[\![X|Y]\!]^*_\delta$, is the largest family of distinct sets covering $[\![X]\!]$, such that*

1. *Each set in the family is $\delta$-connected and contains at least one singly $\delta$-connected set of the form $[\![X|y]\!]$.*
2. *The measure of overlap between any two distinct sets in the family is at most $\delta m_{\mathscr{X}}([\![X]\!])$, namely for all $\mathscr{S}_1, \mathscr{S}_2 \in [\![X|Y]\!]^*_\delta$, such that $\mathscr{S}_1 \neq \mathscr{S}_2$; we also have $m_{\mathscr{X}}(\mathscr{S}_1 \cap \mathscr{S}_2) \leq \delta m_{\mathscr{X}}([\![X]\!])$.*
3. *For every singly $\delta$-connected set, there exist a set in the family containing it.*

The first property of the $\delta$-overlap family ensures that points in the same set of the family *cannot* be distinguished with confidence of at least $(1 - \delta)$, while also ensuring that each set cannot be arbitrarily small. The second and third properties ensure that points that are not covered by the same set of the family *can* be distinguished with confidence of at least $(1 - \delta)$. It follows that the cardinality of the covering family represents the most refined knowledge at a given level of confidence $(1 - \delta)$ that we can have about $X$, given the knowledge of $Y$. This also corresponds to the most refined quantization of the set $[\![X]\!]$ induced by $Y$. This interpretation is analogous to the one in [2], extending the concept of overlap partition introduced there to a $\delta$-overlap family in this work. The stage is now set to introduce the $\delta$-mutual information in terms of the $\delta$-overlap family.

**Definition 7.** *The $\delta$-mutual information provided by $Y$ about $X$ is*

$$I_\delta(X;Y) = \log_2 |[\![X|Y]\!]^*_\delta| \text{ bits}, \tag{38}$$

*if a $[\![X|Y]\!]$ $\delta$-overlap family of $[\![X]\!]$ exists; otherwise, it is zero.*

We now show that when variables are associated at level $(\delta, \delta_2)$, there exists a $\delta$-overlap family, so that the mutual information is well defined.

**Theorem 1.** *If $(X, Y) \overset{a}{\leftrightarrow} (\delta, \delta_2)$, then there exists a $\delta$-overlap family $[\![X|Y]\!]^*_\delta$.*

**Proof.** We show that

$$[\![X|Y]\!] = \{[\![X|y]\!] : y \in [\![Y]\!]\} \tag{39}$$

satisfies all the three properties of $\delta$-overlap family in Definition 6. First, note that $[\![X|Y]\!]$ is a cover of $[\![X]\!]$, since $[\![X]\!] = \cup_{y \in [\![Y]\!]} [\![X|y]\!]$, even though $[\![X|y]\!]$ for different $y$ may overlap with each other. Second, each set in the family $[\![X|Y]\!]$ is singly $\delta$-connected via $[\![X|Y]\!]$, since

trivially, any two points $x_1, x_2 \in [\![X|y]\!]$ are singly $\delta$-connected via the same set. It follows that Property 1 of Definition 6 holds.

Now, since $(X, Y) \overset{a}{\leftrightarrow} (\delta, \delta_2)$, then by Lemma 1, for all $y_1, y_2 \in [\![Y]\!]$, we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!])}{m_{\mathscr{X}}([\![X]\!])} \leq \delta, \tag{40}$$

which shows that Property 2 of Definition 6 holds. Finally, it is also easy to see that Property 3 of Definition 6 holds, since $[\![X|Y]\!]$ contains all sets $[\![X|y]\!]$. Hence, $[\![X|Y]\!]$ satisfies all the three properties in Definition 6, which implies that there exists at least one set satisfying these conditions. Hence, the maximum over these sets is defined and the claim follows. □

Next, we show that a $\delta$-overlap family also exists when variables are disassociated at level $(\delta, \delta_2)$. In this case, we also characterize the mutual information in terms of a partition of $[\![X]\!]$.

**Definition 8.** *$\delta$-isolated partition.*
*A $[\![X|Y]\!]$ $\delta$-isolated partition of $[\![X]\!]$, denoted by $[\![X|Y]\!]_\delta$, is a partition of $[\![X]\!]$ such that any two sets in the partition are $\delta$-isolated via $[\![X|Y]\!]$.*

**Theorem 2.** *If $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, then the following holds:*

1. *There exists a unique $\delta$-overlap family $[\![X|Y]\!]^*_\delta$.*
2. *The $\delta$-overlap family is the $\delta$-isolated partition of largest cardinality, in that, for any $[\![X|Y]\!]_\delta$, we have*
$$|[\![X|Y]\!]_\delta| \leq |[\![X|Y]\!]^*_\delta|, \tag{41}$$
*where the equality holds if and only if $[\![X|Y]\!]_\delta = [\![X|Y]\!]^*_\delta$.*

**Proof.** First, we show the existence of a $\delta$-overlap family. For all $x \in [\![X]\!]$, let $\mathscr{C}(x)$ be the set of points that are $\delta$-connected to $x$ via $[\![X|Y]\!]$, namely

$$\mathscr{C}(x) = \{x_1 \in [\![X]\!] : x \overset{\delta}{\rightsquigarrow} x_1\}. \tag{42}$$

Then, we let

$$\mathcal{C} = \{\mathscr{C}(x) : x \in [\![X]\!]\}, \tag{43}$$

and show that this is a $\delta$-overlap family. First, note that since $[\![X]\!] = \cup_{\mathscr{S} \in \mathcal{C}} \mathscr{S}$, we know that $\mathcal{C}$ is a cover of $[\![X]\!]$. Second, for all $\mathscr{C}(x) \in \mathcal{C}$, there exists a $y \in [\![Y]\!]$ such that $x \in [\![X|y]\!]$, and since any two points $x_1, x_2 \in [\![X|y]\!]$ are singly $\delta$-connected via $[\![X|Y]\!]$, we understand that $[\![X|y]\!] \subseteq \mathscr{C}(x)$. It follows that every set in the family $\mathcal{C}$ contains at least one singly $\delta$-connected set. For all $x_1, x_2 \in \mathscr{C}(x)$, we also have $x_1 \overset{\delta}{\rightsquigarrow} x$ and $x \overset{\delta}{\rightsquigarrow} x_2$. Since $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, by Lemma A2 in Appendix C, this implies that $x_1 \overset{\delta}{\rightsquigarrow} x_2$. It follows that every set in the family $\mathcal{C}$ is $\delta$-connected and contains at least one singly $\delta$-connected set, and we conclude that Property 1 of Definition 6 is satisfied.

We now claim that for all $x_1, x_2 \in [\![X]\!]$, if

$$\mathscr{C}(x_1) \neq \mathscr{C}(x_2), \tag{44}$$

then

$$m_{\mathscr{X}}(\mathscr{C}(x_1) \cap \mathscr{C}(x_2)) = 0. \tag{45}$$

This can be proven by contradiction. Let $\mathscr{C}(x_1) \neq \mathscr{C}(x_2)$ and assume that $m_{\mathscr{X}}(\mathscr{C}(x_1) \cap \mathscr{C}(x_2)) \neq 0$. By (9), this implies that $\mathscr{C}(x_1) \cap \mathscr{C}(x_2) \neq \emptyset$. We can then select

$z \in \mathscr{C}(x_1) \cap \mathscr{C}(x_2)$, such that we have $z \overset{\delta}{\rightsquigarrow} x_1$ and $z \overset{\delta}{\rightsquigarrow} x_2$. Since $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, by Lemma A2 in Appendix C, this also implies that $x_1 \overset{\delta}{\rightsquigarrow} x_2$ , and, therefore, $\mathscr{C}(x_1) = \mathscr{C}(x_2)$, which is a contradiction. It follows that if $\mathscr{C}(x_1) \neq \mathscr{C}(x_2)$, then we must have $m_{\mathscr{X}}(\mathscr{C}(x_1) \cap \mathscr{C}(x_2)) = 0$, and, therefore,

$$\frac{m_{\mathscr{X}}(\mathscr{C}(x_1) \cap \mathscr{C}(x_2))}{m_{\mathscr{X}}([\![X]\!])} = 0 \leq \delta. \tag{46}$$

We conclude that Property 2 of Definition 6 is satisfied.

Finally, we understand that for any singly $\delta$-connected set $[\![X|y]\!]$, there exist an $x \in [\![X]\!]$ such that $x \in [\![X|y]\!]$, which by, (42), implies that $[\![X|y]\!] \subseteq \mathscr{C}(x)$. Namely, for every singly $\delta$-connected set, there exist a set in the family containing it. We can then conclude that $\mathcal{C}$ satisfies all the properties of a $\delta$-overlap family.

Next, we show that $\mathcal{C}$ is a unique $\delta$-overlap family, which implies that this is also the largest set satisfying the three conditions in Definition 6. By contradiction, consider another $\delta$-overlap family $\mathcal{D}$. For all $x \in [\![X]\!]$, let $\mathscr{D}(x)$ denote a set in $\mathcal{D}$ containing $x$. Then, using the definition of $\mathscr{C}(x)$ and the fact that $\mathscr{D}(x)$ is $\delta$-connected, it follows that

$$\mathscr{D}(x) \subseteq \mathscr{C}(x). \tag{47}$$

Next, we show that for all $x \in [\![X]\!]$, we also have

$$\mathscr{C}(x) \subseteq \mathscr{D}(x), \tag{48}$$

from which, we conclude that $\mathcal{D} = \mathcal{C}$.

The proof of (48) is also obtained by contradiction. Assume there exists a point $\tilde{x} \in \mathscr{C}(x) \setminus \mathscr{D}(x)$. Since both $x$ and $\tilde{x}$ are contained in $\mathscr{C}(x)$, $\tilde{x} \overset{\delta}{\rightsquigarrow} x$. Let $x^*$ be a point in a singly connected set that is contained in $\mathscr{D}(x)$, namely $x^* \in [\![X|y^*]\!] \subseteq \mathscr{D}(x)$. Since both $x$ and $x^*$ are in $\mathscr{D}(x)$, we understand that $x \overset{\delta}{\rightsquigarrow} x^*$. Since $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, we can apply Lemma A2 in Appendix C to conclude that $\tilde{x} \overset{\delta}{\rightsquigarrow} x^*$. It follows that there exists a sequence of conditional ranges $\{[\![X|y_i]\!]\}_{i=1}^N$ such that $\tilde{x} \in [\![X|y_1]\!]$ and $x^* \in [\![X|y_N]\!]$, which satisfies (35). Since $x^*$ is in both $[\![X|y_N]\!]$ and $[\![X|y^*]\!]$, we obtain $[\![X|y_N]\!] \cap [\![X|y^*]\!] \neq \emptyset$, and since $(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2)$, we obtain

$$\frac{m_{\mathscr{X}}([\![X|y_N]\!] \cap [\![X|y^*]\!])}{m_{\mathscr{X}}([\![X]\!])} > \delta. \tag{49}$$

Without loss of generality, we can then assume that the last element of our sequence is $[\![X|y^*]\!]$. By Property 3 of Definition 6, every conditional range in the sequence must be contained in some set of the $\delta$-overlap family $\mathcal{D}$. Since $[\![X|y^*]\!] \subseteq \mathscr{D}(x)$ and $[\![X|y_1]\!] \not\subseteq \mathscr{D}(x)$, it follows that there exist two consecutive conditional ranges along the sequence and two sets of the $\delta$-overlap family covering them, such that $[\![X|y_{i-1}]\!] \subseteq \mathscr{D}(x_{i-1})$, $[\![X|y_i]\!] \subseteq \mathscr{D}(x_i)$, and $\mathscr{D}(x_{i-1}) \neq \mathscr{D}(x_i)$. Then, we have

$$\begin{aligned} & m_{\mathscr{X}}(\mathscr{D}(x_{i-1}) \cap \mathscr{D}(x_i)) \\ &= m_{\mathscr{X}}(([\![X|y_{i-1}]\!] \cap [\![X|y_i]\!]) \cup (\mathscr{D}(x_{i-1}^*) \cap \mathscr{D}(x_i^*))) \\ &\overset{(a)}{\geq} m_{\mathscr{X}}([\![X|y_{i-1}]\!] \cap [\![X|y_i]\!]) \\ &\overset{(b)}{>} \delta m_{\mathscr{X}}([\![X]\!]), \end{aligned} \tag{50}$$

where $(a)$ follows from (10) and $(b)$ follows from (35). It follows that

$$\frac{m_{\mathscr{X}}(\mathscr{D}(x_{i-1}) \cap \mathscr{D}(x_i))}{m_{\mathscr{X}}([\![X]\!])} > \delta, \tag{51}$$

and Property 2 of Definition 6 is violated. Thus, $\tilde{x}$ does not exists, which implies $\mathscr{C}(x) \subseteq \mathscr{D}(x)$. Combining (47) and (48), we conclude that the $\delta$-overlap family $\mathcal{C}$ is unique.

We now turn to the proof of the second part of the theorem. Since by (46), the uncertainty associated with the overlap between any two sets of the $\delta$-overlap family $\mathcal{C}$ is zero, it follows that $\mathcal{C}$ is also a partition.

