Technical University of Munich
TUM School of Computation, Information and Technology

Deterministic Identification For Molecular Communications

Mohammad Javad Salariseddigh

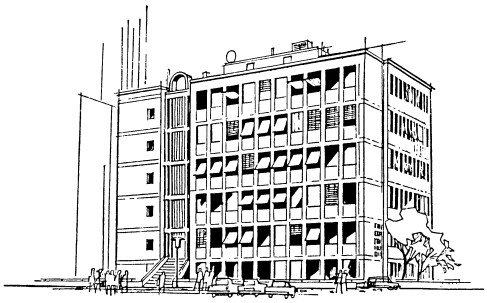
Complete reprint of the dissertation approved by the TUM School of Computation, Information and Technology of the Technical University of Munich for the award of the Doktor der Ingenieurwissenschaften (Dr.-Ing).

Chair: Prof. Dr. Holger Boche

Examiners:

1. Prof. Dr. Gerhard Kramer
2. Prof. Dr. Robert Schober

The dissertation was submitted to the Technical University of Munich on 26.06.2023 and accepted by the TUM School of Computation, Information and Technology on 28.11.2023



Declaration by Doctoral Students

(a) Authenticity of Dissertation

I hereby declare that I am the legitimate author of this Dissertation and that it is my original work. No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

(b) Research Code of Practice and Ethics Review Procedures

I declare that I have abided by the University's Research Ethics Review Procedures.

Department	Department of Electrical and Computer Engineering
School	School of Computation, Information and Technology
Institute	Institute For Communications Engineering
Degree	Doctor of Engineering
Title	Deterministic Identification For Molecular Communications
Doctoral Candidate	Mohammad Javad Salariseddigh

Submitted Date	26.06.2023
Accepted Date	28.11.2023
Examination Date	02.01.2024

Statement of The Originality

I, Mohammad Javad Salariseddigh, declare that this dissertation is my own original work unless where otherwise acknowledged and referenced. Furthermore, I acknowledge that I have clearly referenced in accordance with the standard guidelines of the Technical University of Munich requirements. I confirm that this work has been composed by me without additional assistance. The entire found results, data, figures, and other finding have not been falsified or embellished. This work has not been previously, or concurrently, used either for other university as an thesis or dissertation. Parts of the work has been published in international conferences and journals that are listed in section **extracted publications**. I appreciate that any false claim in respect of this work will result in disciplinary action in accordance with university or departmental regulations.

To Imam Jafar al-Sadiq

For His Knowledge Upon 12000 Universes

Acknowledgements

Without the assistance and direction of many people, this dissertation would not have been possible to complete. To all of them, I express my heartfelt gratitude and appreciation.

I would like to express specific gratitude to my supervisor; **Prof. Dr. Sc. Tech. Gerhard Kramer** for his support, patience, enthusiasm and the knowledge that I was able to acquire. I am also grateful to my mentor; **Dr. Math. Christian Deppe**, at Institute for Communications Engineering (ICE), for his support, and flexibility. My supervisor and mentor provided me excellent atmosphere at the ICE and were patient and flexible during my hard times for research. Their guidance, motivation and constructive feedback helped me throughout this challenging work and enabled me to improve my capacities. My doctoral studies at ICE was a great opportunity to go through the world of information theory and learn various techniques in the course of addressing technical problems. I extend my gratitude to my wonderful colleagues at ICE for their support and the excellent research environment.

I would like to greatly thank the members of my doctoral committee, for their time, feedback, and interest in my work. I appreciate **Prof. Dr. Robert Schober** from Friedrich-Alexander-Universität Erlangen-Nürnberg for serving as the second examiner of my dissertation, and **Prof. Dr. Holger Boche** for chairing my doctoral examination committee.

In the course of my doctoral studies, I had the pleasure of collaborating with many inspiring scientists on my joint research projects. I express my gratitude to my co-authors **Prof. Dr. Uzi Pereg**, **Prof. Dr. Vahid Jamali**, **Prof. Dr. Holger Boche**, and **Prof. Dr. Robert Schober**.

Finally, I would like to express my profound gratitude to my family for providing me with kind support and continuous encouragement throughout my research. This accomplishment would not have been possible without them.

Abstract / Zusammenfassung

Several applications of molecular communications (MC) feature an event trigger behavior for which the prevalent Shannon capacity may not be the appropriate measure for performance assessment. Thus, we motivate and establish the identification capacity as an alternative metric for such systems. In particular, within the context of MC systems, the Poisson and Binomial channel serves as a fundamental model for the MC systems employing molecule-counting receivers. We addressed the deterministic identification (DI) for the discrete-time Poisson and Binomial channels (DTPC & DTBC), subject to an average and a peak constraint on the molecule release rate. It is established that the number of different messages that can be reliably identified for the DTPC and DTBC scales as $2^{(n\log n)R}$, where n and R are the codeword length and coding rate, respectively. Lower and upper bounds on the DI capacity of the DTPC and DTBC are developed.

In addition, we study the DI for the DTPC with inter-symbol interference (ISI) where the transmitter is restricted to an average and a peak molecule release rate constraint. Such a channel serves as a model for diffusive MC systems featuring long channel impulse responses and employing molecule counting receivers. We derive lower and upper bounds on the DI capacity of the DTPC with ISI when the number of ISI channel taps K may grow with the codeword length n (e.g., due to increasing symbol rate). As a key finding, we establish that for deterministic encoding, the codebook size scales as $2^{(n\log n)R}$ assuming that the number of ISI channel taps scales as $K = 2^{\kappa \log n}$, where R is the coding rate and κ is the ISI rate.

Moreover, we determine bounds on the DI capacity of Gaussian channel with slow and fast fading subject to average power constraints. It is found that the correct size of the codebook for fading channels scale super exponentially in the codeword length, i.e., $\sim 2^{(n\log n)R}$. Furthermore, we fill a long standing gap in information theory by establishing full characterization of the DI capacity for discrete memoryless channel subject to average power constraint. In addition, generalized scheme of DI problem called deterministic K-identification (DKI) for the the binary symmetric (BSC) and Gaussian channel with slow fading (GSF) are developed. Specifically, we establish full characterization of the DKI capacity for the BSC subject to a Hamming weight constraint and also obtain bounds on the DKI capacity for the GSF with average power constraint.

Abstract / Zusammenfassung

Mehrere Anwendungen der molekularen Kommunikation (MK) zeichnen sich durch ein ereignisauslösendes Verhalten aus, für das die vorherrschende Shannon-Kapazität möglicherweise nicht das geeignete Maß für die Leistungsbewertung ist. Daher motivieren und etablieren wir die Identifikationskapazität als alternative Metrik für solche Systeme. Insbesondere im Kontext von MC-Systemen dient der Poisson- und Binomialkanal als grundlegendes Modell für MC-Systeme, die molekülzählende Empfänger verwenden. Wir befassten uns mit der deterministischen Identifikation (DI) für die zeitdiskreten Poisson- und Binomialkanäle (ZDPK & ZDBK), vorbehaltlich einer Durchschnitts- und einer Spitzenbeschränkung der Molekülfreisetzungsrate. Es wurde festgestellt, dass die Anzahl verschiedener Nachrichten, die für die ZDPK- und ZDBK zuverlässig identifiziert werden können, $2^{(n\log n)R}$ beträgt, wobei n und R die Codewortlänge und die Codierungsrate sind. jeweils. Es werden Unter- und Obergrenzen für die DI-Kapazität von ZDPK und ZDBK entwickelt.

Darüber hinaus untersuchen wir den DI für den ZDPK mit Intersymbolinterferenz (ISI), bei dem der Sender auf eine durchschnittliche und eine maximale Molekülfreisetzungsratenbeschränkung beschränkt ist. Ein solcher Kanal dient als Modell für diffusive MC-Systeme mit langen Kanalimpulsantworten und dem Einsatz von Molekülzählempfängern. Wir leiten Unter- und Obergrenzen für die DI-Kapazität des ZDPK mit ISI ab, wenn die Anzahl der ISI-Kanalabgriffe K mit der Codewortlänge n wachsen kann (z. B. aufgrund einer zunehmenden Symbolrate). Als wichtigste Erkenntnis stellen wir fest, dass die Codebuchgröße für die deterministische Codierung $2^{(n\log n)R}$ beträgt, vorausgesetzt, dass die Anzahl der ISI-Kanalabgriffe $K = 2^{\kappa\log n}$, wobei R die Kodierungsrate und κ die ISI-Rate ist.

Darüber hinaus bestimmen wir Grenzen für die DI-Kapazität des Gaußschen Kanals mit langsamem und schnellem Fading vorbehaltlich durchschnittlicher Leistungsbeschränkungen. Es wurde festgestellt, dass die korrekte Größe des Codebuchs für Fading-Kanäle superexponentiell in der Codewortlänge skaliert, d. h. $\sim 2^{(n\log n)R}$. Darüber hinaus schließen wir eine seit langem bestehende Lücke in der Informationstheorie, indem wir eine vollständige Charakterisierung der DI-Kapazität für diskrete speicherlose Kanäle unter durchschnittlicher Leistungsbeschränkung erstellen. Darüber hinaus wird ein verallgemeinertes Schema des DI-Problems namens deterministische K-Identifikation (DKI) für den binärsymmetrischen (BSK) und den Gaußschen Kanal mit langsamem Fading (GLF) entwickelt. Insbesondere erstellen wir eine vollständige Charakterisierung der DKI-Kapazität für den BSK, der einer Hamming-Gewichtungsbeschränkung unterliegt, und erhalten außerdem Grenzen für die DKI-Kapazität für den GSF mit einer durchschnittlichen Leistungsbeschränkung.

TABLE OF CONTENTS

LIST OF FIGURES

EXTRACTED ARXIV PREPRINTS

[1] **M. J. Salariseddigh**, Uzi Pereg, Holger Boche, and Christian Deppe, Deterministic Identification Over Channels With Power Constraints, arXiv:2010.04239, 2021, arXiv Link

[2] **M. J. Salariseddigh**, Uzi Pereg, Holger Boche, Christian Deppe, and Robert Schober, Deterministic Identification Over Fading Channels, arXiv:2010.10010, 2020, arXiv Link

[3] **M. J. Salariseddigh**, Uzi Pereg, Holger Boche, Christian Deppe, and Robert Schober, Deterministic Identification Over Poisson Channels, arXiv:2107.06061, 2021, arXiv Link

[4] **M. J. Salariseddigh**, Uzi Pereg, Holger Boche, Christian Deppe, and Robert Schober, Deterministic Identification For Molecular Communications Over The Poisson Channel, arXiv:2203.02784, 2022, arXiv Link

[5] **M. J. Salariseddigh**, Uzi Pereg, Holger Boche, Christian Deppe, and Robert Schober, Deterministic Identification For MC ISI-Poisson Channel, arXiv:2211.11024, 2022, arXiv Link

[6] **M. J. Salariseddigh**, Muris Spahovic, and Christian Deppe, Deterministic K-Identification For Slow Fading Channel, arXiv:2212.02732, 2022, arXiv Link

[7] **M. J. Salariseddigh**, V. Jamali, Holger Boche, Christian Deppe, and Robert Schober, Deterministic Identification For MC Binomial Channel, arXiv:2304.12493, 2023, arXiv Link

[8] O. Dabbabi, **M. J. Salariseddigh**, Christian Deppe, and Holger Boche, Deterministic K-Identification For Binary Symmetric Channel, arXiv:2305.04260, 2023, arXiv Link

EXTRACTED JOURNALS PUBLICATIONS

[1] **M. J. Salariseddigh**, U. Pereg, H. Boche and C. Deppe, Deterministic Identification Over Channels With Power Constraints, Published in IEEE Transactions on Information Theory, vol. 68, no. 1, pp. 1-24, Jan. 2022, doi: 10.1109/TIT.2021.3122811, IEEE Link

[2] **M. J. Salariseddigh**, V. Jamali, U. Pereg, H. Boche, C. Deppe, and R. Schober, Deterministic Identification For Molecular Communications Over The Poisson Channel, Published in Transactions on Molecular, Biological and Multi-Scale Communications, vol. 9, no. 4, pp. 408-424, Dec. 2023, doi: 10.1109/TMBMC.2023.3324487, IEEE Link

[3] **M. J. Salariseddigh**, V. Jamali, U. Pereg, H. Boche and C. Deppe, and R. Schober, Deterministic Identification For MC Poisson Channel With Inter-symbol Interference, Published in IEEE Open Journal of the Communications Society 2023, vol. 5, pp. 1101-1122, Jan. 2024, doi: 10.1109/OJCOMS.2024.3359186, IEEE Link

[4] **M. J. Salariseddigh**, O. Dabbabi, C. Deppe and H. Boche, Deterministic K–Identification For Future Communication Networks; The Binary Symmetric Channel Results, Published in MDPI Future Internet Journal 2023, vol. 16, issue. 3, no. 78, Feb. 2024, doi: 10.3390/fi16030078, MDPI Link

[5] **M. J. Salariseddigh**, V. Jamali, H. Boche and C. Deppe, and R. Schober, Deterministic K–Identification For Binomial Channel, In Preparation For Submission to IEEE Transactions on Information Theory 2024.

[6] **M. J. Salariseddigh**, U. Pereg, C. Deppe, and H. Boche, Deterministic Identification For Broadcast Channel, In Preparation For Submission to IEEE Transactions on Information Theory 2024.

EXTRACTED CONFERENCES PUBLICATIONS

[1] **M. J. Salariseddigh**, U. Pereg, H. Boche and C. Deppe, Deterministic Identification Over Channels With Power Constraints, Published in ICC 2021 - IEEE International Conference on Communications, Montreal, QC, Canada, 2021, pp. 1-6, doi: 10.1109/ICC42927.2021.9500406, IEEE Link

[2] **M. J. Salariseddigh**, U. Pereg, H. Boche and C. Deppe, Deterministic Identification Over Fading Channels, Published in 2020 IEEE Information Theory Workshop (ITW), Riva del Garda, Italy, 2021, pp. 1-5, doi: 10.1109/ITW46852.2021.9457587, IEEE Link

[3] **M. J. Salariseddigh**, U. Pereg, H. Boche, C. Deppe and R. Schober, Deterministic Identification Over Poisson Channels, Published in 2021 IEEE Globecom Workshops (GC Wkshps), Madrid, Spain, 2021, pp. 1-6, doi: 10.1109/GCWkshps52748.2021.9682110, IEEE Link

[4] **M. J. Salariseddigh**, U. Pereg, H. Boche, C. Deppe and R. Schober, Deterministic Identification For MC ISI-Poisson Channel, Published in ICC 2023 - IEEE International Conference on Communications, Rome, Italy, 2023, pp. 6108-6113, doi: 10.1109/ICC45041.2023.10278856, IEEE Link

[5] **M. J. Salariseddigh**, Muris Spahovic, and Christian Deppe, Deterministic K-Identification For Slow Fading Channel, Published in 2023 IEEE Information Theory Workshop (ITW), Saint-Malo, France, 2023, pp. 353-358, doi: 10.1109/ITW55543.2023.10161643, IEEE Link

[6] **M. J. Salariseddigh**, V. Jamali, Holger Boche, Christian Deppe, and Robert Schober, Deterministic Identification For MC Binomial Channel, Published in 2023 IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, 2023, pp. 448-453, doi: 10.1109/ISIT54713.2023.10206627, IEEE Link

[7] O. Dabbabi, **M. J. Salariseddigh**, Christian Deppe, and Holger Boche, Deterministic K-Identification For Binary Symmetric Channel, Published in 2023 IEEE Globecom Workshops (GC Wkshps), IEEE Link

CHAPTER 1

INTRODUCTION

“ A Journey of a Thousand Miles Begins With a Single Step. ”

Laozi,

1.1 | Molecular Communication

Molecular communication (MC) is a new communication strategy where information carriers are signaling molecules [1–3]. The MC is deemed as a bio-inspired promising paradigm for communication between nanomachines or different biological entities, such as cells and organs [4] and realizes the exchange of information via the transmission, propagation, and reception of signaling molecules [1, 3]. Over the past decade, synthetic MC has been investigated in a number of different directions including channel modeling [5, 6], modulation and detection design [7], biological building blocks for transceiver design [8], and information-theoretical performance characterization and relevant mathematical foundations [9–11]. Furthermore, several proof-of-concept implementations for synthetic MC systems have been reported in the literature, see, e.g., [12–14]. Furthermore, the ongoing progress in synthetic biology [8, 15] is expected to enable sophisticated MC systems in the future, capable of performing the complex computation and communication tasks required for the realization of the Internet of Bio-nano Things [16–19]. Also, the authentication problem [20] which exhibit affinity to the identification problem is considered in [21].

Recently, there have been significant advances in molecular communication for complex nano-networks. The interconnection of nanothings with the Internet is known as the Internet of NanoThings (IoNT) and is the basis for various future healthcare and military applications [22]. Furthermore, the concept of the Internet of Bio-NanoThings (IoBNT) has been introduced in [16], where nanothings are biological cells that are cre-

ated using tools from synthetic biology and nanotechnology. For the communication between cells, molecular communication is well suited, since the natural exchange of information between cells is already based on this paradigm. Molecular communication in cells is based on signal pathways (chains of chemical reactions) that process information that is modulated into chemical characteristics, such as molecule concentration.

1.2 | Post Shannon Communication

The identification problem [23] can be regarded as a *Post Shannon* [24] model where the decoder does not perform an *reproduction* of the original message, but rather a binary hypothesis test to decide between the hypotheses 'sent' or 'not sent', based on the observation of the channel output. As the sender has no knowledge of the desired message that the receiver is interested in, the identification problem can be regarded as a test of many hypotheses occurring simultaneously. The scenario where the receiver misses and does not identify his message is called a type I error, or 'missed identification', whereas the event where the receiver accepts a false message is called a type II error, or 'false identification'.

In particular, for object-finding or event-detection scenarios, where the receiver aims to determine the presence of an object or determine the occurrence of an specific event in terms of a reliable Yes / No answer, the so-called identification capacity is the key applicable performance measure [23].

1.3 | XG Wireless Networks

Several applications in the context of Post Shannon communications [25–27] for the future-generation wireless networks (XG) horizon are either based on or give rise to the event-triggered communication settings. Further discussions of the potentials of MC and Post Shannon communication for the 6G can be found in [25]. In such systems, Shannon's message transmission capacity, as studied early by Shannon in [28] and by others in [9–11, 29–37], may not be the appropriate performance metric, instead, the identification capacity is regarded to be a key quantitative measure. In particular, for event-recognition, alarm-prompt or object-finding problems, where the receiver aims to recognize the occurrence of a specific event, determine an alarm, or realize the presence

of an object in terms of a *reliable* Yes / No final decision, the so-called identification capacity is the key applicable performance measure [23].

1.4 | Deterministic Identification

While in the Shannon's communication paradigm [28], sender, encodes its message in a manner such that the receiver can perform a reliable *reproduction*, in the identification setting [23, 38], the coding scheme is designed to accomplish a different objective, namely, to determine whether a *particular* message was sent or not. The identification problem in communication theory is initiated[1] by Ahlswede and Dueck [23] featuring randomization, that is, a randomized encoder is employed, to select the codewords. Therein, the codebook consists of distributions and as a salient property[2], it was established that providing local randomness at the encoder, reliable identification is yielded a remarkable attribute regarding the codebook size, namely, the codebook size exhibit a double-exponentially growth in the codeword length n, i.e., $\sim 2^{2^{nR}}$ [23], where R is the coding rate. This observation is extremely different from the conventional message transmission problem, which has an exponential codebook size in the codeword length, i.e., $\sim 2^{nR}$. The realization of explicitly constructed randomized identification (RI) codes entails extra complexity and is challenging for the applications; cf. [51, see Sec. 1] for further details. The motivation of Ahlswede and Dueck to develop the RI problem [23] is traced back to the work of JáJá [52] who considered deterministic identification (DI)[3], from a communication complexity[4] perspective, that is, where the codewords are determined by a deterministic function from the messages. Further, it

[1] The identification problem has been studied in various setting of deterministic or randomized protocols, in the context of communication complexity; see [39–42]

[2] It is known that employing such resource (distributions) does not bring advantage in terms of a gain in the Shannon's message transmission (TR) capacity [43] or codebook size for the DMCs [43]. Beyond the exponential gain on the codebook size in the RI, the extension of the problem to more advanced scenarios reveals that the RI capacity stands different compared to the TR capacity [44–49]. For instance, the feedback *can* increase the RI capacity [44] of a memoryless channel, as opposed to the TR capacity [50].

[3] The DI capacity in the literature is also referred to as the non-randomized identification (NRI) capacity [53], the dID capacity [46], or also the identification without randomization [53].

[4] An important observation regarding the behavior of the identification function has been well studied in communication complexity where the out-performance of randomized protocols over the deterministic protocols (exponential gap between the two class) for computing such function is established. For instance, while the error-free deterministic complexity of the identification function is lower bounded by $\log m$, where m is the length of message, for the randomized protocol and when ε error is allowed in computation of the identification function, only $\mathcal{O}(\log\log m + \frac{1}{\varepsilon})$ bits suffices; see [39, 54] for further details.

seems that Ahlswede and Dueck were inspired to show that employing randomness similar to what has been accomplished in the communication complexity field, yield an advantage of exponential gap over the DI problem[5] in terms of the codebook size. DI may be favored over the randomized identification (RI) [23] in the context of the complexity-constrained applications. For instance, in the MC systems where development and deploying of a huge number of random number generators may not be clearly established. The construction of RI codes is considered in [55–61]. RI for the Gaussian channels is studied in [46, 62–65]. Deterministic codes often benefit the advantage of simpler implementation and simulation [66, 67], explicit construction [68], and single-block reliable performance. DI for the compound channels is studied by Ahlswede and Cai in [53]. In [69] the DI problem with non-discrete additive white noise and noiseless feedback under both average and peak power constraints, is analyzed, where the DI capacity is shown to be infinite regardless of the scaling for the codebook size.

1.4.1 | Applications

Motivated by this discussion, in this paper, we investigate the fundamental performance limits of identification problem in MC systems, which can be modelled by the discrete time Poisson channel (DTPC) with inter-symbol interference (ISI).

- **Molecular Communications:** It is not clear how the RI codes can be incorporated into the MC systems. It is unclear how much power is required for the random number generators in the synthetic materials on an extremely small scales in the range of micro / nano meters. In the case of Bio-NanoThings, it is uncertain whether natural biological processes can be controlled or reinforced by the local randomness at this level. Therefore, for the design of synthetic IoNT, or for the analysis and utilization of IoBNT, identification with deterministic encoding seems to be a more appropriate and applicable candidate. Concrete examples of the identification problem within the MC context include health monitoring [4, 70, 71], where, e.g., one may desire in whether or not the pH value of the cerebrospinal fluid of brain exceeds a crisis level; targeted drug delivery [4, 72] and cancer treatment [70, 73–75], where, e.g., a nano-device's purpose is to identify whether or not an specific cancer biomarker exist in the vicinity of the target

[5] A detailed comparison of codebook sizes in DI and RI problem over various channel models can be found in [51]

tissue, whether a specific drug is released or not, whether another nano-device has replicated itself, whether a certain molecule was detected, whether a target location in the vessels is identified, or whether the molecular storage is empty, etc. Moreover, identification problems can also be found in various natural MC systems. For instance, in bionic nose setting [76], or in natural pheromone communications [77, 78] where, e.g., animals involved in mating seeks sexual pheromones to realize the presence of an opposite sex. In fact, the olfactory systems of animals have the capability of *recognizing* the presence of extremely large numbers of different molecule mixtures (e.g., pheromones, odors, etc.) [79, 80], which has inspired researchers to regard them as role models for the design of bio-inspired synthetic MC systems [81].

- □ **Vehicle-to-X Communications:** A second application for the identification scheme is the vehicle-to-X communications, where a vehicle that collects sensor data may ask whether a certain alert message concerning the future movement of an adjacent vehicle was transmitted or not [82, Sec. VII].

- □ **Other Fields:** Identification find as well overlapping applications with a number other fields, including identification plus transmission (point-to-multi-point communication) [83], communication complexity [84], private interrogation theory [85], the tactile internet [86], vehicle-to-X communications [87, 88], digital watermarking [89–91], online sales [92, 93], industry 4.0 [94–96], health care [97], and other event-triggered systems.

1.5 | Main Contributions

Motivated by the mentioned applications, we establish the fundamental performance limits of DI and DKI for various channel models. As our main objective, a full characterization or capacity bounds are determined. In particular, we make the following contributions:

- ◊ **Discrete Memoryless Channel**: The DI problem for discrete memoryless channel (DMC) where codewords are restricted by an average power constraint, is studied in [98, 99]. Therein, employing the *method of types*[6] and standard techniques,

[6] The method of type developed and promoted by Csiszar and Körner and treated in depth in [100, 101].

it is established that the codebook size grows exponentially as a function of the codeword length, i.e., $\sim 2^{nR}$ [98, 99]. This result is early reported without a complete proof in [23, 53]. This observation acknowledge that the codebook size of DI over DMCs behaves similar to that of the message transmission problem [28], however, the achievable identification rates are significantly higher compared to the transmission rates [98, 99].

◊ **Gaussian Channels With Fading**: In [99, 105, 106], the DI problem for Gaussian channels with slow and fast fading subject to the average power constraint is addressed. A generalization of DI problem, called deterministic K-identification (DKI) for the Gaussian channel with slow fading is studied in [107] where K-depending bounds on the DI capacity are established.

◊ **Discrete Time Poisson Channel**: In [108, 109], the discrete time Poisson channel (DTPC) is studied and bounds on the DI capacity with a codebook size of super exponentially large in the codeword length, are derived.

◊ **Discrete Memoryless Channel With Memory**: Inter-symbol interference (ISI)-aware discrete time Poisson channel (DTPC) is addressed in [110] where ISI-dependent bounds on the DI capacity are calculated.

◊ **Discrete Time Binomial Channel**: We derive lower and upper bounds on the DI capacity of a discrete time Binomial channel (DTBC) in [111].

◊ **Binary Symmetric Channel**: Generalized DI model called K-Identification for the binary symmetric channel (BSC) with and without the Hamming weight constraints is developed and a full characterization of the deterministic K-identification (DKI) is established in [112].

For all the continuous alphabet works, i.e., the Gaussian, Poisson (with/out ISI), and Binomial models [99, 105, 108–110], a new observation regarding the codebook size is obtained, namely, the codebook size scales *super-exponentially* in the codeword length, i.e., $\sim 2^{(n \log n)R}$ which is different than the standard exponential [98] and double exponential [23] behavior for DI and RI problems, respectively.

Such a method is regarded as a widely used and fundamental technique to obtain capacity results for different source and channel coding settings within the context of mathematical information theory; see [102] for further details with related topics on multi-user models. A survey of recent developments with applications in statistics can found in [103]. Further examples and in-depth mathematical details with applications to large deviation theory can be found in [104, Ch. 13].

1.6 | Organizations

The remainder of this dissertation is structured as follows. In Chapter 2, results for the DMC are introduced and a full characterization on the DI capacity is established. Chapter 3 provides the main contributions and results on the DI capacity of the standard Gaussian channel without fading. Chapter 4 and Chapter 5 presents the capacity results for the Gaussian channels with slow and fast fading, respectively. Chapter 6 brings forward the results for the DI capacity of the DTPC. In Chapter 7, extended results including lower and upper bounds on the DI capacity of the DTPC with inter-symbol interference (ISI) are obtained. Chapter 8 represents the results for the discrete time Binomial channel (DTBC). In the last two chapters, i.e., Chapter 9 and Chapter 10, generalized DI scheme are developed and bounds and closed form expressions on the DKI capacity are established. Finally, Chapter 11 concludes with a summary and possible directions for the future research.

▷ In the following chapters, we follow the notations provided in Appendix A

CHAPTER 2

DI FOR DISCRETE MEMORYLESS CHANNEL

> *All Models Are Wrong, But Some Are Useful*
>
> George E. P. Box,

2.1 | Introduction

Ahlswede and Dueck [23] required randomized coding for their identification coding scheme. This means that a randomized source is available to the sender. The sender can make his encoding dependent on the output of this source. It is known that this resource cannot be used to increase the transmission capacity of discrete memoryless channels [43]. A remarkable result of identification theory is that given local randomness at the encoder, reliable identification can be attained such that the code size, i.e., the number of messages, grows double exponentially in the block length n, i.e., $\sim 2^{2^{nR}}$ [23]. This differs sharply from the traditional transmission setting where the code size scales only exponentially, i.e., $\sim 2^{nR}$. Beyond the exponential gain in identification, the extension of the problem to more complex scenarios reveals that the identification capacity has a very different behavior compared to the transmission capacity [44–49]. For instance, feedback *can* increase the identification capacity [44] of a memoryless channel, as opposed to the transmission capacity [50]. Nevertheless, it is difficult to implement randomized-encoder identification (RI) codes that will achieve such performance, because it requires the encoder to process a bit string of exponential length. The construction of identification codes is considered in [55–59]. Identification for Gaussian channels is considered in [46, 62, 63, 65].

In the deterministic setup for a DMC, the number of messages scales exponentially in the blocklength [23, 52, 53, 113], as in the traditional setting of transmission. Nevertheless, the achievable identification rates are significantly higher than those of trans-

mission. In addition, deterministic codes often have the advantage of simpler implementation and simulation [67], explicit construction [68], and single-block reliable performance. In particular, JáJá [52] showed that the deterministic identification (DI) capacity[1] of a binary symmetric channel is lower bounded (achievability) by 1 bit per channel use. Ahlswede et al. [23, 53] stated that the DI capacity of a discrete memoryless channel (DMC) with a stochastic matrix W is given by the logarithm of the number of distinct row vectors of W (see Section IV. in [23] and abstract of [53]). Nonetheless, an explicit proof for this result was not provided in [23, 53]. Instead, Ahlswede and Cai [53] referred the reader to a paper [114] which does not include identification and addresses a completely different model of an arbitrarily varying channel [114]. Since then, the problem of proving this result has remained unsolved, since a straightforward extension of the methods in [114], using decoding territories, does not seem to yield the desired result on the DI capacity [115].