Now, we show that $\mathcal{C}$ is also a $\delta$-isolated partition. This can be proven by contradiction. Assume that $\mathcal{C}$ is not a $\delta$-isolated partition. Then, there exists two distinct sets $\mathscr{C}(x_1), \mathscr{C}(x_2) \in \mathcal{C}$ such that $\mathscr{C}(x_1)$ and $\mathscr{C}(x_2)$ are not $\delta$-isolated. This implies that there exists a point $\bar{x}_1 \in \mathscr{C}(x_1)$ and $\bar{x}_2 \in \mathscr{C}(x_2)$ such that $\bar{x}_1 \overset{\delta}{\leftrightsquigarrow} \bar{x}_2$. Using the fact that $\mathscr{C}(x_1)$ and $\mathscr{C}(x_2)$ are $\delta$-connected and Lemma A2 in Appendix C, this implies that all points in the set $\mathscr{C}(x_1)$ are $\delta$-connected to all points in the set $\mathscr{C}(x_2)$. Now, let $x_1^*$ and $x_2^*$ be points in a singly $\delta$-connected set contained in $\mathscr{C}(x_1)$ and $\mathscr{C}(x_2)$, respectively: $x_1^* \in [\![X|y_1^*]\!] \subseteq \mathscr{C}(x_1)$ and $x_2^* \in [\![X|y_2^*]\!] \subseteq \mathscr{C}(x_2)$. Since $x_1^* \overset{\delta}{\leftrightsquigarrow} x_2^*$, there exists a sequence of conditional ranges $\{[\![X|y_i]\!]\}_{i=1}^N$ satisfying (35), such that $x_1 \in [\![X|y_1]\!]$ and $x_2 \in [\![X|y_N]\!]$. Without loss of generality, we can assume $[\![X|y_1]\!] = [\![X|y_1^*]\!]$ and $[\![X|y_2]\!] = [\![X|y_2^*]\!]$. Since $\mathcal{C}$ is a partition, we understand that $[\![X|y_1^*]\!] \subseteq \mathscr{C}(x_1)$ and $[\![X|y_2^*]\!] \nsubseteq \mathscr{C}(x_1)$. It follows that there exist two consecutive conditional ranges along the sequence $\{[\![X|y_i]\!]\}_{i=1}^N$ and two sets of the $\delta$-overlap family $\mathcal{C}$ covering them, such that $[\![X|y_{i-1}]\!] \subseteq \mathscr{C}(x_{i-1})$ and $[\![X|y_i]\!] \subseteq \mathscr{C}(x_i)$ and that $\mathscr{C}(x_{i-1}) \neq \mathscr{C}(x_i)$. Similarly to (50), we hold that

$$\frac{m_{\mathscr{X}}(\mathscr{C}(x_{i-1}) \cap \mathscr{C}(x_i))}{m_{\mathscr{X}}([\![X]\!])} > \delta, \tag{52}$$

and Property 2 of Definition 6 is violated. Thus, $\mathscr{C}(x_1)$ and $\mathscr{C}(x_2)$ do not exist, which implies that $\mathcal{C}$ is a $\delta$-isolated partition.

Let $\mathcal{P}$ be any other $\delta$-isolated partition. We wish to show that $|\mathcal{C}| \geq |\mathcal{P}|$ and that the equality holds if and only if $\mathcal{P} = \mathcal{C}$. First, note that every set $\mathscr{C}(x) \in \mathcal{C}$ can intersect, at most, one set in $\mathcal{P}$; otherwise, the sets in $\mathcal{P}$ would not be $\delta$-isolated. Second, since $\mathcal{C}$ is a cover of $[\![X]\!]$, every set in $\mathcal{P}$ must be intersected by at least one set in $\mathcal{C}$. It follows that

$$|\mathcal{C}| \geq |\mathcal{P}|. \tag{53}$$

Now, assume the equality holds. In this case, there is a one-to-one correspondence $\mathscr{P} : \mathcal{C} \to \mathcal{P}$, such that for all $x \in [\![X]\!]$, we have $\mathscr{C}(x) \subseteq \mathscr{P}(\mathscr{C}(x))$, and since both $\mathcal{C}$ and $\mathcal{P}$ are partitions of $[\![X]\!]$, it follows that $\mathcal{C} = \mathcal{P}$. Conversely, assuming that $\mathcal{C} = \mathcal{P}$, then $|\mathcal{C}| = |\mathcal{P}|$ follows trivially. □

We have introduced the notion of mutual information from $Y$ to $X$ in terms of the conditional range $[\![X|Y]\!]$. Since, in general, we have $[\![X|Y]\!] \neq [\![Y|X]\!]$, one may expect the definition of mutual information to be asymmetric in its arguments. Namely, the amount of information provided about $X$ by the knowledge of $Y$ may not be the same as the amount of information provided about $Y$ by the knowledge of $X$. Although this is true in general, we show that for disassociated UVs, symmetry is retained, provided that when swapping $X$ with $Y$, one also rescales $\delta$ appropriately. The following theorem establishes the symmetry in the mutual information under the appropriate scaling of the parameters $\delta_1$ and $\delta_2$. The proof requires the introduction of the notions of taxicab connectedness, taxicab family, and taxicab partition, which are given in Appendix C.1, along with the proof of the theorem.

**Theorem 3.** *If* $(X, Y) \overset{d}{\leftrightarrow} (\delta_1, \delta_2)$ *and a* $(\delta_1, \delta_2)$*-taxicab family of* $[\![X, Y]\!]$ *exists, then we have*

$$I_{\delta_1}(X;Y) = I_{\delta_2}(Y;X). \tag{54}$$

## 4. ($\epsilon$,$\delta$)-Capacity

We now give a definition of the capacity of a communication channel and relate it to the notion of mutual information between the UVs introduced above. We consider a normed space $\mathscr{X}$ to be totally bounded if, for every $\epsilon > 0$, $\mathscr{X}$ can be covered by a finite number of open balls of radius $\epsilon$. We let $\mathscr{X}$ be a totally bounded, normed space such that for all $x \in \mathscr{X}$, we have $\|x\| \leq 1$, where $\|.\|$ represents the norm. This normalization is for convenience in the notation process, and all results can easily be extended to metric spaces of any bounded norm. Let $\mathcal{X} \subseteq \mathscr{X}$ be a discrete set of points in the space, which represents a codebook.

**Definition 9.** *$\epsilon$-perturbation channel.*
*A channel is called $\epsilon$-perturbation if for any transmitted codeword $x \in \mathcal{X}$, $x$ is received with noise perturbation at most $\epsilon$. Namely, we receive a point in the set*

$$S_\epsilon(x) = \{y \in \mathscr{X} : \|x - y\| \leq \epsilon\}. \tag{55}$$

Given the codebook $\mathcal{X}$ is transmitted over an $\epsilon$-perturbation channel, all received codewords lie in the set $\mathcal{Y} = \bigcup_{x \in \mathcal{X}} S_\epsilon(x)$, where $\mathcal{Y} \subseteq \mathscr{Y} = \mathscr{X}$. Transmitted codewords can be decoded correctly as long as the corresponding uncertainty sets at the receiver do not overlap. This can be achieved by simply associating the received codeword to the point in the codebook that is closest to it.

For any $x_1, x_2 \in \mathcal{X}$, we now let

$$e_\epsilon(x_1, x_2) = \frac{m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2))}{m_{\mathscr{Y}}(\mathscr{Y})}, \tag{56}$$

where $m_{\mathscr{Y}}(.)$ is an uncertainty function defined over the space $\mathscr{Y}$. We also assume without loss of generality that the uncertainty associated with the whole space $\mathscr{Y}$ of received codewords is $m_{\mathscr{Y}}(\mathscr{Y}) = 1$. Finally, we let $V_\epsilon \subseteq \mathscr{Y}$ be the smallest uncertainty set corresponding to a transmitted codeword, namely $V_\epsilon = S_\epsilon(x^*)$, where $x^* = \operatorname{argmin}_{x \in \mathscr{X}} m_{\mathscr{Y}}(S_\epsilon(x))$. The quantity $1 - e_\epsilon(x_1, x_2)$ can be viewed as the confidence we have in not confusing $x_1$ and $x_2$ in any transmission or, equivalently, as the amount of adversarial effort required to induce a confusion between the two codewords. For example, if the uncertainty function is constructed using a measure, then all the erroneous codewords generated by an adversary to decode $x_2$ instead of $x_1$ must lie inside the equivocation set depicted in Figure 3, whose relative size is given by (56). The smaller the equivocation set is, the larger the effort required by the adversary to induce an error must be. If the uncertainty function represents the diameter of the set, then all the erroneous codewords generated by an adversary to decode $x_2$ instead of $x_1$ will be close to each other in the sense of (56). Once again, the closer the possible erroneous codewords are, the harder it must be for the adversary to generate an error, since any small deviation allows the decoder to correctly identify the transmitted codeword.

We now introduce the notion of a *distinguishable codebook*, ensuring that every codeword cannot be confused with any other codeword, rather than with a specific one, at a given level of confidence.

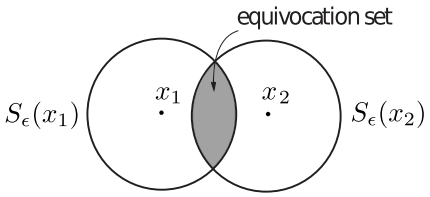



**Figure 3.** The size of the equivocation set is inversely proportional to the amount of adversarial effort required to induce an error.

**Definition 10.** $(\epsilon, \delta)$*-distinguishable codebook.*
*For any* $0 < \epsilon \leq 1$, $0 \leq \delta < m_{\mathscr{Y}}(V_\epsilon)$, *a codebook* $\mathcal{X} \subseteq \mathscr{X}$ *is* $(\epsilon, \delta)$*-distinguishable if for all* $x_1, x_2 \in \mathcal{X}$, *we have* $e_\epsilon(x_1, x_2) \leq \delta/|\mathcal{X}|$.

For any $(\epsilon, \delta)$-distinguishable codebook $\mathcal{X}$ and $x \in \mathcal{X}$, we let

$$e_\epsilon(x) = \sum_{x' \in \mathcal{X}: x' \neq x} e_\epsilon(x, x'). \tag{57}$$

It now follows from Definition 10 that

$$e_\epsilon(x) \leq \delta, \tag{58}$$

and each codeword in an $(\epsilon, \delta)$-distinguishable codebook can be decoded correctly with confidence of at least $1 - \delta$. Definition 10 guarantees even more, namely that the confidence of not confusing any pair of codewords is uniformly bounded by $1 - \delta/|\mathcal{X}|$. This stronger constraint implies that we cannot "balance" the error associated with a codeword transmission by allowing some decoding pair to have a lower confidence and enforcing other pairs to have higher confidence. This is the main difference between our definition and the one used in [7], which bounds the average confidence and allows us to relate the notion of capacity to the mutual information between pairs of codewords.

**Definition 11.** $(\epsilon, \delta)$*-capacity.*
*For any totally bounded, normed metric space* $\mathscr{X}$, $0 < \epsilon \leq 1$, $0 \leq \delta < m_{\mathscr{Y}}(V_\epsilon)$, *and the* $(\epsilon, \delta)$*-capacity of* $\mathscr{X}$ *is*

$$C_\epsilon^\delta = \sup_{\mathcal{X} \in \mathscr{X}_\epsilon^\delta} \log_2 |\mathcal{X}| \text{ bits}, \tag{59}$$

*where* $\mathscr{X}_\epsilon^\delta = \{\mathcal{X} : \mathcal{X}$ *is* $(\epsilon, \delta)$*-distinguishable*$\}$ *is the set of* $(\epsilon, \delta)$*-distinguishable codebooks.*

The $(\epsilon, \delta)$-capacity represents the largest number of bits that can be communicated by using any $(\epsilon, \delta)$-distinguishable codebook. The corresponding geometric picture is illustrated in Figure 4. For $\delta = 0$, our notion of capacity reduces to Kolmogorov's $\epsilon$-capacity, which is the logarithm of the packing number of the space with balls of radius $\epsilon$.

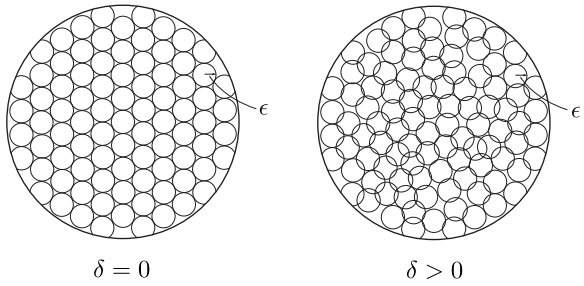



**Figure 4.** Illustration of the $(\epsilon, \delta)$-capacity in terms of packing $\epsilon$-balls with maximum overlap $\delta$.

In the definition of capacity, we have restricted $\delta < m_{\mathscr{Y}}(V_\epsilon)$ to rule out cases when the decoding error can be at least as large as the error introduced by the channel and when the $(\epsilon, \delta)$-capacity is infinite. Also, note that $m_{\mathscr{Y}}(V_\epsilon) \leq 1$ since $V_\epsilon \subseteq \mathscr{Y}$ and (10) holds.

We now relate our operational definition of capacity to the notion of UVs and mutual information introduced in Section 3. Let $X$ be the UV corresponding to the transmitted codeword. This is a map $X : \mathscr{X} \to \mathcal{X}$ and $[\![X]\!] = \mathcal{X} \subseteq \mathscr{X}$. Likewise, let $Y$ be the UV corresponding to the received codeword. This is a map $Y : \mathscr{Y} \to \mathcal{Y}$ and $[\![Y]\!] = \mathcal{Y} \subseteq \mathscr{Y}$. For our $\epsilon$-perturbation channel, these UVs are such that for all $y \in [\![Y]\!]$ and $x \in [\![X]\!]$, we have

$$[\![Y|x]\!] = \{y \in [\![Y]\!] : \|x - y\| \leq \epsilon\}, \tag{60}$$

$$[\![X|y]\!] = \{x \in [\![X]\!] : \|x - y\| \leq \epsilon\}, \tag{61}$$

(see Figure 5). Clearly, the set in (60) is continuous, while the set in (61) is discrete.

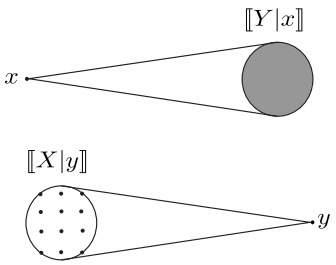



**Figure 5.** Conditional ranges $[\![Y|x]\!]$ and $[\![X|y]\!]$ due to the $\epsilon$-perturbation channel.

To measure the levels of association and disassociation between $X$ and $Y$, we use an uncertainty function $m_{\mathscr{X}}(.)$ defined over $\mathscr{X}$ and $m_{\mathscr{Y}}(.)$ defined over $\mathscr{Y}$. We introduce the feasible set

$$\begin{aligned} \mathscr{F}_\delta = \{X : [\![X]\!] \subseteq \mathscr{X}, \text{ and either} \\ (X, Y) \overset{d}{\leftrightarrow} (0, \delta/|[\![X]\!]|) \\ \text{or } (X, Y) \overset{a}{\leftrightarrow} (1, \delta/|[\![X]\!]|)\}, \end{aligned} \tag{62}$$

representing the set of UVs $X$ such that the marginal range $[\![X]\!]$ is a discrete set representing a codebook, and the UV can either achieve $(0, \delta/|[\![X]\!]|)$ levels of disassociation or $(1, \delta/|[\![X]\!]|)$ levels of association with $Y$. In our channel model, this feasible set also depends on the $\epsilon$-perturbation through (60) and (61).

We can now state the non-stochastic channel coding theorem for our $\epsilon$-perturbation channel.

**Theorem 4.** *For any totally bounded, normed metric space $\mathscr{X}$, $\epsilon$-perturbation channel satisfying (60) and (61), $0 < \epsilon \leq 1$, and $0 \leq \delta < m_{\mathscr{Y}}(V_\epsilon)$, we have*

$$C_\epsilon^\delta = \sup_{X \in \mathscr{F}_{\tilde{\delta}}, \tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])} I_{\tilde{\delta}/|[\![X]\!]|}(Y; X) \text{ bits}. \tag{63}$$

**Proof.** First, we show that there exists a UV $X$ and $\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])$ such that $X \in \mathscr{F}_{\tilde{\delta}}$, which implies that the supremum is well defined. Second, for all $X$ and $\tilde{\delta}$ such that

$$X \in \mathscr{F}_{\tilde{\delta}}, \tag{64}$$

and

$$\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!]), \tag{65}$$

we show that

$$I_{\tilde{\delta}/|[\![X]\!]|}(Y;X) \leq C_{\epsilon}^{\delta}. \tag{66}$$

Finally, we show the existence of $X \in \mathscr{F}_{\tilde{\delta}}$ and $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$ such that $I_{\tilde{\delta}/|[\![X]\!]|}(Y;X) = C_{\epsilon}^{\delta}$.

Let us begin with the first step. Consider a point $x \in \mathscr{X}$. Let $X$ be a UV such that

$$[\![X]\!] = \{x\}. \tag{67}$$

Then, we hold that the marginal range of the UV $Y$ corresponding to the received variable is

$$[\![Y]\!] = [\![Y|x]\!], \tag{68}$$

and, therefore, for all $y \in [\![Y]\!]$, we have

$$[\![X|y]\!] = \{x\}. \tag{69}$$

Using Definition 2 and (67), we hold that

$$\mathscr{A}(Y;X) = \emptyset, \tag{70}$$

because $[\![X]\!]$ consists of a single point, and, therefore, the set in (12) is empty.

On the other hand, using Definition 2 and (69), we have

$$\mathscr{A}(X;Y) = \begin{cases} \{1\} \text{ if } \exists y_1, y_2 \in [\![Y]\!], \\ \emptyset \quad \text{otherwise.} \end{cases} \tag{71}$$

Using (70), and since $\mathscr{A} \preceq \delta$ holds for $\mathscr{A} = \emptyset$, we have

$$\mathscr{A}(Y;X) \preceq \delta/(|[\![X]\!]|m_{\mathscr{Y}}([\![Y]\!])). \tag{72}$$

Similarly, using (71), we have

$$\mathscr{A}(X;Y) \preceq 1. \tag{73}$$

Now, combining (72) and (73), we have

$$(X,Y) \overset{a}{\leftrightarrow} (1, \delta/(|[\![X]\!]|m_{\mathscr{Y}}([\![Y]\!]))). \tag{74}$$

Letting $\tilde{\delta} = \delta/m_{\mathscr{Y}}([\![Y]\!])$, this implies that $X \in \mathscr{F}_{\tilde{\delta}}$ and the first step of the proof is complete.

To prove the second step, we define the set of discrete UVs

$$\begin{aligned} \mathscr{G} = \{X : [\![X]\!] \subseteq \mathscr{X}, \exists \tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!]) \\ \text{such that } \forall \mathscr{S}_1, \mathscr{S}_2 \in [\![Y|X]\!], \\ m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_2)/m_{\mathscr{Y}}([\![Y]\!]) \leq \tilde{\delta}/|[\![X]\!]|\}, \end{aligned} \tag{75}$$

which is a larger set than the one containing all UVs $X$ that are $(1, \tilde{\delta}/|[\![X]\!]|)$ associated with $Y$. Now, we will show that if a UV $X \in \mathscr{G}$, then the corresponding codebook $\mathcal{X} \in \mathscr{X}_{\epsilon}^{\delta}$. If $X \in \mathscr{G}$, then there exists a $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$ such that for all $\mathscr{S}_1, \mathscr{S}_2 \in [\![Y|X]\!]$, we have

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{Y}}([\![Y]\!])} \leq \frac{\tilde{\delta}}{|[\![X]\!]|}. \tag{76}$$

It follows that for all $x_1, x_2 \in [\![X]\!]$, we have

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!])}{m_{\mathscr{Y}}([\![Y]\!])} \leq \frac{\tilde{\delta}}{|[\![X]\!]|}. \tag{77}$$

Using $\mathcal{X} = [\![X]\!]$, (60), $[\![Y]\!] = \mathcal{Y} = \bigcup_{x\in\mathcal{X}} S_\epsilon(x)$ and $m_{\mathscr{Y}}(\mathscr{Y}) = 1$, for all $x_1, x_2 \in \mathcal{X}$, we have

$$\begin{aligned} \frac{m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2))}{m_{\mathscr{Y}}(\mathscr{Y})} &\leq \frac{\tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])}{|\mathcal{X}|}, \\ &\overset{(a)}{\leq} \frac{\delta}{|\mathcal{X}|}, \end{aligned} \tag{78}$$

where $(a)$ follows from $\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])$. Putting things together, it follows that

$$X \in \mathscr{G} \implies \mathcal{X} \in \mathscr{X}_\epsilon^\delta \tag{79}$$

Consider now a pair of $X$ and $\tilde{\delta}$ such that $\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])$ and

$$X \in \mathscr{F}_{\tilde{\delta}}. \tag{80}$$

If $(X, Y) \overset{d}{\leftrightarrow} (0, \tilde{\delta}/|[\![X]\!]|)$, then, using Lemma A1 in Appendix C, there exist two UVs, $\bar{X}$ and $\bar{Y}$ and $\bar{\delta} \leq \delta / m_{\mathscr{Y}}([\![\bar{Y}]\!])$, such that

$$(\bar{X}, \bar{Y}) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}]\!]|), \tag{81}$$

and

$$|[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}| = |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{82}$$

On the other hand, if $(X, Y) \overset{a}{\leftrightarrow} (1, \tilde{\delta}/|[\![X]\!]|)$, then (81) and (82) also trivially hold. It then follows that (81) and (82) hold for all $X \in \mathscr{F}_{\tilde{\delta}}$. We now have

$$\begin{aligned} I_{\tilde{\delta}/|[\![X]\!]|}(Y; X) &= \log(|[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}|) \\ &\overset{(a)}{=} \log(|[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|) \\ &\overset{(b)}{\leq} \log(|[\![\bar{X}]\!]|), \\ &\overset{(c)}{=} \log(|\bar{\mathcal{X}}|), \\ &\overset{(d)}{\leq} C_\epsilon^\delta, \end{aligned} \tag{83}$$

where $(a)$ follows from (81) and (82), $(b)$ follows from Lemma A3 in Appendix C since $\bar{\delta} \leq \delta / m_{\mathscr{Y}}([\![\bar{Y}]\!]) < m_{\mathscr{Y}}(V_\epsilon) / m_{\mathscr{Y}}([\![\bar{Y}]\!])$, $(c)$ follows by defining the codebook $\bar{\mathcal{X}}$ corresponding to the UV $\bar{X}$, and $(d)$ follows from the fact that using (81) and Lemma 1 allows $\bar{X} \in \mathcal{G}$, which implies for (79) that $\bar{\mathcal{X}} \in \mathscr{X}_\epsilon^\delta$.

Finally, let

$$\mathcal{X}^* = \text{argsup}_{\mathcal{X}\in\mathscr{X}_\epsilon^\delta} \log(|\mathcal{X}|), \tag{84}$$

which achieves the capacity $C_\epsilon^\delta$. Let $X^*$ be the UV whose marginal range corresponds to the codebook $\mathcal{X}^*$. It follows that for all $\mathscr{S}_1, \mathscr{S}_1 \in [\![Y^*|X^*]\!]$, we have

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_1)}{m_{\mathscr{Y}}(\mathscr{Y})} \leq \frac{\delta}{|[\![X^*]\!]|}, \tag{85}$$

which implies that $m_{\mathscr{Y}}(\mathscr{Y}) = 1$,

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_1)}{m_{\mathscr{Y}}([\![Y^*]\!])} \leq \frac{\delta}{|[\![X^*]\!]| m_{\mathscr{Y}}([\![Y^*]\!])}. \tag{86}$$

Letting $\delta^* = \delta / m_{\mathscr{Y}}([\![Y^*]\!])$, and using Lemma 1, we hold that $(X^*, Y^*) \overset{a}{\leftrightarrow} (1, \delta^*/|[\![X^*]\!]|)$, which implies that $X^* \in \cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y^*]\!])} \mathscr{F}_{\tilde{\delta}}$, and the proof is complete. □

Theorem 4 characterizes the capacity as the supremum of the mutual information over all UVs in the feasible set. The following theorem shows that the same characterization is obtained if we optimize the right-hand side in (63) over all UVs in the space. It follows by Theorem 4 that rather than optimizing over all UVs representing all the codebooks in the space, a capacity-achieving codebook can be found within the smaller class $\cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}(V_\epsilon)} \mathscr{F}_{\tilde{\delta}}$ of feasible sets with error at most $\delta / m_{\mathscr{Y}}(V_\epsilon)$, since for all $[\![Y]\!] \subseteq \mathscr{Y}$, $m_{\mathscr{Y}}(V_\epsilon) \leq m_{\mathscr{Y}}([\![Y]\!])$.

**Theorem 5.** *The* $(\epsilon, \delta)$*-capacity in* (63) *can also be written as*

$$C_\epsilon^\delta = \sup_{\substack{X:[\![X]\!] \subseteq \mathscr{X}, \\ \tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])}} I_{\tilde{\delta}/|[\![X]\!]|}(Y;X) \text{ bits}. \tag{87}$$

**Proof.** Consider a UV $X \notin \cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])} \mathscr{F}_{\tilde{\delta}}$, where $Y$ is the corresponding UV at the receiver. The idea of the proof is to show the existence of a UV $\bar{X} \in \cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![\bar{Y}]\!])} \mathscr{F}_{\tilde{\delta}}$ and the corresponding UV $\bar{Y}$ at the receiver, and

$$\bar{\delta} = \tilde{\delta} m_{\mathscr{Y}}([\![Y]\!]) / m_{\mathscr{Y}}([\![\bar{Y}]\!]) \leq \delta / m([\![\bar{Y}]\!]), \tag{88}$$

such that the cardinality of the overlap partitions

$$|[\![\bar{Y}|\bar{X}]\!]^*_{\tilde{\delta}/|[\![\bar{X}]\!]|}| = |[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}|. \tag{89}$$

Let the cardinality

$$|[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}| = K. \tag{90}$$

By Property 1 of Definition 6, we hold that for all $\mathscr{S}_i \in [\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}$, there exists an $x_i \in [\![X]\!]$ such that $[\![Y|x_i]\!] \subseteq \mathscr{S}_i$. Now, consider another UV $\bar{X}$ whose marginal range is composed of $K$ elements of $[\![X]\!]$, namely

$$[\![\bar{X}]\!] = \{x_1, \dots x_K\}. \tag{91}$$

Let $\bar{Y}$ be the UV corresponding to the received variable. Using the fact that for all $x \in \mathscr{X}$, we have $[\![\bar{Y}|x]\!] = [\![Y|x]\!]$ since (60) holds, and using Property 2 of Definition 6, for all $x, x' \in [\![\bar{X}]\!]$, we obtain

$$\begin{aligned} \frac{m_{\mathscr{Y}}([\![\bar{Y}|x]\!] \cap [\![\bar{Y}|x']\!])}{m_{\mathscr{Y}}([\![Y]\!])} &\leq \frac{\tilde{\delta}}{|[\![X]\!]|}, \\ &\overset{(a)}{\leq} \frac{\tilde{\delta}}{|[\![\bar{X}]\!]|}, \end{aligned} \tag{92}$$

where $(a)$ follows from the fact that $[\![\bar{X}]\!] \subseteq [\![X]\!]$ using (91). Then, for all $x, x' \in [\![\bar{X}]\!]$, we hold that

$$\begin{aligned} \frac{m_{\mathscr{Y}}([\![\bar{Y}|x]\!] \cap [\![\bar{Y}|x']\!])}{m_{\mathscr{Y}}([\![\bar{Y}]\!])} &\leq \frac{\tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])}{|[\![\bar{X}]\!]| m_{\mathscr{Y}}([\![\bar{Y}]\!])} \\ &= \frac{\bar{\delta}}{|[\![\bar{X}]\!]|}, \end{aligned} \tag{93}$$

since $\bar{\delta} = \tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])/m_{\mathscr{Y}}([\![\bar{Y}]\!])$. Then, by Lemma 1, it follows that

$$(\bar{X}, \bar{Y}) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}]\!]|). \tag{94}$$

Since $\tilde{\delta} \leq \delta/m_{\mathscr{Y}}([\![Y]\!])$, we have

$$\bar{\delta} \leq \delta/m_{\mathscr{Y}}([\![\bar{Y}]\!]) < m_{\mathscr{Y}}(V_\epsilon)/m_{\mathscr{Y}}([\![\bar{Y}]\!]). \tag{95}$$

Therefore, $\bar{X} \in \mathscr{F}_{\bar{\delta}}$ and $\bar{\delta} \leq \delta/m_{\mathscr{Y}}([\![\bar{Y}]\!])$. We now hold that

$$\begin{aligned} |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}| &\overset{(a)}{=} |[\![\bar{X}]\!]| \\ &\overset{(b)}{=} |[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}|, \end{aligned} \tag{96}$$

where $(a)$ follows by applying Lemma A4 in Appendix C using (94) and (95) and $(b)$ follows from (90) and (91). Combining (96) with Theorem 4, the proof is complete. □

We now make some considerations with respect to previous results in the literature. First, we note that for $\delta = 0$, all of our definitions reduce to Nair's ones and Theorem 4 recovers Nair's coding theorem ([2] (Theorem 4.1)) for the zero-error capacity of an additive $\epsilon$-perturbation channel.