In this chapter, we establish the DI capacity of channels subject to an input constraint. Such a constraint is often associated with a limited power supply or regulation, as in the case of the Gaussian channel. We consider the settings of a DMC and of Gaussian channels with fast fading and slow fading, with CSI available at the decoder. For a DMC, one may assume without loss of generality that the rows of the channel matrix are distinct (see Remarks 2.3.1 and 2.3.3). Our first result is that the DI capacity of a DMC $\mathcal{W}$ under this assumption, subject to the input constraint $\frac{1}{n}\sum_{t=1}^{n}\phi(x_t) \leq A$, is given by

$$C_{DI}(\mathcal{W}) = \max_{p_X \,:\, \mathbb{E}\{\phi(X)\} \leq A} H(X)\,. \tag{2.1}$$

We note that the DI capacity does not depend on the specific values of the transition probabilities, as long as the rows of the channel matrix are distinct. This result has the following geometric interpretation. At first glance, it may seem reasonable that for the purpose of identification, one codeword could represent two messages. While identification allows overlap between decoding regions [116, 117], it turns out that overlap at the encoder is not allowed for deterministic codes. However, if two messages are represented by the same codeword, then the low probability of a type I error comes at the expense of the high probability of a type II error, and vice versa. Thus, DI coding

[1] The DI capacity in the literature is also referred to as the non-randomized identification (NRI) capacity [53] or the dID capacity [46].

imposes the restriction that every message must have a distinct codeword. The converse proof follows from this property in a straightforward manner, since the volume of the input subset of sequences that satisfy the input constraint is $\approx 2^{nC_{DI}(\mathcal{W})}$. A similar principle guides the direct part as well. The input space is covered such that each codeword is surrounded by a sphere of radius $n\epsilon$ to separate the codewords.

By providing a detailed proof for the DMC, we thus fill the gap in the previous analysis [23, 53] as well. In the proof, we use the method of types, while the derivation is based on ideas that are analogous to the combinatoric analysis of Hamming distances by JáJá [52]. Although the codebook construction is similar to that of Ahlswede's coding scheme [114], the decoder is significantly different. In particular, we do *not* use decoding territories as in [114], but rather perform a typicality check. Nonetheless, the type-class intersection lemma and the message-set analysis in [114] turn out to be useful in our analysis as well. Hence, our proof combines techniques and ideas from both works, by JáJá [52] and by Ahlswede [114], to derive the DI capacity both with and without an input constraint. The analysis for Gaussian channels also relies on geometric considerations, using sphere packing. Based on fundamental properties of packing arrangements [118], the optimal packing of non-overlapping spheres of radius $\sqrt{n\epsilon}$ contains an exponential number of spheres, and by decreasing the radius of the codeword spheres, the exponential rate can be made arbitrarily large. However, in the derivation of our lower bound in the $2^{n\log(n)R}$-scale, we pack spheres of a sub-linear radius $\sqrt{n\epsilon_n} \sim n^{1/4}$, which results in $\sim 2^{\frac{1}{4}n\log(n)}$ codewords.

This chapter is organized as follows: In Section 2.2 we give the definitions and a brief review of related work. In Section 2.3 we address deterministic identification for the DMC with and without an input constraint. In Subsection 2.3.1, a channel reduction procedure is described such that high-cost identical rows are removed from the channel matrix. The capacity theorem is stated in Subsection 2.3.2. The direct part is proved in Subsection 2.3.3, and the converse proof is in Subsection 2.3.4. Section 2.4 is dedicated to summary and discussion.

2.2 | Definitions and Related Works

In this section we introduce the channel model and coding definitions. Here we only consider the discrete memoryless channel (DMC). The channel description and coding definition for the Gaussian channel will be presented in Section 5.0.1.

2.2.1 | Channel Description

A DMC $(\mathcal{X}, \mathcal{Y}, W)$ consists of finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, respectively, and a conditional pmf $W(y|x)$. The channel is memoryless without feedback, and therefore $W^n(y^n|x^n) = \prod_{t=1}^n W(y_t|x_t)$. We denote a DMC by $\mathcal{W} = (\mathcal{X}, \mathcal{Y}, W)$. Next, we consider an input constraint. Let $\phi : \mathcal{X} \to [0, \infty)$ be some given bounded cost function, and define

$$\phi^n(x^n) = \frac{1}{n} \sum_{t=1}^{n} \phi(x_t) \, . \tag{2.2}$$

Given an input constraint $A > 0$ corresponding to the cost function $\phi^n(x^n)$, the channel input x^n must satisfy

$$\phi^n(x^n) \leq A \, . \tag{2.3}$$

We may assume without loss of generality that $0 \leq A \leq \phi_{\max}$, where $\phi_{\max} = \max_{x \in \mathcal{X}} \phi(x)$. It is also assumed that for some $x_0 \in \mathcal{X}$, $\phi(x_0) = 0$.

2.2.2 | Coding

The definitions for DI codes, achievable rates, and capacity are given below. In this chapter we consider codes with different size orders. For instance when we discuss the exponential scale, we refer to a code size that scales as $L(n, R) = 2^{nR}$. On the other hand, in the double exponential scale, the code size is $L(n, R) = 2^{2^{nR}}$. Later, in Section 5.0.1 where we consider Gaussian channels, we will see that the appropriate scale turns out to be neither exponential nor double exponential, but in between. Throughout the paper we use the mathematical convention that $L : \mathbb{N} \times \mathbb{R}^+ \to \mathbb{N}$ denotes a map, and $L(n, R) \in \mathbb{N}$ is its value for a given blocklength n and rate R.

Definition 2.2.1. *Let $L_1(n, R)$ and $L_2(n, R)$ be two coding scales. We say that L_1* dominates *L_2 if*

$$\lim_{n \to \infty} \frac{L_2(n, b)}{L_1(n, a)} = 0 \, , \tag{2.4}$$

for all $a, b > 0$. We will denote this relation by $L_2 \prec L_1$.

In complexity theory of computer science, the relation above is denoted by the 'small o-notation', $L_2(n, 1) = o(L_1(n, 1))$ [119]. Beyond exponential, other orders that commonly appear in complexity theory are the linear, logarithmic, and polynomial scales,

nR, $\log(nR)$, and $(nR)^k$. The corresponding ordering is

$$\log(nR) \prec nR \prec (nR)^k \prec 2^{nR} \prec 2^{n\log(n)R} \prec 2^{2^{nR}} \,. \tag{2.5}$$

Definition 2.2.2 (DMC DI Code). *An $(L(n,R),n)$ DI code for a DMC $\mathcal{W}$ under input constraint A, assuming $L(n,R)$ is an integer, is defined as a system $(\mathcal{U},\mathscr{D})$ that consists of a codebook $\mathcal{U} = \{u_i\}_{i\in[\![L(n,R)]\!]}$, $\mathcal{U} \subset \mathcal{X}^n$, such that*

$$\phi^n(u_i) \leq A \,, \text{ for all } i \in [\![L(n,R)]\!] \,, \tag{2.6}$$

and a collection of decoding regions $\mathscr{D} = \{\mathcal{D}_i\}_{i\in[\![L(n,R)]\!]}$ with $\bigcup_{i=1}^{L(n,R)} \mathcal{D}_i \subset \mathcal{Y}^n$. Given a message $i \in [\![L(n,R)]\!]$, the encoder transmits u_i. The decoder's aim is to answer the following question: Was a desired message j sent or not? Two types of errors may occur: Rejecting of the true message, or accepting a false message. Those error events are often referred to as type I and type II errors, respectively. Specifically, $P_{e,1}^{(n)}(i)$ is the type I error probability for rejecting the true message i, while $P_{e,2}^{(n)}(i,j)$ is the type II error probability for accepting the false message j, given that the message i was sent.

The error probabilities of the identification code $(\mathcal{U},\mathscr{D})$ are given by

$$P_{e,1}(i) = W^n(\mathcal{D}_i^c|u_i) \;\; \textit{(missed-identification error)}, \tag{2.7}$$

$$P_{e,2}(i,j) = W^n(\mathcal{D}_j|u_i) \;\; \textit{(false identification error)}. \tag{2.8}$$

An $(L(n,R),n,\lambda_1,\lambda_2)$ DI code further satisfies

$$P_{e,1}(i) \leq \lambda_1 \,, \tag{2.9}$$

$$P_{e,2}(i,j) \leq \lambda_2 \,, \tag{2.10}$$

for all $i,j \in [\![L(n,R)]\!]$ such that $i \neq j$.

A rate $R > 0$ is called achievable if for every $\lambda_1,\lambda_2 > 0$ and sufficiently large n there exists an $(L(n,R),n,\lambda_1,\lambda_2)$ DI code. The operational DI capacity is defined as the supremum of achievable rates and will be denoted by $\mathbb{C}_{DI}(\mathcal{W},L)$.

As mentioned earlier, Ahlswede and Dueck [23] needed randomized encoding for their identification-coding scheme. This means that a randomized source is available to the sender. The sender can make his encoding dependent on the output of this source. Therefore, a randomized-encoder identification (RI) code is defined in a similar manner where the encoder is allowed to select a codeword U_i at random, according to some

conditional input distribution $Q(x^n|i)$. The RI capacity is then denoted by $\mathbb{C}_{RI}(\mathcal{W}, L)$. Given local randomness at the encoder, reliable identification can be attained such that the number of messages grows double exponentially in the block length n, i.e., $L(n,R) = L_{\text{double}}(n,R) \triangleq 2^{2^{nR}}$ [23]. This differs sharply from the traditional transmission setting where the code size scales only exponentially, i.e., $L(n,R) = L_{\exp}(n,R) \triangleq 2^{nR}$. Remarkably, in [23] it was shown that $\mathbb{C}_{RI}(\mathcal{W}, L_{\text{double}}) = \mathbb{C}_T(\mathcal{W}, L_{\exp})$, where $\mathbb{C}_T(\mathcal{W}, L_{\exp})$ denotes the transmission capacity of the channel in the exponential scale.

Remark 2.2.1. *The code scale can also be thought of as a sequence of monotonically increasing functions $L_n(R)$ of the rate. Hence, given a code of size $M = L_n(R)$, the coding rate can be obtained from the inverse relation $R = L_n^{-1}(M)$. In particular, for the transmission setting [28], or DI coding for a DMC [52], the coding rate is defined as*

$$R = \frac{1}{n}\log(M)\,. \tag{2.11}$$

Whereas for RI coding [23], the rate was defined as

$$R = \frac{1}{n}\log\log(M)\,. \tag{2.12}$$

On the other hand, using the scale $L(n,R) = 2^{n\log(n)R}$ as for Gaussian channels stated in Theorem 5.0.1, the coding rate is

$$R = \frac{\log M}{n\log n}\,. \tag{2.13}$$

Remark 2.2.2. *It can be readily shown that in general, if the capacity in an exponential scale is finite, then it is zero in the double exponential scale. Conversely, if the capacity in a double exponential scale is positive, then the capacity in the exponential scale is $+\infty$. This principle can be generalized to any pair of scales L_1 and L_2, where L_2 is dominated by L_1. We come back to this in Subsection 5.0.4.*

A geometric illustration for the type I and II error probabilities is given in Figure 2.1. When the encoder sends the message i but the channel output is outside $\mathcal{D}_i$, then type I error occurs. This kind of error is also considered in traditional transmission. In identification, the decoding sets can overlap. A type II error covers the case where the output sequence belongs to the intersection of $\mathcal{D}_i$ and $\mathcal{D}_j$ for $j \neq i$.

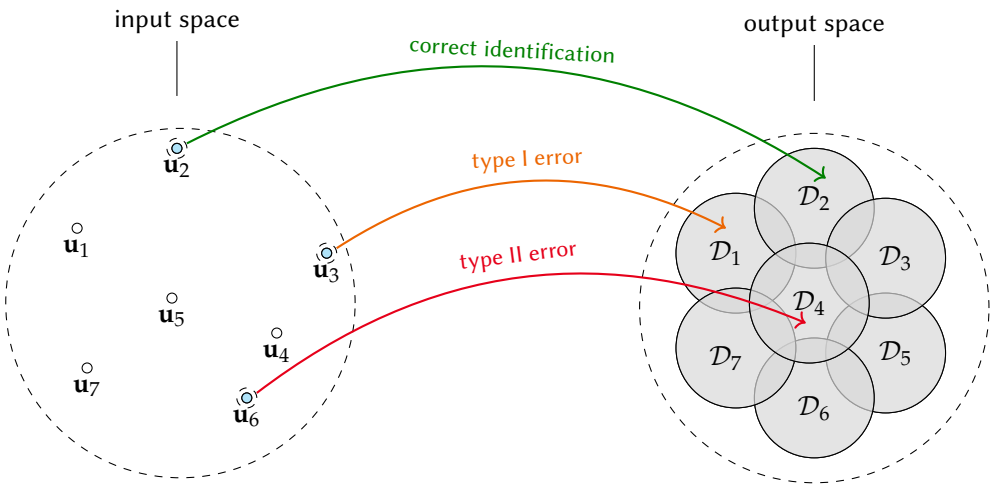


Figure 2.1: Geometric illustration of identification errors in the deterministic setting. The arrows indicate three scenarios for the channel output, given that the encoder transmitted the codeword u_1 corresponding to $i = 1$. If we focus on the decoder $\mathcal{D}_2$, then the green arrow event corresponds to the correct identification scenario since the output is observed in a decoder whose index is identical to the index of sent message, i.e., 2. Now if we assume that the decoder $\mathcal{D}_4$ is the decision maker entity, then since the channel output for orange arrow event is outside of $\mathcal{D}_4$, then a type I error has occurred for the decoder $\mathcal{D}_4$. However, for the red arrow event, since the output is observed in the decoder $\mathcal{D}_4$ region, then it declares that its index, i.e., 4 was the sent message which is different than the actual sent message, i.e., 6, therefore, we refer to this event as the Type II error.

2.2.3 | Related Work

We briefly review Ahlswede and Dueck's result [23] on the RI capacity, i.e., when the encoder uses a stochastic mapping. As mentioned above, using RI codes, it is possible to identify a double exponential number of messages in the block length n. That is, given a rate $R < \mathbb{C}_{RI}(\mathcal{W}, L)$, there exists a sequence of $(L(n,R) = 2^{2^{nR}}, n)$ RI codes with vanishing error probabilities. Despite the significant difference between the definitions in the identification setting and in the transmission setting, it was shown that the value of the RI capacity in the double exponential scale equals the Shannon capacity of transmission.

Theorem 2.2.1 (see [23,83]). *The RI capacity in the double exponential scale of a DMC $\mathcal{W}$ is given by*

$$\mathbb{C}_{RI}(\mathcal{W}, L) = \max_{p_X \,:\, \mathbb{E}\{\phi(X)\} \le A} I(X;Y)\,, \quad \textit{for } L(n,R) = 2^{2^{nR}}\,. \tag{2.14}$$

Hence, the RI capacity in the exponential scale is infinite, i.e.,

$$\mathbb{C}_{RI}(\mathcal{W}, L) = \infty\,, \quad \textit{for } L(n,R) = 2^{nR}\,. \tag{2.15}$$

In the next sections, we will consider the identification setting when the encoder does not have access to randomization.

Theorem 2.2.2 (see [23, 63]). *Let $\mathcal{G}$ denote the standard Gaussian channel, where the input-ouptut relation is given by $Y = gX + Z$, with $Z \sim \mathcal{N}(0, \sigma_Z^2)$ and a fixed known gain $g > 0$. Then, the RI capacity in the double exponential scale is given by*

$$\mathbb{C}_{RI}(\mathcal{G}, L) = \frac{1}{2}\log\left(1 + \frac{g^2 A}{\sigma_Z^2}\right)\,, \quad \textit{for } L(n,R) = 2^{2^{nR}}\,. \tag{2.16}$$

Hence, the RI capacity in the exponential scale is infinite, i.e.,

$$\mathbb{C}_{RI}(\mathcal{G}, L) = \infty\,, \quad \textit{for } L(n,R) = 2^{nR}\,. \tag{2.17}$$

2.3 | Main Result - DMC

We give our main results on the DI capacity of the DMC. For a DI code, as opposed to the randomized case, the number of messages 2^{nR} is only exponential in the blocklength. In this sense, DI codes are similar to transmission codes. However, the achievable rates for identification are significantly higher, as the DI capacity is given in terms of the input entropy instead of the mutual information.

2.3.1 | Channel Reduction

We begin with a procedure of channel reduction where we remove identical rows from the channel matrix, so that the remaining input letters have a lower cost compared to the deleted letters. As will be seen below, the DI capacity remains the same following this reduction. The characterization of the DI capacity will be given in the next section in terms of the reduced input alphabet.

We begin with the definition of the reduced channel.

Definition 2.3.1 (Reduced channel). *Given a DMC $\mathcal{W}$ with a stochastic matrix $W : \mathcal{X} \to \mathcal{Y}$, we define the reduced DMC W_r as follows. Let $\{\mathcal{X}(\ell)\}$ be a partition of $\mathcal{X}$ into equivalent classes, so that two letters x and x' belong to the same equivalent class if and only if the corresponding rows are identical, namely*

$$x, x' \in \mathcal{X}(\ell) \;\Leftrightarrow\; W(y|x) = W(y|x') \quad \forall y \in \mathcal{Y}\,. \tag{2.18}$$

For every class $\mathcal{X}(\ell)$, assign a representative element

$$z(\ell) = \arg \min_{x \in \mathcal{X}(\ell)} \phi(x) \,, \tag{2.19}$$

which is associated with the lowest input cost. If there is more than one letter that is associated with the lowest input cost in $\mathcal{X}(\ell)$, then choose one of them arbitrarily. Then the reduced input alphabet is defined as

$$\mathcal{X}_r = \{z(\ell)\} \,, \tag{2.20}$$

and the reduced DMC $\mathcal{W}_r$ is defined by a channel matrix $W_r : \mathcal{X}_r \to \mathcal{Y}$, consisting of the rows in $\mathcal{X}_r$, i.e.,

$$W_r(y|x) = W(y|x) \,, \tag{2.21}$$

for $x \in \mathcal{X}_r$ and $y \in \mathcal{Y}$.

Lemma 2.3.1. *The operational capacities of the reduced channel $\mathcal{W}_r$ and the original channel $\mathcal{W}$ are the same:*

$$\mathbf{C}_{DI}(\mathcal{W}, L) = \mathbf{C}_{DI}(\mathcal{W}_r, L) \,, \text{ for } L(n,R) = 2^{nR} \,. \tag{2.22}$$

We give the proof of Lemma 2.3.1 in Appendix E. As we will see shortly, the DI capacity of a DMC $\mathcal{W}$ depends on W only through $\mathcal{X}_r$. That is, the DI capacity does not depend on the individual values of the channel matrix and depends solely on the distinctness of its rows.

Remark 2.3.1. *Based on Lemma 2.3.1, it is sufficient to consider a channel with distinct rows. That is, if we establish the DI capacity for channels with distinct rows, we can then determine the DI capacity for a general channel. In other words, in order to derive a capacity result, we may assume without loss of generality that the channel rows are distinct.*

2.3.2 | Capacity Theorem

In this section, we give our main result on the DI capacity of a channel subject to input constraint. The capacity result is stated in terms of the reduced channel as defined in the previous section. Let $\mathcal{W}$ be a DMC channel with input cost function $\phi(x)$ and input constraint A as specified in (2.3). Define

$$\mathbf{C}_{DI}(\mathcal{W}) = \max_{p_X \,:\, \mathbb{E}\{\phi(X)\} \le A} H(X) \,, \tag{2.23}$$

for $X \sim p_X$.

Theorem 2.3.1. *The DI capacity of a DMC $\mathcal{W}$ under input constraint is given by*

$$\mathsf{C}_{DI}(\mathcal{W}, L) = \mathsf{C}_{DI}(\mathcal{W}_r)\,, \quad \text{for } L(n,R) = 2^{nR}\,, \tag{2.24}$$

where W_r denotes the reduced channel (see Definition 2.3.1). Hence, the DI capacity in the double exponential scale is zero.

We prove the direct part in Subsection 2.3.3 and the converse part in Subsection 2.3.4. As can be seen in Subsection 2.3.4, the *strong converse property* holds for the DI capacity [38, see Def. 3.5.1]. Notice that we have characterized the DI capacity of the DMC $\mathcal{W}$ in terms of its reduced version, as specified in Lemma 2.3.1.

Corollary 2.3.1.1 (also in [23, 53]). *The DI capacity of a DMC $\mathcal{W}$ without constraints, i.e., with $A = \phi_{\max}$, is given by*

$$\mathsf{C}_{DI}(\mathcal{W}, L) = \log\left(n_{row}(W)\right)\,, \tag{2.25}$$

for $L(n,R) = 2^{nR}$ where $n_{row}(W)$ is the number of distinct rows of W.

The corollary above is an immediate consequence of Theorem 2.3.1. Indeed, for $A = \phi_{\max}$, we have

$$\begin{aligned}\mathsf{C}_{DI}(\mathcal{W}_r) &= \max_{p_{X_0} \in \mathcal{P}(\mathcal{X}_r)} H(X_0) \\ &= \log|\mathcal{X}_r| \\ &= \log\left(n_{row}(W)\right)\,,\end{aligned} \tag{2.26}$$

where X_0 is a random variable, the support of which is in the reduced input alphabet $\mathcal{X}_r$ as defined in Definition 2.3.1. The second equality holds since the maximal value of $H(X)$ is $\log|\mathcal{X}|$, and the last equality because the size of the reduced input alphabet is $|\mathcal{X}_r| = n_{row}(W)$.

Remark 2.3.2. *Ahlswede et al. [23, 53] stated the result in Corollary 2.3.1.1 on the DI capacity of a DMC without constraints (see Section IV. [23] and abstract of [53]), without providing an explicit proof. A straightforward extension of the methods in [114], using decoding territories, does not seem to yield the desired result on the DI capacity. Thereafter, the proof has remained an open problem.*

Remark 2.3.3. *An alternative expression for the DI capacity is as follows,*

$$\mathsf{C}_{DI}(\mathcal{W}, L) = \max_{p_{X_0} \in \mathcal{P}(\mathcal{X}_r)\,:\, \mathbb{E}\{\phi(X_0)\} \le A} H(X_0)\,, \tag{2.27}$$

where X_0 is as in (2.26). As explained in Remark 2.3.1, one may assume without loss of generality that the channel has distinct rows. Under this assumption, the DI capacity formula reduces to the formula in (2.1), i.e.,

$$\mathsf{C}_{DI}(\mathcal{W}, L) = \max_{p_X : \mathbb{E}\{\phi(X)\} \le A} H(X) \,, \tag{2.28}$$

for a channel $\mathcal{W}$ with distinct rows and $L(n,R) = 2^{nR}$.

To illustrate our results, we give the following example.

Example 2.3.1. *Consider the binary symmetric channel (BSC),*

$$Y = X + Z \mod 2 \,, \tag{2.29}$$

where $\mathcal{X} = \mathcal{Y} = \{0,1\}$, $Z \sim$ Bernoulli(ϵ), with crossover probability $0 \le \epsilon \le \frac{1}{2}$. Suppose that the channel is subject to a Hamming weight input constraint,

$$\frac{1}{n}\sum_{t=1}^{n} x_t \le A \,, \tag{2.30}$$

with $\phi(x) = x$. Observe that for $\epsilon = \frac{1}{2}$, the rows of the channel matrix are identical. Hence, the reduced input alphabet consists of one letter, and the DI capacity is zero (see Definition 2.3.1).

Now, suppose that $\epsilon < \frac{1}{2}$. Then the rows of the channel matrix $W = \begin{pmatrix} 1-\epsilon & \epsilon \\ \epsilon & 1-\epsilon \end{pmatrix}$ are distinct, hence $\mathcal{W}_r = \mathcal{W}$. By Theorem 2.3.1, the DI capacity is given by

$$\mathsf{C}_{DI}(\mathcal{W}, L) = \mathsf{C}_{DI}(\mathcal{W}) = \max_{0 \le p \le A} H_2(p) \,, \tag{2.31}$$

since the channel input is binary, where $H_2(p) = -(1-p)\log(1-p) - p\log(p)$ is the binary entropy function. Therefore, the DI capacity of the BSC with Hamming weight constraint is

$$\mathsf{C}_{DI}(\mathcal{W}, L) = \begin{cases} H_2(A) & \text{if } A < \frac{1}{2} \\ 1 & \text{if } A \ge \frac{1}{2} \end{cases} \,, \quad \text{for } L(n,R) = 2^{nR} \,. \tag{2.32}$$

(See Figure 2.2). To show the direct part, set $X \sim$ Bernoulli(A) if $A < \frac{1}{2}$ and $X \sim$ Bernoulli$(\frac{1}{2})$, otherwise. The converse part follows as the binary entropy function $H_2(p)$ is strictly increasing on $0 \le p \le \frac{1}{2}$, attaining its maximum value $H_2(\frac{1}{2}) = 1$, and strictly decreasing on $\frac{1}{2} < p \le 1$ (see Figure 2.2). The geometric interpretation is that the binary

Hamming ball of radius np can be covered with codewords. As the volume of the Hamming ball is approximately $2^{nH_2(p)}$, one can achieve rates that are arbitrarily close to $H_2(p)$. Without an input constraint, i.e., for $A = 1$, we recover the result of JáJá [52],

$$\mathbb{C}_{DI}(\mathcal{W}, L) = 1 \,. \tag{2.33}$$

This example demonstrates that the DI capacity is discontinuous in the channel statistics, as $\mathbb{C}_{DI}(\mathcal{W}, L) = 1$ for $\epsilon < \frac{1}{2}$ and $\mathbb{C}_{DI}(\mathcal{W}, L) = 0$ for $\epsilon = \frac{1}{2}$.

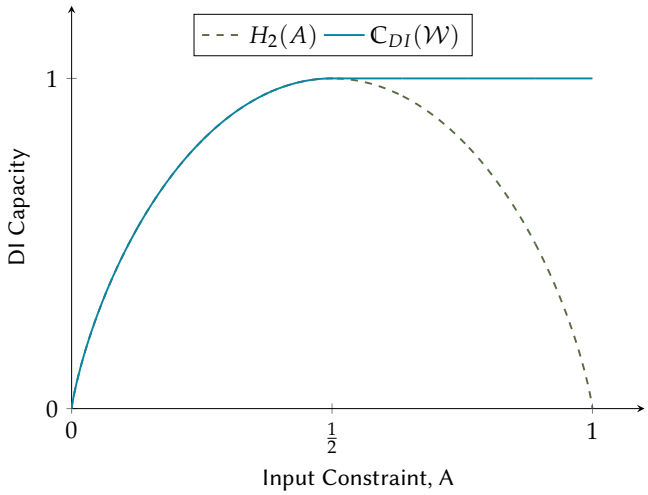


Figure 2.2: The deterministic identification (DI) capacity of the BSC as a function of the input constraint A. The dashed red line indicates the binary entropy function, which is maximized in (2.31). The solid blue line indicates the DI capacity.

2.3.3 | Achievability proof

Consider a DMC $\mathcal{W}$. By Lemma 2.3.1 we can assume without loss of generality that the channel matrix $W : \mathcal{X} \to \mathcal{Y}$ has distinct row vectors. To prove achievability of the DI capacity, we combine methods and ideas from the work of JáJá [52] as well as techniques by Ahlswede [114]. The analysis for the type II error is based on ideas that are analogous to the combinatoric analysis of Hamming distances in [52]. The codebook construction is similar to that of Ahlswede's coding scheme [114], yet the decoder is significantly different. Nonetheless, the type-class intersection lemma and the message-set analysis in [114] are useful in our analysis for the type II error.

We extensively use the method of types [100, Ch. 2]. Here a brief review of the definitions for type classes and δ-typical sets is given. The type $\hat{P}_{x^n}$ of a given se-

quence x^n is defined as the empirical distribution $\hat{P}_{x^n}(a) = N(a|x^n)/n$ for $a \in \mathcal{X}$, where $N(a|x^n)$ is the number of occurrences of the symbol $a \in \mathcal{X}$ in the sequence x^n. The space of all types over $\mathcal{X}$ of sequences of length n is denoted by $\mathcal{P}_n(\mathcal{X})$. The δ-typical set $\mathcal{T}_\delta(p_X)$ is defined as the set of sequences $x^n \in \mathcal{X}^n$ such that for every $a \in \mathcal{X}$: $|\hat{P}_{x^n}(a) - p_X(a)| \leq \delta$ if $p_X(a) > 0$, and $\hat{P}_{x^n}(a) = 0$ if $p_X(a) = 0$. A type class is denoted by $\mathcal{T}(\hat{P}) = \{x^n : \hat{P}_{x^n} = \hat{P}\}$. Similarly, a joint type is denoted by $\hat{P}_{x^n,y^n}(a,b) = N(a,b|x^n,y^n)/n$ for $(a,b) \in \mathcal{X} \times \mathcal{Y}$, where $N(a,b|x^n,y^n)$ is the number of occurrences of the symbol pair (a,b) in the sequence $(x_i,y_i)_{i=1}^n$, and as a conditional type by $\hat{P}_{y^n|x^n}(b|a) = N(a,b|x^n,y^n)/N(a|x^n)$. The conditional δ-typical set $\mathcal{T}_\delta(p_{Y|X}|x^n)$ is defined as the set of sequences $y^n \in \mathcal{Y}^n$ such that for every $b \in \mathcal{Y}$: $|\hat{P}_{y^n|x^n}(b|a) - p_{Y|X}(b|a)| \leq \delta$ if $p_{X,Y}(a,b) > 0$, and $p_{X,Y}(a,b) = 0$ if $p_X(a) = 0$.

The Codebook

First, we show that there exists a code such that the codewords are separated by a distance of $n\epsilon$. Let $p_X(x)$ be an input distribution on $\mathcal{X}$, such that

$$\mathbb{E}\left\{\phi(X)\right\} = \sum_{x\in\mathcal{X}} p_X(x)\phi(x) \leq A - \epsilon'(\delta) \tag{2.34}$$

for $X \sim p_X(x)$, where $\epsilon'(\delta) \to 0$ as $\delta \to 0$. We may assume without loss of generality that p_X is a type, due to the entropy continuity lemma [100, Lem. 2.7].

Lemma 2.3.2. *Let $R < H(X)$. Then, for sufficiently small $\epsilon \in (0,1)$ and sufficiently large n, there exists a codebook $\mathcal{U}^* = \{v_i,\ i \in \mathcal{M}\}$, which consists of $|\mathcal{M}|$ sequences in $\mathcal{X}^n$, such that the following hold:*

1. *All the codewords belong to the type class $\mathcal{T}(p_X)$, namely*

$$v_i \in \mathcal{T}(p_X) \text{ for all } i \in \mathcal{M}\ . \tag{2.35}$$

2. *The codewords are distanced by $n\epsilon$, i.e.,*

$$d_H(v_i,v_j) \geq n\epsilon \text{ for all } i \neq j\ . \tag{2.36}$$

3. *The codebook size is at least $\frac{1}{2}\cdot 2^{nR}$, that is, $|\mathcal{M}| \geq 2^{n(R-\frac{1}{n})}$.*

Proof of Lemma 2.3.2. Denote

$$M \triangleq 2^{nR}\ . \tag{2.37}$$

Let $U_1, ..., U_M$ be independent random sequences, each uniformly distributed over the type class of p_X, i.e.,

$$\Pr\left(U_i = x^n\right) = \begin{cases} \frac{1}{|\mathcal{T}(p_X)|} & x^n \in \mathcal{T}(p_X)\,, \\ 0 & x^n \notin \mathcal{T}(p_X)\,. \end{cases} \tag{2.38}$$

Next, define a new collection of sequences $V_1, ..., V_M$ as follows,

$$V_i = \begin{cases} U_i & \text{if } d_H(U_i, U_j) \geq n\epsilon \quad \forall i \neq j\,, \\ \varnothing & \text{otherwise}\,, \end{cases} \tag{2.39}$$

where $d_H(\cdot,\cdot)$ denotes the Hamming distance, and $\varnothing$ represents an idle sequence of no interest. The assignment $V_i = \varnothing$ is interpreted as "dropping the ith word U_i." Consider the following message set,

$$\widetilde{\mathcal{M}} = \left\{i : V_i \neq \varnothing, i \in [\![M]\!]\right\}\,, \tag{2.40}$$

corresponding to words that were not dropped, where we use the notation $\widetilde{\mathcal{M}}$ to indicate that the set is random.