Second, we point out that the $(\epsilon, \delta)$-capacity considered in [7] defines the set of $(\epsilon, \delta)$-distinguishable codewords such that the *average* overlap among all codewords is at most $\delta$. In contrast, our definition requires the overlap for *each* pair of codewords to be at most $\delta/|\mathcal{X}|$. The following theorem provides the relationship between our $C_\epsilon^\delta$ and the capacity $\tilde{C}_\epsilon^\delta$ considered in [7], which is defined using the Euclidean norm.

**Theorem 6.** *Let $\tilde{C}_\epsilon^\delta$ be the $(\epsilon, \delta)$-capacity defined in [7]. We have*

$$C_\epsilon^\delta \leq \tilde{C}_\epsilon^{\delta/(2m_{\mathscr{Y}}(V_\epsilon))}, \tag{97}$$

*and*

$$\tilde{C}_\epsilon^\delta \leq C_\epsilon^{\delta m_{\mathscr{Y}}(V_\epsilon) 2^{2\tilde{C}_\epsilon^\delta + 1}}. \tag{98}$$

**Proof.** For every codebook $\mathcal{X} \in \mathscr{X}_\epsilon^\delta$ and $x_1, x_2 \in \mathcal{X}$, we have

$$e_\epsilon(x_1, x_2) \leq \delta/|\mathcal{X}|. \tag{99}$$

Since $m_{\mathscr{Y}}(\mathscr{Y}) = 1$, this implies that for all $x_1, x_2 \in \mathcal{X}$, we have

$$m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2)) \leq \delta/|\mathcal{X}|. \tag{100}$$

For all $\mathcal{X} \in \mathscr{X}$, the average overlap defined in ([7] (53)) is

$$\Delta = \frac{1}{|\mathcal{X}|} \sum_{x \in \mathcal{X}} \frac{e_\epsilon(x)}{2m_{\mathscr{Y}}(V_\epsilon)}. \tag{101}$$

Then, we have

$$\begin{aligned}\Delta &= \frac{1}{2|\mathcal{X}|m_{\mathscr{Y}}(V_\epsilon)} \sum_{x_1,x_2\in\mathcal{X}} m_{\mathscr{Y}}(S_\epsilon(x_1)\cap S_\epsilon(x_2)),\\ &\overset{(a)}{\leq} \frac{\delta|\mathcal{X}|^2}{2|\mathcal{X}|^2 m_{\mathscr{Y}}(V_\epsilon)},\\ &\leq \frac{\delta}{2m_{\mathscr{Y}}(V_\epsilon)},\end{aligned} \tag{102}$$

where $(a)$ follows from (100). Thus, we have

$$C_\epsilon^\delta \leq \tilde{C}_\epsilon^{\delta/(2m_{\mathscr{Y}}(V_\epsilon))}, \tag{103}$$

and (97) follows.

Now, let $\mathcal{X}$ be a codebook with average overlap at most $\delta$, namely

$$\frac{1}{2|\mathcal{X}|m_{\mathscr{Y}}(V_\epsilon)} \sum_{x_1,x_2\in\mathcal{X}} m_{\mathscr{Y}}(S_\epsilon(x_1)\cap S_\epsilon(x_2)) \leq \delta. \tag{104}$$

This implies that for all $x_1, x_2 \in \mathcal{X}$, we have

$$\begin{aligned}\frac{|\mathcal{X}|m_{\mathscr{Y}}(S_\epsilon(x_1)\cap S_\epsilon(x_2))}{m_{\mathscr{Y}}(\mathscr{Y})} &\leq \frac{2\delta|\mathcal{X}|^2 m_{\mathscr{Y}}(V_\epsilon)}{m_{\mathscr{Y}}(\mathscr{Y})},\\ &\overset{(a)}{=} 2\delta|\mathcal{X}|^2 m_{\mathscr{Y}}(V_\epsilon),\\ &\leq \delta 2^{2\tilde{C}_\epsilon^\delta+1} m_{\mathscr{Y}}(V_\epsilon),\end{aligned} \tag{105}$$

where $(a)$ follows from the fact that $m_{\mathscr{Y}}(\mathscr{Y}) = 1$. Thus, we have

$$\tilde{C}_\epsilon^\delta \leq C_\epsilon^{\delta 2^{2\tilde{C}_\epsilon^\delta+1} m_{\mathscr{Y}}(V_\epsilon)}, \tag{106}$$

and (98) follows. □

To better understand the relationship between the two capacities and show how they can be distinct, consider the case in which the output space is the union of the three $\epsilon$-balls depicted in Figure 6; this is the only feasible output configuration.

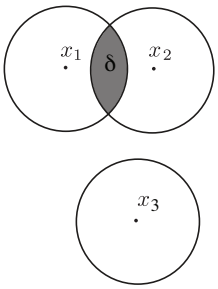



**Figure 6.** Output configuration for the computation of $C_\epsilon^{\delta'}$ and $\tilde{C}_\epsilon^{\delta'}$.

We now compute the two capacities $C_\epsilon^{\delta'}$ and $\tilde{C}_\epsilon^{\delta'}$ in this case. We have

$$m_{\mathscr{Y}}(S_\epsilon(x_1)\cap S_\epsilon(x_2)) = \delta, \tag{107}$$

and the average overlap (101) is

$$\Delta = \frac{1}{3}\frac{\delta}{2} = \frac{\delta}{6}. \tag{108}$$

It follows that

$$\tilde{C}_{\epsilon}^{\delta'} = \begin{cases} \log_2 3, & \text{if } \delta' \geq \delta/6. \\ \log_2 2, & \text{otherwise.} \end{cases} \tag{109}$$

On the other hand, the worst case overlap is

$$m_{\mathscr{Y}}(S_\epsilon(x_1) \cap S_\epsilon(x_2)) = \delta = \frac{3\delta}{|\mathscr{X}|}, \tag{110}$$

and it follows that

$$C_{\epsilon}^{\delta'} = \begin{cases} \log_2 3, & \text{if } \delta' \geq 3\delta. \\ \log_2 2, & \text{otherwise.} \end{cases} \tag{111}$$

## 5. $(N, \delta)$-Capacity of General Channels

We now extend our results to more general channels where the noise can be different across codewords and is not necessarily contained within a ball of radius $\epsilon$.

Let $\mathcal{X} \subseteq \mathscr{X}$ be a discrete set of points in the space, which represents a codebook. Any point $x \in \mathcal{X}$ represents a codeword that can be selected at the transmitter, sent over the channel, and received with perturbation. A channel with transition mapping $N : \mathscr{X} \to \mathscr{Y}$ associates with any point in $\mathscr{X}$ a set in $\mathscr{Y}$, such that the received codeword lies in the set

$$S_N(x) = \{y \in \mathscr{Y} : y \in N(x)\}. \tag{112}$$

Figure 7 illustrates possible uncertainty sets associated with three different codewords.

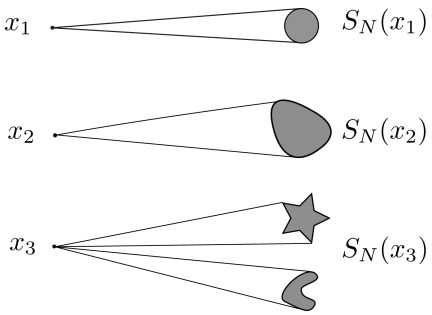



**Figure 7.** Uncertainty sets associated with three different codewords. Sets are not necessarily balls; they can be different across codewords and can also be composed of disconnected subsets.

All received codewords lie in the set $\mathcal{Y} = \bigcup_{x \in \mathcal{X}} S_N(x)$, where $\mathcal{Y} \subseteq \mathscr{Y}$. For any $x_1, x_2 \in \mathcal{X}$, we now let

$$e_N(x_1, x_2) = \frac{m_{\mathscr{Y}}(S_N(x_1) \cap S_N(x_2))}{m_{\mathscr{Y}}(\mathscr{Y})}, \tag{113}$$

where $m_{\mathscr{Y}}(.)$ is an uncertainty function defined over $\mathscr{Y}$. We also assume without loss of generality that the uncertainty associated with the space $\mathscr{Y}$ of received codewords is $m_{\mathscr{Y}}(\mathscr{Y}) = 1$. We also let $V_N = N(x^*)$, where $x^* = \text{argmin}_{x \in \mathscr{X}} m_{\mathscr{Y}}(N(x))$. Thus, $V_N$ is the set corresponding to the minimum uncertainty introduced by the noise mapping $N$.

**Definition 12.** $(N, \delta)$*-distinguishable codebook.*
*For any* $0 \leq \delta < m_{\mathscr{Y}}(V_N)$*, a codebook* $\mathcal{X} \subseteq \mathscr{X}$ *is* $(N, \delta)$*-distinguishable if for all* $x_1, x_2 \in \mathcal{X}$*, we have* $e_N(x_1, x_2) \leq \delta/|\mathcal{X}|$.

**Definition 13.** $(N, \delta)$*-capacity.*
*For any totally bounded, normed metric space* $\mathscr{X}$*, channel with transition mapping* $N$*, and* $0 \leq \delta < m_{\mathscr{Y}}(V_N)$*, the* $(N, \delta)$*-capacity of* $\mathscr{X}$ *is*

$$C_N^{\delta} = \sup_{\mathcal{X} \in \mathscr{X}_N^{\delta}} \log_2 |\mathcal{X}| \text{ bits}, \tag{114}$$

*where* $\mathscr{X}_N^{\delta} = \{\mathcal{X} : \mathcal{X} \text{ is } (N, \delta)\text{-distinguishable}\}$.

We now relate our definition of capacity to the notion of UVs and mutual information introduced in Section 3. As usual, let $X$ be the UV corresponding to the transmitted codeword and $Y$ be the UV corresponding to the received codeword. For a channel with transition mapping $N$, these UVs are such that for all $y \in [\![Y]\!]$ and $x \in [\![X]\!]$, we have

$$[\![Y|x]\!] = \{y \in [\![Y]\!] : y \in N(x)\}, \tag{115}$$

$$[\![X|y]\!] = \{x \in [\![X]\!] : y \in N(x)\}. \tag{116}$$

To measure the levels of association and disassociation between UVs $X$ and $Y$, we use an uncertainty function $m_{\mathscr{X}}(.)$ defined over $\mathscr{X}$, and $m_{\mathscr{Y}}(.)$ is defined over $\mathscr{Y}$. The definition of the feasible set is the same as the one given in (62). In our channel model, this feasible set depends on the transition mapping $N$ through (115) and (116).

We can now state the non-stochastic channel coding theorem for channels with transition mapping $N$.

**Theorem 7.** *For any totally bounded, normed metric space* $\mathscr{X}$*, channel with transition mapping* $N$ *satisfying* (115) *and* (116)*, and* $0 \leq \delta < m_{\mathscr{Y}}(V_N)$*, we have*

$$C_N^{\delta} = \sup_{X \in \mathscr{F}_{\tilde{\delta}}, \tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])} I_{\tilde{\delta}/|[\![X]\!]|}(Y; X) \text{ bits}. \tag{117}$$

The proof is along the same lines as the one of Theorem 4 and is omitted.

Theorem 7 characterizes the capacity as the supremum of the mutual information over all codebooks in the feasible set. The following theorem shows that the same characterization is obtained if we optimize the right hand side in (117) over all codebooks in the space. It follows by Theorem 7 that rather than optimizing over all codebooks, a capacity-achieving codebook can be found within the smaller class $\cup_{\tilde{\delta} \leq \delta / m_{\mathscr{Y}}(V_N)} \mathscr{F}_{\tilde{\delta}}$ of feasible sets with error at most $\delta / m_{\mathscr{Y}}(V_N)$.

**Theorem 8.** *The* $(N, \delta)$*-capacity in* (117) *can also be written as*

$$C_N^{\delta} = \sup_{\substack{X : [\![X]\!] \subseteq \mathscr{X}, \\ \tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])}} I_{\tilde{\delta}/|[\![X]\!]|}(Y; X) \text{ bits}. \tag{118}$$

The proof is along the same lines as the one of Theorem 5 and is omitted.

## 6. Capacity of Stationary Memoryless Uncertain Channels

In this section, we consider the special case of stationary, memoryless, uncertain channels.

Let $\mathscr{X}^{\infty}$ be the space of $\mathscr{X}$-valued discrete-time functions $x : \mathbb{Z}_{>0} \to \mathscr{X}$, where $\mathbb{Z}_{>0}$ is the set of positive integers denoting the time step. Let $x(a : b)$ denote the function $x \in \mathscr{X}^{\infty}$ restricted over the time interval $[a, b]$. Let $\mathcal{X} \subseteq \mathscr{X}^{\infty}$ be a discrete set which represents a codebook. Also, let $\mathcal{X}(1 : n) = \cup_{x \in \mathcal{X}} x(1 : n)$ denote the set of all codewords up to time $n$ and $\mathcal{X}(n) = \cup_{x \in \mathcal{X}} x(n)$ denote the set of all codeword symbols in the codebook at time $n$. The codeword symbols can be viewed as the coefficients representing a continuous signal in an infinite-dimensional space. For example, transmitting one symbol per time step can be viewed as transmitting a signal of unit spectral support over time. Any discrete-time

function $x \in \mathcal{X}$ can be selected at the transmitter, sent over a channel, received with noise perturbation, and introduced by the channel. The perturbation of the signal at the receiver due to the noise can be described as a displacement experienced by the corresponding codeword symbols $x(1), x(2), \ldots$. To describe this perturbation, we consider the set-valued map $N^\infty : \mathscr{X}^\infty \to 2^{\mathscr{Y}^\infty}$, associating any point in $\mathscr{X}^\infty$ to a set in $\mathscr{Y}^\infty$, where $\mathscr{Y}^\infty$ is the space of $\mathscr{Y}$-values discrete-time functions. For any transmitted codeword $x \in \mathcal{X} \subseteq \mathscr{X}^\infty$, the corresponding received codeword lies in the set

$$S_{N^\infty}(x) = \{y \in \mathscr{Y}^\infty : y \in N^\infty(x)\}. \tag{119}$$

Also, the noise set associated with $x(1:n) \in \mathcal{X}(1:n)$ is

$$S_{N^\infty}(x(1:n)) = \{y(1:n) \in \mathscr{Y}^n : y \in N^\infty(x)\}, \tag{120}$$

where $\mathscr{Y}^n = \underbrace{\mathscr{Y} \times \mathscr{Y} \times \cdots \times \mathscr{Y}}_{n}$. We are now ready to define stationary, memoryless, uncertain channels.