We show that even though we removed words from the original collection $\{U_i\}_{i\in[\![M]\!]}$ (of size M), the rate decrease can be made negligible. Following the lines of [114], we derive an upper-bound on $\Pr(|\widetilde{\mathcal{M}}| \leq \frac{1}{2}M)$ where $\widetilde{\mathcal{M}}$ defined in (2.40) is the operational message set. To this end, we will use the following concentration lemma,

Lemma 2.3.3 (also in [114]). *Let $A_1, \ldots, A_K$ be a sequence of discrete random variables. Then,*

$$\Pr\left(\frac{1}{K}\sum_{i=1}^{K} A_i \geq c\right) \leq 2^{-cK} \prod_{i=1}^{K} \max_{a^{i-1}} \mathbb{E}\left(2^{A_i} \,\middle|\, A^{i-1} = a^{i-1}\right)\,. \tag{2.41}$$

Now, define an indicator for dropping the ith word by

$$\hat{V}_i = \begin{cases} 1 & V_i = \varnothing\,, \\ 0 & V_i \neq \varnothing\,, \end{cases} \tag{2.42}$$

and notice the equivalence between the following events,

$$\left\{\left|\widetilde{\mathcal{M}}\right| \leq \frac{1}{2}M\right\} = \left\{\sum_{i=1}^{M} \hat{V}_i > \frac{1}{2}M\right\}\,. \tag{2.43}$$

Observe that $\hat{V}_i = 1$, if and only if U_i is inside an ϵ-sphere of some other U_j. Namely, $\hat{V}_i = 1$ iff $U_i \in \bigcup_{j \neq i} \mathcal{S}_\epsilon(U_j)$. The selection of codewords can be viewed as an iterative procedure. Specifically, define

$$A_i = \begin{cases} 1 & U_i \in \bigcup_{j<i} \mathcal{S}_\epsilon(U_j) \,, \\ 0 & \text{otherwise} \,, \end{cases} \tag{2.44}$$

$$B_i = \begin{cases} 1 & U_i \in \bigcup_{j>i} \mathcal{S}_\epsilon(U_j) \,, \\ 0 & \text{otherwise} \,. \end{cases} \tag{2.45}$$

Now, since $\hat{V}_i = 1$ implies that either $A_i = 1$ or $B_i = 1$, it follows that the number of dropped messages is bounded by

$$\begin{aligned} M - \left|\widetilde{\mathcal{M}}\right| &= \sum_{i=1}^{M} \hat{V}_i \\ &\leq \sum_{i=1}^{M} A_i + \sum_{i=1}^{M} B_i \,. \end{aligned} \tag{2.46}$$

Consider the event that

$$\sum_{i=1}^{M} \hat{V}_i > \frac{1}{2} M \,. \tag{2.47}$$

If this holds, then the two sums in the right hand side of (2.46) cannot be smaller than $\frac{1}{4}M$ together, that is, either $\sum_{i=1}^{M} A_i \geq \frac{1}{4}M$, or $\sum_{i=1}^{M} B_i \geq \frac{1}{4}M$, or both. Hence,

$$\left\{ \sum_{i=1}^{M} \hat{V}_i > \frac{1}{2} M \right\} \subseteq \left\{ \sum_{i=1}^{M} A_i \geq \frac{1}{4} M \right\} \cup \left\{ \sum_{i=1}^{M} B_i \geq \frac{1}{4} M \right\} \,, \tag{2.48}$$

and by the union bound,

$$\begin{aligned} \Pr\left(\sum_{i=1}^{M} \hat{V}_i > \frac{1}{2} M \right) &\leq \Pr\left(\sum_{i=1}^{M} A_i \geq \frac{1}{4} M \right) + \Pr\left(\sum_{i=1}^{M} B_i \geq \frac{1}{4} M \right) \\ &= 2 \Pr\left(\sum_{i=1}^{M} A_i \geq \frac{1}{4} M \right) \,, \end{aligned} \tag{2.49}$$

where the last line follows by symmetry, as the random variables $\bar{A} = \sum_{i=1}^{M} A_i$ and $\bar{B} = \sum_{i=1}^{M} B_i$ have the same probability distribution.

Next we apply Lemma 2.3.3,

$$\Pr\left(\sum_{i=1}^{M} A_i \geq \frac{1}{4}M\right) \leq 2^{-\frac{1}{4}M} \prod_{i=1}^{M} \max_{a^{i-1}} \mathbb{E}\left(2^{A_i} | A^{i-1} = a^{i-1}\right) . \tag{2.50}$$

Consider the conditional expectation above. Using the law of total expectation, we can add conditioning on U^{i-1} as well, i.e.,

$$\begin{aligned}
&\mathbb{E}\left(2^{A_i} | A^{i-1} = a^{i-1}\right) \\
&= \sum_{u^{i-1}} \Pr(U^{i-1} = u^{i-1} | A^{i-1} = a^{i-1}) \cdot \mathbb{E}(2^{A_i} | U^{i-1} = u^{i-1}, A^{i-1} = a^{i-1}) \\
&= \sum_{u^{i-1}} \Pr(U^{i-1} = u^{i-1} | A^{i-1} = a^{i-1}) \cdot \mathbb{E}(2^{A_i} | U^{i-1} = u^{i-1}) \\
&\leq \max_{u^{i-1}} \mathbb{E}(2^{A_i} | U^{i-1} = u^{i-1}) ,
\end{aligned} \tag{2.51}$$

where the second equality holds since A_i, is a deterministic function of U^{i-1} (see (2.44)). Hence, by (2.50)-(2.51),

$$\begin{aligned}
&\Pr\left(\sum_{i=1}^{M} A_i \geq \frac{1}{4}M\right) \\
&\leq 2^{-\frac{1}{4}M} \prod_{i=1}^{M} \max_{u^{i-1}} \mathbb{E}\left(2^{A_i} | U^{i-1} = u^{i-1}\right) \\
&= 2^{-\frac{1}{4}M} \prod_{i=1}^{M} \max_{u^{i-1}} \left(\Pr\left\{A_i = 0 | U^{i-1} = u^{i-1}\right\} + 2\Pr\left\{A_i = 1 | U^{i-1} = u^{i-1}\right\}\right) \\
&\leq 2^{-\frac{1}{4}M} \prod_{i=1}^{M} \left(1 + 2\max_{u^{i-1}} \Pr(A_i = 1 | U^{i-1} = u^{i-1})\right) .
\end{aligned} \tag{2.52}$$

We bound the probability term $\Pr(A_i = 1 | U^{i-1} = u^{i-1})$, as follows. For a Hamming sphere of radius $n\epsilon$,

$$\left|S_\epsilon(x^n)\right| \leq \binom{n}{n\epsilon} \cdot |\mathcal{X}|^{n\epsilon} \leq 2^{n\theta(\epsilon)} , \tag{2.53}$$

for sufficiently large n, where

$$\theta(\epsilon) = H_2(\epsilon) + \epsilon \log |\mathcal{X}| , \tag{2.54}$$

tends to zero as $\epsilon \to 0$. The first inequality holds by a simple combinatoric argument. Namely, counting the number of sequences with up to $n\epsilon$ different entries compared to a given x^n, we have $\binom{n}{n\epsilon}$ optional choices for the locations of those entries, and $|\mathcal{X}|$

possible values for each of those entries. The last inequality follows from Stirling's approximation [120, Example 11.1.3]. Hence,

$$\begin{aligned}\left|\bigcup_{j=1}^{M}\mathcal{S}_\epsilon(u_j)\right| &\leq M2^{n\theta(\epsilon)}\\ &= 2^{n\left(R+\theta(\epsilon)\right)}\,,\end{aligned} \tag{2.55}$$

for every given collection of sequences, $u_1,\ldots,u_M \in \mathcal{T}(p_X)$. Consider a random sequence $\bar{X}^n$ that is uniformly distributed over the type class $\mathcal{T}(p_X)$, and statistically independent of $U_1,\ldots,U_M$. We use this external sequence as an auxiliary in the derivation below. Then,

$$\begin{aligned}\Pr\left(A_i=1\,\middle|\,U^{i-1}=u^{i-1}\right) &= \Pr\left(U_i\in\bigcup_{j<i}\mathcal{S}_\epsilon(u_j)\right)\\ &= \Pr\left(\bar{X}^n\in\bigcup_{j<i}\mathcal{S}_\epsilon(u_j)\right)\\ &\leq \Pr\left\{\bar{X}^n\in\bigcup_{j=1}^{M}\mathcal{S}_\epsilon(u_j)\right\}\,.\end{aligned} \tag{2.56}$$

The first equality follows from the definition of A_i in (2.44) and because $U_1,\ldots,U_M$ are statistically independent. The second equality holds because U_i and $\bar{X}^n$ are both uniformly distributed over the type class of p_X. The inequality follows as $\Pr(\mathcal{F}_1)\leq \Pr(\mathcal{F}_1\cup\mathcal{F}_2)$ for every pair $\mathcal{F}_1$, F_2 of probabilistic events. Since $\bar{X}^n$ is uniformly distributed over $\mathcal{T}(p_X)$, we have

$$\begin{aligned}\Pr\left\{\bar{X}^n\in\bigcup_{j=1}^{M}\mathcal{S}_\epsilon(u_j)\right\} &= \sum_{x^n\in\mathcal{T}(p_X)\,\cap\,\bigcup_{j=1}^{M}\mathcal{S}_\epsilon(u_j)}\frac{1}{\left|\mathcal{T}(p_X)\right|}\\ &= \frac{1}{\left|\mathcal{T}(p_X)\right|}\cdot\left|\mathcal{T}(p_X)\,\cap\,\bigcup_{j=1}^{M}\mathcal{S}_\epsilon(u_j)\right|\\ &\leq \frac{2^{n(R+\theta(\epsilon))}}{\left|\mathcal{T}(p_X)\right|}\\ &\leq (n+1)^{|\mathcal{X}|}\cdot\frac{2^{n(R+\theta(\epsilon))}}{2^{nH(X)}}\\ &\leq 2^{-n(H(X)-R-2\theta(\epsilon))}\,,\end{aligned} \tag{2.57}$$

for sufficiently large n, where the first inequality follows from (2.55), and the second is due to standard type class properties [120, Th. 11.1.3]. The last expression tends to zero as $n \to \infty$, provided that

$$R < H(X) - 3\theta(\epsilon) \, . \tag{2.58}$$

Together with (2.56)-(2.57), this implies

$$\Pr\left(A_i = 1 \,\middle|\, U^{i-1} = u^{i-1} \right) \leq 2^{-n\theta(\epsilon)} \, . \tag{2.59}$$

Now plugging (2.59) into (2.52) yields

$$\begin{aligned} \Pr\left(\sum_{i=1}^{M} A_i \geq \frac{1}{4} M \right) &\leq 2^{-\frac{1}{4}M} \left(1 + 2 \cdot 2^{-n\theta(\epsilon)} \right)^M \\ &= \left(2^{-\frac{1}{4}} + 2^{\frac{3}{4}} \cdot 2^{-n\theta(\epsilon)} \right)^M \, , \end{aligned} \tag{2.60}$$

for sufficiently large n, we have $2^{\frac{3}{4}} \cdot 2^{-n\theta(\epsilon)} \leq 2^{-5}$ hence,

$$\begin{aligned} 2^{-\frac{1}{4}} + 2^{\frac{3}{4}} \cdot 2^{-n\theta(\epsilon)} &\leq 2^{-\frac{1}{4}} + 2^{-5} \\ &= 0.8721 \\ &< 1 \, . \end{aligned} \tag{2.61}$$

Thus we have a double exponential bound

$$\begin{aligned} \Pr\left(\left| \widetilde{\mathcal{M}} \right| \leq \frac{1}{2} M \right) &\leq 2^{-\alpha_1 M} \\ &= 2^{-\alpha_1 2^{nR}} \, , \end{aligned} \tag{2.62}$$

for some $\alpha_1 > 0$. We deduce that there exists at least one codebook with the desired properties. This completes the proof of Lemma 2.3.2. □

We continue to the main part of the achievability proof. Let $\mathcal{U}^* = \{v_i \, , \, i \in \mathcal{M}\}$ be a codebook of size $2^{n(R-\frac{1}{n})}$ as in Lemma 2.3.2. Consider the following DI coding scheme for $\mathcal{W}$.

2.3.3.1 | Encoding

Given a message $i \in \mathcal{M}$ at the sender, transmit $x^n = v_i$.

2.3.3.2 | Decoding

Let $\delta > 0$, such that $\delta \to 0$ as $\epsilon \to 0$. Let $j \in \mathcal{M}$ be the message that the decoder wishes to identify. To do so, the decoder checks whether the channel output y^n belongs to the corresponding decoding set $\mathcal{D}_j$ or not, where

$$\mathcal{D}_j = \left\{ y^n \,:\, (v_j, y^n) \in \mathcal{T}_\delta(p_X W) \right\} . \tag{2.63}$$

Namely, given the channel output $y^n \in \mathcal{Y}^n$, if $(v_j, y^n) \in \mathcal{T}_\delta(p_X W)$, then the decoder declares that the message j was sent. On the other hand, if $(v_j, y^n) \notin \mathcal{T}_\delta(p_X W)$, it declares that j was not sent.

Error Analysis

First, consider the error of type I, i.e., the event that $Y^n \notin \mathcal{D}_i$. For every $i \in \mathcal{M}$, the probability of identification error of type I, $P_{e,1}(i) = \Pr((v_i, Y^n) \notin \mathcal{T}_\delta(p_X W))$ tends to zero by standard type class considerations [121, Th. 1.2].

We move to the error of type II, i.e., when $Y^n \in \mathcal{D}_j$ for $j \neq i$. To bound the probability of error $P_{e,2}(i,j)$, we use the conditional type-class intersection lemma, due to Ahlswede [114], as stated below.

Lemma 2.3.4 (see [114, Lem. I_1]). *Let $W : \mathcal{X} \to \mathcal{Y}$ be a channel matrix of a DMC $\mathcal{W}$ with distinct rows. Then, for every $x^n, x'^n \in \mathcal{T}_\delta(p_X)$ with $d_H(x^n, x'^n) \geq n\epsilon$,*

$$\frac{|\mathcal{T}_\delta(p_{Y|X}|x^n) \cap \mathcal{T}_\delta(p_{Y|X}|x'^n)|}{|\mathcal{T}_\delta(p_{Y|X}|x^n)|} \leq 2^{-nL(\epsilon)} , \tag{2.64}$$

with $p_{Y|X} \equiv W$, for sufficiently large n and some positive function $L(\epsilon) > 0$ which is independent of n.

Now, for short notation, denote the conditional δ-typical set in $\mathcal{Y}^n$, given $x^n \in \mathcal{T}(p_X)$, by

$$\mathcal{G}(x^n) \equiv \mathcal{T}_\delta(W|x^n) = \left\{ y^n \,:\, (x^n, y^n) \in \mathcal{T}_\delta(p_X W) \right\} . \tag{2.65}$$

Then, for every $i \neq j$,

$$\begin{aligned} P_{e,2}(i,j) &= \Pr(\mathcal{D}_j | x^n = v_i) \\ &= \sum_{y^n \in \mathcal{G}(v_j)} W^n(y^n | v_i) \end{aligned}$$

$$= \sum_{y^n \in \mathcal{G}(v_j) \cap \mathcal{G}(v_i)} W^n(y^n|v_i) + \sum_{y^n \in \mathcal{G}(v_j) \cap (\mathcal{G}(v_i))^c} W^n(y^n|v_i) \,. \tag{2.66}$$

Observe that the second sum in the last line is bounded by the probability $\Pr(Y^n \notin \mathcal{T}_\delta(W|v_i)|x^n = v_i)$, which in turn is bounded by $2^{-\alpha_1(\delta)n}$ as before, and tends to zero as well.

To bound the first sum in (2.66), we first consider the cardinality of the set that the sum acts upon (the domain). We note that since v_i and v_j belong to the type class $\mathcal{T}(p_X)$ by the first property of Lemma 2.3.2, it follows that they also belong to the δ-typical set, i.e., $v_i, v_j \in \mathcal{T}_\delta(p_X)$. Further, according to the second property of Lemma 2.3.2, every pair of codewords v_i and v_j satisfy $d_H(v_i, v_j) \geq n\epsilon$. Finally, having assumed that the rows of W are distinct, we have by Lemma 2.3.4,

$$\begin{aligned} |\mathcal{G}(v_j) \cap \mathcal{G}(v_i)| &\leq 2^{-nL(\epsilon)} |\mathcal{G}(v_j)| \\ &\leq 2^{n[H(Y|X) - L(\epsilon)]} \,, \end{aligned} \tag{2.67}$$

where $X \sim p_X$, as we explained below. The second inequality in (2.67) holds since the size of the conditional type class $\mathcal{G}(x^n) = \mathcal{T}_\delta(W|x^n)$ is bounded by $2^{nH(Y|X)}$ [100, Lem. 2.5], as the type of v_i and v_j is p_X. Furthermore, by standard type class properties [121, Th. 1.2],

$$W^n(y^n|v_i) \leq 2^{-n[H(Y|X) - \delta \log |\mathcal{Y}|]} \,. \tag{2.68}$$

Now by Equation (2.67) and (2.68),

$$\sum_{y^n \in \mathcal{G}(v_j) \cap \mathcal{G}(v_i)} W^n(y^n|v_i) \leq 2^{-n[L(\epsilon) - \delta \log |\mathcal{Y}|]} \,, \tag{2.69}$$

which tends to zero as $n \to \infty$ for sufficiently small $\delta > 0$, such that $\delta \log |\mathcal{Y}| < L(\epsilon)$. Thus, by (2.66) and (2.69), the probability of type II error is bounded by

$$P_{e,2}(i,j) \leq 2^{-n\alpha_2(\epsilon,\delta)} \,, \tag{2.70}$$

for sufficiently large n, where $\alpha_2(\epsilon, \delta) = \min\{\alpha_1(\delta), L(\epsilon) - \delta \log |\mathcal{Y}|\}$. The proof follows by taking the limits $n \to \infty$, and $\epsilon, \delta \to 0$.

2.3.4 | Converse Proof

To prove the converse part, we will use the following observation. Let $R > 0$ be an achievable rate. We will assume to the contrary that there exist two different messages

i_1 and i_2 that are represented by the same codeword, i.e., $u_{i_1} = u_{i_2} = x^n$, and show that this leads to error probabilities such that

$$P_{e,1}(i_1) + P_{e,2}(i_2, i_1) = 1 \,. \tag{2.71}$$

Hence the assumption is false. The number of messages 2^{nR} is thus bounded by the size of the subset of input sequences that satisfy the input constraint $\phi^n(x^n) \leq A$. Then we notice that the average cost of a codeword depends only on its type, and hence this subset is in fact a union of type classes. This also implies that we have a strong converse for the DI capacity.

Consider a sequence of $(2^{nR}, n, \lambda_1^{(n)}, \lambda_2^{(n)})$ codes $(\mathcal{U}^{(n)}, \mathcal{D}^{(n)})$ such that $\lambda_1^{(n)}$ and $\lambda_2^{(n)}$ tend to zero as $n \to \infty$.

Lemma 2.3.5. *Consider a sequence of codes as described above. Then, given a sufficiently large n, the codebook $\mathcal{U}^{(n)}$ satisfies the following property. There cannot be two distinct messages that are represented by the same codeword, i.e.,*

$$i_1 \neq i_2 \quad \Rightarrow \quad u_{i_1} \neq u_{i_2} \,, \tag{2.72}$$

where $i_1, i_2 \in [\![2^{nR}]\!]$.

Proof. Assume to the contrary that there exist two messages i_1 and i_2, where $i_1 \neq i_2$, such that

$$u_{i_1} = u_{i_2} = x^n \,, \tag{2.73}$$

for some $x^n \in \mathcal{X}^n$. Since $(\mathcal{U}^{(n)}, \mathcal{D}^{(n)})$ form a $(2^{nR}, n, \lambda_1^{(n)}, \lambda_2^{(n)})$ code, we have

$$\begin{aligned} P_{e,1}(i_1) &= W^n(\mathcal{D}_{i_1}^c | x^n) \leq \lambda_1^{(n)} \\ P_{e,2}(i_2, i_1) &= W^n(\mathcal{D}_{i_1} | x^n) \leq \lambda_2^{(n)} \,. \end{aligned} \tag{2.74}$$

This leads to a contradiction as

$$\begin{aligned} 1 &= W^n(\mathcal{D}_{i_1}^c | x^n) + W^n(\mathcal{D}_{i_1} | x^n) \\ &= P_{e,1}(i_1) + P_{e,2}(i_2, i_1) \\ &\leq \lambda_1^{(n)} + \lambda_2^{(n)} \,. \end{aligned} \tag{2.75}$$

Hence, the assumption is false, and i_1 and i_2 cannot have the same codeword. □

By Lemma 2.3.5, each message has a distinct codeword. Hence, the number of messages is bounded by the number of input sequences that satisfy the input constraint. That is, the size of the codebook is upper-bounded as follows:

$$2^{nR} \leq \left| \left\{ x^n \ : \ \frac{1}{n} \sum_{t=1}^{n} \phi(x_t) \leq A \right\} \right| . \tag{2.76}$$

Notice that the input cost of a given sequence x^n depends only on the type of the sequence, since

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^{n} \phi(x_t) &= \sum_{a \in \mathcal{X}} \hat{P}_{x^n}(a) \phi(a) \\ &= \mathbb{E}\Big\{ \phi(X') \Big\} , \end{aligned} \tag{2.77}$$

where the random variable X' is distributed according to the type of x^n, i.e., $p_{X'} = \hat{P}_{x^n}$. Therefore, the subset on the right hand side of (2.76) can be written as a union of type classes:

$$\begin{aligned} \left| \left\{ x^n : \frac{1}{n} \sum_{t=1}^{n} \phi(x_t) \leq A \right\} \right| &= \left| \bigcup_{\substack{p_{X'} \in \mathcal{P}_n(\mathcal{X}): \\ \mathbb{E}\{\phi(X')\} \leq A}} \mathcal{T}(p_{X'}) \right| \\ &\leq \left| \mathcal{P}_n(\mathcal{X}) \right| \max_{\substack{p_{X'} \in \mathcal{P}_n(\mathcal{X}): \\ \mathbb{E}\{\phi(X')\} \leq A}} \left| \mathcal{T}(p_{X'}) \right| \\ &\leq \left| \mathcal{P}_n(\mathcal{X}) \right| \cdot 2^{nH(X')} \\ &\leq 2^{n\left(H(X') + \alpha_n\right)} \\ &\leq 2^{n\left(\mathsf{C}_{DI}(\mathcal{W}) + \alpha_n\right)} , \end{aligned} \tag{2.78}$$

where $\alpha_n \to 0$ as $n \to \infty$, where $\mathcal{P}_n(\mathcal{X})$ denotes the space of all types over $\mathcal{X}$ of sequences of length n. The second inequality holds since the size of a type class $\mathcal{T}(p_{X'})$ is bounded by $|\mathcal{T}(p_{X'})| \leq 2^{nH(X')}$ [120, Th. 11.1.3]. The third inequality holds since the number of types on $\mathcal{X}$ is polynomial in n [120, Th. 11.1.1]. Thus, by (2.76) and (2.78), the code rate is bounded by $R \leq \mathsf{C}_{DI}(\mathcal{W}) + \alpha_n$, which completes the proof of Theorem 2.3.1. □

2.4 | Summary and Discussion

We have established the deterministic identification (DI) capacity of a DMC subject to an input constraint. For the DMC, the DI capacity formula is given in terms of the

entropy of the reduced channel input (see Definition 2.3.1). The DI capacity characterization does not depend on the specific transition probabilities corresponding to the reduced input alphabet.

In the following, we compare and discuss different results from the literature on the DI capacity. For the double exponential scale, or equivalently, when the rate is defined as $R = \frac{1}{n} \log\log (\#\text{ of messages})$, the DI capacity is

$$\mathbb{C}_{DI}(\mathcal{W}, L) = 0 \,, \tag{2.79}$$

for $L(n,R) = 2^{2^{nR}}$, since the code size of DI codes scales only exponentially in block length. On the other hand, as observed in [49], if one considers an average error criterion instead of the maximal error, then the double exponential performance of randomized-encoder codes can also be achieved using deterministic codes.

By providing a detailed proof for the DI capacity theorem with and without an input constraint, we have filled the gap in the previous analysis [23, 53] as well. In particular, in [53], Ahlswede and Cai asserted that the DI capacity for a compound channel is given by

$$\mathbb{C}_{DI}(\mathcal{W}_{\text{compound}}, L) = \max_{p_X} \min_{s} H(\hat{X}(s)) \,, \tag{2.80}$$

for $L(n,R) = 2^{nR}$, where $s \in \mathcal{S}$ is the channel state, and the map $\hat{X}(s)$ is induced from X by a partition of the input alphabet to equivalent classes as specified in [53, Sec. I.F]. This result immediately yields Corollary 2.3.1.1, since the DMC is a special case of a compound channel with a single state value. Indeed, taking $|\mathcal{S}| = 1$ and considering the reduced channel W_r (see Definition 2.3.1), it can be readily shown that $\hat{X}(s) = X$. Nonetheless, a significant part of the proof in [53] is missing. At the beginning of Sec. VII in [53], the following claim is given: "It was shown in [A'80] that for any channel $\tilde{V} : \mathcal{X} \to \mathcal{Y}$ without two identical rows, any $u_1, u_2, \epsilon > 0$, sufficiently large n and any $\mathcal{U} \subset \mathcal{X}^n$ such that for all $u, u' \in \mathcal{U}, d_H(u,u') > n\epsilon$, there exists a family of subsets of $\mathcal{Y}^n$, say $\mathcal{D}_u, u \in \mathcal{U}$, such that $\tilde{V}^n(\mathcal{D}_u|u) > 1 - u_1$ and $\tilde{V}^n(\mathcal{D}_u|u') < u_2$ for all $u \neq u'$, where d_H is the Hamming distance.", where [A'80] refers to a paper by Ahlswede [114] on the arbitrarily varying channel, and does not include identification.

Alternatively, one may consider the ϵ-capacity for a fixed $0 < \epsilon < 1$. A rate R is called ϵ-achievable in an L-scale if there exists an $(L(n,R), n, \epsilon, \epsilon)$ code for sufficiently large n (see Definition 2.2.2). The DI ϵ-capacity $\mathbb{C}_{DI}^{\epsilon}(\mathcal{W}, L)$ is then defined as the supremum of ϵ-achievable rates. As the RI capacity in the double exponential scale has a

strong converse [23, 122, 123], for $L(n,R) = L_{\text{double}}(n,R)$ and $0 < \epsilon < \frac{1}{2}$,

$$\mathbb{C}_{RI}^{\epsilon}(\mathcal{W}, L_{\text{double}}) = \mathbb{C}_{RI}(\mathcal{W}, L_{\text{double}}) = \max_{p_X} I(X;Y) \,. \tag{2.81}$$

Based on Subsection 2.3.4, a strong converse holds for the DI capacity as well, hence

$$\mathbb{C}_{DI}^{\epsilon}(\mathcal{W}, L_{\text{double}}) = \mathbb{C}_{DI}(\mathcal{W}, L_{\text{double}}) = 0 \,. \tag{2.82}$$

On the other hand, for $\epsilon \geq \frac{1}{2}$ we have

$$\mathbb{C}_{DI}^{\epsilon}(\mathcal{W}, L_{\text{double}}) = \mathbb{C}_{RI}^{\epsilon}(\mathcal{W}, L_{\text{double}}) = \infty \,. \tag{2.83}$$

To understand (2.83), suppose $\epsilon > \frac{1}{2}$, and consider an arbitrary set of codewords with a stochastic decoder that makes a decision for the identification hypothesis by flipping a fair coin [23]. Both error probabilities of type I and II equal $\frac{1}{2}$, and are thus smaller than ϵ. Similarly, for the Gaussian channel, [46, 62],

$$\mathbb{C}_{RI}^{\epsilon}(\mathscr{G}, L_{\text{double}}) = \mathbb{C}_{RI}(\mathscr{G}, L_{\text{double}}) = \frac{1}{2}\log\left(1 + \frac{A}{\sigma_Z^2}\right) \quad , \quad \text{for } 0 < \epsilon < \frac{1}{2} \,, \tag{2.84}$$

$$\mathbb{C}_{DI}^{\epsilon}(\mathscr{G}, L_{\text{double}}) = \mathbb{C}_{RI}^{\epsilon}(\mathscr{G}, L_{\text{double}}) = \infty \quad , \quad \text{for } \epsilon \geq \frac{1}{2} \,. \tag{2.85}$$

Based on our result in Theorem 5.0.1, with $G_t \equiv 1$ for all t, we have that

$$\frac{1}{4} \leq \mathbb{C}_{DI}(\mathscr{G}, L^*) \leq \mathbb{C}_{DI}^{\epsilon}(\mathscr{G}, L^*) \leq 1 \,, \tag{2.86}$$

with

$$L^*(n,R) = 2^{n\log(n)R} = n^{nR} \,. \tag{2.87}$$

Hence, in the double exponential scale, $\mathbb{C}_{DI}^{\epsilon}(\mathscr{G}, L_{\text{double}}) = \mathbb{C}_{DI}(\mathscr{G}, L_{\text{double}}) = 0$.

CHAPTER 3

DI FOR STANDARD GAUSSIAN CHANNEL

> *It is Not Knowledge, But The Act of Learning, Not Possession But The Act of Getting There, Which Grants The Greatest Enjoyment.*
>
> Carl Friedrich Gauss,

3.1 | Introduction

Modern communications require the transfer of enormous amounts of data in wireless systems, for cellular communication [124], sensor networks [125], smart appliances [126], and the internet of things [127], etc. Wireless communication is often modelled by fading channels with additive white Gaussian noise [128–136].