**Definition 14.** *A stationary, memoryless, uncertain channel is a transition mapping $N^\infty : \mathscr{X}^\infty \to 2^{\mathscr{Y}^\infty}$ that can be factorized into identical terms describing the noise experienced by the codeword symbols. Namely, there exists a set-valued map $N : \mathscr{X} \to \mathscr{Y}$ such that for all $n \in \mathbb{Z}_{>0}$ and $x(1:n) \in \mathscr{X}^\infty$, we have*

$$S_{N^\infty}(x(1:n)) = N(x(1)) \times \ldots \times N(x(n)). \tag{121}$$

According to the definition, a stationary, memoryless, uncertain channel maps the $n$th input symbol into the $n$th output symbol in a way that does not depend on the symbols at other time steps, and the mapping is the same at all time steps. Since the channel can be characterized by the mapping $N$, to simplify the notation, we will use $S_N(.)$ instead of $S_{N^\infty}(.)$.

Another important observation is that the $\epsilon$-perturbation channel in Definition 9 may not admit a factorization like the one in (121). For example, consider the space to be equipped with the $L^2$ norm, the codeword symbols to represent the coefficients of an orthogonal representation of a transmitted signal, and the noise experienced by any codeword to be within a ball of radius $\epsilon$. In this case, if a codeword symbol is perturbed by a value close to $\epsilon$, the perturbation of all other symbols must be close to zero.

For stationary, memoryless, uncertain channels, all received codewords lie in the set $\mathcal{Y} = \cup_{x \in \mathcal{X}} S_N(x)$, and the received codewords up to time $n$ lie in the set $\mathcal{Y}(1:n) = \cup_{x \in \mathcal{X}} S_N(x(1:n))$. Then, for any $x_1(1:n), x_2(1:n) \in \mathcal{X}(1:n)$, we let

$$\begin{aligned} &e_N(x_1(1:n), x_2(1:n)) \\ &= \frac{m_{\mathscr{Y}}(S_N(x_1(1:n)) \cap S_N(x_2(1:n)))}{m_{\mathscr{Y}}(\mathscr{Y}^n)}, \end{aligned} \tag{122}$$

where $m_{\mathscr{Y}}(.)$ is an uncertainty function defined over the space of the received codewords. We also assume without loss of generality that at any time step $n$, the uncertainty associated with the space $\mathscr{Y}^n$ of received codewords is $m_{\mathscr{Y}}(\mathscr{Y}^n) = 1$. We also let $V_N = N(x^*)$, where $x^* = \text{argmin}_{x \in \mathscr{X}} m_{\mathscr{Y}}(N(x))$. Thus, $V_N$ is the set corresponding to the minimum uncertainty introduced by the noise mapping at a single time step. Finally, we let $V_N^n = \underbrace{V_N \times V_N \times \cdots \times V_N}_{n}$. The quantity $1 - e_\epsilon(x_1(1:n), x_2(1:n))$ can be viewed as the confidence we have of not confusing $x_1(1:n)$ and $x_2(1:n)$ in any transmission or, equivalently, as the amount of adversarial effort required to induce a confusion between the two codewords. For example, if the uncertainty function is constructed using a measure,

then all the erroneous codewords generated by an adversary to decode $x_2(1:n)$ instead of $x_1(1:n)$ must lie inside the equivocation set $S_N(x_1(1:n)) \cap S_N(x_2(1:n))$ whose relative size is given by (122). The smaller the equivocation set is, the larger the effort required by the adversary to induce an error must be. If the uncertainty function represents the diameter of the set, then all the erroneous codewords generated by an adversary to decode $x_2(1:n)$ instead of $x_1(1:n)$ will be close to each other, in the sense of (122).

We now introduce the notion of a *distinguishable codebook*, ensuring that every codeword cannot be confused with any other codeword, rather than with a specific one, at a given level of confidence.

**Definition 15.** $(N, \delta_n)$*-distinguishable codebook.*
*For all* $n \in \mathbb{Z}_{>0}$ *and* $0 \leq \delta_n < m_{\mathscr{Y}}(V_N^n)$*, a codebook* $\mathcal{X}_n = \mathcal{X}(1:n)$ *is* $(N, \delta_n)$*-distinguishable if for all* $x_1(1:n), x_2(1:n) \in \mathcal{X}_n$*, we have*

$$e_N(x_1(1:n), x_2(1:n)) \leq \delta_n / |\mathcal{X}_n|. \tag{123}$$

It immediately follows that for any $(N, \delta_n)$-distinguishable codebook $\mathcal{X}_n$, we have

$$e_N(x(1:n)) = \sum_{\substack{x'(1:n) \in \mathcal{X}_n: \\ x'(1:n) \neq x(1:n)}} e_N(x(1:n), x'(1:n)) \leq \delta_n, \tag{124}$$

so that each codeword in $\mathcal{X}_n$ can be decoded correctly with confidence at least $1 - \delta_n$. Definion 15 guarantees even more, namely that the confidence of not confusing any pair of codewords is at least $1 - \delta_n / |\mathcal{X}_n|$.

We now associate with any sequence $\{\delta_n\}$ the largest distinguishable rate sequence $\{R_{\delta_n}\}$, whose elements represent the largest rates that satisfy that confidence sequence.

**Definition 16.** *Largest* $\{\delta_n\}$*-distinguishable rate sequence.*
*For any sequence* $\{\delta_n\}$*, the largest* $\{\delta_n\}$*-distinguishable rate sequence* $\{R_{\delta_n}\}$ *is such that for all* $n \in \mathbb{Z}_{>0}$*, we have*

$$R_{\delta_n} = \sup_{\mathcal{X}_n \in \mathscr{X}_N^{\delta_n}(n)} \frac{\log_2 |\mathcal{X}_n|}{n} \text{ bits per symbol}, \tag{125}$$

*where* $\mathscr{X}_N^{\delta_n}(n) = \{\mathcal{X}_n : \mathcal{X}_n \text{ is } (N, \delta_n)\text{-distinguishable}\}$.

We say that any constant rate $R$ that lays below the largest $\{\delta_n\}$-distinguishable rate sequence is $\{\delta_n\}$-distinguishable. Such a $\{\delta_n\}$-distinguishable rate ensures the existence of a sequence of distinguishable codes that, for all $n \in \mathbb{Z}_{>0}$, have a rate of at least $R$ and confidence of at least $1 - \delta_n$.

**Definition 17.** $\{\delta_n\}$*-distinguishable rate.*
*For any sequence* $\{\delta_n\}$*, a constant rate* $R$ *is said to be* $\{\delta_n\}$*-distinguishable if for all* $n \in \mathbb{Z}_{>0}$*, we have*

$$R \leq R_{\delta_n}. \tag{126}$$

We now give our first definition of capacity for stationary, memoryless, uncertain channels as the supremum of the $\{\delta_n\}$-distinguishable rates. Using this definition, transmitting at a constant rate below capacity ensures the existence of a sequence of codes that, for all $n \in \mathbb{Z}_{>0}$, have confidence of at least $1 - \delta_n$.

**Definition 18.** $(N, \{\delta_n\})_*$ *capacity.*
*For any stationary, memoryless, uncertain channel with transition mapping* $N$*, and any given sequence* $\{\delta_n\}$*, we let*

$$C_N(\{\delta_n\})_* = \sup\{R : R \text{ is } \{\delta_n\}\text{-distinguishable}\} \tag{127}$$

$$= \inf_{n \in \mathbb{Z}_{>0}} R_{\delta_n} \quad \text{bits per symbol.} \tag{128}$$

Another definition of capacity arises if, rather than the largest lower bound to the sequence of rates, one considers the least upper bound for which we can transmit, satisfying a given confidence sequence. Using this definition, transmitting at a constant rate below capacity ensures the existence of a finite-length code (rather than a sequence of codes) that satisfies at least one confidence value along the sequence $\{\delta_n\}$.

**Definition 19.** $(N, \{\delta_n\})^*$ *capacity.*
*For any stationary, memoryless, uncertain channel with transition mapping* $N$*, and any given sequence* $\{\delta_n\}$*, we define*

$$C_N(\{\delta_n\})^* = \sup_{n \in \mathbb{Z}_{>0}} R_{\delta_n} \quad \text{bits per symbol.} \tag{129}$$

Next, consider Definition 19 in the case of $\{\delta_n\}$ as a constant sequence; namely, for all $n \in \mathbb{Z}_{>0}$, we have $\delta_n = \delta \geq 0$. In this case, transmitting below capacity ensures the existence of a finite-length code that has confidence of at least $1 - \delta$. This is a generalization of the zero-error capacity,

**Definition 20.** $(N, \delta)^*$ *capacity.*
*For any stationary, memoryless, uncertain channel with transition mapping* $N$ *and any sequence* $\{\delta_n\}$*, where for all* $n \in \mathbb{Z}_{>0}$ *we have* $\delta_n = \delta \geq 0$*, we define*

$$C_N^{\delta *} = \sup_{n \in \mathbb{Z}_{>0}} R_{\delta_n} \quad \text{bits per symbol.} \tag{130}$$

Letting $\delta = 0$, we obtain the zero-error capacity. In this case, below capacity, there exists a code with which we can transmit with full confidence.

Finally, to give a definition of a non-stochastic analog of Shannon's probabilistic capacity, we first say that any constant rate $R$ is achievable if there exists a sequence $\delta_n \to 0$ as $n \to \infty$ such that $R$ lays below $\limsup_n R_{\delta_n}$. An achievable rate $R$ then ensures that for all $\epsilon > 0$, there exists an infinite sequence of distinguishable codes of rate of at least $R - \epsilon$ whose confidence tends towards one as $n \to \infty$. It follows that in this case, we can achieve communication at rate $R$ with arbitrarily high confidence by choosing a sufficiently large codebook.

**Definition 21.** *Achievable rate.*
*A constant rate* $R$ *is achievable if there exists a sequence* $\{\delta_n\}$ *such that* $\delta_n \to 0$ *as* $n \to \infty$ *and*

$$R \leq \limsup_{n \to \infty} R_{\delta_n}. \tag{131}$$

We now introduce the non-stochastic analog of Shannon's probabilistic capacity as the supremum of the achievable rates. This means that we can pick any confidence sequence such that $\delta_n$ tends towards zero as $n \to \infty$. In this way, $\delta_n$ plays the role of the probability of error and the capacity is the largest rate that can be achieved by a sequence of codebooks with an arbitrarily high confidence level. Using this definition, transmitting at a rate below

capacity ensures the existence of a sequence of codes achieving arbitrarily high confidence by increasing the codeword size.

**Definition 22.** $(N, \{\downarrow 0\})^*$ *capacity.*
*For any stationary, memoryless, uncertain channel with transition mapping* $N$*, we define the* $(N, \{\downarrow 0\})^*$ *capacity as*

$$C_N(\{\downarrow 0\})^* = \sup\{R : R \text{ is achievable}\} \tag{132}$$

$$= \sup_{\{\delta_n\}:\delta_n = o(1)} \limsup_{n\to\infty} R_{\delta_n}. \tag{133}$$

We point out the key difference between Definitions 20 and 22. Transmitting below the $(N, \delta)^*$ capacity ensures the existence of a fixed codebook that has confidence of at least $1 - \delta$. In contrast, transmitting below the $(N, \{\downarrow 0\})^*$ capacity allows us to achieve arbitrarily high confidence by increasing the codeword size.

To give a visual illustration of the different definitions of capacity, we refer to Figure 8.

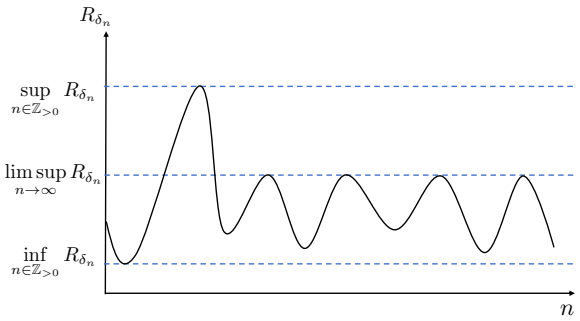



**Figure 8.** Illustration of capacities: This figure plots the sequence $R_{\delta_n}$ for a given sequence of $\delta_n$ with respect to $n > 0$.