We consider deterministic identification for Gaussian channels without the fading. Applying discretization to our capacity result above; cf. (2.1), we obtain that the DI capacity of the standard Gaussian channel is infinite in the exponential scale (as we have recently observed in [98, 137]). However, for a finite blocklength n, the number of codewords must be finite. Thereby, the meaning of the infinite capacity result is that the number of messages scales super-exponentially. This raises the question: What is the true order of the code size. In mathematical terms, what is the scale $L(n, R)$ for which the DI capacity is positive yet finite. We address the exact derivation of such a scale, in the next chapters which are dedicated for the fast and slow fading channels. In this chapter, we aim to focus on the standard Gaussian channel which may be deemed as a preliminary model for studying more advanced scenarios such as the slow or fast fading channels. In this section, we consider the Gaussian channel $\mathscr{G}$, specified by the input-output relation

$$\mathbf{Y} = \mathbf{x} + \mathbf{Z} \,. \tag{3.1}$$

with additive white Gaussian noise, i.e., when the noise sequence $\mathbf{Z}$ is i.i.d. $\sim \mathcal{N}(0, \sigma^2)$. The transmission power is limited to $\|\mathbf{x}\|^2 \leq nA$.

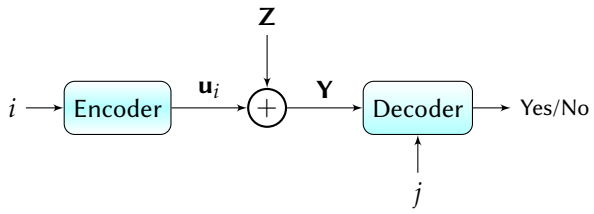


Figure 3.1: Deterministic identification for the standard Gaussian channel.

3.1.1 | Coding For The Gaussian Channel

The definition of a DI code for the Gaussian channel is given below.

Definition 3.1.1 (Gaussian DI Code). *A $(2^{nR}, n)$ DI code for a Gaussian channel $\mathscr{G}$ under input constraint A, assuming 2^{nR} is an integer, is defined as a system $(\mathcal{U}, \mathscr{D})$ consisting of a codebook $\mathcal{U} = \{\mathbf{u}_i\}_{i \in [\![2^{nR}]\!]}$, $\mathcal{U} \subset \mathcal{X}^n$, such that*

$$\|\mathbf{u}_i\|^2 \leq nA \,, \tag{3.2}$$

for all $i \in [\![2^{nR}]\!]$ and a collection of decoding regions $\mathscr{D} = \{\mathcal{D}_i\}_{i \in [\![2^{nR}]\!]}$ with

$$\bigcup_{i=1}^{2^{nR}} \mathcal{D}_i \subset \mathbb{R}^n \,. \tag{3.3}$$

Given a message $i \in [\![2^{nR}]\!]$, the encoder transmits $\mathbf{u}_i$. The decoder's aim is to answer the following question: Was a desired message j sent or not? There are two types of errors that may occur: Rejecting of the true message, or accepting a false message. Those are referred to as type I and type II errors, respectively.

The error probabilities of the identification code $(\mathcal{U}, \mathscr{D})$ are given by

$$P_{e,1}(i) = 1 - \int_{\mathcal{D}_i} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{u}_i)\, d\mathbf{y} \quad \textit{correctness property} \,, \tag{3.4}$$

$$P_{e,2}(i,j) = \int_{\mathcal{D}_j} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{u}_i)\, d\mathbf{y} \qquad \textit{disjointedness property} \,. \tag{3.5}$$

with the noise formula given by

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\|\mathbf{z}\|^2/2\sigma^2} \,, \tag{3.6}$$

(see Figure 2.1). A $(2^{nR}, n, \lambda_1, \lambda_2)$ *DI code further satisfies*

$$P_{e,1}(i) \leq \lambda_1 \,, \tag{3.7}$$

$$P_{e,2}(i,j) \leq \lambda_2 \,, \tag{3.8}$$

for all $i, j \in [\![2^{nR}]\!]$, *such that* $i \neq j$.

A rate $R > 0$ *is called achievable if for every* $\lambda_1, \lambda_2 > 0$ *and sufficiently large* n, *there exists a* $(2^{nR}, n, \lambda_1, \lambda_2)$ *DI code. The operational DI capacity of the Gaussian channel is defined as the supremum of achievable rates, and will be denoted by* $\mathbb{C}_{DI}(\mathscr{G})$.

3.1.2 | Main Result - Standard Gaussian Channel

Our DI capacity theorem for the Gaussian channel is stated below.

Theorem 3.1.1. *The DI capacity of the Gaussian channel* $\mathscr{G}$ *subject to average power constraint of* $\|\mathbf{u}_i\|^2 \leq nA$ *is given by*

$$\mathbb{C}_{DI}(\mathscr{G}) = \infty \,. \tag{3.9}$$

The proof of Theorem 3.1.1 is given below.

Proof. Consider the Gaussian channel $\mathscr{G}$. To show that the capacity is infinite, it suffices to prove the direct part. We show here that the DI capacity of the Gaussian channel can be achieved using a simple distance-decoder. A DI code for the Gaussian channel $\mathscr{G}$ is constructed as follows. Since the decoder can normalize the output symbols by $\frac{1}{\sqrt{n}}$, we have an equivalent input-output relation,

$$\bar{\mathbf{Y}} = \bar{\mathbf{x}} + \bar{\mathbf{Z}} \,, \tag{3.10}$$

where the noise sequence $\bar{\mathbf{Z}}$ is i.i.d. $\sim \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$, and an input power constraint

$$\|\bar{\mathbf{x}}\| \leq \sqrt{A} \,, \tag{3.11}$$

with $\bar{\mathbf{x}} = \frac{1}{\sqrt{n}}\mathbf{x}$, $\bar{\mathbf{Z}} = \frac{1}{\sqrt{n}}\mathbf{Z}$, and $\bar{\mathbf{Y}} = \frac{1}{\sqrt{n}}\mathbf{Y}$.

3.1.2.1 | Codebook construction

Let $\mathscr{S}$ denote a sphere packing, i.e., an arrangement of L non-overlapping spheres $\mathcal{S}_{\mathbf{u}_i}(n, r_0)$, $i \in [\![L]\!]$, that cover a bigger sphere $\mathcal{S}_{\mathbf{0}}(n, r_1)$, with $r_1 > r_0$. As opposed to

standard sphere packing coding techniques, the small spheres are not necessarily entirely contained within the bigger sphere (see Figure 3.2). That is, we only require that the spheres are disjoint from each other and have a non-empty intersection with $\mathcal{S}_{\mathbf{0}}(n, r_1)$. The packing density $\Delta_n(\mathscr{S})$ is defined as the fraction of the big sphere volume $\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, r_1)\right)$ that is covered by the small spheres, i.e.

$$\Delta_n(\mathscr{S}) \triangleq \frac{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, r_1) \cap \bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n, r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n, r_1))}, \tag{3.12}$$

(see [138, Ch. 1]). A sphere packing is called *saturated* if no spheres can be added to the arrangement without overlap. sphere packing is called *saturated* if no spheres can be added to the arrangement without overlap.

We use a packing argument that has a similar flavor as in the Minkowski–Hlawka theorem in lattice theory [138]. We use the property that there exists an arrangement $\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n, \sqrt{\epsilon_n})$ of non-overlapping spheres inside $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A})$ with a density of $\Delta_n(\mathscr{S}) \geq 2^{-n}$ [118, Lem. 2.1]. Specifically, consider a saturated packing arrangement of $L(n, R) = 2^{nR}$ spheres of radius $r_0 = \sqrt{\epsilon}$ covering the big sphere $\mathcal{S}_{\mathbf{0}}(n, r_1 = \sqrt{A} - \sqrt{\epsilon})$, i.e., such that no spheres can be added without overlap. Then, for such an arrangement, there cannot be a point in the big sphere $\mathcal{S}_{\mathbf{0}}(n, r_1)$ with a distance of more than $2r_0$ from all sphere centers. Otherwise, a new sphere could be added. As a consequence, if we double the radius of each sphere, the $2r_0$-radius spheres cover the whole sphere

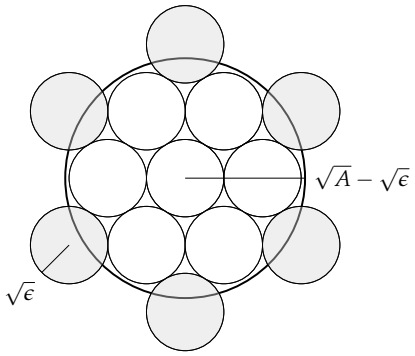


Figure 3.2: Illustration of a sphere packing, where small spheres of radius $r_0 = \sqrt{\epsilon}$ cover a bigger sphere of radius $r_1 = \sqrt{A} - \sqrt{\epsilon}$. The small spheres are disjoint from each other and have a non-empty intersection with the big sphere. Some of the small spheres, marked in gray, are not entirely contained within the bigger sphere, and yet they are considered to be a part of the packing arrangement. As we assign a codeword to each small sphere center, the norm of a codeword is bounded by $\sqrt{A}$ as required.

of radius r_1. In general, the volume of a hyper-sphere of radius r is given by

$$\mathrm{Vol}\left(\mathcal{S}_\epsilon(\mathbf{x}, r)\right) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r^n \,, \tag{3.13}$$

(see Eq. (16) in [138]). Hence, doubling the radius multiplies the volume by 2^n. Since the $2r_0$-radius spheres cover the entire sphere of radius r_1, it follows that the original r_0-radius packing has density at least 2^{-n}, i.e.,

$$\Delta_n(\mathscr{S}) \geq 2^{-n} \,. \tag{3.14}$$

We assign a codeword to the center $\mathbf{u}_i$ of each small sphere. The codewords satisfy the input constraint as $\|\mathbf{u}_i\| \leq r_0 + r_1 = \sqrt{A}$. Since the small spheres have the same volume, the total number of spheres is bounded from below by

$$\begin{aligned} L &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n, r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n, r_0))} \\ &\geq \frac{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, r_1) \cap \bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n, r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n, r_0))} \\ &= \frac{\Delta_n(\mathscr{S}) \cdot \mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n, r_1)))}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n, r_0))} \\ &\geq 2^{-n} \cdot \frac{\mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n, r_1)))}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n, r_0))} \\ &= 2^{-n} \cdot \frac{r_1^n}{r_0^n} \,, \end{aligned} \tag{3.15}$$

where the second equality is due to (3.12), the inequality that follows holds by (3.14), and the last equality follows from (3.13). That is, the codebook size satisfies

$$\begin{aligned} L(n, R) &= 2^{nR} \\ &\geq 2^{-n} \cdot \left(\frac{\sqrt{A} - \sqrt{\epsilon}}{\sqrt{\epsilon}}\right)^n \,. \end{aligned} \tag{3.16}$$

Hence,

$$R \geq \frac{1}{2} \log\left(\frac{A}{\epsilon}\right) - 1 \,. \tag{3.17}$$

3.1.2.2 | Encoding

Given a message $i \in [\![2^{nR}]\!]$, transmit $\bar{\mathbf{x}} = \bar{\mathbf{u}}_i$.

3.1.2.3 | Decoding

Let $\delta > 0$. To identify whether a message $j \in \mathcal{M}$ was sent, the decoder checks whether the channel output $\mathbf{y}$ belongs to the following decoding set,

$$\mathcal{D}_j = \left\{ \bar{\mathbf{y}} \in \mathbb{R}^n \,:\, \left\| \bar{\mathbf{y}} - \bar{\mathbf{u}}_j \right\| \leq \sqrt{\sigma_Z^2 + \delta} \right\} . \tag{3.18}$$

3.1.2.4 | Error Analysis

Consider the type I error, i.e., when the transmitter sends $\bar{\mathbf{u}}_i$, yet $\bar{\mathbf{Y}} \notin \mathcal{D}_i$. For every $i \in [\![2^{nR}]\!]$, the type I error probability is bounded by

$$\begin{aligned} P_{e,1}(i) &= \Pr\left(\left\| \bar{\mathbf{Y}} - \bar{\mathbf{u}}_i \right\|^2 > \sigma_Z^2 + \delta \,\middle|\, \bar{\mathbf{x}} = \bar{\mathbf{u}}_i \right) \\ &= \Pr\left(\left\| \bar{\mathbf{Z}} \right\|^2 > \sigma_Z^2 + \delta \right) \\ &= \Pr\left(\sum_{t=1}^{n} \bar{Z}_t{}^2 > \sigma_Z^2 + \delta \right) \\ &\leq \frac{3\sigma_Z^4}{n\delta^2} \\ &\leq \lambda_1 \,, \end{aligned} \tag{3.19}$$

which tends to zero as $n \to \infty$, where the last inequality holds by Chebyshev's inequality.

Next, we address the type II error, i.e., when $\bar{\mathbf{Y}} \in \mathcal{D}_j$ while the transmitter sent $\bar{\mathbf{u}}_i$. Then, for every $i, j \in [\![2^{nR}]\!]$, where $i \neq j$, the type II error probability is given by

$$\begin{aligned} P_{e,2}(i,j) &= \Pr\left(\left\| \bar{\mathbf{Y}} - \bar{\mathbf{u}}_j \right\|^2 \leq \sigma_Z^2 + \delta \,\middle|\, \bar{\mathbf{x}} = \bar{\mathbf{u}}_i \right) \\ &= \Pr\left(\left\| \bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j + \bar{\mathbf{Z}} \right\|^2 \leq \sigma_Z^2 + \delta \right) . \end{aligned} \tag{3.20}$$

Observe that the square norm can be expressed as

$$\left\| \bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j + \bar{\mathbf{Z}} \right\|^2 = \left\| \bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j \right\|^2 + \left\| \bar{\mathbf{Z}} \right\|^2 + 2 \sum_{t=1}^{n} (\bar{u}_{i,t} - \bar{u}_{j,t}) Z_t \,. \tag{3.21}$$

Then, define the event

$$\mathcal{E}_0 = \left\{ \left| \sum_{t=1}^{n} (\bar{u}_{i,t} - \bar{u}_{j,t}) \bar{Z}_t \right| > \frac{\delta}{2} \right\} , \tag{3.22}$$

By Chebyshev's inequality, the probability of this event vanishes,

$$\begin{aligned}\Pr(\mathcal{E}_0) &\leq \frac{\sigma_Z^2 \sum_{t=1}^{n} (\bar{u}_{i,t} - \bar{u}_{j,t})^2}{n \left(\frac{\delta}{2}\right)^2} \\ &= \frac{4\sigma_Z^2 \left\| \bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j \right\|^2}{n\delta^2} \\ &\leq \frac{16\sigma_Z^2 A}{n\delta^2} \\ &\leq \zeta \,, \end{aligned} \tag{3.23}$$

for sufficiently large n, where $\zeta > 0$ is arbitrary constant, where the first inequality holds since the sequence $\{\bar{Z}_t\}$ is i.i.d. $\sim \mathcal{N}\left(0, \frac{\sigma_Z^2}{n}\right)$, and the second inequality follows as

$$\begin{aligned}\left\| \bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j \right\|^2 &\leq (\|\bar{\mathbf{u}}_i\| + \left\| \bar{\mathbf{u}}_j \right\|)^2 \\ &\leq (\sqrt{A} + \sqrt{A})^2 \\ &= 4A \,, \end{aligned} \tag{3.24}$$

by the triangle inequality. Now let us define following event

$$\mathcal{A}_{i,j}\left(\sigma_Z^2 + \delta\right) \equiv \left\{ \bar{\mathbf{Z}} \in \mathbb{R}^n \,:\, \left\| \bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j + \bar{\mathbf{Z}} \right\|^2 \leq \sigma_Z^2 + \delta \right\} \,, \tag{3.25}$$

Observe that given the complementary event $\mathcal{E}_0^c$, we have

$$2\sum_{t=1}^{n} (\bar{u}_{i,t} - \bar{u}_{j,t}) \bar{Z}_t \geq -\delta \,, \tag{3.26}$$

hence, by (3.21), the event $\mathcal{A}_{i,j}\left(\sigma_Z^2 + \delta\right)$ implies following event

$$\mathcal{E}_1 = \left\{ \bar{\mathbf{Z}} \in \mathbb{R}^n \,:\, \left\| \bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j \right\|^2 + \left\| \bar{\mathbf{Z}} \right\|^2 \leq \sigma_Z^2 + 2\delta \right\} \,. \tag{3.27}$$

Applying the law of total probability to (3.20), we have

$$\begin{aligned} P_{e,2}(i,j) &\overset{(a)}{=} \Pr\left(\left\{ \mathcal{A}_{i,j}\left(\sigma_Z^2 + \delta\right) \right\} \cap \mathcal{E}_0 \right) + \Pr\left(\left\{ \mathcal{A}_{i,j}\left(\sigma_Z^2 + \delta\right) \right\} \cap \mathcal{E}_0^c \right) \\ &\overset{(b)}{\leq} \Pr(\mathcal{E}_0) + \Pr\left(\left\{ \mathcal{A}_{i,j}\left(\sigma_Z^2 + \delta\right) \right\} \cap \mathcal{E}_0^c \right) \\ &\overset{(c)}{\leq} \zeta + \Pr\left(\mathcal{E}_1\right) \,, \end{aligned} \tag{3.28}$$

where (a) is due to (3.25), (b) holds since each probability is bounded by 1 and (c) follows from (3.27). Based on the codebook construction, each codeword is surrounded by a sphere of radius $\sqrt{\epsilon}$, which implies

$$\left\|\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j\right\| \geq \sqrt{\epsilon}\,, \tag{3.29}$$

Hence,

$$-\left\|\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j\right\|^2 \leq -\epsilon\,. \tag{3.30}$$

Thus, choosing $\delta = \frac{\epsilon}{3}$, we obtain

$$\begin{aligned} P_{e,2}(i,j) &\leq \Pr\left(\left\|\bar{\mathbf{Z}}\right\|^2 \leq \sigma_Z^2 - \delta\right) + \zeta \\ &= \Pr\left(\sum_{t=1}^{n} \bar{Z}_t^2 - \sigma_Z^2 \leq -\delta\right) + \zeta \\ &\leq \frac{\sum_{t=1}^{n} \mathrm{Var}(\bar{Z}_t^2)}{\delta^2} + \zeta \\ &\leq \frac{n \cdot \mathbb{E}\{\bar{Z}_t^4\}}{\delta^2} + \zeta \\ &= \frac{3\sigma_Z^4}{n\delta^2} + \zeta \\ &\leq \lambda_2\,, \end{aligned} \tag{3.31}$$

for sufficiently large n, where $\lambda_2 > 0$ is arbitrary constant, since the fourth moment of a Gaussian variable $V \sim \mathcal{N}(0, \sigma_V^2)$ is $\mathbb{E}\{V^4\} = 3\sigma_V^4$.

We have thus shown that for every $\lambda_1, \lambda_2 > 0$ and sufficietnly large n, there exists a $(2^{nR}, n, \lambda_1, \lambda_2)$ code. The proof follows by taking the limits $n \to \infty$, then γ, $\delta \to 0$, hence ϵ, $\beta \to 0$ and $R \to \infty$ by (3.17). □

3.1.3 | Alternative Proof: Discretization

In this subsection, we give a second proof for the DI capacity theorem of the Gaussian channel, Theorem 3.1.1. We show that the theorem can be obtained from our result on the DMC in Theorem 2.3.1, using discretization. We show that given a Gaussian random variable $X \sim \mathcal{N}(0, A)$, the entropy of the discretized variable is approximately

$$\frac{1}{2}\log(2\pi e A) - \frac{2\Delta}{\sqrt{2\pi A}} + \log\frac{1}{\Delta}\,, \tag{3.32}$$

where $\Delta > 0$ is the discretization step. Therefore, as Δ tends to zero, the discretized entropy grows to infinity.

Our discretization procedure is similar to the one presented in [139, see Sec. 3.4.1]. Consider a Gaussian random variable $X \sim \mathcal{N}(0, A)$, hence

$$h(X) = \frac{1}{2}\log(2\pi e A)\,. \tag{3.33}$$

Let $J > 0$ be arbitrarily large and $\Delta > 0$ be arbitrarily small. Consider the discretized variable

$$\widehat{X} \in \left\{-J\Delta, -(J-1)\Delta, \cdots, -\Delta, 0, \Delta, \cdots, (J-1)\Delta, J\Delta\right\}\,, \tag{3.34}$$

obtained by mapping X to the closest discretization point $\widehat{X} = g_{J,\Delta}(X)$, such that $|\widehat{X}| \leq |X|$. Clearly, $\mathbb{E}(\widehat{X}^2) \leq \mathbb{E}(X^2) = A$. More specifically,

$$g_{J,\Delta}(x) = \begin{cases} k\Delta & k\Delta \leq x < (k+1)\Delta\,, \\ -k\Delta & -(k+1)\Delta < x \leq -k\Delta\,, \\ J\Delta & x \geq J\Delta\,, \\ -J\Delta & x \leq -J\Delta\,. \end{cases} \tag{3.35}$$

Let $\widehat{Y} = \widehat{X} + Z$ be the output corresponding to the input $\widehat{X}$ and let $\widetilde{Y} = g_{J',\Delta}(\widehat{Y})$ be a discretized version of $\widehat{Y}$ defined in the same manner. Observe that the rows of the discretized DMC from $\widehat{X}$ to $\widetilde{Y}$ are distinct for sufficiently large J and small Δ, since for every pair of inputs $x_1, x_2 \in \mathbb{R}$, $x_1 \neq x_2$, we have

$$f_Z(y - x_1) \neq f_Z(y - x_2)\,, \tag{3.36}$$

for some $y \in \mathbb{R}$ (e.g. $y = x_1$). Thus, based on Theorem 2.3.1, any rate

$$R = H(\widehat{X}) - \epsilon\,, \tag{3.37}$$

is achievable for the DMC with input $\widehat{X}$ and output $\widetilde{Y}$ under power constraint A, where $\epsilon > 0$ is arbitrarily small. By (3.35), the probability distribution of the discretized variable is specified by

$$\Pr(\widehat{X} = \pm k\Delta) = \begin{cases} p_k & k \in [\![J-1]\!]\,, \\ \sum_{k=J}^{\infty} p_k & k \in \{J, J+1, \cdots\}\,, \\ 2p_0 & k = 0\,, \end{cases} \tag{3.38}$$

where

$$p_k = \int_{k\Delta}^{(k+1)\Delta} f_X(x)\,dx\,, \tag{3.39}$$

for $k \in \{0, 1, \cdots\}$ and for the case $k = 0$ we have

$$\begin{aligned} \Pr(\widehat{X} = 0) &= \int_0^{\Delta} f_X(x)\, dx \\ &= \frac{1}{2}\int_{-\Delta}^{\Delta} f_X(x)\, dx\,, \end{aligned} \tag{3.40}$$

Thus, the corresponding entropy is bounded by

$$\begin{aligned} R + \epsilon = H(\widehat{X}) & \\ &= -\sum_{k=-J}^{J} \Pr(\widehat{X} = k\Delta) \log \Pr(\widehat{X} = k\Delta) \\ &\geq -\sum_{k=-(J-1)}^{J-1} \Pr(\widehat{X} = k\Delta) \log \Pr(\widehat{X} = k\Delta) \\ &= -2p_0 \log(2p_0) - 2\sum_{k=1}^{J-1} p_k \log p_k\,. \end{aligned} \tag{3.41}$$

Since the Gaussian density function f_X is continuous, then, by the *mean value theorem*, there exists a value x_k within each discretization interval such that

$$\begin{aligned} f_X(x_k)\Delta &= \int_{k\Delta}^{(k+1)\Delta} f_X(x)\, dx \\ &= p_k\,, \end{aligned} \tag{3.42}$$

where the last equality holds by the definition of p_k in (3.39). Plugging this into (3.41), we obtain

$$\begin{aligned} H(\widehat{X}) &\geq -2f_X(x_0)\Delta \log(2f_X(x_0)\Delta) - 2\sum_{k=1}^{J-1} f_X(x_k)\Delta \log(f_X(x_k)\Delta) \\ &= -2f_X(x_0)\Delta \log(2f_X(x_0)) - 2\sum_{k=1}^{J-1} f_X(x_k)\Delta \log(f_X(x_k)) \\ &= -2f_X(x_0)\Delta \log\Delta - 2\sum_{k=1}^{J-1} f_X(x_k)\Delta \log\Delta\,. \end{aligned} \tag{3.43}$$

Then, taking J to infinity, we have

$$\begin{aligned} &\lim_{J\to\infty} H(\widehat{X}) \\ &\geq -2f_X(x_0)\Delta \log(2f_X(x_0)) - 2\sum_{k=1}^{\infty} f_X(x_k)\Delta \log(f_X(x_k)) - \log\Delta\left(2\sum_{k=0}^{\infty} f_X(x_k)\Delta\right) \end{aligned}$$

$$= -2f_X(x_0)\Delta - 2\sum_{k=0}^{\infty} f_X(x_k)\Delta \log(f_X(x_k)) - \log\Delta\,, \tag{3.44}$$

since

$$\begin{aligned} 2\sum_{k=0}^{\infty} f_X(x_k)\Delta &= \sum_{k=-\infty}^{\infty} \Pr(X = k\Delta) \\ &= 1\,, \end{aligned} \tag{3.45}$$

As the Gaussian pdf is bounded by $f_X(x) \leq (2\pi A)^{-1/2}$, the last bound, (3.44), implies

$$\lim_{J\to\infty} H(\widehat{X}) \geq -2\sum_{k=0}^{\infty} \Delta f_X(x_k)\log(f_X(x_k)) - \frac{2}{\sqrt{2\pi A}}\Delta + \log\frac{1}{\Delta}\,. \tag{3.46}$$

At last, we take the limit $\Delta \to 0^+$. First, consider the sum. Since $f_X(x)\log f_X(x)$ is Riemann integrable,

$$\begin{aligned} \lim_{\Delta\to 0^+}\left(-2\sum_{k=0}^{\infty}\Delta f_X(x_k)\log(f_X(x_k))\right) &= -2\int_0^{\infty} f_X(x)\Delta\log(f_X(x))\,dx \\ &= -\int_{-\infty}^{\infty} f_X(x)\log f_X(x)\,dx \\ &= h(X) \\ &= \frac{1}{2}\log(2\pi eA)\,. \end{aligned} \tag{3.47}$$

The second term in the right hand side of (3.46) tends to zero as $\delta \to 0^+$. Hence, as $J \to \infty$ and $\delta \to 0^+$, we obtain $R + \epsilon = H(\widehat{X})$ converges to

$$\frac{1}{2}\log(2\pi eA) + \lim_{\Delta\to 0^+}\log\frac{1}{\Delta}\,, \tag{3.48}$$

which tends to ∞. This completes the proof. □

3.2 | Summary and Discussion

We have observed that the DI capacity of the standard memoryless Gaussian channel $Y = X + Z$ is infinite in the exponential scale, i.e.,

$$\mathbb{C}_{DI}(\mathscr{G}, L) = \infty\,, \tag{3.49}$$

for $L(n,R) = 2^{nR}$. However, for a finite blocklength n, the number of codewords must be finite. Thereby, the meaning of the infinite capacity result is that the number of

messages scales super-exponentially. This raises the question: What is the true order of the code size. In mathematical terms, what is the scale $L(n,R)$ for which the DI capacity is positive yet finite. To answer this question, we will address the general Gaussian channels with fading in the next two chapters and will realize that the number of messages scales as $\mathbf{2}^{(n\log n)R}$. As a consequence, we have deduced that the DI capacity of a standard Gaussian channel without fading is infinite only in the exponential scale of the codebook, and zero in the double exponential scale, regardless of the channel noise. Further discussions and analysis are provided in Chapter 4 and Chapter 5.

CHAPTER 4

DI FOR SLOW FADING GAUSSIAN CHANNELS

“ Slow is Smooth, Smooth is Fast ”

Mitsuyo Maeda,

In this section we consider Gaussian channels with slow fading. We will see that the capacity characterization is inherently different in the sense that for the Gaussian channel, the code size scales $L(n,R) = 2^{n\log(n)R} = n^{nR}$. We note that the scale of the DI capacity can be viewed as a special case of a tetration function, as ${}^2n = n^n = 2^{n\log(n)}$ [140, 141]. To prove this property, we establish lower and upper bounds in this scale, both positive and finite. As a consequence, it follows that the capacity is infinite in the exponential scale $L(n,R) = 2^{nR}$ and zero in the double exponential scale $L(n,R) = 2^{2^{nR}}$.

4.0.1 | Fading Channels

Consider the Gaussian channel $\mathscr{G}_{\text{fast}}$ with fast fading, specified by the input-output relation

$$\mathbf{Y} = \mathbf{G} \circ \mathbf{x} + \mathbf{Z}\,, \tag{4.1}$$

where $\mathbf{G}$ is a random sequence of fading coefficients and $\mathbf{Z}$ is an additive white Gaussian process (see Figure 4.1). Specifically, $\mathbf{G}$ is a sequence of i.i.d. continuous random variables $\sim f_G$ with finite moments, while the noise sequence $\mathbf{Z}$ is i.i.d. $\sim \mathcal{N}(0,\sigma_Z^2)$. It is assumed that the noise sequence $\mathbf{Z}$ and the sequence of fading coefficients $\mathbf{G}$ are statistically independent, and that the values of the fading coefficients belong to a bounded set $\mathcal{G}$, either countable or uncountable. The transmission power is limited to $\|\mathbf{x}\|^2 \leq nA$.

In this section, we consider the Gaussian channel $\mathscr{G}_{\text{slow}}$ with slow fading, specified by the input-output relation

$$Y_t = Gx_t + Z_t\,, \tag{4.2}$$

where G is a continuous random variable $\sim f_G(g)$. Suppose that the values of G belong to a set $\mathcal{G}$, and that G has finite expectation, and finite variance $\mathrm{var}(G) > 0$. with additive white Gaussian noise, i.e., where the noise sequence $\mathbf{Z}$ is i.i.d. $\sim \mathcal{N}(0, \sigma_Z^2)$. The transmission power is limited to $\|\mathbf{x}\|^2 \leq nA$.

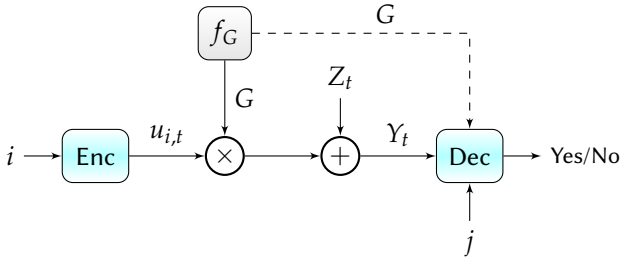


Figure 4.1: End-to-end transmission chain for DI communication in a generic wireless communication system modelled as a slow fading channel. For fast fading, $\mathbf{G} = (G_t)_{t=1}^{\infty}$ is a sequence of i.i.d. fading coefficients $\sim f_G$. For slow fading, the fading sequence remains constant throughout the transmission block, i.e., $G_t = G$.