For a given sequence $\{\delta_n\}$, the figure sketches the largest $\{\delta_n\}$-distinguishable rate sequence $R_{\delta_n}$. According to definitions 18 and 19, the capacities $C_N(\{\delta_n\})^*$ and $C_N(\{\delta_n\})_*$ are given by the supremum and infimum of this sequence, respectively. On the other hand, according to Definition 22, the capacity $C_N(\{\downarrow 0\})^*$ is the largest limsup over all vanishing sequences $\{\delta_n\}$. Assuming the figure refers to a vanishing sequence $\{\delta_n\}$ that achieves the supremum in (133), we have

$$C_N(\{\delta_n\})_* \geq C_N(\{\downarrow 0\})^* \geq C_N(\{\delta_n\})^*. \tag{134}$$

We now relate our notions of capacity to the *mutual information rate* between transmitted and received codewords. Let $X$ be the UV corresponding to the transmitted codeword. This is a map of $X : \mathscr{X}^\infty \to \mathcal{X}$ and $[\![X]\!] = \mathcal{X} \subseteq \mathscr{X}^\infty$. Restricting this map to a finite time $n \in \mathbb{Z}_{>0}$ yields another UV $X(n)$ and $[\![X(n)]\!] = \mathcal{X}(n) \subseteq \mathscr{X}$. Likewise, a codebook segment is a UV $X(a:b) = \{X(n)\}_{a\leq n\leq b}$ of marginal range $[\![X(a:b)]\!] = \mathcal{X}(a:b) \subseteq \mathscr{X}^{b-a+1}$. Likewise, let $Y$ be the UV corresponding to the received codeword. It is a map of $Y : \mathscr{Y}^\infty \to \mathcal{Y}$ and $[\![Y]\!] = \mathcal{Y} \subseteq \mathscr{Y}^\infty$. $Y(n)$ and $Y(a:b)$ are UVs, and $[\![Y(n)]\!] = \mathcal{Y} \subseteq \mathscr{Y}^\infty$ and $[\![Y(a:b)]\!] = \mathcal{Y}(a:b) \subseteq \mathscr{Y}^{b-a+1}$. For a stationary, memoryless, uncertain channel with transition mapping $N$, these UVs are such that for all $n \in \mathbb{Z}_{>0}$, $y(1:n) \in [\![Y(1:n)]\!]$ and $x(1:n) \in [\![X(1:n)]\!]$, and we have

$$\begin{aligned}[\![Y(1:n)|x(1:n)]\!] = &\{y(1:n) \in [\![Y(1:n)]\!] : \\ &y(1:n) \in S_N(x(1:n))\},\end{aligned} \tag{135}$$

$$\begin{aligned}\llbracket X(1:n)|y(1:n)\rrbracket =&\{x(1:n)\in\llbracket X(1:n)\rrbracket:\\ &y(1:n)\in S_N(x(1:n))\}.\end{aligned}\tag{136}$$

Now, we define the largest $\delta_n$-*mutual information rate* as the supremum mutual information per unit-symbol transmission that a codeword $X(1:n)$ can provide about $Y(1:n)$ with confidence of at least $1-\delta_n/|\llbracket X(1:n)\rrbracket|$.

**Definition 23.** *Largest $\delta_n$-information rate.*
*For all $n\in\mathbb{Z}_{>0}$, the largest $\delta_n$-information rate from $X(1:n)$ to $Y(1:n)$ is*

$$R^I_{\delta_n}=\sup_{\substack{X(1:n):\llbracket X(1:n)\rrbracket\subseteq\mathscr{X}^n,\\ \tilde{\delta}\leq\delta_n/m_{\mathscr{Y}}(\llbracket Y(1:n)\rrbracket)}}\frac{I_{\tilde{\delta}/|\llbracket X(1:n)\rrbracket|}(Y(1:n);X(1:n))}{n}.\tag{137}$$

In the following theorem. we establish the relationship between $R_{\delta_n}$ and $R^I_{\delta_n}$.

**Theorem 9.** *For any totally bounded, normed metric space $\mathscr{X}$, disrete-time space $\mathscr{X}^\infty$, stationary, memoryless, uncertain channel with transition mapping $N$ satisfying (135) and (136), and sequence $\{\delta_n\}$ such that, for all $n\in\mathbb{Z}_{>0}$, we have $0\leq\delta_n<m_{\mathscr{Y}}(V_N^n)$, we have*

$$R_{\delta_n}=\sup_{\substack{X(1:n)\in\mathscr{F}_{\tilde{\delta}}(n),\\ \tilde{\delta}\leq\delta_n/m_{\mathscr{Y}}(\llbracket Y(1:n)\rrbracket)}}\frac{I_{\tilde{\delta}/|\llbracket X(1:n)\rrbracket|}(Y(1:n);X(1:n))}{n}.\tag{138}$$

*We also have*

$$R_{\delta_n}=R^I_{\delta_n}.\tag{139}$$

**Proof.** The proof of the theorem is similar to the one of Theorem 4 and is given in Appendix B. □

The following coding theorem is now an immediate consequence of Theorem 9 and of our capacity definitions.

**Theorem 10.** *For any totally bounded, normed metric space $\mathscr{X}$, disrete-time space $\mathscr{X}^\infty$, stationary, memoryless, uncertain channel with transition mapping $N$ satisfying (135) and (136), and sequence $\{\delta_n\}$ such that for all $n\in\mathbb{Z}_{>0}$, $0\leq\delta_n<m_{\mathscr{Y}}(V_N^n)$ and $0\leq\delta<m_{\mathscr{Y}}(V_N^n)$, we have*

$$(1)\quad C_N(\{\delta_n\})_*=\inf_{n\in\mathbb{Z}_{>0}}R^I_{\delta_n},\tag{140}$$

$$(2)\quad C_N(\{\delta_n\})^*=\sup_{n\in\mathbb{Z}_{>0}}R^I_{\delta_n},\tag{141}$$

$$(3)\quad C_N(\{\downarrow 0\})^*=\sup_{\{\delta_n\}:\delta_n=o(1)}\limsup_{n\to\infty}R^I_{\delta_n},\tag{142}$$

$$(4)\quad C_N^{\delta*}=\sup_{n\in\mathbb{Z}_{>0}}R^I_{\delta_n}:\forall n\in\mathbb{Z}_{>0},\delta_n=\delta.\tag{143}$$

Theorem 10 provides multi-letter expressions of capacity, since $R^I_{\delta_n}$ depends on $I_{\tilde{\delta}/|\llbracket X(1:n)\rrbracket|}(Y(1:n);X(1:n))$ according to (137). In the Supplementary Materials, we establish some special cases of uncertainty functions, confidence sequences, and classes of stationary, memoryless, uncertain channels, leading to the factorization of the mutual information and to single-letter expressions.

## 7. Conclusions and Future Directions

We presented a non-stochastic notion of information with worst-case confidence and related it to the capacity of a communication channel subject to unknown noise. Using the non-stochastic variables framework of Nair [5] and a generalization of the Kolmogorov capacity allowing some amount of overlap in the packing sets [7], we showed that the capacity equals the largest amount of information conveyed by the transmitter to the receiver, with a given level of confidence. These results are the natural generalization of Nair's ones, obtained in a zero-error framework, and provide an information-theoretic interpretation of the geometric problem of sphere packing with overlap, as studied in [7].

Non-stochastic approaches to information and their use to quantify the performance of various engineering systems have recently received attention in the context of estimation, control, security, communication over non-linear optical channels, and learning systems [8–14]. We hope that the theory developed here can be useful in the future in some of these contexts. While refinements and extensions of the theory are certainly of interest, explorations of application domains are of paramount importance. There is evidence in the literature regarding the need for a non-stochastic approach to study the flow of information in complex systems, and there is a certain tradition in computer science, especially in the field of online learning, to study various problems in both a stochastic and a non-stochastic setting [15–17]. Nevertheless, it seems that only a few isolated efforts have been made towards the formal development of a non-stochastic information theory. Wider involvement of the community in developing alternative, even competing, theories is certainly advisable to eventually fulfill the need of these application areas.

**Supplementary Materials:** The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/e27050472/s1, Kolmogorov capacity with overlap.

**Author Contributions:** Conceptualization, M.F. and A.R.; methodology, M.F. and A.R.; investigation, M.F. and A.R.; writing—original draft preparation, A.R.; writing—review and editing, M.F. and A.R.; supervision, M.F.; project administration, M.F.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially supported by NSF Award Number: 2127605.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Data is contained within the article.

**Acknowledgments:** This article is a revised and expanded version of a paper entitled Insights into [Towards a Non-Stochastic Information Theory], which was presented at [2019 IEEE International Symposium of Information Theory, Paris, France, 7–12 July 2019] and [Channel Coding Theorems in Non-Stochastic Information Theory], which was presented at [2021 IEEE International Symposium of Information Theory, Melbourne, VIC, Australia, 12–20 July 2021] [18,19].

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## Appendix A. Proof of Lemma 1

**Proof.** Let $(X, Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$. Then,

$$\mathscr{A}(X;Y) \preceq \delta_1, \tag{A1}$$

$$\mathscr{A}(Y;X) \preceq \delta_2. \tag{A2}$$

Let

$$\mathscr{S}_1 = \left\{ (y_1, y_2) : \frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} = 0 \right\}. \tag{A3}$$

Then, for all $(y_1, y_2) \in \mathscr{S}_1$, we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} = 0 \leq \delta_1. \tag{A4}$$

Also, if $(y_1, y_2) \in \mathscr{S}_1$, then

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} \notin \mathscr{A}(X;Y), \tag{A5}$$

and if $(y_1, y_2) \notin \mathscr{S}_1$, then using (9), we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} \in \mathscr{A}(X;Y). \tag{A6}$$

This along with (A1) and (A4) implies that (17) follows.

Likewise, let

$$\mathscr{S}_2 = \left\{ (x_1, x_2) : \frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} = 0 \right\}. \tag{A7}$$

Then, for all $(x_1, x_2) \in \mathscr{S}_2$,

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} = 0 \leq \delta_2. \tag{A8}$$

Also, if $(x_1, x_2) \in \mathscr{S}_2$, then

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} \notin \mathscr{A}(Y;X), \tag{A9}$$

and if $(y_1, y_2) \notin \mathscr{S}_2$, then using (9), we have

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} \in \mathscr{A}(Y;X). \tag{A10}$$

This along with (A2) and (A8) implies that (18) follows.

Now, we prove the opposite direction of the statement. Given that for all $y_1, y_2 \in [\![Y]\!]$, we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!]|)}{m_{\mathscr{X}}([\![X]\!])} \leq \delta_1, \tag{A11}$$

and for all $x_1, x_2 \in [\![X]\!]$, we have

$$\frac{m_{\mathscr{Y}}([\![Y|x_1]\!] \cap [\![Y|x_2]\!]|)}{m_{\mathscr{Y}}([\![Y]\!])} \leq \delta_2. \tag{A12}$$

Then, using the definition of $\mathscr{A}(X;Y)$ and $\mathscr{A}(Y;X)$, we have

$$\mathscr{A}(X;Y) \preceq \delta_1, \tag{A13}$$

$$\mathscr{A}(Y;X) \preceq \delta_2. \tag{A14}$$

The statement of the lemma follows. □

## Appendix B. Proof of Theorem 9

**Proof.** We will show (138). Then, using Lemma A4 in Appendix C, (139) follows using the same argument as in the proof of Theorem 8.

We proceed in three steps. First, we show that for all $n > 0$, there exists a UV $X(1:n)$ and $\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])$ such that $X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n)$, which implies that $\mathscr{F}_{\tilde{\delta}}(n)$ is not empty, and so the supremum is well defined. Second, for all $n > 0$, $X(1:n)$, and $\tilde{\delta}$ such that

$$X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n), \tag{A15}$$

and

$$\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!]), \tag{A16}$$

we show that

$$\frac{I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n);X(1:n))}{n} \leq R_{\delta_n}.$$

Finally, for all $n > 0$, we show the existence of $X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n)$ and $\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])$ such that

$$\frac{I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n);X(1:n))}{n} = R_{\delta_n}. \tag{A17}$$

Let us begin with the first step. Consider a point $x(1:n) \in \mathscr{X}^n$. Let $X(1:n)$ be a UV such that

$$[\![X(1:n)]\!] = \{x(1:n)\}. \tag{A18}$$

Then, we hold that the marginal range of the UV $Y(1:n)$ corresponding to the received variable is

$$[\![Y(1:n)]\!] = [\![Y(1:n)|x(1:n)]\!], \tag{A19}$$

and therefore, for all $y(1:n) \in [\![Y(1:n)]\!]$, we have

$$[\![X(1:n)|y(1:n)]\!] = \{x(1:n)\}. \tag{A20}$$

Using Definition 2 and (A18), we have

$$\mathscr{A}(Y(1:n);X(1:n)) = \emptyset, \tag{A21}$$

because $[\![X(1:n)]\!]$ consists of a single point, and therefore, the set in (12) is empty.

On the other hand, using Definition 2 and (A20), we have

$$\mathscr{A}(X(1:n);Y(1:n)) = \begin{cases} \{1\} \text{ if } \exists y_1(1:n), y_2(1:n) \\ \qquad \in [\![Y(1:n)]\!], \\ \emptyset \quad \text{otherwise.} \end{cases} \tag{A22}$$

Using (A21), and since $\mathscr{A} \preceq \delta$ holds for $\mathscr{A} = \emptyset$, we have

$$\mathscr{A}(Y(1:n);X(1:n)) \preceq \delta_n / (|[\![X(1:n)]\!]| m_{\mathscr{Y}}([\![Y(1:n)]\!])). \tag{A23}$$

Similarly, using (A22), we have

$$\mathscr{A}(X(1:n);Y(1:n)) \preceq 1. \tag{A24}$$

Now, combining (A23) and (A24), we have

$$(X(1:n),Y(1:n)) \overset{a}{\leftrightarrow} (1, \delta_n / (|[\![X(1:n)]\!]| m_{\mathscr{Y}}([\![Y(1:n)]\!]))). \tag{A25}$$

Letting $\tilde{\delta} = \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])$, this implies that $X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n)$ and that the first step of the proof is complete.

To prove the second step, we define

$$\begin{aligned}\mathscr{G}(n) = \Big\{ & X(1:n) : [\![X(1:n)]\!] \subseteq \mathscr{X}^n, \\ & \exists \tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!]) \text{ such that} \\ & \forall \mathscr{S}_1, \mathscr{S}_2 \in [\![Y(1:n)|X(1:n)]\!], \\ & \frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{Y}}([\![Y(1:n)]\!])} \leq \frac{\tilde{\delta}}{|[\![X(1:n)]\!]|} \Big\},\end{aligned} \tag{A26}$$

which is a larger set than the one containing all UVs $X(1:n)$ that are $(1, \tilde{\delta}/|[\![X(1:n)]\!]|)$ associated with $Y(1:n)$. Similarly to (79), it can be shown that

$$X(1:n) \in \mathscr{G}(n) \implies \mathcal{X}(1:n) \in \mathscr{X}_N^{\delta_n}(n) \tag{A27}$$

Consider now a pair, $X(1:n)$ and $\tilde{\delta}$, such that $\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y(1:n)]\!])$ and

$$X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n). \tag{A28}$$

If $(X(1:n), Y(1:n)) \overset{d}{\leftrightarrow} (0, \tilde{\delta}/|[\![X(1:n)]\!]|)$, then, using Lemma A1 in Appendix C, there exist UVs $\bar{X}(1:n)$ and $\bar{Y}(1:n)$ and $\bar{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!])$ such that

$$(\bar{X}(1:n), \bar{Y}(1:n)) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}(1:n)]\!]|), \tag{A29}$$

and

$$\begin{aligned}&|[\![Y(1:n)|X(1:n)]\!]^*_{\tilde{\delta}/|[\![X(1:n)]\!]|}| \\ &= |[\![\bar{Y}(1:n)|\bar{X}(1:n)]\!]^*_{\bar{\delta}/|[\![\bar{X}(1:n)]\!]|}|.\end{aligned} \tag{A30}$$

On the other hand, if $(X(1:n), Y(1:n)) \overset{a}{\leftrightarrow} (1, \tilde{\delta}/|[\![X(1:n)]\!]|)$, then (A29) and (A30) also trivially hold. It then follows that (A29) and (A30) hold for all $X(1:n) \in \mathscr{F}_{\tilde{\delta}}(n)$. We now have

$$\begin{aligned}&I_{\tilde{\delta}/|[\![X(1:n)]\!]|}(Y(1:n); X(1:n)) \\ &= \log(|[\![Y(1:n)|X(1:n)]\!]^*_{\tilde{\delta}/|[\![X(1:n)]\!]|}|) \\ &\overset{(a)}{=} \log(|[\![\bar{Y}(1:n)|\bar{X}(1:n)]\!]^*_{\bar{\delta}/|[\![\bar{X}(1:n)]\!]|}|) \\ &\overset{(b)}{\leq} \log(|[\![\bar{X}(1:n)]\!]|), \\ &\overset{(c)}{=} \log(|\bar{\mathcal{X}}(1:n)|), \\ &\overset{(d)}{\leq} nR_{\delta_n},\end{aligned} \tag{A31}$$

where $(a)$ follows from (A29) and (A30), $(b)$ follows from Lemma A3 in Appendix C since $\bar{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!]) < m_{\mathscr{Y}}(V_N^n) / m_{\mathscr{Y}}([\![\bar{Y}(1:n)]\!])$, $(c)$ follows by defining the codebook $\bar{\mathcal{X}}(1:n)$ corresponding to the UV $[\![\bar{X}(1:n)]\!]$, and $(d)$ follows from the fact that using (A29) and Lemma 1, we have $\bar{X}(1:n) \in \mathcal{G}(n)$, which implies by (A27) that $\bar{\mathcal{X}}(1:n) \in \mathscr{X}_N^{\delta_n}(n)$.

For any $n \in \mathbb{Z}_{>0}$, let

$$\mathcal{X}_n^* = \operatorname{argsup}_{\mathcal{X}_n \in \mathscr{X}_N^{\delta_n}(n)} \frac{\log(|\mathcal{X}_n|)}{n}, \tag{A32}$$

which achieves the rate $R_{\delta_n}$. Let $X^*$ be the UV whose marginal range corresponds to the codebook $\mathcal{X}_n^*$. It follows that for all $\mathscr{S}_1, \mathscr{S}_1 \in [\![Y^*|X^*]\!]$, we have

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_1)}{m_{\mathscr{Y}}(\mathscr{Y}^n)} \leq \frac{\delta_n}{|[\![X^*]\!]|}, \tag{A33}$$

which implies that $m_{\mathscr{Y}}(\mathscr{Y}^n) = 1$,

$$\frac{m_{\mathscr{Y}}(\mathscr{S}_1 \cap \mathscr{S}_1)}{m_{\mathscr{Y}}([\![Y^*]\!])} \leq \frac{\delta_n}{(|[\![X^*]\!]| m_{\mathscr{Y}}([\![Y^*]\!]))}. \tag{A34}$$

Letting $\delta^* = \delta_n / m_{\mathscr{Y}}([\![Y^*]\!])$, and using Lemma 1, we hold that $(X^*, Y^*) \overset{a}{\leftrightarrow} (1, \delta^*/|[\![X^*]\!]|)$, which implies

$$X^* \in \cup_{\tilde{\delta} \leq \delta_n / m_{\mathscr{Y}}([\![Y^*]\!])} \mathscr{F}_{\tilde{\delta}}(n), \tag{A35}$$

and (138) follows. □

## Appendix C. Auxiliary Results

**Lemma A1.** *Given a $\delta < m_{\mathscr{Y}}(V_N)$, two UVs $X$ and $Y$ satisfying (60) and (61), and a $\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])$ such that*

$$(X, Y) \overset{d}{\leftrightarrow} (0, \tilde{\delta}/|[\![X]\!]|). \tag{A36}$$

*Then, there exist two UVs $\bar{X}$ and $\bar{Y}$ satisfying (60) and (61), and there exists a $\bar{\delta} \leq \delta / m_{\mathscr{Y}}([\![\bar{Y}]\!])$ such that*

$$(\bar{X}, \bar{Y}) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}]\!]|), \tag{A37}$$

*and*

$$|[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}| = |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}|. \tag{A38}$$

**Proof.** Let the cardinality

$$|[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}| = K. \tag{A39}$$

By Property 1 of Definition 6, we hold that for all $\mathscr{S}_i \in [\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}$, there exists a $x_i \in [\![X]\!]$ such that $[\![Y|x_i]\!] \subseteq \mathscr{S}_i$. Now, consider a new UV $\bar{X}$ whose marginal range is composed of $K$ elements of $[\![X]\!]$, namely

$$[\![\bar{X}]\!] = \{x_1, x_2, \ldots, x_K\}. \tag{A40}$$

Let $\bar{Y}$ be the UV corresponding to the received variable. Using the fact that for all $x \in \mathscr{X}$, we have $[\![\bar{Y}|x]\!] = [\![Y|x]\!]$ since (60) holds, and using Property 2 of Definition 6, for all $x, x' \in [\![\bar{X}]\!]$, we have

$$\begin{aligned} \frac{m_{\mathscr{Y}}([\![\bar{Y}|x]\!] \cap [\![\bar{Y}|x']\!])}{m_{\mathscr{Y}}([\![Y]\!])} &\leq \frac{\tilde{\delta}}{|[\![X]\!]|}, \\ &\overset{(a)}{\leq} \frac{\tilde{\delta}}{|[\![\bar{X}]\!]|}, \end{aligned} \tag{A41}$$

where $(a)$ follows from the fact that $[\![\bar{X}]\!] \subseteq [\![X]\!]$ using (A40). Then, for all $x, x' \in [\![\bar{X}]\!]$, we hold that

$$\begin{aligned} \frac{m_{\mathscr{Y}}([\![\bar{Y}|x]\!] \cap [\![\bar{Y}|x']\!])}{m_{\mathscr{Y}}([\![\bar{Y}]\!])} &\leq \frac{\tilde{\delta} m_{\mathscr{Y}}([\![Y]\!])}{|[\![X]\!]| m_{\mathscr{Y}}([\![\bar{Y}]\!])}, \\ &\overset{(a)}{\leq} \frac{\bar{\delta}}{|[\![\bar{X}]\!]|}, \end{aligned} \tag{A42}$$

where $\bar{\delta} = \tilde{\delta} m_{\mathscr{Y}}([\![Y]\!]) / m_{\mathscr{Y}}([\![\bar{Y}]\!])$. Then, by Lemma 1, it follows that

$$(\bar{X}, \bar{Y}) \overset{a}{\leftrightarrow} (1, \bar{\delta}/|[\![\bar{X}]\!]|). \tag{A43}$$

Since $\tilde{\delta} \leq \delta / m_{\mathscr{Y}}([\![Y]\!])$, we have

$$\bar{\delta} \leq \delta / m_{\mathscr{Y}}([\![\bar{Y}]\!]) < m_{\mathscr{Y}}(V_\epsilon)/m_{\mathscr{Y}}([\![\bar{Y}]\!]). \tag{A44}$$

Using (A43) and (A44), we now hold that

$$\begin{aligned} |[\![\bar{Y}|\bar{X}]\!]^*_{\bar{\delta}/|[\![\bar{X}]\!]|}| &\overset{(a)}{=} |[\![\bar{X}]\!]| \\ &\overset{(b)}{=} |[\![Y|X]\!]^*_{\tilde{\delta}/|[\![X]\!]|}|, \end{aligned} \tag{A45}$$

where $(a)$ follows from Lemma A4 in Appendix C and $(b)$ follows from (A39) and (A40). Hence, the statement of the lemma follows. □

**Lemma A2.** *Let*

$$(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2). \tag{A46}$$

*If* $x \overset{\delta}{\rightsquigarrow} x_1$ *and* $x \overset{\delta}{\rightsquigarrow} x_2$, *then we hold that* $x_1 \overset{\delta}{\rightsquigarrow} x_2$.

**Proof.** Let $\{[\![X|y_i]\!]\}_{i=1}^{N}$ be the sequence of conditional range connecting $x$ and $x_1$. Likewise, let $\{[\![X|\tilde{y}_i]\!]\}_{i=1}^{\tilde{N}}$ be the sequence of conditional range connecting $x$ and $x_2$.

Now, by Definition 5, we have

$$x_1 \in [\![X|y_N]\!], \tag{A47}$$

$$x_2 \in [\![X|\tilde{y}_{\tilde{N}}]\!], \tag{A48}$$

$$x \in [\![X|y_1]\!], \tag{A49}$$

and

$$x \in [\![X|\tilde{y}_1]\!] \tag{A50}$$

Then, using (9), we hold that

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|\tilde{y}_1]\!]|}{m_{\mathscr{X}}([\![X]\!])} > 0, \tag{A51}$$

which implies that

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|\tilde{y}_1]\!]|}{m_{\mathscr{X}}([\![X]\!])} \in \mathscr{A}(X;Y). \tag{A52}$$

Using the fact that

$$(X, Y) \overset{d}{\leftrightarrow} (\delta, \delta_2), \tag{A53}$$

we will now show that

$$\{[\![X|y_N]\!], [\![X|y_{N-1}]\!], \ldots [\![X|y_1]\!], [\![X|\tilde{y}_1]\!], \ldots [\![X|\tilde{y}_{\tilde{N}}]\!]\}, \tag{A54}$$

is a sequence of conditional ranges connecting $x_1$ and $x_2$. Using (A52) and (A53), we hold that

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|\tilde{y}_1]\!]|}{m_{\mathscr{X}}([\![X]\!])} > \delta. \tag{A55}$$

Also, for all $1 < i \leq N$ and $1 < j \leq \tilde{N}$, we have

$$\frac{m_{\mathscr{X}}([\![X|y_i]\!] \cap [\![X|y_{i-1}]\!]|}{m_{\mathscr{X}}([\![X]\!])} > \delta, \tag{A56}$$

and

$$\frac{m_{\mathscr{X}}([\![X|\tilde{y}_j]\!] \cap [\![X|\tilde{y}_{j-1}]\!]|}{m_{\mathscr{X}}([\![X]\!])} > \delta. \tag{A57}$$

Also, we have

$$x_1 \in [\![X|y_N]\!], \text{ and } x_2 \in [\![X|\tilde{y}_{\tilde{N}}]\!]. \tag{A58}$$

Hence, combining (A55), (A56), (A57) and (A58), we hold that $x_1 \overset{\delta}{\leftrightsquigarrow} x_2$ via (A54). □

**Lemma A3.** *Consider two UIVs X and Y. Let*

$$\delta^* = \frac{\min_{y \in [\![Y]\!]} m_{\mathscr{X}}([\![X|y]\!])}{m_{\mathscr{X}}([\![X]\!])}. \tag{A59}$$

*If* $\delta_1 < \delta^*$*, then we have*

$$|[\![X|Y]\!]^*_{\delta_1}| \leq |[\![Y]\!]|. \tag{A60}$$

**Proof.** We will prove this by contradiction. Let

$$|[\![X|Y]\!]^*_{\delta_1}| > |[\![Y]\!]|. \tag{A61}$$

Then, by Property 1 of Definition 6, there exists two sets $\mathscr{S}_1, \mathscr{S}_2 \in [\![X|Y]\!]^*_{\delta_1}$ and one singly $\delta_1$-connected set $[\![X|y]\!]$ such that

$$[\![X|y]\!] \subseteq \mathscr{S}_1, \text{ and } [\![X|y]\!] \subseteq \mathscr{S}_2. \tag{A62}$$

Then, we have

$$\begin{aligned} \frac{m_{\mathscr{X}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{X}}([\![X]\!])} &\overset{(a)}{\geq} \frac{m_{\mathscr{X}}([\![X|y]\!])}{m_{\mathscr{X}}([\![X]\!])}, \\ &\overset{(b)}{\geq} \delta^*, \\ &\overset{(c)}{>} \delta_1, \end{aligned} \tag{A63}$$

where $(a)$ follows from (A62) and (10), $(b)$ follows from (A59), and $(c)$ follows from the fact that $\delta_1 < \delta^*$. However, by Property 2 of Definition 6, we have

$$\frac{m_{\mathscr{X}}(\mathscr{S}_1 \cap \mathscr{S}_2)}{m_{\mathscr{X}}([\![X]\!])} \leq \delta_1. \tag{A64}$$

Hence, we hold that (A63) and (A64) contradict each other, which implies that (A61) does not hold. Hence, the statement of the theorem follows. □

**Lemma A4.** *Consider two UIVs X and Y. Let*

$$\delta^* = \frac{\min_{y \in [\![Y]\!]} m_{\mathscr{X}}([\![X|y]\!])}{m_{\mathscr{X}}([\![X]\!])}. \tag{A65}$$

*For all* $\delta_1 < \delta^*$ *and* $\delta_2 \leq 1$*, if* $(X, Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$*, then we have*

$$|[\![X|Y]\!]^*_{\delta_1}| = |[\![Y]\!]|. \tag{A66}$$

*Additionally, $[\![X|Y]\!]$ is a $\delta_1$-overlap family.*

**Proof.** We show that

$$[\![X|Y]\!] = \{[\![X|y]\!] : y \in [\![Y]\!]\} \tag{A67}$$

is a $\delta_1$-overlap family. First, note that $[\![X|Y]\!]$ is a cover of $[\![X]\!]$, since $[\![X]\!] = \cup_{y\in[\![Y]\!]}[\![X|y]\!]$. Second, each set in the family $[\![X|Y]\!]$ is singly $\delta_1$-connected via $[\![X|Y]\!]$, since trivially any two points $x_1, x_2 \in [\![X|y]\!]$ are singly $\delta_1$-connected via the same set. It follows that Property 1 of Definition 6 holds.