4.0.2 | Coding with Slow Fading

We move to the Gaussian channel with slow fading. In the compound channel model, we consider the worst-case channel and the error is maximized over the set of the values of the fading coefficients (see [139, Sec. 23.3.1]). As a result, we will show that the capacity is infinite as long as the set of fading values $\mathcal{G}$ does not include zero.

The definition of DI codes with CSI available at the decoder is given below.

Definition 4.0.1 (Slow fading DI code). *An $(L(n,R), n)$ DI code for a Gaussian channel $\mathscr{G}_{slow}$ with CSI at the decoder, assuming $L(n,R)$ is an integer, is defined as a system $(\mathcal{U}, \mathscr{D})$ which consists of a codebook $\mathcal{U} = \{\mathbf{u}_i\}_{i \in [\![L(n,R)]\!]}$, $\mathcal{U} \subset \mathbb{R}^n$, such that*

$$\|\mathbf{u}_i\|^2 \leq nA \,, \tag{4.3}$$

for all $i \in [\![L(n,R)]\!]$ and a collection of decoding regions

$$\mathscr{D} = \{\mathcal{D}_{i,g}\} \,, \tag{4.4}$$

for $i \in [\![L(n,R)]\!]$ and $g \in \mathcal{G}^n$ with

$$\bigcup_{i=1}^{L(n,R)} \mathcal{D}_{i,g} \subset \mathbb{R}^n \,. \tag{4.5}$$

The error probabilities of the identification code $(\mathcal{U}, \mathscr{D})$ *are given by*

$$P_{e,1}(i) = \sup_{g\in\mathcal{G}} \left[1 - \int_{\mathcal{D}_{i,g}} \left(\prod_{t=1}^{n} f_Z(y_t - gu_{i,t})\right) d\mathbf{y}\right] , \tag{4.6}$$

$$P_{e,2}(i,j) = \sup_{g\in\mathcal{G}} \left[\int_{\mathcal{D}_{j,g}} \left(\prod_{t=1}^{n} f_Z(y_t - gu_{i,t})\right) d\mathbf{y}\right] , \tag{4.7}$$

with $f_Z(z) = \frac{1}{(2\pi\sigma_Z^2)^{1/2}} e^{-z^2/2\sigma_Z^2}$ *(see Lemma 4.1). An* $(L(n,R), n, \lambda_1, \lambda_2)$ *DI code is defined in a similar manner as for fast fading (see Section 4.0.1)*

A rate $R > 0$ *is called achievable if for every* $\lambda_1, \lambda_2 > 0$ *and sufficiently large* n*, there exists an* $(L(n,R), n, \lambda_1, \lambda_2)$ *DI code. The operational DI capacity of the Gaussian channel is defined as the supremum of achievable rates, and will be denoted by* $\mathbb{C}_{DI}(\mathscr{G}_{slow}, L)$.

Remark 4.0.1. *Consider the Gaussian channel* $\mathscr{G}_{slow}$ *with slow fading. When* $0 \in cl(\mathcal{G})$*, it immediately follows that the DI capacity is zero. To see this, observe that if* $0 \in cl(\mathcal{G})$*, then by (4.6)-(4.7),*

$$\begin{aligned}
&P_{e,1}(i) + P_{e,2}(j,i) \\
&= \sup_{g\in\mathcal{G}} \left[1 - \int_{\mathcal{D}_{i,g}} \left(\prod_{t=1}^{n} f_Z(y_t - gu_{i,t})\right) d\mathbf{y}\right] + \sup_{g\in\mathcal{G}} \left[\int_{\mathcal{D}_{j,g}} \left(\prod_{t=1}^{n} f_Z(y_t - gu_{i,t})\right) d\mathbf{y}\right] \\
&\geq \left[1 - \int_{\mathcal{D}_{i,g}} \left(\prod_{t=1}^{n} f_Z(y_t - gu_{i,t})\right) d\mathbf{y}\right]_{g=0} + \left[\int_{\mathcal{D}_{i,g}} \left(\prod_{t=1}^{n} f_Z(y_t - gu_{j,t})\right) d\mathbf{y}\right]_{g=0} \\
&= 1 .
\end{aligned} \tag{4.8}$$

Hence, in the sequel we suppose that $0 \notin cl(\mathcal{G})$.

4.0.3 | Main Result - Slow Fading

Our DI capacity theorem for the Gaussian channel with slow fading is stated below.

Theorem 4.0.1. *The DI capacity of the Gaussian channel* $\mathscr{G}_{slow}$ *with slow fading in the super-exponential scale, i.e., for* $L(n,R) = 2^{n\log(n)R}$ *is bounded by*

$$\begin{aligned}
&\tfrac{1}{4} \leq \mathbb{C}_{DI}(\mathscr{G}_{slow}, L) \leq 1 \quad &&\text{if } 0 \notin cl(\mathcal{G}) , \\
&\mathbb{C}_{DI}(\mathscr{G}_{slow}, L) = 0 \quad &&\text{if } 0 \in cl(\mathcal{G}) .
\end{aligned} \tag{4.9}$$

Hence, the DI capacity is infinite in the exponential scale, if $0 \notin cl(\mathcal{G})$,

$$\mathbb{C}_{DI}(\mathscr{G}_{slow}, L) = \begin{cases} 0 & \text{if } 0 \in cl(\mathcal{G}) \,, \\ \infty & \text{if } 0 \notin cl(\mathcal{G}) \,, \end{cases} \tag{4.10}$$

and zero in the double exponential scale, i.e., for $L(n,R) = 2^{2^{nR}}$, *we have*

$$\mathbb{C}_{DI}(\mathscr{G}_{slow}, L) = 0 \,. \tag{4.11}$$

The derivation of the above result is similar to that of the proof for the fast fading cases which will be given in Chapter 5. The proof of Theorem 4.0.1 is given in the following.

4.0.4 | Lower Bound (Achievability Proof)

Consider the Gaussian channel $\mathscr{G}_{\text{slow}}$ with slow fading. Based on Remark 4.0.1, when $0 \in \text{cl}(\mathcal{G})$, it immediately follows that the DI capacity is zero. Now, suppose that $0 \notin \text{cl}(\mathcal{G})$. We show here that the DI capacity of the Gaussian channel with slow fading can be achieved using a dense packing arrangement and a simple distance-decoder.

A DI code for the Gaussian channel $\mathscr{G}_{\text{slow}}$ with slow fading is constructed as follows. Since the decoder can normalize the output symbols by $\frac{1}{\sqrt{n}}$, we have an equivalent input-output relation,

$$\bar{Y}_t = G\bar{x}_t + \bar{Z}_t \,, \tag{4.12}$$

where $G_t = G \sim f_G$, and the noise sequence $\bar{\mathbf{Z}}$ is i.i.d. $\sim \mathcal{N}\left(0, \frac{\sigma_Z^2}{n}\right)$, with an input power constraint

$$\|\bar{\mathbf{x}}\| \leq \sqrt{A} \,, \tag{4.13}$$

with $\bar{\mathbf{x}} = \frac{1}{\sqrt{n}}\mathbf{x}$, $\bar{\mathbf{Z}} = \frac{1}{\sqrt{n}}\mathbf{Z}$, and $\bar{\mathbf{Y}} = \frac{1}{\sqrt{n}}\mathbf{Y}$.

4.0.4.1 | Codebook Construction

As in our achievability proof for the fast fading setting (see Subsection 4.0.4.1), we use a packing arrangement of non-overlapping hyper-spheres of radius $\sqrt{\epsilon_n}$ over a hyper-sphere of radius $(\sqrt{A} - \sqrt{\epsilon_n})$, with

$$\epsilon_n = \frac{A}{n^{\frac{1}{2}(1-b)}} \,, \tag{4.14}$$

where $b > 0$ is an arbitrary small. As observed in Subsection 5.0.5, there exists an arrangement

$$\bigcup_{i=1}^{2^{n\log(n)R}} \mathcal{S}_{\mathbf{u}_i}(n, \sqrt{\epsilon_n}),$$

over $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\epsilon_n})$ with a density of $\Delta_n \geq 2^{-n}$ [118, Lem. 2.1]. We assign a codeword to the center of each small sphere $\mathbf{u}_i$. Since the small spheres have the same volume, the total number of spheres, i.e., the codebook size, satisfies

$$\begin{aligned} 2^{n\log(n)R} &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{2^{n\log(n)R}} \mathcal{S}_{\mathbf{u}_i}(n, \sqrt{\epsilon_n})\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n, \sqrt{\epsilon_n}))} \\ &\geq 2^{-n} \cdot \frac{\mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\epsilon_n}))}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n, \sqrt{\epsilon_n}))} \\ &= 2^{-n} \cdot \left(\frac{\sqrt{A} - \sqrt{\epsilon_n}}{\sqrt{\epsilon_n}}\right)^n , \end{aligned} \tag{4.15}$$

in a similar manner as in Subsection 4.0.4.1, hence,

$$R \geq \frac{1}{4}(1-b) - \frac{2}{\log(n)} , \tag{4.16}$$

which tends to $\frac{1}{4}$ when $n \to \infty$ and $b \to 0$.

4.0.4.2 | Encoding

Given a message $i \in [\![L(n,R)]\!]$, transmit $\bar{\mathbf{x}} = \bar{\mathbf{u}}_i$.

4.0.4.3 | Decoding

Let

$$\begin{aligned} \delta_n &= \frac{\gamma^2 \epsilon_n}{3} \\ &= \frac{A\gamma^2}{3n^{\frac{1}{2}(1-b)}} , \end{aligned} \tag{4.17}$$

where $b > 0$ is an arbitrary small. To identify whether a message $j \in [\![L(n,R)]\!]$ was sent, given the fading coefficient g, the decoder checks whether the channel output $\bar{\mathbf{y}}$ belongs to the following decoding set,

$$\mathcal{D}_{j,g} = \left\{ \bar{\mathbf{y}} \in \mathbb{R}^n \ : \ \sum_{t=1}^{n} (\bar{y}_t - g\bar{u}_{j,t})^2 \leq \sigma_Z^2 + \delta_n \right\} . \tag{4.18}$$

4.0.4.4 | Error Analysis

Consider the type I error, i.e., when the transmitter sends $\bar{\mathbf{u}}_i$, yet $\bar{\mathbf{Y}} \notin \mathcal{D}_{i,G}$. For every $i \in [\![L(n,R)]\!]$, the type I error probability is given by

$$P_{e,1}(i) = \sup_{g \in \mathcal{G}} \left[P_{e,1}\left(i \,|g\,\right) \right] \,, \tag{4.19}$$

where we have defined

$$\begin{aligned} P_{e,1}\left(i \,|g\,\right) &\equiv \Pr\left(\sum_{t=1}^{n} (\bar{Y}_t - G\bar{u}_{i,t})^2 > \sigma_Z^2 + \delta_n \,\middle|\, \bar{\mathbf{x}} = (\bar{u}_{i,t})_{t=1}^n \,, G = g \right) \\ &= \Pr\left(\sum_{t=1}^{n} \bar{Z}_t{}^2 > \sigma_Z^2 + \delta_n \right) \,, \end{aligned} \tag{4.20}$$

for $g \in \mathcal{G}$, as the fading coefficient G and the noise vector $\bar{\mathbf{Z}}$ are statistically independent.

Now we can bound the type I error probability by

$$\begin{aligned} P_{e,1}\left(i \,|g\,\right) &= \Pr\left(\sum_{t=1}^{n} \bar{Z}_t{}^2 - \sigma_Z^2 > \delta_n \right) \\ &\leq \frac{3\sigma_Z^4}{n\delta_n^2} \\ &= \frac{27\sigma_Z^4}{A^2\gamma^4 n^b} \\ &\leq \lambda_1 \,, \end{aligned} \tag{4.21}$$

for sufficiently large n and arbitrarily small $\lambda_1 > 0$, where the first inequality holds by Chebyshev's inequality and since the fourth moment of a Gaussian variable $V \sim \mathcal{N}(0, \sigma_V^2)$ is $\mathbb{E}\{V^4\} = 3\sigma_V^4$. Thus we have $P_{e,1}\left(i \,|g\,\right) \leq \lambda_1$ for all $g \in \mathcal{G}$. Hence, the type I error probability satisfies $P_{e,1}\left(i\right) \leq \lambda_1$ (see (4.19)).

Next we address the type II error, i.e., when $\bar{\mathbf{Y}} \in \mathcal{D}_{j,G}$ while the transmitter sent $\bar{\mathbf{u}}_i$. Then, for every $i, j \in [\![L(n,R)]\!]$, where $i \neq j$, the type II error probability is given by

$$P_{e,2}(i,j) = \sup_{g \in \mathcal{G}} \left[P_{e,2}\left(i,j \,|g\,\right) \right] \,, \tag{4.22}$$

where we have defined

$$P_{e,2}\left(i,j \,|g\,\right) \equiv \Pr\left(\sum_{t=1}^{n} (\bar{Y}_t - G\bar{u}_{j,t})^2 \leq \sigma_Z^2 + \delta_n \,\middle|\, \bar{\mathbf{x}} = (\bar{u}_{i,t})_{t=1}^n \,, G = g \right)$$

$$= \Pr\left(\sum_{t=1}^{n}(g(\bar{u}_{i,t}-\bar{u}_{j,t})+\bar{Z}_t)^2 \le \sigma_Z^2+\delta_n\right) \tag{4.23}$$

for $g \in \mathcal{G}$, as the fading coefficient G and the noise vector $\bar{\mathbf{Z}}$ are statistically independent. Now we bound the probability within the square brackets.

We divide into two cases. First, consider $g \in \mathcal{G}$ such that $\left\|g\left(\bar{\mathbf{u}}_i-\bar{\mathbf{u}}_j\right)\right\| > 2\sqrt{\sigma_Z^2+\delta_n}$. Therefore, by the reverse triangle inequality, $\|\mathbf{a}-\mathbf{b}\| \ge \big|\|\mathbf{a}\|-\|\mathbf{b}\|\big|$, we have

$$\begin{aligned}\sqrt{\sum_{t=1}^{n}\left(\left(g\left(\bar{u}_{i,t}-\bar{u}_{j,t}\right)\right)+\bar{Z}_t\right)^2} &\ge \left\|g\left(\bar{\mathbf{u}}_i-\bar{\mathbf{u}}_j\right)\right\|-\|\bar{\mathbf{Z}}\| \\ &\ge 2\sqrt{\sigma_Z^2+\delta_n}-\|\bar{\mathbf{Z}}\| \ . \end{aligned} \tag{4.24}$$

Hence, for every g such that $\left\|g\left(\bar{\mathbf{u}}_\mathbf{i}-\bar{\mathbf{u}}_\mathbf{j}\right)\right\| > 2\sqrt{\sigma_Z^2+\delta_n}$, we can bound the type II error probability by

$$\begin{aligned}P_{e,2}\left(i,j\,\middle|\,g\right) &\le \Pr\left(\|\bar{\mathbf{Z}}\| \ge \sqrt{\sigma_Z^2+\delta_n}\right) \\ &= \Pr\left(\sum_{t=1}^{n}{\bar{Z}_t}^2 > \sigma_Z^2+\delta_n\right) \\ &\le \frac{3\sigma_Z^4}{n\delta_n^2} \\ &= \frac{27\sigma_Z^4}{n^b A^2\gamma^4} \\ &\le \lambda_2 \ , \end{aligned} \tag{4.25}$$

for sufficiently large n and arbitrarily small $\lambda_2 > 0$, where the second inequality follows by Chebyshev inequality.

Now, we turn to the second case, i.e., when

$$\left\|g\left(\bar{\mathbf{u}}_i-\bar{\mathbf{u}}_j\right)\right\| \le 2\sqrt{\sigma_Z^2+\delta_n} \ . \tag{4.26}$$

Observe that for every given $g \in \mathcal{G}$,

$$\sum_{t=1}^{n}(g(\bar{u}_{i,t}-\bar{u}_{j,t})+\bar{Z}_t)^2 = \sum_{t=1}^{n}g^2(\bar{u}_{i,t}-\bar{u}_{j,t})^2+\sum_{t=1}^{n}\bar{Z}_t^2+2\sum_{t=1}^{n}g(\bar{u}_{i,t}-\bar{u}_{j,t})Z_t \ . \tag{4.27}$$

Then define the event

$$\mathcal{E}_0(g) = \left\{\left|\sum_{t=1}^{n}g(\bar{u}_{i,t}-\bar{u}_{j,t})\bar{Z}_t\right| > \frac{\delta_n}{2}\right\} \ . \tag{4.28}$$

By Chebyshev's inequality, the probability of this event vanishes,

$$\begin{aligned}\Pr\left(\mathcal{E}_0(g)\right) &\leq \frac{g^2\sum_{t=1}^{n}(\bar{u}_{i,t}-\bar{u}_{j,t})^2\mathbb{E}\{\bar{Z}_t^2\}}{\left(\frac{\delta_n}{2}\right)^2} \\ &= \frac{4\sigma_Z^2\left\|g\left(\bar{\mathbf{u}}_i-\bar{\mathbf{u}}_j\right)\right\|^2}{n\delta_n^2} \\ &\leq \frac{16\sigma_Z^2\left(\sigma_Z^2+\delta_n\right)}{n\delta_n^2} \\ &\leq \tau_0\ , \end{aligned} \tag{4.29}$$

for sufficiently large n and arbitrarily small $\tau_0 > 0$, where the first inequality holds since the sequence $\{\bar{Z}_t\}$ is i.i.d. $\sim \mathcal{N}\left(0, \frac{\sigma_Z^2}{n}\right)$, and the second inequality follows from (4.26). Furthermore, observe that given the complementary event $\mathcal{E}_0^c(g)$, we have

$$2\sum_{t=1}^{n} g\left(\bar{u}_{i,t}-\bar{u}_{j,t}\right)\bar{Z}_t \geq -\delta_n\ , \tag{4.30}$$

Therefore, the event $\mathcal{E}_0^c$, the type II error event in (4.23), and the identity in (4.27) together imply that the following event occurs,

$$\mathcal{E}_1(g) = \left\{\sum_{t=1}^{n} g^2(\bar{u}_{i,t}-\bar{u}_{j,t})^2 + \sum_{t=1}^{n}\bar{Z}_t^2 \leq \sigma_Z^2 + 2\delta_n\right\}\ . \tag{4.31}$$

Now lets define

$$\mathcal{H}_{i,j}^n = \left\{\mathbf{G}\in\mathcal{G}^n\ :\ \sum_{t=1}^{n}(g(\bar{u}_{i,t}-\bar{u}_{j,t})+\bar{Z}_t)^2 \leq \sigma_Z^2+\delta_n\right\}\ . \tag{4.32}$$

Therefore, applying the law of total probability to (4.33), we have

$$\begin{aligned}P_{e,2}\left(i,j\,\middle|\,g\right) &= \Pr\left(\mathcal{H}_{i,j}^n\cap\mathcal{E}_0(g)\right)+\Pr\left(\mathcal{H}_{i,j}^n\cap\mathcal{E}_0^c(g)\right) \\ &\leq \Pr(\mathcal{E}_0(g))+\Pr\left(\mathcal{E}_1(g)\right) \\ &\leq \tau_0+\Pr\left(\mathcal{E}_1(g)\right)\ , \end{aligned} \tag{4.33}$$

where the last inequality holds by (4.29).

Now we focus on the second term in (4.33), i.e., $\Pr(\mathcal{E}_1(g))$. To this end, observe that based on the codebook construction, each codeword is surrounded by a sphere of radius $\sqrt{\epsilon_n}$, which implies that

$$\left\|\bar{\mathbf{u}}_i-\bar{\mathbf{u}}_j\right\| \geq \sqrt{\epsilon_n}\ . \tag{4.34}$$

Thus, we have

$$g^2\left\|\bar{\mathbf{u}}_i-\bar{\mathbf{u}}_j\right\|^2\geq\gamma^2\epsilon_n\,, \tag{4.35}$$

where γ is the minimal value in $\mathcal{G}$. Hence, according to (4.33),

$$\begin{aligned}P_{e,2}\left(i,j\,\middle|g\right)&\leq\Pr\left(\left\|\bar{\mathbf{Z}}\right\|^2\leq\sigma_Z^2+2\delta_n-\gamma^2\epsilon_n\right)+\tau_0\\&=\Pr\left(\left\|\bar{\mathbf{Z}}\right\|^2-\sigma_Z^2\leq-\delta_n\right)+\tau_0\,,\end{aligned} \tag{4.36}$$

(see (4.17)). Therefore, by Chebyshev's inequality,

$$\begin{aligned}P_{e,2}\left(i,j\,\middle|g\right)&\leq\Pr\left(\sum_{t=1}^{n}\bar{Z}_t^2-\sigma_Z^2\leq-\delta_n\right)+\tau_0\\&\leq\frac{\sum_{t=1}^{n}\mathrm{var}(\bar{Z}_t^2)}{\delta_n^2}+\tau_0\\&\leq\frac{\sum_{t=1}^{n}\mathbb{E}\{\bar{Z}_t^4\}}{\delta_n^2}+\tau_0\\&=\frac{3n\left(\frac{\sigma_Z^2}{n}\right)^2}{\delta_n^2}+\tau_0\\&=\frac{3\sigma_Z^4}{n\delta_n^2}+\tau_0\\&=\frac{27\sigma_Z^4}{n^bA^2\gamma^4}+\tau_0\\&\leq\lambda_2\,,\end{aligned} \tag{4.37}$$

for sufficiently large n. Based on (4.25) and (4.37), we have $P_{e,2}\left(i,j\,|g\right)\leq\lambda_2$ for all $g\in\mathcal{G}$. Hence, the type II error probability satisfies $P_{e,2}\left(i,j\right)\leq\lambda_2$ (see (4.22)).

We have thus shown that for every $\lambda_1,\lambda_2>0$ and sufficiently large n, there exists a $(2^{n\log(n)R},n,\lambda_1,\lambda_2)$ code. As we take the limits of $n\to\infty$, and then $b\to0$, the lower bound on the achievable rate tends to $\frac{1}{4}$, by (4.16). This completes the achievability proof for Theorem 4.0.1.

4.0.5 | Upper Bound (Converse Proof)

Suppose that R is an achievable rate in the L-scale for the Gaussian channel with slow fading. Consider a sequence of $(L(n,R),n,\lambda_1^{(n)},\lambda_2^{(n)})$ codes $(\mathcal{U}^{(n)},\mathcal{D}^{(n)})$, such that $\lambda_1^{(n)}$ and $\lambda_2^{(n)}$ tend to zero as $n\to\infty$. We begin with the following lemma.

Lemma 4.0.1. *Consider a sequence of codes as described above. Let $b > 0$ be an arbitrarily small constant that does not depend on n. Then there exists $n_0(b)$, such that for all $n > n_0(b)$, every pair of codewords in the codebook $\mathcal{U}^{(n)}$ are distanced by at least $\sqrt{n\epsilon_n}$, i.e.,*

$$\left\|\mathbf{u}_{i_1} - \mathbf{u}_{i_2}\right\| \geq \sqrt{n\epsilon_n}\,, \tag{4.38a}$$

where

$$\epsilon_n = \frac{A}{n^{2(1+b)}}\,, \tag{4.38b}$$

for all $i_1, i_2 \in [\![L(n,R)]\!]$, such that $i_1 \neq i_2$.

Proof. Fix λ_1 and λ_2. Let $\kappa, \theta, \zeta > 0$ be arbitrarily small. Assume to the contrary that there exist two messages i_1 and i_2, where $i_1 \neq i_2$, such that

$$\left\|\mathbf{u}_{i_1} - \mathbf{u}_{i_2}\right\| < \sqrt{n\epsilon_n} = \alpha_n\,, \tag{4.39}$$

where

$$\alpha_n \equiv \frac{\sqrt{A}}{n^{\frac{1}{2}(1+2b)}}\,. \tag{4.40}$$

Now let us define two subsets as follows

$$\mathfrak{B}_{i_1,i_2} = \left\{\mathbf{y} \in \mathcal{D}_{i_1,g} \,:\, \sum_{t=1}^{n} \left(y_t - g u_{i_2,t}\right)^2 \leq n\left(\sigma_Z^2 + \zeta\right)\right\} \tag{4.41}$$

$$\mathfrak{C}_{i_1,i_2} = \left\{\mathbf{y} \in \mathcal{Y}^n \,:\, \sum_{t=1}^{n} \left(y_t - g u_{i_2,t}\right)^2 \leq n\left(\sigma_Z^2 + \zeta\right)\right\}\,. \tag{4.42}$$

Observe that for every $g \in \mathcal{G}$,

$$\begin{aligned}
&\int_{\mathcal{D}_{i_1,g}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i_1,t}\right)\right) d\mathbf{y} \\
&= \int_{\mathfrak{B}_{i_1,i_2}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i_1,t}\right)\right) d\mathbf{y} + \int_{\mathcal{D}_{i,1,g} \setminus \mathfrak{B}_{i_1,i_2}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i_1,t}\right)\right) d\mathbf{y} \\
&\leq \int_{\mathfrak{B}_{i_1,i_2}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i_1,t}\right)\right) d\mathbf{y} + \int_{\mathfrak{C}^c_{i_1,i_2}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i_1,t}\right)\right) d\mathbf{y}\,.
\end{aligned} \tag{4.43}$$

where the last inequality holds since

$$\mathfrak{C}^c_{i_1,i_2} \supset \mathcal{D}_{i_1,g} \setminus \mathfrak{B}_{i_1,i_2}\,, \tag{4.44}$$

with $\setminus$ being the set minus operation. Consider the second integral, for which the domain is $\mathfrak{C}^c_{i_1,i_2}$ and where we denote

$$\mathbf{g} \equiv (g, g, \ldots, g) \,. \tag{4.45}$$

Then, by the triangle inequality,

$$\begin{aligned} \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,1}\| &\geq \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,2}\| - \|\mathbf{g} \circ (\mathbf{u}_{i,1} - \mathbf{u}_{i,2})\| \\ &= \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,2}\| - g\|\mathbf{u}_{i,1} - \mathbf{u}_{i,2}\| \\ &> \sqrt{n(\sigma_Z^2 + \zeta)} - g\|\mathbf{u}_{i,1} - \mathbf{u}_{i,2}\| \\ &\geq \sqrt{n(\sigma_Z^2 + \zeta)} - g\alpha_n \,. \end{aligned} \tag{4.46}$$

For sufficiently large n, this implies the following subset

$$\mathfrak{F}^c_{i_1,i_2} = \left\{ y^n \in \mathcal{Y}^n \,:\, \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,1}\| > \sqrt{n\left(\sigma_Z^2 + \eta\right)} \right\} \,, \tag{4.47}$$

for $\eta < \frac{\zeta}{2}$. That is,

$$\left\{ \mathbf{y} : \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,2}\| \geq \sqrt{n\left(\sigma_Z^2 + \zeta\right)} \right\} \overset{\text{implies}}{\longrightarrow} \left\{ \mathbf{y} : \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,1}\| \geq \sqrt{n\left(\sigma_Z^2 + \eta\right)} \right\}. \tag{4.48}$$

Thus we deduce that

$$\mathfrak{F}^c_{i_1,i_2} \supset \mathfrak{C}^c_{i_1,i_2} \,, \tag{4.49}$$

Hence, the second integral in the right hand side of (4.43) is bounded by

$$\begin{aligned} \int_{\mathfrak{F}^c_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}) d\mathbf{y} &= \Pr\left(\|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,1}\| \geq \sqrt{n\left(\sigma_Z^2 + \eta\right)} \right) \\ &= \Pr(\|\mathbf{Z}\|^2 - n\sigma_Z^2 > n\eta) \\ &\leq \frac{3\sigma_Z^4}{n\eta^2} \\ &\leq \kappa \,, \end{aligned} \tag{4.50}$$

for sufficiently large n, where the third line is due to Chebyshev's inequality, followed by the substitution of $\mathbf{z} \equiv \mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}$. Thus, by (4.43),

$$\int_{\mathcal{D}_{i_1,g}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i_1,t}\right) \right) d\mathbf{y} \leq \int_{\mathfrak{B}_{i_1,i_2}} f_{\mathbf{Z}}\left(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}\right) d\mathbf{y} + \kappa \,. \tag{4.51}$$

Now, we can focus on the first integral with domain of $\mathfrak{B}_{i_1,i_2}$, i.e., when

$$\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i,2}\right\| \leq \sqrt{n(\sigma_Z^2+\zeta)}\,. \tag{4.52}$$

Observe that

$$\begin{aligned} &f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}) - f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}) \\ &= f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})\left[1-e^{-\frac{1}{2\sigma_Z^2}\left(\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2-\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right\|^2\right)}\right]\,. \end{aligned} \tag{4.53}$$

By the triangle inequality,

$$\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right\| \leq \left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\| + g\left\|\mathbf{u}_{i_1}-\mathbf{u}_{i_2}\right\|\,. \tag{4.54}$$

Taking the square of both sides, we have

$$\begin{aligned} \left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right\|^2 &\leq \left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2 + g^2\left\|\mathbf{u}_{i_2}-\mathbf{u}_{i_1}\right\|^2 + 2\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|\cdot g\left\|\mathbf{u}_{i_2}-\mathbf{u}_{i_1}\right\| \\ &\leq \left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2 + g^2\alpha_n^2 + 2g\alpha_n\sqrt{n(\sigma_Z^2+\zeta)} \\ &= \left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2 + g^2\alpha_n^2 + 2g\frac{\sqrt{A(\sigma_Z^2+\zeta)}}{n^b}\,, \end{aligned} \tag{4.55}$$

where the second line follows from (4.39) and (4.52), and the line is due to (4.40). Thus, for sufficiently large n,

$$\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right\|^2 - \left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2 \leq \theta\,. \tag{4.56}$$