Now, since $(X, Y) \overset{a}{\leftrightarrow} (\delta_1, \delta_2)$, then by Lemma 1 for all $y_1, y_2 \in [\![Y]\!]$ we have

$$\frac{m_{\mathscr{X}}([\![X|y_1]\!] \cap [\![X|y_2]\!])}{m_{\mathscr{X}}([\![X]\!])} \leq \delta_1, \tag{A68}$$

which shows that Property 2 of Definition 6 holds. Finally, it is also easy to see that Property 3 of Definition 6 holds, since $[\![X|Y]\!]$ contains all sets $[\![X|y]\!]$. Hence, $[\![X|Y]\!]$ satisfies all the properties of $\delta_1$-overlap family, which implies that

$$|[\![X|Y]\!]| \leq |[\![X|Y]\!]^*_{\delta_1}|. \tag{A69}$$

Since $|[\![X|Y]\!]| = |[\![Y]\!]|$, using Lemma A3, we also have

$$|[\![X|Y]\!]| \geq |[\![X|Y]\!]^*_{\delta_1}|. \tag{A70}$$

Combining (A69), (A70) and the fact that $[\![X|Y]\!]$ satisfies all the properties of $\delta_1$-overlap family, the statement of the lemma follows. □

*Appendix C.1. Taxicab Symmetry of the Mutual Information*

**Definition A1.** *$(\delta_1, \delta_2)$-taxicab connectedness and $(\delta_1, \delta_2)$-taxicab isolation.*

- *Points $(x, y), (x', y') \in [\![X, Y]\!]$ are $(\delta_1, \delta_2)$-taxicab connected via $[\![X, Y]\!]$ and are denoted by $(x, y) \overset{\delta_1,\delta_2}{\rightsquigarrow} (x', y')$, if there exists a finite sequence $\{(x_i, y_i)\}_{i=1}^N$ of points in $[\![X, Y]\!]$ such that $(x, y) = (x_1, y_1)$, $(x', y') = (x_N, y_N)$, and for all $2 < i \leq N$, we have either*

$$A_1 = \{x_i = x_{i-1} \text{ and } \frac{m_{\mathscr{X}}([\![X|y_i]\!] \cap [\![X|y_{i-1}]\!])}{m_{\mathscr{X}}([\![X]\!])} > \delta_1\},$$

*or*

$$A_2 = \{y_i = y_{i-1} \text{ and } \frac{m_{\mathscr{Y}}([\![Y|x_i]\!] \cap [\![Y|x_{i-1}]\!])}{m_{\mathscr{Y}}([\![Y]\!])} > \delta_2\}.$$

*If $(x, y) \overset{\delta_1,\delta_2}{\rightsquigarrow} (x', y')$ and $N = 2$, then we say that $(x, y)$ and $(x', y')$ are singly $(\delta_1, \delta_2)$-taxicab connected, i.e., either $y = y'$ and $x, x' \in [\![X|y]\!]$ or $x = x'$ and $y, y' \in [\![Y|x]\!]$.*
- *A set $\mathscr{S} \subseteq [\![X, Y]\!]$ is (singly) $(\delta_1, \delta_2)$-taxicab connected via $[\![X, Y]\!]$ if every pair of points in the set is (singly) $(\delta_1, \delta_2)$-taxicab connected in $[\![X, Y]\!]$.*
- *Two sets $\mathscr{S}_1, \mathscr{S}_2 \subseteq [\![X, Y]\!]$ are $(\delta_1, \delta_2)$-taxicab isolated via $[\![X, Y]\!]$ if no point in $\mathscr{S}_1$ is $(\delta_1, \delta_2)$-taxicab connected to any point in $\mathscr{S}_2$.*

**Definition A2.** *Projection of a set*

- *The projection $\mathscr{S}^+_x$ of a set $\mathscr{S} \subseteq [\![X, Y]\!]$ on the x-axis is defined as*

$$\mathscr{S}^+_x = \{x : (x, y) \in \mathscr{S}\}. \tag{A71}$$

- *The projection $\mathscr{S}_y^+$ of a set $\mathscr{S} \subseteq [\![X, Y]\!]$ on the y-axis is defined as*

$$\mathscr{S}_y^+ = \{y : (x, y) \in \mathscr{S}\}. \tag{A72}$$

**Definition A3.** *$(\delta_1, \delta_2)$-taxicab family*
*A $(\delta_1, \delta_2)$-taxicab family of $[\![X, Y]\!]$, denoted by $[\![X, Y]\!]^*_{(\delta_1,\delta_2)}$, is a largest family of distinct sets covering $[\![X, Y]\!]$ such that*

1. *Each set in the family is $(\delta_1, \delta_2)$-taxicab connected and contains at least one singly $\delta_1$-connected set of form $[\![X|y]\!] \times \{y\}$ and at least one singly $\delta_2$-connected set of the form $[\![Y|x]\!] \times \{x\}$.*
2. *The measure of overlap between the projections on the x-axis and y-axis of any two distinct sets in the family are at most $\delta_1 m_{\mathscr{X}}([\![X]\!])$ and $\delta_2 m_{\mathscr{Y}}([\![Y]\!])$, respectively.*
3. *For every singly $(\delta_1, \delta_2)$-connected set, there exists a set in the family containing it.*

We now show that when $(X, Y) \overset{d}{\leftrightarrow} (\delta_1, \delta_2)$ hold, the cardinality of the $(\delta_1, \delta_2)$-taxicab family is same as the cardinality of the $[\![X|Y]\!]$ $\delta_1$-overlap family and $[\![Y|X]\!]$ $\delta_2$-overlap family.

**Proof of Theorem 3.** We will show that $|[\![X, Y]\!]^*_{(\delta_1,\delta_2)}| = |[\![X|Y]\!]^*_{\delta_1}|$. Then, $|[\![X, Y]\!]^*_{(\delta_1,\delta_2)}| = |[\![Y|X]\!]^*_{\delta_2}|$ can be derived along the same lines. Hence, the statement of the theorem follows.

First, we will show that

$$\mathcal{D} = \{\mathscr{S}_x^+ : \mathscr{S} \in [\![X, Y]\!]^*_{(\delta_1,\delta_2)}\}, \tag{A73}$$

satisfies all the properties of $[\![X|Y]\!]^*_{\delta_1}$.

Since $[\![X, Y]\!]^*_{(\delta_1,\delta_2)}$ is a covering of $[\![X, Y]\!]$, we have

$$\cup_{\mathscr{S}_x^+ \in \mathcal{D}} \mathscr{S}_x^+ = [\![X]\!], \tag{A74}$$

which implies that $\mathcal{D}$ is a covering of $[\![X]\!]$.

Consider a set $\mathscr{S} \in [\![X, Y]\!]^*_{(\delta_1,\delta_2)}$. For all $(x, y), (x', y') \in \mathscr{S}$, $(x, y)$ and $(x', y')$ are $(\delta_1, \delta_2)$-taxicab connected. Then, there exists a taxicab sequence of the form

$$(x, y), (x_1, y), (x_1, y_1), \ldots (x_{n-1}, y'), (x', y').$$

such that either $A_1$ or $A_2$ in Definition A1 is true. Then, the sequence $\{y, y_1, \ldots, y_{n-1}, y'\}$ yields a sequence of conditional range $\{[\![X|\tilde{y}_j]\!]\}_{j=1}^{n+1}$ such that for all $1 < j \leq n + 1$,

$$\frac{m_{\mathscr{X}}([\![X|\tilde{y}_j]\!] \cap [\![X|\tilde{y}_{j-1}]\!])}{m_{\mathscr{X}}([\![X]\!])} > \delta_1, \tag{A75}$$

$$x \in [\![X|\tilde{y}_1]\!], \text{ and } x' \in [\![X|\tilde{y}_{n+1}]\!]. \tag{A76}$$

Hence, $x \overset{\delta_1}{\rightsquigarrow} x'$ via $[\![X|Y]\!]$. Hence, $\mathscr{S}_x^+$ is $\delta_1$-connected via $[\![X|Y]\!]$. Also, $\mathscr{S}$ contains at least one singly $\delta_1$-connected set of the form $[\![X|y]\!] \times \{y\}$, which implies $[\![X|y]\!] \subseteq \mathscr{S}_x^+$. Hence, $\mathscr{S}_x^+$ contains at least one singly $\delta_1$-connected set of the form $[\![X|y]\!]$. Hence, $\mathcal{D}$ satisfies Property 1 in Definition 6.

For all $\mathscr{S}_1, \mathscr{S}_2 \in [\![X, Y]\!]^*_{(\delta_1,\delta_2)}$, we have

$$m_{\mathscr{X}}(\mathscr{S}_{1,x}^+ \cap \mathscr{S}_{2,x}^+) \leq \delta_1 m_{\mathscr{X}}([\![X]\!]), \tag{A77}$$

using Property 2 in Definition A3. Hence, $\mathcal{D}$ satisfies Property 2 in Definition 6.

Using Property 3 in Definition A3, we hold that for all $[\![X|y]\!] \times \{y\}$, there exists a set $\mathscr{S}(y) \in [\![X,Y]\!]^*_{(\delta_1,\delta_2)}$ containing it. This implies that for all $[\![X|y]\!] \in [\![X|Y]\!]$, we have

$$[\![X|y]\!] \subseteq \mathscr{S}(y)^+_x. \tag{A78}$$

Hence, $\mathcal{D}$ satisfies Property 3 in Definition 6.

Thus, $\mathcal{D}$ satisfies all the three properties of $[\![X|Y]\!]^*_{\delta_1}$. This implies, along with Theorem 2, that

$$|\mathcal{D}| = |[\![X|Y]\!]^*_{\delta_1}|, \tag{A79}$$

which implies that

$$|[\![X,Y]\!]^*_{(\delta_1,\delta_2)}| = |[\![X|Y]\!]^*_{\delta_1}|. \tag{A80}$$

Hence, the statement of the theorem follows. □

## References


1. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
2. Nair, G.N. A nonstochastic information theory for communication and state estimation. *IEEE Trans. Autom. Control* **2013**, *58*, 1497–1510. [CrossRef]
3. Rosenfeld, M. On a problem of C.E. Shannon in graph theory. *Proc. Am. Math. Soc.* **1967**, *18*, 315–319. [CrossRef]
4. Shannon, C. The zero error capacity of a noisy channel. *IRE Trans. Inf. Theory* **1956**, *2*, 8–19. [CrossRef]
5. Nair, G.N. A nonstochastic information theory for feedback. In Proceedings of the 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), Maui, HI, USA, 10–13 December 2012; IEEE: Maui, HI, USA, 2012; pp. 1343–1348.
6. Tikhomirov, V.M.; Kolmogorov, A.N. $\epsilon$-entropy and $\epsilon$-capacity of sets in functional spaces. *Uspekhi Mat. Nauk* **1959**, *14*, 3–86.
7. Lim, T.J.; Franceschetti, M. Information without rolling dice. *IEEE Trans. Inf. Theory* **2017**, *63*, 1349–1363. [CrossRef]
8. Borujeny, R.R.; Kschischang, F.R. A Signal-Space Distance Measure for Nondispersive Optical Fiber. *IEEE Trans. Inf. Theory* **2021**, *67*, 5903–5921. [CrossRef]
9. Ferng, C.S.; Lin, H.T. Multi-label classification with error-correcting codes. In Proceedings of the Asian Conference on Machine Learning, Taoyuan, Taiwan, 13–15 November 2011; pp. 281–295.
10. Saberi, A.; Farokhi, F.; Nair, G. Estimation and Control over a Nonstochastic Binary Erasure Channel. *IFAC-PapersOnLine* **2018**, *51*, 265–270. [CrossRef]
11. Saberi, A.; Farokhi, F.; Nair, G.N. State Estimation via Worst-Case Erasure and Symmetric Channels with Memory. In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; IEEE: Paris, France, 2019; pp. 3072–3076.
12. Verma, G.; Swami, A. Error correcting output codes improve probability estimation and adversarial robustness of deep neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 8646–8656.
13. Weng, T.W.; Zhang, H.; Chen, P.Y.; Yi, J.; Su, D.; Gao, Y.; Hsieh, C.J.; Daniel, L. Evaluating the robustness of neural networks: An extreme value theory approach. *arXiv* **2018**, arXiv:1801.10578.
14. Wiese, M.; Johansson, K.H.; Oechtering, T.J.; Papadimitratos, P.; Sandberg, H.; Skoglund, M. Uncertain wiretap channels and secure estimation. In Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT), Barcelona, Spain, 10–15 July 2016; IEEE: Barcelona, Spain, 2016; pp. 2004–2008.
15. Agrawal, R. Sample mean based index policies with O (log n) regret for the multi-armed bandit problem. *Adv. Appl. Probab.* **1995**, 27, 1054–1078. [CrossRef]
16. Auer, P.; Cesa-Bianchi, N.; Freund, Y.; Schapire, R.E. The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* **2002**, *32*, 48–77. [CrossRef]
17. Rangi, A.; Franceschetti, M. Online learning with feedback graphs and switching costs. In Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics, Naha, Japan, 16–18 April 2019; pp. 2435–2444.

18. Rangi, A.; Franceschetti, M. Towards a non-stochastic information theory. In Proceedings of the 2019 IEEE International Symposium on Information Theory (ISIT), Paris, France, 7–12 July 2019; IEEE: Paris, France, 2019; pp. 997–1001.
19. Rangi, A.; Franceschetti, M. Channel Coding Theorems in Non-stochastic Information Theory. In Proceedings of the 2021 IEEE International Symposium on Information Theory (ISIT), Melbourne, Australia, 12–20 July 2021; IEEE: Melbourne, Australia, 2021; pp. 2295–2300.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.