Hence,

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}) - f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}) &\leq f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})\left(1-e^{-\frac{\theta}{2\sigma_Z^2}}\right) \\ &\leq \kappa f_{\mathbf{Z}}\left(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right)\,, \end{aligned} \tag{4.57}$$

for sufficiently small $\theta > 0$, such that $1-e^{-\frac{\theta}{2\sigma_Z^2}} \leq \kappa$. Now by (4.51), we get,

$$\begin{aligned} &\lambda_1+\lambda_2 \\ &\geq P_{e,1}(i_1) + P_{e,2}(i_2,i_1) \\ &\overset{(a)}{\geq} \sup_{g\in\mathcal{G}}\left[P_{e,1}(i_1|g)\right] + \sup_{g\in\mathcal{G}}\left[P_{e,2}(i_2,i_1|g)\right] \end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\geq} \sup_{g\in\mathcal{G}} \left[P_{e,1}(i_1|g) + P_{e,2}(i_2,i_1|g)\right] \\
&\geq \sup_{g\in\mathcal{G}} \left[1 - \int_{\mathcal{D}_{i_1,g}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i_1,t}\right)\right) d\mathbf{y} + \int_{\mathcal{D}_{i_1,g}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i_2,t}\right)\right) d\mathbf{y}\right] \\
&\overset{(c)}{\geq} \sup_{g\in\mathcal{G}} \left[1 - \kappa - \int_{\mathfrak{B}_{i_1,i_2}} f_{\mathbf{Z}}\left(\mathbf{y} - \mathbf{g}\circ\mathbf{u}_{i_1}\right) d\mathbf{y} + \int_{\mathcal{D}_{i_1,g}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g}\circ\mathbf{u}_{i_2})\, d\mathbf{y}\right] \\
&\geq \sup_{g\in\mathcal{G}} \left[1 - \kappa - \int_{\mathfrak{B}_{i_1,i_2}} f_{\mathbf{Z}}\left(\mathbf{y} - \mathbf{g}\circ\mathbf{u}_{i_1}\right) d\mathbf{y} + \int_{\mathfrak{B}_{i_1,i_2}} f_{\mathbf{Z}}\left(\mathbf{y} - \mathbf{g}\circ\mathbf{u}_{i_2}\right) d\mathbf{y}\right] \\
&= \sup_{g\in\mathcal{G}} \left[1 - \kappa - \int_{\mathfrak{B}_{i_1,i_2}} \left(f_{\mathbf{Z}}\left(\mathbf{y} - \mathbf{g}\circ\mathbf{u}_{i_1}\right) - f_{\mathbf{Z}}\left(\mathbf{y} - \mathbf{g}\circ\mathbf{u}_{i_2}\right)\right) d\mathbf{y}\right] , \qquad (4.58)
\end{aligned}
$$

where (a) follows by definitions given in (4.19) and (4.22), (b) holds since supremum is sub-additive and (c) is due to (4.51). Hence, by (4.57),

$$
\begin{aligned}
\lambda_1 + \lambda_2 &\geq 1 - \kappa - \kappa \inf_{g\in\mathcal{G}} \left[\int_{\mathfrak{B}_{i_1,i_2}} f_{\mathbf{Z}}\left(\mathbf{y} - \mathbf{g}\circ\mathbf{u}_{i_1}\right) d\mathbf{y}\right] \\
&\geq 1 - 2\kappa , \qquad (4.59)
\end{aligned}
$$

which leads to a contradiction for sufficiently small κ such that $2\kappa < 1 - \lambda_1 - \lambda_2$. This completes the proof of Lemma 4.0.1. □

By Lemma 4.0.1 we can define an arrangement of non-overlapping spheres $\mathcal{S}_{\mathbf{u}_i}(n, \sqrt{n\epsilon_n})$ of radius $\sqrt{n\epsilon_n}$ centered at the codewords $\mathbf{u}_i$. Since the codewords all belong to a sphere $\mathcal{S}_{\mathbf{0}}(n, \sqrt{nA})$ of radius $\sqrt{nA}$ centered at the origin, it follows that the number of packed spheres, i.e., the number of codewords $2^{n\log(n)R}$, is bounded by

$$
\begin{aligned}
2^{n\log(n)R} &\leq \frac{\mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n, \sqrt{nA} + \sqrt{n\epsilon_n}))}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_i}(n, \sqrt{n\epsilon_n}))} \\
&= \left(\frac{\sqrt{A} + \sqrt{\epsilon_n}}{\sqrt{\epsilon_n}}\right)^n . \qquad (4.60)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
R &\leq \frac{1}{\log n} \log\left(\frac{\sqrt{A} + \sqrt{\epsilon_n}}{\sqrt{\epsilon_n}}\right) \\
&= 1 + b + \frac{\log\left(1 + \frac{1}{n^{1+b}}\right)}{\log n} , \qquad (4.61)
\end{aligned}
$$

by the same arguments as in the proof for the Gaussian channel with fast fading (see (5.75)), which tends to $1+b$ as $n \to \infty$. Now, since $b > 0$ is arbitrarily small, an achievable rate must satisfy $R \leq 1$. This completes the proof of Theorem 4.0.1. □

4.1 | Summary and Discussion

We have developed lower and upper bounds on the DI capacity of Gaussian channels with slow fading, where CSI is available at the decoder, in the scale $L(n,R) = 2^{n\log(n)R} = n^{nR}$, where n is the blocklength. We have thus established that the super-exponential scale n^{nR} is the appropriate scale for the DI capacity of the slow fading Gaussian channels. That is, the DI capacity can only be positive and finite in this coding scale. This scale is sharply different from the usual scales in the transmission and randomized-identification settings, where the codebook size scales exponentially and double exponentially, respectively. Different non-standard scales are also observed in other communication models, such as *covert identification* [142, 143], where the identification message is of size $2^{2^{\sqrt{n}R}}$. We observed that for the slow fading channels, the DI capacity in the exponential scale is infinite, unless the fading gain can be zero or arbitrarily close to zero (with positive probability), in which case the DI capacity is zero. Note, however, that this scale is comparably lower than the double exponential scale of RI coding.

CHAPTER 5

DI FOR FAST FADING GAUSSIAN CHANNELS

> *He (Grigori Perelman) Was not Fast. Speed Means Nothing.*
> *Math Doesn't Depend on Speed. It is About Depth.*
>
> Yuri Burago,

5.0.1 | Introduction

Modern communications require the transfer of enormous amounts of data in wireless systems, for cellular communication [124], sensor networks [125], smart appliances [126], and the internet of things [127], etc. Wireless communication is often modelled by fading channels with additive white Gaussian noise [128–136]. In the fast fading regime, the transmission spans over a large number of coherence time intervals [144], hence the signal attenuation is characterized by a stochastic process or a sequence of random parameters [145–148]. In some applications, the receiver may acquire channel side information (CSI) by instantaneous estimation of the channel parameters [149–151]. On the other hand, in the slow fading regime, the latency is short compared to the coherence time [144], and the behaviour is that of a compound channel [100, 152–155].

In chapter 2; see [98, 137], we addressed deterministic identification for the DMC subject to an input constraint and have also shown that the DI capacity of the standard Gaussian channel, without fading, is infinite in the exponential scale. Our previous results [98, 137] reveal a gap of knowledge in the following sense. For a finite block-length n, the number of codewords must be finite. Thereby, the meaning of the infinite capacity result is that the number of messages scales super-exponentially. The question remains what is the true order of the code size. In mathematical terms, what is the scale L for which the DI capacity is positive yet finite. Here, we will answer this question.

In this chapter, we consider deterministic identification for Gaussian channels with fast fading and slow fading, where channel side information (CSI) is available at the de-

coder. We show that for Gaussian channels, the number of messages scales as $2^{n\log(n)R}$, and develop lower and upper bounds on the DI capacity in this scale. As a consequence, we deduce that the DI capacity of a Gaussian Channel with fast fading is infinite in the exponential scale, and zero in the double-exponential scale, regardless of the channel noise. For slow fading, the DI capacity in the exponential scale is infinite, unless the fading gain can be zero or arbitrarily close to zero (with positive probability), in which case the DI capacity is zero. In comparison with the double exponential scale in RI coding, the scale here is significantly lower.

The results have the following geometric interpretation. At first glance, it may seem reasonable that for the purpose of identification, one codeword could represent two messages. While identification allows overlap between decoding regions [117], overlap at the encoder is not allowed for deterministic codes. We observe that when two messages are represented by codewords that are close to one another, then identification fails. Thus, deterministic coding imposes the restriction that the codewords need to be distanced from each other.

Based on fundamental properties of packing arrangements [118], [138], the optimal packing of non-overlapping spheres of radius $\sqrt{n\epsilon}$ contains an exponential number of spheres, and by decreasing the radius of the codeword spheres, the exponential rate can be made arbitrarily large. However, in the derivation of our lower bound in the $2^{n\log(n)R}$-scale, we pack spheres of a sub-linear radius $\sqrt{n\epsilon_n} \sim n^{1/4}$, which results in $\sim 2^{\frac{1}{4}n\log(n)}$ codewords. In this section we consider Gaussian channels with either fast fading or slow fading. We will see that the capacity characterization is inherently different in the sense that for the Gaussian channel, the code size scales $L(n,R) = 2^{n\log(n)R} = n^{nR}$. We note that the scale of the DI capacity can be viewed as a special case of a tetration function, as ${}^2n = n^n = 2^{n\log(n)}$ [140, 141]. To prove this property, we establish lower and upper bounds in this scale, both positive and finite. As a consequence, it follows that the capacity is infinite in the exponential scale $L(n,R) = 2^{nR}$ and zero in the double exponential scale $L(n,R) = 2^{2^{nR}}$.

5.0.2 | Fading Channels

Consider the Gaussian channel $\mathscr{G}_{\text{fast}}$ with fast fading, specified by the input-output relation

$$\mathbf{Y} = \mathbf{G} \circ \mathbf{x} + \mathbf{Z}\,, \tag{5.1}$$

where $\mathbf{G}$ is a random sequence of fading coefficients and $\mathbf{Z}$ is an additive white Gaussian process (see Figure 4.1). Specifically, $\mathbf{G}$ is a sequence of i.i.d. continuous random variables $\sim f_G$ with finite moments, while the noise sequence $\mathbf{Z}$ is i.i.d. $\sim \mathcal{N}(0, \sigma_Z^2)$. It is assumed that the noise sequence $\mathbf{Z}$ and the sequence of fading coefficients $\mathbf{G}$ are statistically independent, and that the values of the fading coefficients belong to a bounded set $\mathcal{G}$, either countable or uncountable. The transmission power is limited to $\|\mathbf{x}\|^2 \leq nA$.

Similarly, the Gaussian channel $\mathscr{G}_{\text{slow}}$ with slow fading is specified by the input-output relation

$$Y_t = Gx_t + Z_t \,, \tag{5.2}$$

where G is a continuous random variable $\sim f_G(g)$. Suppose that the values of G belong to a set $\mathcal{G}$, and that G has finite expectation and finite variance $\text{var}(G) > 0$ with additive white Gaussian noise, i.e., where the noise sequence $\mathbf{Z}$ is i.i.d. $\sim \mathcal{N}(0, \sigma_Z^2)$. The transmission power is limited to $\|\mathbf{x}\|^2 \leq nA$.

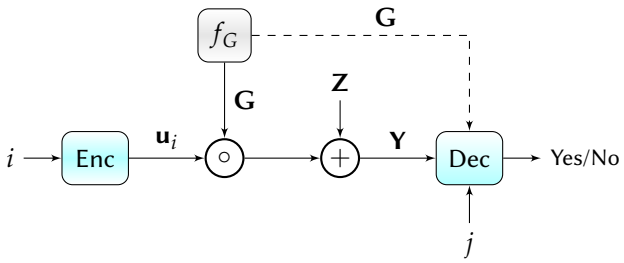


Figure 5.1: Deterministic identification for the Gaussian channel with fast fading, where $\mathbf{G}$ is a sequence of i.i.d. fading coefficients $\sim f_G$, and the noise sequence $\mathbf{Z}$ is i.i.d. $\sim \mathcal{N}(0, \sigma_Z^2)$.

5.0.3 | Coding For Fast Fading Channels

A code for the Gaussian channel with fast fading is defined below.

Definition 5.0.1 (Fast fading DI Code). *An $(L(n,R), n)$ DI code with channel side information (CSI) at the decoder for a Gaussian channel $\mathscr{G}_{fast}$ under input constraint A, assuming $L(n,R)$ is an integer, is defined as a system $(\mathcal{U}, \mathscr{D})$ which consists of a codebook $\mathcal{U} = \{\mathbf{u}_i\}_{i \in [\![L(n,R)]\!]}, \mathcal{U} \subset \mathbb{R}^n$, such that*

$$\|\mathbf{u}_i\|^2 \leq nA \,, \text{ for all } i \in [\![L(n,R)]\!] \,, \tag{5.3}$$

and a collection of decoding regions $\mathscr{D} = \{\mathcal{D}_{i,\mathbf{g}}\}_{i\in[\![L(n,R)]\!],\,\mathbf{g}\in\mathcal{G}^n}$ *with*

$$\bigcup_{i=1}^{L(n,R)} \mathcal{D}_{i,\mathbf{g}} \subset \mathbb{R}^n \,. \tag{5.4}$$

Given a message $i \in [\![L(n,R)]\!]$*, the encoder transmits* $\mathbf{u}_i$*. The decoder's aim is to answer the following question: Was a desired message* j *sent or not? There are two types of errors that may occur: Rejecting the true message, or accepting a false message. Those are referred to as type I and type II errors, respectively.*

The error probabilities of the identification code $(\mathcal{U}, \mathscr{D})$ *are given by*

$$P_{e,1}(i) = 1 - \int_{\mathcal{G}^n} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{D}_{i,\mathbf{g}}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_i) d\mathbf{y} \right] d\mathbf{g} \,, \tag{5.5}$$

$$P_{e,2}(i,j) = \int_{\mathcal{G}^n} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{D}_{j,\mathbf{g}}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_i)\, d\mathbf{y} \right] d\mathbf{g} \,, \tag{5.6}$$

with $f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi\sigma_Z^2)^{n/2}} e^{-\|\mathbf{z}\|^2/2\sigma_Z^2}$ *(see Figure 4.1). An* $(L(n,R), n, \lambda_1, \lambda_2)$ *DI code further satisfies*

$$P_{e,1}(i) \le \lambda_1 \,, \tag{5.7}$$

$$P_{e,2}(i,j) \le \lambda_2 \,, \tag{5.8}$$

for all $i, j \in [\![L(n,R)]\!]$*, such that* $i \neq j$*. A rate* $R > 0$ *is called achievable if for every* $\lambda_1, \lambda_2 > 0$ *and sufficiently large* n*, there exists an* $(L(n,R), n, \lambda_1, \lambda_2)$ *DI code. The operational DI capacity in the L-scale is defined as the supremum of achievable rates, and will be denoted by* $\mathbf{C}_{DI}(\mathscr{G}_{fast}, L)$.

Coding for the Gaussian channel with slow fading is defined as in the compound channel model, considering the worst-case channel. Thus, the error is maximized over the set of values of the fading coefficients. A code for the Gaussian channel with slow fading is defined in a similar manner as in Definition 5.0.1. However, the errors are defined with a supremum over the values of the fading coefficient $G \in \mathcal{G}$, namely,

$$P_{e,1}(i) = \sup_{g\in\mathcal{G}} \left[1 - \int_{\mathcal{D}_{i,g}} \left(\prod_{t=1}^{n} f_Z(y_t - g u_{i,t}) \right) d\mathbf{y} \right] \,, \tag{5.9}$$

$$P_{e,2}(i,j) = \sup_{g\in\mathcal{G}} \left[\int_{\mathcal{D}_{j,g}} \left(\prod_{t=1}^{n} f_Z\left(y_t - g u_{i,t}\right) \right) d\mathbf{y} \right] \,. \tag{5.10}$$

The capacity of the Gaussian channel with slow fading is denoted by $\mathbf{C}_{DI}(\mathscr{G}_{\text{slow}}, L)$.

5.0.4 | Main Results

We determine the coding scale for the DI capacity of Gaussian channels with fading. Before we give our results, we make the following observation. Recall that we use the notation of $L_2 \prec L_1$ for a coding scale L_1 that dominates L_2 (see Definition 2.2.1). The following property readily follows from the definition. Suppose that the capacity in a given scale is positive yet finite, i.e., $0 < \mathbb{C}_{DI}(\mathcal{G}_{\text{fast}}, L_0) < \infty$ for a given L_0. Then, for every $L^- \prec L_0$,

$$\mathbb{C}_{DI}(\mathcal{G}_{\text{fast}}, L^-) = \infty \,, \tag{5.11}$$

and for every $L^+ \succ L_0$,

$$\mathbb{C}_{DI}(\mathcal{G}_{\text{fast}}, L^+) = 0 \,. \tag{5.12}$$

Our DI capacity theorem for the Gaussian channel with fast fading is stated below.

Theorem 5.0.1. *Assume that the fading coefficients are positive and bounded away from zero, i.e., $0 \notin cl(\mathcal{G})$. The DI capacity of the Gaussian channel $\mathcal{G}_{fast}$ with fast fading in the $2^{n\log(n)}$-scale, i.e., for $L(n,R) = 2^{(n\log n)R}$ is bounded by*

$$\frac{1}{4} \leq \mathbb{C}_{DI}(\mathcal{G}_{fast}, L) \leq 1 \,. \tag{5.13}$$

Hence, the DI capacity is infinite in the exponential scale and zero in the double exponential scale, i.e.,

$$\mathbb{C}_{DI}(\mathcal{G}_{fast}, L) = \begin{cases} \infty & \text{for } L(n,R) = 2^{nR} \,, \\ 0 & \text{for } L(n,R) = 2^{2^{nR}} \,. \end{cases} \tag{5.14}$$

The proofs for the lower and upper bounds in the first part of Theorem 5.0.1 are given in Subsection 5.0.5 and Subsection 5.0.6, respectively. The second part of the theorem is a direct consequence of the observation given at the beginning of this subsection (see (5.11)-(5.12)).

Remark 5.0.1. *Observe that Theorem 5.0.1 assumes that the fading coefficients are positive and bounded away from zero, i.e., $0 \notin cl(\mathcal{G})$. In practice, however, communication with fading may involve small gain that can be close to zero. For instance, if the fading distribution f_G is Gaussian, then the probability $\Pr(|G_t| < \epsilon)$ is positive for arbitrarily small $\epsilon > 0$. Hence, Gaussian fading does not meet the assumption in our theorem. Unfortunately, the case*

where the fading coefficients can be arbitrarily close (or equal) to zero remains unsolved. We give a rough explanation of how we use the assumption in the analysis. In the achievability proof in Subsection 5.0.5, we assume that there exists $\gamma > 0$ such that $G_t > \gamma$ for all t with probability 1. The codebook is constructed such that the codewords are distanced from each other by $\sqrt{n\epsilon_n}$. Then, in the analysis for the type II error, we consider an error event of the form

$$\|\mathbf{Z}\|^2 \leq n(\sigma_Z^2 + 2\delta_n) - \left\|\mathbf{G} \circ (\mathbf{u}_i - \mathbf{u}_j)\right\|^2 , \tag{5.15}$$

(see (5.41)). Given our assumption, we have $\left\|\mathbf{G} \circ (\mathbf{u}_i - \mathbf{u}_j)\right\|^2 \geq \gamma^2 \left\|\mathbf{u}_i - \mathbf{u}_j\right\|^2$. By choosing $\epsilon_n = \frac{3\delta_n}{\gamma^2}$, the error event in (5.15) implies

$$\|\mathbf{Z}\|^2 \leq n(\sigma_Z^2 + 2\delta_n - \gamma^2\epsilon_n) = n(\sigma_Z^2 - \delta_n) , \tag{5.16}$$

or equivalently,

$$\frac{1}{n}\sum_{i=1} Z_i^2 - \sigma_Z^2 \leq -\delta_n . \tag{5.17}$$

As the random sequence Z_i^2 is i.i.d. with $\mathbb{E}(Z_i^2) = \sigma_Z^2$, we show that this probability tends to zero using large deviations arguments. Further details are given in Subsection 5.0.5.

5.0.5 | Lower Bound (Achievability Proof for Theorem 5.0.1)

Consider the Gaussian channel $\mathscr{G}_{\text{fast}}$ with fast fading. We show that the DI capacity is bounded by $\mathbb{C}_{DI}(\mathscr{G}_{\text{fast}}, L) \geq \frac{1}{4}$ for $L(n,R) = 2^{n\log(n)R}$. Achievability is established using a dense packing arrangement and a simple distance-decoder. A DI code for the Gaussian channel $\mathscr{G}_{\text{fast}}$ with fast fading is constructed as follows. Consider the normalized input-output relation,

$$\bar{\mathbf{Y}} = \mathbf{G} \circ \bar{\mathbf{x}} + \bar{\mathbf{Z}} , \tag{5.18}$$

where the noise sequence $\bar{\mathbf{Z}}$ is i.i.d. $\sim \mathcal{N}(0, \frac{\sigma_Z^2}{n})$, and an input power constraint

$$\|\bar{\mathbf{x}}\| \leq \sqrt{A} , \tag{5.19}$$

with $\bar{\mathbf{x}} = \frac{1}{\sqrt{n}}\mathbf{x}$, $\bar{\mathbf{Z}} = \frac{1}{\sqrt{n}}\mathbf{Z}$, and $\bar{\mathbf{Y}} = \frac{1}{\sqrt{n}}\mathbf{Y}$. Assuming $0 \notin \text{cl}(\mathcal{G})$, there exists a positive number γ such that

$$|G_t| > \gamma , \tag{5.20}$$

for all t with probability 1.

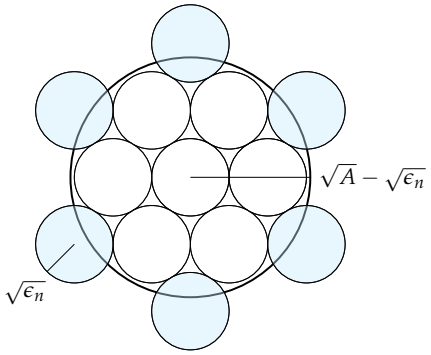


Figure 5.2: Illustration of a sphere packing, where small spheres of radius $r_0 = \sqrt{\epsilon_n}$ cover a bigger sphere of radius $r_1 = \sqrt{A} - \sqrt{\epsilon_n}$. The small spheres are disjoint from each other and have a non-empty intersection with the large sphere. Some of the small spheres, marked in yellow, are not entirely contained within the bigger sphere, and yet they are considered to be a part of the packing arrangement. As we assign a codeword to each small sphere center, the norm of a codeword is bounded by $\sqrt{A}$ as required.

5.0.5.1 | Codebook Construction

We use a packing arrangement of non-overlapping hyper-spheres of radius $\sqrt{\epsilon_n}$ in a hyper-sphere of radius $(\sqrt{A} - \sqrt{\epsilon_n})$, with

$$\epsilon_n = \frac{A}{n^{\frac{1}{2}(1-b)}}\,, \tag{5.21}$$

where $b > 0$ is arbitrarily small.

Let $\mathscr{S}$ denote a sphere packing, i.e., an arrangement of L non-overlapping spheres $\mathcal{S}_{\mathbf{u}_i}(n, r_0)$, $i \in [\![L(n,R)]\!]$ that cover a bigger sphere $\mathcal{S}_{\mathbf{0}}(n, r_1)$, with $r_1 > r_0$. As opposed to standard sphere packing coding techniques, the small spheres are not necessarily entirely contained within the bigger sphere. That is, we only require that the spheres are disjoint from each other and have a non-empty intersection with $\mathcal{S}_{\mathbf{0}}(n, r_1)$. See illustration in Figure 5.2. The packing density $\Delta_n(\mathscr{S})$ is defined as the fraction of the large sphere volume $\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, r_1)\right)$ that is covered by the small spheres, i.e.

$$\Delta_n(\mathscr{S}) \triangleq \frac{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, r_1) \cap \bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n, r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n, r_1))}\,, \tag{5.22}$$

([138, see Ch. 1]). A sphere packing is called *saturated* if no spheres can be added to the arrangement without overlap.

We use a packing argument that has a similar flavor as in the Minkowski–Hlawka theorem in lattice theory [138]. We use the property that there exists an arrangement $\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n, \sqrt{\epsilon_n})$ of non-overlapping spheres inside $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A})$ with a density

of $\Delta_n(\mathscr{S}) \geq 2^{-n}$ [118, Lem. 2.1]. Specifically, consider a saturated packing arrangement of $L(n,R) = 2^{n\log(n)R}$ spheres of radius $r_0 = \sqrt{\epsilon_n}$ covering the large sphere $\mathcal{S}_{\mathbf{0}}(n, r_1 = \sqrt{A} - \sqrt{\epsilon_n})$, i.e., such that no spheres can be added without overlap. Then, for such an arrangement, there cannot be a point in the large sphere $\mathcal{S}_{\mathbf{0}}(n,r_1)$ with a distance of more than $2r_0$ from all sphere centers. Otherwise, a new sphere could be added. As a consequence, if we double the radius of each sphere, the $2r_0$-radius spheres cover the whole sphere of radius r_1. In general, the volume of a hyper-sphere of radius r is given by

$$\mathrm{Vol}\left(\mathcal{S}_{\mathbf{x}}(n,r)\right) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r^n \,, \tag{5.23}$$

(see Eq. (16) in [138]). Hence, doubling the radius multiplies the volume by 2^n. Since the $2r_0$-radius spheres cover the entire sphere of radius r_1, it follows that the original r_0-radius packing has a density of at least 2^{-n}, i.e.,

$$\Delta_n(\mathscr{S}) \geq 2^{-n} \,. \tag{5.24}$$

We assign a codeword to the center $\mathbf{u}_i$ of each small sphere. The codewords satisfy the input constraint as

$$\begin{aligned} \|\mathbf{u}_i\| &\leq r_0 + r_1 \\ &= \sqrt{A} \,. \end{aligned} \tag{5.25}$$

Since the small spheres have the same volume, the total number of spheres is bounded from below by

$$\begin{aligned} L &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n,r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n,r_0))} \\ &\geq \frac{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n,r_1) \cap \bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n,r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n,r_0))} \\ &= \frac{\Delta_n(\mathscr{S}) \cdot \mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n,r_1)))}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n,r_0))} \\ &\geq 2^{-n} \cdot \frac{\mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n,r_1)))}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n,r_0))} \\ &= 2^{-n} \cdot \frac{r_1^n}{r_0^n} \,, \end{aligned} \tag{5.26}$$

where the second equality is due to (5.22), the inequality that follows holds by (5.24), and the last equality follows from (5.23). That is, the codebook size satisfies

$$L(n,R) = 2^{n\log(n)R}$$

$$\geq 2^{-n} \cdot \left(\frac{\sqrt{A} - \sqrt{\epsilon_n}}{\sqrt{\epsilon_n}} \right)^n . \tag{5.27}$$

Hence,

$$\begin{aligned} R &\geq \frac{1}{\log(n)} \log \left(\frac{\sqrt{A} - \sqrt{\epsilon_n}}{\sqrt{\epsilon_n}} \right) - \frac{1}{\log(n)} \\ &= \frac{1}{\log(n)} \log \left(n^{\frac{1}{4}(1-b)} - 1 \right) - \frac{1}{\log(n)} \\ &\geq \frac{1}{\log(n)} \left(\log n^{\frac{1}{4}(1-b)} - 1 \right) - \frac{1}{\log(n)} \\ &= \frac{1}{4}(1-b) - \frac{2}{\log(n)} , \end{aligned} \tag{5.28}$$

which tends to $\frac{1}{4}$ when $n \to \infty$ and $b \to 0$, where the second inequality holds, since

$$\log(t-1) \geq \log t - 1 \text{ for } t \geq 2 \tag{5.29}$$

5.0.5.2 | Encoding

Given a message $i \in [\![L(n,R)]\!]$, transmit $\bar{\mathbf{x}} = \bar{\mathbf{u}}_i$.

5.0.5.3 | Decoding

Let

$$\delta_n = \frac{\gamma^2 \epsilon_n}{3} = \frac{\gamma^2 A}{3n^{\frac{1}{2}(1-b)}} . \tag{5.30}$$

To identify whether a message $j \in [\![L(n,R)]\!]$ was sent, given the sequence $\mathbf{g}$, the decoder checks whether the channel output $\bar{\mathbf{y}}$ belongs to the following decoding set,

$$\mathcal{D}_{j,\mathbf{g}} = \left\{ \bar{\mathbf{y}} \in \mathbb{R}^n \, : \, \left\| \bar{\mathbf{y}} - \mathbf{g} \circ \bar{\mathbf{u}}_j \right\| \leq \sqrt{\sigma_Z^2 + \delta_n} \right\} . \tag{5.31}$$

5.0.5.4 | Error Analysis

Consider the type I error, i.e., when the transmitter sends $\bar{\mathbf{u}}_i$, yet $\bar{\mathbf{Y}} \notin \mathcal{D}_{i,\mathbf{G}}$. For every $i \in [\![L(n,R)]\!]$, the type I error probability is bounded by

$$\begin{aligned} P_{e,1}(i) &= \Pr \left(\left\| \bar{\mathbf{Y}} - \mathbf{G} \circ \bar{\mathbf{u}}_i \right\|^2 > \sigma_Z^2 + \delta_n \,\middle|\, \bar{\mathbf{x}} = \bar{\mathbf{u}}_i \right) \\ &= \Pr \left(\left\| \bar{\mathbf{Z}} \right\|^2 > \sigma_Z^2 + \delta_n \right) \end{aligned}$$

$$
\begin{aligned}
&= \Pr\left(\sum_{t=1}^{n} \bar{Z}_t{}^2 > \sigma_Z^2 + \delta_n\right) \\
&\leq \Pr\left(\sum_{t=1}^{n} \bar{Z}_t{}^2 > \sigma_Z^2 + \delta_n\right) \\
&\leq \frac{3\sigma_Z^4}{n\delta_n^2} \\
&= \frac{27\sigma_Z^4}{n^b A^2 \gamma^4} \\
&\leq \lambda_1 \,, \qquad (5.32)
\end{aligned}
$$

for sufficiently large n and arbitrarily small $\lambda_1 > 0$, where the second inequality follows by Chebyshev's inequality, and since the fourth moment of a Gaussian variable $V \sim \mathcal{N}(0, \sigma_V^2)$ is $\mathbb{E}\{V^4\} = 3\sigma_V^4$.

Next we address the type II error, i.e., when $\bar{\mathbf{Y}} \in \mathcal{D}_{j,\mathbf{G}}$ while the transmitter sent $\bar{\mathbf{u}}_i$. Then, for every $i, j \in [\![L(n,R)]\!]$, where $i \neq j$, the type II error probability is given by

$$
\begin{aligned}
P_{e,2}(i,j) &= \Pr\left(\left\|\bar{\mathbf{Y}} - \mathbf{G} \circ \bar{\mathbf{u}}_j\right\|^2 \leq \sigma_Z^2 + \delta_n \,\middle|\, \bar{\mathbf{x}} = \bar{\mathbf{u}}_i\right) \\
&= \Pr\left(\left\|\mathbf{G} \circ (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j) + \bar{\mathbf{Z}}\right\|^2 \leq \sigma_Z^2 + \delta_n\right) \,. \qquad (5.33)
\end{aligned}
$$

Observe that the square norm can be expressed as

$$
\left\|\mathbf{G} \circ (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j) + \bar{\mathbf{Z}}\right\|^2 = \left\|\mathbf{G} \circ (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j)\right\|^2 + \left\|\bar{\mathbf{Z}}\right\|^2 + 2\sum_{t=1}^{n} G_t(\bar{u}_{i,t} - \bar{u}_{j,t})\bar{Z}_t \,. \qquad (5.34)
$$

Then, define the event

$$
\mathcal{E}_0 = \left\{\left|\sum_{t=1}^{n} G_t(\bar{u}_{i,t} - \bar{u}_{j,t})\bar{Z}_t\right| > \frac{\delta_n}{2}\right\} \,. \qquad (5.35)
$$

By Chebyshev's inequality, the probability of this event vanishes,

$$
\begin{aligned}
\Pr(\mathcal{E}_0) &\leq \frac{\sum_{t=1}^{n}(\bar{u}_{i,t} - \bar{u}_{j,t})^2 \mathbb{E}\{G_t^2\}\mathbb{E}\{\bar{Z}_t^2\}}{\left(\frac{\delta_n}{2}\right)^2} \\
&= \frac{\sigma_Z^2(\sigma_G^2 + \mu_G^2)\sum_{t=1}^{n}(\bar{u}_{i,t} - \bar{u}_{j,t})^2}{n\left(\frac{\delta_n}{2}\right)^2} \\
&= \frac{4\sigma_Z^2(\sigma_G^2 + \mu_G^2)\left\|\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j\right\|^2}{n\delta_n^2} \,, \qquad (5.36)
\end{aligned}
$$

where the first inequality holds since the sequences $\{\bar{Z}_t\}$ and $\{G_t\}$ are i.i.d. $\sim \mathcal{N}\left(0, \frac{\sigma_Z^2}{n}\right)$ and $\sim f_G$ with

$$\mathbb{E}\{G_t\} = \mu_G \quad \text{and} \quad \mathbb{E}\{G_t^2\} = \sigma_G^2 + \mu_G^2 \,. \tag{5.37}$$

By the triangle inequality,

$$\begin{aligned} \left\|\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j\right\|^2 &\leq \left(\|\bar{\mathbf{u}}_i\| + \left\|\bar{\mathbf{u}}_j\right\|\right)^2 \\ &\leq (\sqrt{A} + \sqrt{A})^2 \\ &= 4A \,, \end{aligned} \tag{5.38}$$

hence

$$\begin{aligned} \Pr(\mathcal{E}_0) &\leq \frac{16A\sigma_Z^2(\sigma_G^2 + \mu_G^2)}{n\delta_n^2} \\ &= \frac{144\sigma_Z^2(\sigma_G^2 + \mu_G^2)}{\gamma^4 A n^b} \\ &\leq \eta_0 \,, \end{aligned} \tag{5.39}$$

for sufficiently large n, where $\eta_0 > 0$ is arbitrarily small. Furthermore, observe that given the complementary event $\mathcal{E}_0^c$, we have

$$2\sum_{t=1}^{n} G_t(\bar{u}_{i,t} - \bar{u}_{j,t})\bar{Z}_t \geq -\delta_n \,, \tag{5.40}$$

Therefore, the event $\mathcal{E}_0^c$, the type II error event in (5.33), and the identity in (5.34) together imply that the following event occurs,

$$\mathcal{E}_1 = \left\{\left\|\mathbf{G} \circ (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j)\right\|^2 + \|\bar{\mathbf{Z}}\|^2 \leq \sigma_Z^2 + 2\delta_n\right\} \,. \tag{5.41}$$

Now lets define

$$\mathcal{G}_{i,j}^n = \left\{\mathbf{G} \in \mathcal{G}^n \,:\, \left\|\mathbf{G} \circ (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j) + \bar{\mathbf{Z}}\right\|^2 \leq \sigma_Z^2 + \delta_n\right\} \,. \tag{5.42}$$

Therefore, applying the law of total probability to (5.33), we obtain

$$\begin{aligned} P_{e,2}(i,j) &= \Pr\left(\mathcal{G}_{i,j}^n \cap \mathcal{E}_0\right) + \Pr\left(\mathcal{G}_{i,j}^n \cap \mathcal{E}_0^c\right) \\ &\leq \Pr\left(\mathcal{E}_0\right) + \Pr(\mathcal{E}_1) \\ &\leq \eta_0 + \Pr(\mathcal{E}_1) \,, \end{aligned} \tag{5.43}$$

where the last inequality holds by (5.39).

Now we focus on the second term in (5.43), i.e., $\Pr(\mathcal{E}_1)$. To this end, observe that based on the codebook construction, each codeword is surrounded by a sphere of radius $\sqrt{\epsilon_n}$, which implies that

$$\left\|\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j\right\| \geq \sqrt{\epsilon_n}\,. \tag{5.44}$$

Then, by (5.20),

$$\begin{aligned}\left\|\mathbf{G}\circ(\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j)\right\|^2 &\geq \gamma^2\left\|\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_j\right\|^2\\ &\geq \gamma^2\epsilon_n\,,\end{aligned} \tag{5.45}$$

where γ is the minimal value in $\mathcal{G}$. Hence, according to (5.43),

$$\begin{aligned}P_{e,2}(i,j) &\leq \Pr\left(\left\|\bar{\mathbf{Z}}\right\|^2 \leq \sigma_Z^2 + 2\delta_n - \gamma^2\epsilon_n\right) + \eta_0\\ &\leq \Pr\left(\left\|\bar{\mathbf{Z}}\right\|^2 \leq \sigma_Z^2 - \delta_n\right) + \eta_0\,,\end{aligned} \tag{5.46}$$

where the last line holds, since $2\delta_n - \gamma^2\epsilon_n = -\delta_n$ by (5.30). Therefore, by Chebyshev's inequality,

$$\begin{aligned}P_{e,2}(i,j) &\leq \Pr\left(\sum_{t=1}^{n}\bar{Z}_t^2 - \sigma_Z^2 \leq -\delta_n\right) + \eta_0\\ &\leq \frac{\sum_{t=1}^{n}\mathrm{var}(\bar{Z}_t^2)}{\delta_n^2} + \eta_0\\ &\leq \frac{\sum_{t=1}^{n}\mathbb{E}\{\bar{Z}_t^4\}}{\delta_n^2} + \eta_0\\ &= \frac{3n\left(\frac{\sigma_Z^2}{n}\right)^2}{\delta_n^2} + \eta_0\\ &= \frac{27\sigma_Z^4}{\gamma^4 A^2 n^b} + \eta_0\\ &\leq \lambda_2\,,\end{aligned} \tag{5.47}$$

for sufficiently large n, where λ_2 is arbitrarily small.

We have thus shown that for every $\lambda_1, \lambda_2 > 0$ and sufficiently large n, there exists a $(2^{n\log(n)R}, n, \lambda_1, \lambda_2)$ code. As we take the limits of $n \to \infty$, and then $b \to 0$, the lower bound on the achievable rate tends to $\frac{1}{4}$, by (5.39). This completes the achievability proof for Theorem 5.0.1. □

5.0.6 | Upper Bound (Converse Proof for Theorem 5.0.1)

We show that the capacity is bounded by $\mathbb{C}_{DI}(\mathscr{G}_{\text{fast}}, L) \leq 1$. We note that in the converse proof, we do *not* normalize the sequences. Suppose that R is an achievable rate in the L-scale for the Gaussian channel with fast fading. Consider a sequence of $(L(n,R), n, \lambda_1^{(n)}, \lambda_2^{(n)})$ codes $(\mathcal{U}^{(n)}, \mathcal{D}^{(n)})$ such that $\lambda_1^{(n)}$ and $\lambda_2^{(n)}$ tend to zero as $n \to \infty$.

We begin with the following lemma.

Lemma 5.0.1. *Consider a sequence of codes as described above. Let $b > 0$ be an arbitrarily small constant that does not depend on n. Then there exists $n_0(b)$, such that for all $n > n_0(b)$, every pair of codewords in the codebook $\mathcal{U}^{(n)}$ are distanced by at least $\sqrt{n\epsilon_n}$, i.e.,*

$$\left\| \mathbf{u}_{i_1} - \mathbf{u}_{i_2} \right\| \geq \sqrt{n\epsilon_n} \,, \tag{5.48a}$$

where

$$\epsilon_n = \frac{A}{n^{2(1+b)}} \,, \tag{5.48b}$$

for all $i_1, i_2 \in [\![L(n,R)]\!]$ such that $i_1 \neq i_2$.

Proof. Fix λ_1 and λ_2. Let $\kappa, \theta, \zeta > 0$ be arbitrarily small. Assume to the contrary that there exist two messages i_1 and i_2, where $i_1 \neq i_2$, such that

$$\left\| \mathbf{u}_{i_1} - \mathbf{u}_{i_2} \right\| < \sqrt{n\epsilon_n} = \alpha_n \,, \tag{5.49}$$

where

$$\alpha_n = \frac{\sqrt{A}}{n^{\frac{1}{2}(1+2b)}} \,. \tag{5.50}$$

Observe that

$$\begin{aligned} \mathbb{E}\left\{ \left\| \mathbf{G} \circ (\mathbf{u}_{i_1} - \mathbf{u}_{i_2}) \right\|^2 \right\} &= \sum_{t=1}^{n} \mathbb{E}\left\{ G_t^2 \right\} \left(\mathbf{u}_{i_1,t} - \mathbf{u}_{i_2,t} \right)^2 \\ &= \mathbb{E}\left\{ G^2 \right\} \left\| \mathbf{u}_{i_1} - \mathbf{u}_{i_2} \right\|^2 \,, \end{aligned} \tag{5.51}$$

and consider the subset

$$\mathcal{A}_{i_1,i_2} = \left\{ \mathbf{g} \in \mathcal{G}^n \,:\, \left\| \mathbf{g} \circ (\mathbf{u}_{i_1} - \mathbf{u}_{i_2}) \right\| > \delta_n \right\} \,, \tag{5.52}$$

where

$$\delta_n = \frac{\sqrt{A}}{n^{\frac{1}{2}(1+b)}} \,. \tag{5.53}$$

By Markov's inequality, the probability that the fading sequence $\mathbf{G}$ belongs to this set is bounded by

$$\begin{aligned}\Pr\left(\mathbf{G}\in\mathcal{A}_{i_1,i_2}\right) &= \Pr\left(\left\|\mathbf{G}\circ(\mathbf{u}_{i_1}-\mathbf{u}_{i_2})\right\|^2 > \delta_n^2\right)\\ &\overset{(a)}{\leq} \frac{\mathbb{E}\{G^2\}\left\|\mathbf{u}_{i_1}-\mathbf{u}_{i_2}\right\|^2}{\delta_n^2}\\ &\overset{(b)}{\leq} \frac{\mathbb{E}\{G^2\}\alpha_n^2}{\delta_n^2}\\ &= \frac{\mathbb{E}\{G^2\}}{n^b}\\ &\leq \kappa\,, \end{aligned} \tag{5.54}$$

for sufficiently large n where (a) holds since the sequence $\{G_t\}_{t=1}^n$ is i.i.d. and (b) is due to (5.49).

Then, observe that

$$\begin{aligned}1-P_{e,1}(i_1) &= \int_{\mathcal{G}^n} f_{\mathbf{G}}(\mathbf{g})\left[\int_{\mathcal{D}_{i_1,\mathbf{g}}} f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})d\mathbf{y}\right]d\mathbf{g}\\ &\leq \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g})\left[\int_{\mathcal{D}_{i_1,\mathbf{g}}} f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})d\mathbf{y}\right]d\mathbf{g}+\Pr(\mathbf{G}\in\mathcal{A}_{i_1,i_2})\\ &\leq \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g})\left[\int_{\mathcal{D}_{i_1,\mathbf{g}}} f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})d\mathbf{y}\right]d\mathbf{g}+\kappa\,. \end{aligned} \tag{5.55}$$

Now let us define two subsets as follows

$$\mathcal{B}_{i_1,i_2} = \left\{\mathbf{y}\in\mathcal{D}_{i_1,\mathbf{g}} : \|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i,2}\| \leq \sqrt{n\left(\sigma_Z^2+\zeta\right)}\right\}\,, \tag{5.56}$$

$$\mathcal{C}_{i_1,i_2} = \left\{\mathbf{y}\in\mathcal{Y}^n : \|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i,2}\| \leq \sqrt{n(\sigma_Z^2+\zeta)}\right\}\,. \tag{5.57}$$

Hence,

$$\begin{aligned}&1-\kappa-P_{e,1}(i_1)\\ &\leq \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g})\left[\int_{\mathcal{D}_{i_1,\mathbf{g}}} f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})d\mathbf{y}\right]d\mathbf{g}\\ &= \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g})\left[\int_{\mathcal{B}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})d\mathbf{y}+\int_{\mathcal{D}_{i_1,\mathbf{g}}\setminus\mathcal{B}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})d\mathbf{y}\right]d\mathbf{g}\end{aligned}$$

$$\leq \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{B}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}) d\mathbf{y} + \int_{\mathcal{C}^c_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}) d\mathbf{y} \right] d\mathbf{g} \,. \tag{5.58}$$

where the last inequality holds since

$$\mathcal{C}^c_{i_1,i_2} \supset \mathcal{D}_{i_1,\mathbf{g}} \setminus \mathcal{B}_{i_1,i_2} \,, \tag{5.59}$$

with $\setminus$ being the set minus operation. Consider the second integral, for which the domain is $\mathcal{C}^c_{i_1,i_2}$. Then, by the triangle inequality

$$\begin{aligned} \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,1}\| &\geq \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,2}\| - \|\mathbf{g} \circ (\mathbf{u}_{i,1} - \mathbf{u}_{i,2})\| \\ &> \sqrt{n(\sigma_Z^2 + \zeta)} - \|\mathbf{g} \circ (\mathbf{u}_{i,1} - \mathbf{u}_{i,2})\| \\ &\geq \sqrt{n(\sigma_Z^2 + \zeta)} - \delta_n \,, \end{aligned} \tag{5.60}$$

for every $\mathbf{g} \in \mathcal{A}^c_{i_1,i_2}$ (see (5.49)). For sufficiently large n, this implies the following subset

$$\mathcal{F}^c_{i_1,i_2} = \left\{ y^n \in \mathcal{Y}^n \,:\, \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,1}\| > \sqrt{n\left(\sigma_Z^2 + \eta\right)} \right\} \,, \tag{5.61}$$

for $\eta < \frac{\zeta}{2}$. That is,

$$\left\{ \mathbf{y} \,:\, \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,2}\| \geq \sqrt{n(\sigma_Z^2 + \zeta)} \right\} \overset{\text{implies}}{\Longrightarrow} \left\{ \mathbf{y} \,:\, \|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,1}\| \geq \sqrt{n(\sigma_Z^2 + \eta)} \right\} \,. \tag{5.62}$$

Thus, we deduce that for every $\mathbf{g} \in \mathcal{A}^c_{i_1,i_2}$,

$$\mathcal{F}^c_{i_1,i_2} \supset \mathcal{C}^c_{i_1,i_2} \,, \tag{5.63}$$

Hence, the second integral in the right hand side of (5.58) is bounded by

$$\begin{aligned} \int_{\mathcal{F}^c_{i_1,i_2}} f_{\mathbf{Z}}\left(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}\right) d\mathbf{y} &= \Pr\left(\|\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,1}\| > \sqrt{n\left(\sigma_Z^2 + \eta\right)} \right) \\ &= \Pr\left(\|\mathbf{Z}\|^2 - n\sigma_Z^2 > n\eta \right) \\ &= \Pr\left(\|\mathbf{Z}\|^2 - n\sigma_Z^2 > n\eta \right) \\ &\leq \frac{3\sigma_Z^4}{n\eta^2} \\ &\leq \kappa \,, \end{aligned} \tag{5.64}$$

for sufficiently large n, where the third line is due to Chebyshev's inequality, followed by the substitution of $\mathbf{z} \equiv \mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}$. Thus, by (5.58),

$$1 - 2\kappa - P_{e,1}(i_1) \leq \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{B}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}) d\mathbf{y} \right] d\mathbf{g} \,. \tag{5.65}$$

Now, we can focus on the inner integral with domain of $\mathcal{B}_{i_1,i_2}$, i.e., when

$$\left\| \mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i,2} \right\| \leq \sqrt{n(\sigma_Z^2 + \zeta)} \,. \tag{5.66}$$

Observe that

$$f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})-f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2})=f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})\left[1-e^{-\frac{1}{2\sigma_Z^2}\left(\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2-\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right\|^2\right)}\right] . \tag{5.67}$$

By the triangle inequality,

$$\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right\|\le\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|+\left\|\mathbf{g}\circ(\mathbf{u}_{i_1}-\mathbf{u}_{i_2})\right\| . \tag{5.68}$$

Taking the square of both sides, we have

$$\begin{aligned}\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right\|^2&\le\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2+\left\|\mathbf{g}\circ(\mathbf{u}_{i_2}-\mathbf{u}_{i_1})\right\|^2+2\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|\cdot\left\|\mathbf{g}\circ(\mathbf{u}_{i_2}-\mathbf{u}_{i_1})\right\|\\&\le\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2+\delta_n^2+2\delta_n\sqrt{n(\sigma_Z^2+\zeta)}\\&=\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2+\delta_n^2+\frac{2\sqrt{A(\sigma_Z^2+\zeta)}}{n^{\frac{b}{2}}} ,\end{aligned} \tag{5.69}$$

where the last inequality follows from the definition of $\mathcal{A}_{i_1,i_2}$ and $\mathcal{B}_{i_1,i_2}$ according to (5.52) and (5.56), respectively. Thus, for sufficiently large n,

$$\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}\right\|^2-\left\|\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2}\right\|^2\le\theta . \tag{5.70}$$

Hence,

$$\begin{aligned}f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})-f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_2})&\le f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1})(1-e^{-\frac{\theta}{2\sigma_Z^2}})\\&\le\kappa f_{\mathbf{Z}}(\mathbf{y}-\mathbf{g}\circ\mathbf{u}_{i_1}) ,\end{aligned} \tag{5.71}$$

for sufficiently small $\theta>0$ such that $1-e^{-\frac{\theta}{2\sigma_Z^2}}\le\kappa$. Now by (5.65) we get

$$\lambda_1+\lambda_2$$

$$
\begin{aligned}
&\geq P_{e,1}(i_1) + P_{e,2}(i_2, i_1) \\
&\geq 1 - 2\kappa - \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{B}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}) d\mathbf{y} \right] d\mathbf{g} + \int_{\mathcal{G}^n} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{D}_{i_1,\mathbf{g}}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_2})\, d\mathbf{y} \right] d\mathbf{g} \\
&\geq 1 - 2\kappa - \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{B}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}) d\mathbf{y} \right] d\mathbf{g} + \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{B}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_2})\, d\mathbf{y} \right] d\mathbf{g} \\
&\geq 1 - 2\kappa - \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{B}_{i_1,i_2}} (f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}) - f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_2})) d\mathbf{y} \right] d\mathbf{g} \,. \qquad (5.72)
\end{aligned}
$$

Hence, by (5.71),

$$
\begin{aligned}
\lambda_1 + \lambda_2 &\geq 1 - 2\kappa - \kappa \int_{\mathcal{A}^c_{i_1,i_2}} f_{\mathbf{G}}(\mathbf{g}) \left[\int_{\mathcal{B}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - \mathbf{g} \circ \mathbf{u}_{i_1}) d\mathbf{y} \right] d\mathbf{g} \\
&\geq 1 - 3\kappa \,, \qquad (5.73)
\end{aligned}
$$

which leads to a contradiction for sufficiently small κ such that $3\kappa < 1 - \lambda_1 - \lambda_2$. This completes the proof of Lemma 5.0.1. □

By Lemma 5.0.1, we can define an arrangement of non-overlapping spheres $\mathcal{S}_{\mathbf{u}_i}(n, \sqrt{n\epsilon_n})$ of radius $\sqrt{n\epsilon_n}$ centered at the codewords $\mathbf{u}_i$. Since the codewords all belong to a sphere $\mathcal{S}_{\mathbf{0}}(n, \sqrt{nA})$ of radius $\sqrt{nA}$ centered at the origin, it follows that the number of packed spheres, i.e., the number of codewords $2^{n\log(n)R}$, is bounded by

$$\begin{aligned} 2^{n\log(n)R} &\leq \frac{\mathrm{Vol}(\mathcal{S}_{\mathbf{0}}(n, \sqrt{nA} + \sqrt{n\epsilon_n}))}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_i}(n, \sqrt{n\epsilon_n}))} \\ &= \left(\frac{\sqrt{A} + \sqrt{\epsilon_n}}{\sqrt{\epsilon_n}}\right)^n . \end{aligned} \tag{5.74}$$

Thus,

$$\begin{aligned} R &\leq \frac{1}{\log n} \log\left(\frac{\sqrt{A} + \sqrt{\epsilon_n}}{\sqrt{\epsilon_n}}\right) \\ &= \frac{\log\left(1 + n^{1+b}\right)}{\log n} \\ &= \frac{\log\left(n^{1+b}\left(1 + \frac{1}{n^{1+b}}\right)\right)}{\log n} \\ &= \frac{(1+b)\log n + \log\left(1 + \frac{1}{n^{1+b}}\right)}{\log n} \\ &= 1 + b + \frac{\log\left(1 + \frac{1}{n^{1+b}}\right)}{\log n} , \end{aligned} \tag{5.75}$$

which tends to $1 + b$ as $n \to \infty$. Therefore, $R \leq 1 + b$. Now, since $b > 0$ is arbitrarily small, an achievable rate must satisfy $R \leq 1$. This completes the proof of Theorem 5.0.1. □

5.1 | Summary and Discussion

We have developed lower and upper bounds on the DI capacity of Gaussian channels with fast fading fading, where CSI is available at the decoder, in the scale $L(n, R) = 2^{n\log(n)R} = n^{nR}$, where n is the blocklength. We have thus established that the super-exponential scale n^{nR} is the appropriate scale for the DI capacity of the fast fading

Gaussian channels. That is, the DI capacity can only be positive and finite in this coding scale. This scale is sharply different from the usual scales in the transmission and randomized-identification settings, where the codebook size scales exponentially and double exponentially, respectively. Different non-standard scales are also observed in other communication models, such as *covert identification* [142, 143], where the identification message is of size $2^{2^{\sqrt{n}R}}$.

CHAPTER 6

DI FOR POISSON CHANNELS

> *Life is Good For Only Two Things:*
> *To Do Mathematics and To Teach it.*
>
> Siméon Denis Poisson,

6.1 | Introduction

Molecular communication (MC) is a new paradigm in communication engineering where information is transmitted via signaling molecules [1, 3]. Over the past decade, synthetic MC has been extensively studied in the literature from different perspectives including channel modeling [5], modulation and detection design [7], biological building blocks for transceiver design [8], and information-theoretical performance characterization [9, 10]. Moreover, several proof-of-concept implementations of synthetic MC systems have been reported in the literature, see, e.g., [12–14]. Moreover, the ongoing progress in synthetic biology [8, 15] is expected to enable sophisticated MC systems in the future, capable of performing the complex computation and communication tasks needed for realizing the internet of bio-nano things [16].

Information-theoretical analysis of MC systems is useful not only for the characterization of their performance limits, but also for guiding MC system design and assessing the efficiency of practical designs against these performance limits. In this context, a mathematical foundation for information-theoretical analysis of diffusion-based MC is established in [11] where a channel coding theorem is proved. The information rate capacity of diffusion-based MC was studied in [30] where both channel memory and molecular noise are taken into account. For diffusion-based MC, the capacity limits of molecular timing channels are investigated in [29] and lower and upper bounds on the corresponding capacity are reported. In [31], a new characterization of capac-

ity limits and capacity achieving distributions for the particle-intensity channel are studied. Capacity bounds for point-to-point communication are studied in [10] and a corresponding mathematical framework is established. A comprehensive overview of mathematical challenges and relevant mathematical tools for studying MC channels is provided in [9]. In particular, one of the basic and widely-accepted abstract models for MC systems with molecule counting receivers is the discrete-time Poisson channel (DTPC) model [9, 156, 157]. The DTPC model has been used to study MC systems in several setups. For example, bounds on the transmission capacity of the DTPC with memory are developed in [32, 33]. An upper bound on the transmission capacity of the compound DTPC is determined in [34, 35]. A lower bound on the transmission capacity of the DTPC is reported in [36]. Analytical lower and upper bounds on the transmission capacity of the DTPC with input constraints and memory are provided in [37].

Various applications of MC within the framework of sixth generation wireless networks (6G) [26, 27] are associated with event-triggered systems, where Shannon's message transmission capacity, as considered in [9–11, 29–31], may not be the appropriate performance metric. In particular, in event-detection scenarios, where the receiver wishes to decides about the occurrence of a specific event in terms of a reliable Yes / No answer, the so-called identification capacity is the relevant performance measure [23]. Specific examples of the identification problem in the context of MC can be found in targeted drug delivery [4, 72] and cancer treatment [73–75], where, e.g., a nano-device's objective may be to identify whether or not a specific cancer biomarker is present around the target tissue; in health monitoring [70, 71] where, e.g., one may be interested in whether or not the pH value of the blood exceeds a critical threshold. Moreover, identification problems can also be found in various natural MC systems. For instance, in natural pheromone communications [77, 78] where, e.g., a male insect searches for sex pheromones indicating the presence of a nearby female insect. In fact, the olfactory systems of animal have the capability of *detecting* the presence of extremely large numbers of different molecule mixtures (e.g., pheromones, odors, etc.) [79, 80], which has motivated researchers to use them as inspiration for the design of synthetic MC systems [81]. Considering the above discussion, in this chapter, we investigate the fundamental performance limits of identification in MC systems, that can be modelled by the DTPC.

6.1.1 | Related Work on Identification Capacity

In Shannon's communication paradigm [28], a sender, Alice, encodes her message in a manner that will allow the receiver, Bob, to reliably recover the message. In other words, the receiver's task is to determine which message was sent. In contrast, in the identification setting, the coding scheme is designed to accomplish a different objective [23]. The decoder's main task is to determine whether a *particular* message was sent or not, while the transmitter does not know which message the decoder is interested in. Ahlswede and Dueck [23] introduced a randomized-encoder identification (RI) scheme, in which the codewords are tailored according to their corresponding random source (distributions). It is well-known that such distributions do not increase the transmission capacity for Shannon's message transmission task [43]. On the other hand, Ahlswede and Dueck [23] established that given local randomness at the encoder, reliable identification is accomplished with a codebook size that is double-exponential in the codeword length n, i.e., $\sim 2^{2^{nR}}$ [23], where R is the coding rate. This behavior differs radically from the conventional message transmission setting, where the codebook size grows only exponentially, with the codeword length, i.e., $\sim 2^{nR}$. Therefore, RI yields an exponential gain in the codebook size compared to the transmission problem.

The construction of RI codes is considered in [60, 61]. For example, in [61], a binary code was constructed based on a three-layer concatenated constant-weight codes. Nevertheless, realizing RI codes can be challenging in practice since they require the implementation of a random mapping function. Therefore, from a practical point of view, it is of interest to consider the case where the codewords are not selected based on a distribution but rather by means of a deterministic mapping from the message set to the input space. In the literature, this approach is also referred to as identification without randomization [53] or deterministic identification (DI) [98, 99, 105]. DI may be preferred over RI in those complexity-constrained applications of MC systems where the generation of random codewords in a controlled manner according to a specific distribution is challenging.[1]

In the deterministic coding setup for identification, for DMCs, the codebook size grows only exponentially in the codeword length, similar to the conventional transmission problem [23, 52, 53, 98, 113]. However, the achievable identification rates are

[1] On the other hand, we note that the biological hardware of MC systems (e.g., using reaction networks) features an inherent stochastic nature [158] which can potentially be exploited for realizing RI.

significantly higher compared to the transmission rates [98, 99]. Deterministic codes often have the advantage of simpler implementation and simulation [66, 67] and explicit construction [68]. In chapter 2, 4 and 5 [98, 99, 105], we have considered DI for channels with an average power constraint, including DMCs and Gaussian channels with fast and slow fading, respectively. In the Gaussian case, we have shown that the codebook size scales as $2^{(n\log n)R}$, by deriving bounds on the DI capacity. Furthermore, DI for Gaussian channels is also studied in [46, 64, 69, 105].

6.1.2 | Contributions

In this chapter, we consider MC systems employing molecule counting receivers, where the received signal has been shown to follow the Poisson distribution[2] when the number of released molecules is large, see [5, Sec. IV], [33, 162] for details. Thereby, our main objective is to investigate the fundamental performance limits of DI over the DTPC. Specifically, this chapter makes the following contributions:

- ♢ We formulate the problem of DI over the DTPC under average and peak power constraints to account for the limited molecule production / release rates of the transmitter. To the best of the authors' knowledge, the DI capacity of the DTPC has not been studied in the literature, yet. Moreover, to model the diverse applications of coded identification in the context of MC, we introduce two approaches to realize the channel uses within each codeword, namely spatial and temporal channel uses. For the latter, different channel uses for each codeword are realized by releasing the same type of molecules in different time instances, whereas, for the former, a different type of signaling molecule is used for each channel use. The spatial channel use can be used to model molecule-mixture communications in mammalian and insect olfactory systems, where a given mixture of different types of molecules represents a codeword [81, 163].

- ♢ We derive lower and upper bounds on the DI capacity of the DTPC, which are the main results of this chapter. In particular, as a key finding, we establish that the codebook size for deterministic encoding scales as $2^{(n\log n)R}$. This result is in contrast with the scaling of the codebook size for conventional transmission (i.e.,

[2] The DTPC has been used to model other communication systems such as optical communication systems with direct-detection receivers [159–161].

2^{nR} [28]) and RI (i.e., $2^{2^{nR}}$ [23]). The enlarged codebook size of the identification problem compared to the transmission problem may have interesting implications for MC system design. For instance, it may help explain the extremely large identification capability of natural olfactory systems and guide the design of olfactory-inspired synthetic MC systems [81] by e.g. determining the maximum number of identifiable molecule mixtures.

♢ The general tools used in this chapter to derive the bounds on the capacity are similar to those for Gaussian channels in [105]. However, the explicit techniques used in the analysis are different. In particular, to obtain the proposed lower bound, we exploit the existence of an appropriate sphere packing within the input space where the distance between the centers of the spheres does not fall below a certain value. However, we consider the packing of hyper spheres with radius $\sim n^{\frac{1}{4}}$ inside a larger hyper cube. While the radius of the small spheres in the Gaussian case [105] tends to zero, here the radius grows in the codeword length, n. Yet, we show that we can pack a super-exponential number of spheres within the larger cube. For the proposed upper bound, we assert a certain minimum distance between the codewords of any given sequence of codes with vanishing error probabilities. However, for the DTPC, the derivation of the upper bound on the capacity is more involved compared to that for the Gaussian channel [105] and leads to a larger upper bound. In fact, instead of establishing a minimum distance between the codewords (codeword-wise distance), as in the Gaussian case [105], we use a criterion imposed on the symbols of every two codewords, namely, we show that for each pair of codewords, there exists at least one index for which the ratio of the corresponding symbols is different from 1 (symbol-wise distance).

♢ We note that our theoretical results target the capacity of the DI channel in the standard asymptotic definition, i.e., as $n \to \infty$, and the explicit construction of identification codes is beyond the scope of this chapter. However, to gain insights into the performance of practical codes, we provide simulation results for a simple heuristic code where the size of the corresponding codebook is super-exponential in n. Surprisingly, our simulation results reveal that the error rate of the considered sub-optimal code decays fast already at small codeword length, e.g., $n = 20 - 30$, which is agreement with our asymptotic theoretical results.

6.1.3 | Organization

The remainder of this chapter is structured as follows. In Section 6.2, scenarios for application of DI in the context of MC are discussed and the required preliminaries regarding DI codes are established. Section 6.2.4 provides the main contributions and results on the message identification capacity of the DTPC. Section 6.3 presents simulation results for the empirical type I and type II error rates. Finally, Section 6.4 of the paper concludes with a summary and directions for future research.

6.2 | System Model and MC Scenarios for DI

In this section, we present the adopted system model, introduce MC scenarios for DI, and establish some preliminaries regarding DI coding.

6.2.1 | System Model

We focus on an identification setup, where the decoder wishes to reliably determine whether or not a particular message was sent by the transmitter, while the transmitter does not know which message the decoder is interested in, see Figure 6.1. To achieve this objective, we establish coded communication between the transmitter and the receiver over n channel uses of an MC channel. We consider a stochastic release model, where for the t-th channel use, the transmitter releases molecules with rate x_t (molecules/second) over a time interval of T_{rls} seconds into the channel [9]. These molecules propagate through the channel via diffusion and/or advection, and may even be degraded in the channel via enzymatic reactions [5]. We assume a counting-type receiver which is able to count the number of received molecules. Examples include the transparent (perfect monitoring or passive) receiver, which counts the molecules at a given time within its sensing volume [14], the fully absorbing (perfect sink) receiver, which absorbs and counts the molecules hitting its surface within a given time interval [164], and the reactive (ligand-based) receiver which counts the number of molecules bound to the ligand proteins on its sensing surface at a given time [165].

Assuming that the release, propagation, and reception of individual molecules are statistically similar but independent of each other, the received signal follows Poisson statistics when the number of released molecules is large, i.e., $x_t T_{\text{rls}} \gg 1$ [5, Sec. IV]. Let $X \in \mathbb{R}_{\geq 0}$ and $Y \in \mathbb{N}_0$ denote random variables (RVs) modeling the rate of molecule release by the transmitter and the number of molecules observed at the receiver, respec-

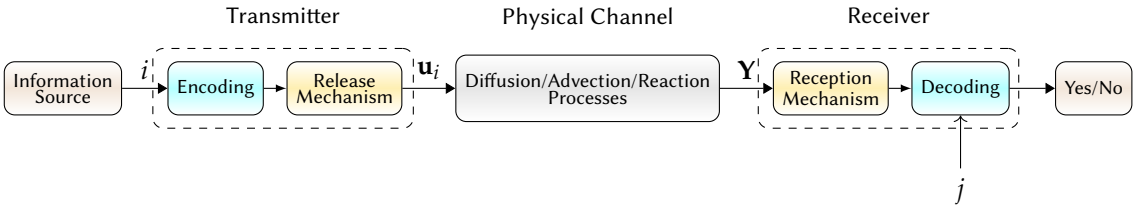


Figure 6.1: End-to-end transmission chain for DI communication in a generic MC system modelled as a DTPC. Relevant processes in the molecular channel include diffusion, advection, and chemical reactions. Transmitter maps message i onto a codeword $\mathbf{u}_i$. Receiver is provided with an arbitrary message j and given the channel output vector $\mathbf{Y}$, asks whether j is identical to i or not.

tively. For the DTPC, the channel output Y is related to the channel input X according to

$$Y = \text{Pois}\left(\rho X + \lambda\right) \ , \tag{6.1}$$

where ρX is the mean number of observed molecules due to their release by the transmitter, $\rho = p_{\text{ch}} T_{\text{rls}}$, and $p_{\text{ch}} \in (0,1]$ denotes the probability that a given molecule released by the transmitter is observed at the receiver. The value of p_{ch} depends on the propagation environment (e.g., diffusion, advection, and reaction processes) and the reception mechanism (e.g., transparent, absorbing, or reactive receiver) as well as the distance between transmitter and receiver, see [5, Sec. III] for the characterization of p_{ch} for various setups. Moreover, $\lambda \in \mathbb{R}_{>0}$ is the mean number of observed interfering molecules originating from external noise sources which employ the same type of molecule as the considered MC system.

The letter-wise conditional distribution of the DTPC output is given by

$$W(y|x) = \frac{e^{-(\rho x+\lambda)}(\rho x+\lambda)^y}{y!} \, . \tag{6.2}$$

Standard transmission schemes employ strings of letters (symbols) of length n, referred to as codewords, that is, the encoding schemes use the channel in n consecutive times to transmit one message. As a consequence, the receiver observes a string of length n, referred to as output vector (received signal). We assume that different channel uses are orthogonal. This assumption is justified for different MC scenarios in Section 6.2.2. Therefore, for n channel uses, the transition probability law reads

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{n} W(y_t|x_t) = \prod_{t=1}^{n} \frac{e^{-(\rho x_t+\lambda)}(\rho x_t+\lambda)^{y_t}}{y_t!} \, , \tag{6.3}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ denote the transmit codeword and the received signal, respectively. The codewords are subject to peak and average power constraints as follows

$$0 \leq x_t \leq P_{\max} \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^{n} x_t \leq P_{\text{avg}}\,, \tag{6.4}$$

respectively, $\forall t \in [\![n]\!]$, where $P_{\max} > 0$ and $P_{\text{avg}} > 0$ constrain the rate of molecule release per channel use and over the entire n channel uses in each codeword, respectively. We note that while the average power constraint for the Gaussian channel is a non-linear (square) function of the symbols (signifying the signal energy), here for the DTPC, it is a linear function (signifying the number of released molecules) [9].

6.2.2 | Spatial vs. Temporal Channel Uses

The coded communication considered in this chapter requires n independent uses of the MC channel; however, how the MC channel is accessed for each channel use may depend on the application of interest. In the following, we introduce two application scenarios, which employ spatial and temporal channel uses, respectively.

- **Spatial Channel Use**: In olfactory-based communications, signalling chemicals are, e.g., odorants and pheromones. Odorants belong to large chemical substance classes with a high diversity of chemical features that share a certain degree of volatility and both polar and nonpolar properties. They can be as diverse as being esters, alcohols, thiols, terpenoids, and aromatic substances to name a few [80]. Pheromones are chemical molecule substances that are used for communication among the individuals of a species and can regulate their behavior [77]. Odorants and pheromones are often not composed of only one type of molecule but a mixture of various molecule types [163]. Therefore, the olfactory-based communication using odorants and pheromones is referred to as "molecule mixture" communication [81]. The *Ahlswede-Dueck* identification problem applies to natural olfaction since each molecular mixture conveys a particular message that the receiver may be interested in. The molecule mixtures can be *interpreted* as codewords since the structural composition of these mixtures enables their reliable identification by the olfactory systems even at very low concentrations [80]. Motivated by this, we consider a communication scenario, where the transmitter

releases a mixture of n different types of molecules to convey a message to the receiver, see Figure 6.2. The receiver is equipped with a dedicated type of receptor for each type of molecule, which ensures the orthogonality of the n channel uses. The receiver's task is to determine whether or not a desired message (molecular mixture) has been sent by the transmitter.

- **Temporal Channel Use**: For spatial channel uses, the complexity of transmitter and receiver may be high as they have to be able to generate and detect n different types of molecules, respectively. To avoid this complexity, one may employ only one type of molecule and access the MC channel at different time instances. Thereby, transmitter and receiver have to be equipped with memory for generation and processing of all n channel uses, respectively. In addition, due to the dispersive nature of the diffusive MC channel, the channel has memory and proper measures have to be taken to ensure the orthogonality of different channel uses. An immediate approach is to make the symbol duration sufficiently large such that the channel response (practically) decays to zero within each symbol interval. However, this may lead to an inefficient design due to the reduction of the rate of channel access. More efficient approaches proposed in the literature include the use of enzymes [166] and reactive cleaning molecules [157] to generate a concentrated channel response for the desired signaling molecules, see, e.g., [5, Fig. 15].

Figure 6.2: Illustration of an olfactory-inspired MC system, where three orthogonal molecule types, namely type square, triangle, and circle, are shown. The transmitter secretes a mixture of these molecules corresponding to a particular message. Each receptor located on the receiver surface is sensitive to only one type of molecule. The receiver's task is to determine whether or not a desired message (molecular mixture) has been sent by the transmitter.

Remark 6.2.1. *We note that in the case of the spatial channel use, each type of molecules may be observed at the receiver over a period of time. Hence, the receiver might use multiple temporal observation samples for each type of molecules, which can be combined to improve the overall quality of the received signal. However, the temporal samples still correspond to one spatial channel use and does not contain new information.*

6.2.3 | DI Coding for the DTPC

The definition of a DI code for the DTPC is given below.

Definition 6.2.1 (Poisson DI Code). *An $\left(L(n,R),n,\lambda_1,\lambda_2\right)$ DI code for a DTPC $\mathcal{W}$ under average and peak power constraints of P_{ave} and P_{max}, respectively, and for integer $L(n,R)$, where n and R are the codeword length and coding rate, respectively, is defined as a system $(\mathcal{U},\mathscr{D})$ which consists of a codebook $\mathcal{U}=\{\mathbf{u}_i\}_{i\in[\![L]\!]}\subset\mathcal{R}^n$, such that*

$$0\leq u_{i,t}\leq P_{max}\quad \textit{and}\quad \frac{1}{n}\sum_{t=1}^{n}u_{i,t}\leq P_{avg}\,, \tag{6.5}$$

$\forall i\in[\![L]\!]$, $\forall t\in[\![n]\!]$, and a collection of decoding regions $\mathscr{D}=\{\mathcal{D}_i\}_{i\in[\![L]\!]}$ with

$$\bigcup_{i=1}^{L(n,R)}\mathcal{D}_i\subset\mathbb{N}_0^n\,.$$

Given a message $i\in[\![L]\!]$, the encoder transmits $\mathbf{u}_i$, see Figure 6.1, the decoder's aim is to answer the following question: Was a desired message j sent or not? There are two types of errors that may occur: Rejection of the true message or acceptance of a false message. These errors are referred to as type I and type II errors, respectively.
The corresponding error probabilities of the identification code $(\mathcal{U},\mathscr{D})$ are given by

$$P_{e,1}(i)=1-\sum_{\mathbf{y}\in\mathcal{D}_i}W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_i\right)\quad \textit{(miss-identification error)}\,, \tag{6.6}$$

$$P_{e,2}(i,j)=\sum_{\mathbf{y}\in\mathcal{D}_j}W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_i\right)\qquad \textit{(false identification error)}\,. \tag{6.7}$$

and satisfy the following bounds

$$P_{e,1}(i)\leq\lambda_1\quad \textit{and}\quad P_{e,2}(i,j)\leq\lambda_2\,, \tag{6.8}$$

$\forall\, i,j\underset{i\neq j}{\in}[\![L]\!]$ and every $\lambda_1,\lambda_2>0$. A rate $R>0$ is called achievable if for every $\lambda_1,\lambda_2>0$ and sufficiently large n, there exists an $\left(L(n,R),n,\lambda_1,\lambda_2\right)$ DI code. The operational DI

capacity of the DTPC is defined as the supremum of all achievable rates, and is denoted by $\mathbb{C}_{DI}(\mathcal{W}, L)$.

In this section, we first present our main results, i.e., lower and upper bounds on the achievable identification rates for the DTPC. Subsequently, we provide the detailed proofs of these bounds.

6.2.4 | Main Results

The DI capacity theorem for the DTPC is stated below.

Theorem 6.2.1. *The DI capacity of the DTPC* $\mathcal{W}$ *subject to average and peak power constraints of the form* $n^{-1}\sum_{t=1}^{n} u_{i,t} \leq P_{ave}$ *and* $0 \leq u_{i,t} \leq P_{max}$*, respectively, in the super-exponential scale, i.e.,* $L(n,R) = 2^{(n\log n)R}$*, is bounded by*

$$\frac{1}{4} \leq \mathbb{C}_{DI}(\mathcal{W}, L) \leq \frac{3}{2}. \tag{6.9}$$

Proof. The proof of Theorem 6.2.1 consists of two parts, namely the achievability and the converse proofs, which are provided in Sections 6.2.5 and 6.2.6, respectively. □

Before we provide the proof, we highlight some insights obtained from Theorem 6.2.1 and its proof.

Scale: Theorem 6.2.1 shows a different behavior compared to the traditional scaling of the codebook size with respect to codeword length n. The bounds given in Theorem 6.2.1 are valid in the super-exponential scale of $L = 2^{(n\log n)R}$ which is in between the conventional exponential and double exponential codebook sizes (see Figure 6.3). In other words, Theorem 6.2.1 reveals that for the capacity to assume informative non-zero finite values, the coding rate should be defined in the following scale in terms of n:

$$R = \frac{\log L}{n\log n}. \tag{6.10}$$

The capacity values in the standard codebook sizes, i.e., exponential and double exponential, are infinite ($\lim_{n\to\infty} \frac{\log L}{n} = \infty$) and zero ($\lim_{n\to\infty} \frac{\log\log L}{n} = 0$), respectively [98, see Rem. 1].

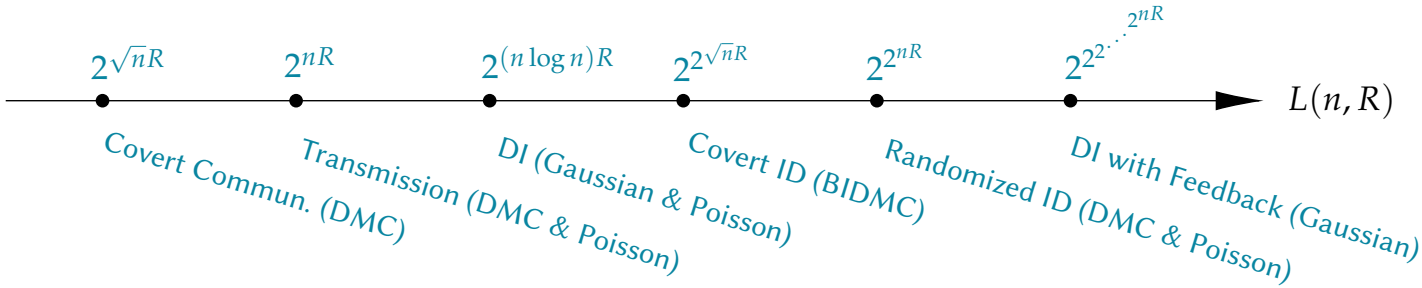


Figure 6.3: Spectrum of codebook sizes for different transmission and identification setups. Apart from the conventional exponential and double exponential codebook sizes for transmission [28] and RI [23], respectively, different non-standard codebook sizes are observed for other communication tasks, such as *covert communication* [167, 168] or *covert identification* [142] for the binary-input DMC (BIDMC), where the codebook size scales as $2^{\sqrt{n}R}$ and $2^{2^{\sqrt{n}R}}$, respectively. For the Gaussian DI channel with feedback [64, 69], the codebook size can be arbitrarily large.

Budget for Molecule Release: The proposed capacity bounds in the super exponential scale are independent of the values of P_{ave} and $P_{\max}$ as long a the codeword length n grows sufficiently large, i.e., $n \to \infty$. However, for finite n, the codebook size is indeed a function of P_{ave} and $P_{\max}$. This can be readily seen from the achievability proof, where the codebook size in its raw form (see (6.23)) before division by the dominant term reads

$$L(n,R) = 2^{(n\log n)R + n(\log \frac{A}{e\sqrt{a}}) + o(n)} \,. \tag{6.11}$$

where $A = \min(P_{\text{ave}}, P_{\max})$ and $a > 0$ is a parameter of the codebook construction, cf. (6.14). In other words, the codebook size increases as A increases; however, since A appears in a term that is exponential in n, i.e., $\sim 2^{n(\log \frac{A}{e\sqrt{a}})}$, the influence of A becomes negligible compared to the dominant super-exponential term, i.e., $2^{(n\log n)R}$ as $n \to \infty$[3]. While the proof of Theorem 6.2.1 mainly concerns the asymptotic regime of $n \to \infty$, we are still able to get some insight for finite n, too. For instance, the error constraints in (6.8) can be met by the proposed achievable scheme even for finite n if A is sufficiently large and $a = \Omega(A^2)$, cf. (6.33), (6.43), and (6.45). A comprehensive study of the achievable DI rates for finite n constitutes an interesting research topic for future work, but is beyond the scope of this chapter.

Adopted Decoder: For the achievability proof, we adopt a decoder that upon observing an output sequence $\mathbf{y}$, declares that the message j was sent if the following condition is met

$$\left| \|\mathbf{y} - \mathbb{E}(\mathbf{Y}|\mathbf{u}_j)\|^2 - \|\mathbf{y}\|_1 \right| \leq n\delta_n \,, \tag{6.12}$$

where $\mathbf{u}_j = [u_{j,1}, \ldots, u_{j,n}]$ is the codeword associated with message j and δ_n is a decoding threshold. In contrast to the popular distance decoder used for the Gaussian channels [105] that includes only the distance term $\|\mathbf{y} - \mathbb{E}(\mathbf{Y}|\mathbf{u}_j)\|$, the proposed decoder in (7.25) comprises the additional correction term $\|\mathbf{y}\|_1$. This choice stems from the fact that the noise in the DTPC is signal dependent [5]. Therefore, the variance of $\|\mathbf{y} - \mathbb{E}(\mathbf{Y}|\mathbf{u}_j)\|$ depend on the adopted codeword $\mathbf{u}_j$ which implies that unlike the Gaussian channel, here the radius of the decoding region is not constant for all the codewords. To account for this fact, we include a correction term of $\|\mathbf{y}\|_1$.

[3] It is interesting to recall that the codebook size for the transmission capacities of both the DTPC [169, see Eq. (5)] and the Gaussian channel [170, 171] scale with $2^{n\log\sqrt{A}}$ in terms of A.

6.2.5 | Achievability

Consider DTPC $\mathcal{W}$. We show achievability of (6.9) using a packing of hyper spheres and a distance decoder. We pack hyper spheres with radius $\sim n^{\frac{1}{4}}$ inside a larger hyper cube. While the radius of the spheres in a similar proof for Gaussian channels vanishes, as n increases [105], the radius here diverges to infinity. Yet, we can obtain a positive rate while packing a super-exponential number of spheres satisfying the power and error constraints in (6.8). A DI code for the DTPC $\mathcal{W}$ is constructed as follows.

6.2.5.1 | Codebook construction

Let

$$A = \min\left(P_{\text{ave}}, P_{\max}\right) \ . \tag{6.13}$$

In the following, we restrict ourselves to codewords that meet the condition $0 \leq x_t \leq A$, $\forall\, t \in [\![n]\!]$. We argue that this condition ensures both the average and the peak power constraints in (6.4). In particular, when $P_{\text{ave}} \geq P_{\max}$, then $A = P_{\max}$ and the constraint $0 \leq x_t \leq A$ automatically implies that the constraint $\frac{1}{n}\sum x_t \leq P_{\text{ave}}$ is met, hence, in this case, the setup with average and peak power constraints simplifies to the case with only a peak power constraint. On the other hand, when $P_{\text{ave}} < P_{\max}$, then $A = P_{\text{ave}}$ and by $0 \leq x_t \leq A$, $\forall\, t \in [\![n]\!]$, both power constraints are met, namely $\frac{1}{n}\sum x_t \leq P_{\text{ave}}$ and $0 \leq x_t \leq P_{\max}$, $\forall\, t \in [\![n]\!]$. Hence, in the following, we restrict our considerations to a hyper cube with edge length A.

We use a packing arrangement of non-overlapping hyper spheres of radius $r_0 = \sqrt{n\epsilon_n}$ in a hyper cube with edge length A, where

$$\epsilon_n = \frac{a}{n^{\frac{1}{2}(1-b)}} \ , \tag{6.14}$$

and $a > 0$ is a non-vanishing fixed constant and $0 < b < 1$ is an arbitrarily small constant[4].

Let $\mathscr{S}$ denote a sphere packing, i.e., an arrangement of L non-overlapping spheres $\mathcal{S}_{\mathbf{u}_i}(n, r_0)$, $i \in [\![L]\!]$, that are packed inside the larger cube $\mathcal{Q}_{\mathbf{0}}(n, A)$ with an edge length A, see Figure 6.4. As opposed to standard sphere packing coding techniques [138], the spheres are not necessarily entirely contained within the cube. That is, we only require

[4] We note that our achievability proof is valid for any $b \in (0,1)$; however, arbitrarily small values of b leads to the tightest lower bound and hence are of interest here.

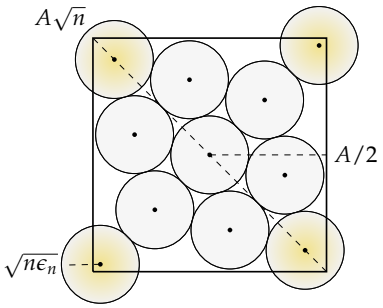


Figure 6.4: Illustration of a saturated sphere packing inside a cube, where small spheres of radius $r_0 = \sqrt{n\epsilon_n}$ cover a larger cube. Yellow colored spheres are not entirely contained within the larger cube, and yet they contribute to the packing arrangement. As we assign a codeword to each sphere center, the 1-norm and arithmetic mean of a codeword are bounded by A as required.

that the centers of the spheres are inside $\mathcal{Q}_{\mathbf{0}}(n, A)$ and are disjoint from each other and have a non-empty intersection with $\mathcal{Q}_{\mathbf{0}}(n, A)$. The packing density $\Delta_n(\mathscr{S})$ is defined as the ratio of the saturated packing volume to the cube volume $\text{Vol}\left[\mathcal{Q}_{\mathbf{0}}(n, A)\right]$, i.e.,

$$\Delta_n(\mathscr{S}) \triangleq \frac{\text{Vol}\left(\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n, r_0)\right)}{\text{Vol}\left[\mathcal{Q}_{\mathbf{0}}(n, A)\right]}. \tag{6.15}$$

Sphere packing $\mathscr{S}$ is called *saturated* if no spheres can be added to the arrangement without overlap. In particular, we use a packing argument that has a similar flavor as that observed in the Minkowski–Hlawka theorem for saturated packing [138]. Specifically, consider a saturated packing arrangement of

$$\bigcup_{i=1}^{L(n,R)} \mathcal{S}_{\mathbf{u}_i}(n, \sqrt{n\epsilon_n}) \tag{6.16}$$

spheres with radius $r_0 = \sqrt{n\epsilon_n}$ embedded within cube $\mathcal{Q}_{\mathbf{0}}(n, A)$. Then, for such an arrangement, we have the following lower [118, Lem. 2.1] and upper bounds [138, Eq. 45] on the packing density

$$2^{-n} \leq \Delta_n(\mathscr{S}) \leq 2^{-0.599n}. \tag{6.17}$$

We use the above lower bound in our subsequent analysis which can be proved as follows: For the saturated packing arrangement given in (6.16), there cannot be a point in the larger cube $\mathcal{Q}_{\mathbf{0}}(n, A)$ with a distance of more than $2r_0$ from all sphere centers. Otherwise, a new sphere could be added which contradicts the assumption that the union of $L(n, R)$ spheres with radius $\sqrt{n\epsilon_n}$, are saturated. Now, if we double the radius of each sphere, the spheres with radius $2r_0$ cover thoroughly the entire volume of $\mathcal{Q}_{\mathbf{0}}(n, A)$,

that is, each point inside the hyper cube $\mathcal{Q}_{\mathbf{0}}(n,A)$ belongs to at least one of the small spheres. In general, the volume of a hyper sphere of radius r is given by [138, Eq. (16)]

$$\mathrm{Vol}\left(\mathcal{S}_{\mathbf{x}}(n,r)\right) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r^n \,. \tag{6.18}$$

Hence, if the radius of small spheres is doubled, the volume of $\bigcup_{i=1}^{L(n,R)} \mathcal{S}_{\mathbf{u}_i}(n,\sqrt{n\epsilon_n})$ is increased by 2^n. Since the spheres with radius $2r_0$ cover the $\mathcal{Q}_{\mathbf{0}}(n,A)$, it follows that the original r_0-radius packing has a density of at least 2^{-n} [5]. We assign a codeword to the center $\mathbf{u}_i$ of each small sphere. The codewords satisfy the input constraint as $0 \le u_{i,t} \le A$, $\forall t \in [\![n]\!]$, $\forall i \in [\![L]\!]$, which is equivalent to

$$\|\mathbf{u}_i\|_\infty \le A \,. \tag{6.19}$$

Since the volume of each sphere is equal to $\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n,r_0))$ and the centers of all spheres lie inside the cube, the total number of spheres is bounded from below by

$$\begin{aligned} L &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n,r_0\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n,r_0))} \\ &= \frac{\Delta_n(\mathscr{S}) \cdot \mathrm{Vol}\left[\mathcal{Q}_{\mathbf{0}}(n,A)\right]}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n,r_0))} \\ &\ge 2^{-n} \cdot \frac{A^n}{\mathrm{Vol}(\mathcal{S}_{\mathbf{u}_1}(n,r_0))} \,, \end{aligned} \tag{6.20}$$

where the first inequality holds by (6.15) and the second inequality holds by (6.17). The above bound can be further simplified as follows

$$\begin{aligned} \log L &\ge \log\left(\frac{P_{\mathrm{ave}}^n}{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{u}_1}(n,r_0)\right)}\right) - n \overset{(a)}{\ge} n\log\left(\frac{P_{\mathrm{ave}}}{\sqrt{\pi}r_0}\right) + \log\left(\Gamma\left(\frac{n}{2}+1\right)\right) - n \\ &\overset{(b)}{\ge} n\log P_{\mathrm{ave}} - n\log r_0 + \left\lfloor\frac{n}{2}\right\rfloor \log\left(\left\lfloor\frac{n}{2}\right\rfloor\right) - \left\lfloor\frac{n}{2}\right\rfloor \log e + o\left(\left\lfloor\frac{n}{2}\right\rfloor\right) - n \,, \end{aligned} \tag{6.21}$$

where (a) exploits (6.18) and (b) follows since

$$\Gamma\left(\frac{n}{2}+1\right) \overset{(a)}{=} \frac{n}{2}\Gamma\left(\frac{n}{2}\right) \overset{(b)}{\ge} \left\lfloor\frac{n}{2}\right\rfloor \Gamma\left(\left\lfloor\frac{n}{2}\right\rfloor\right) \overset{(c)}{\triangleq} \left\lfloor\frac{n}{2}\right\rfloor! \,, \tag{6.22}$$

[5] We note that the proposed proof of the lower bound in (6.17) is non-constructive in the sense that, while the existence of the respective saturated packing is proved, no systematic construction method is provided.

where (a) holds by the recurrence relation of the Gamma function for a positive real argument ($\frac{n}{2} \in \mathbb{R}$) [172] and (b) follows from $\left\lfloor \frac{n}{2} \right\rfloor \leq \frac{n}{2}$ for positive integer n and the monotonicity of the Gamma function for $n \geq 4 \equiv \left\lfloor \frac{n}{2} \right\rfloor \in [z_1, \infty)$ where $z_1 \approx 1.46$ is the first root of the Digamma function [172]; and (b) follows from the Stirling's approximation, that is, $\log n! = n \log n - n \log e + o(n)$ with e being the Euler number, for integer n ($n \leftarrow \left\lfloor \frac{n}{2} \right\rfloor \in \mathbb{Z}$), [173, P. 52]; see Appendix B for detailed elaborations. Now, for $r_0 = \sqrt{n\epsilon_n} = \sqrt{a} n^{\frac{1+b}{4}}$, we obtain

$$
\begin{aligned}
&\log L \\
&\overset{(a)}{\geq} n \log \frac{P_{\text{ave}}}{\sqrt{a}} - \frac{1}{4}(1+b)\, n \log n + \left(\frac{n}{2} - 1\right) \log \left(\frac{n}{2} - 1\right) - \left\lfloor \frac{n}{2} \right\rfloor \log e + o\left(\frac{n}{2}\right) - n \\
&\overset{(b)}{\geq} n \log \frac{P_{\text{ave}}}{\sqrt{a}} - \frac{1}{4}(1+b)\, n \log n + \frac{1}{2} n \log n - 2n - \log n - \frac{n}{2} \log e + o\left(\frac{n}{2}\right) \\
&= \left(\frac{1-b}{4}\right) n \log n + n \left(\log \frac{P_{\text{ave}}}{\sqrt{ae}}\right) - 2n + o\left(\frac{n}{2}\right) \\
&= \left(\frac{1-b}{4}\right) n \log n + n \left(\log \frac{P_{\text{ave}}}{\sqrt{ae}}\right) + \mathcal{O}(n)\,, \qquad (6.23)
\end{aligned}
$$

where (a) follows by $\left\lfloor \frac{n}{2} \right\rfloor > \frac{n}{2} - 1$ for integer n and (b) holds since $\log(t-1) \geq \log t - 1$ for $t \geq 2$ and $\left\lfloor \frac{n}{2} \right\rfloor \leq \frac{n}{2}$ for integer n, and (c) follows since base of the logarithm is 2. Observe that the dominant term in 6.23 is of order $n \log n$. Hence, for obtaining a finite value for the lower bound of the rate, R, (6.23) induces the scaling law of L to be $2^{(n \log n) R}$. Therefore, we obtain

$$
R \geq \frac{1}{n \log n} \left[\left(\frac{1-b}{4}\right) n \log n + n \log \left(\frac{A}{e\sqrt{a}}\right) + o(n) \right]\,, \qquad (6.24)
$$

which tends to $\frac{1}{4}$ when $n \to \infty$ and $b \to 0$.

6.2.5.2 | Encoding

Given message $i \in [\![L]\!]$, transmit $\mathbf{x} = \mathbf{u}_i$.

6.2.5.3 | Decoding

Let

$$
\delta_n = c\rho^2 \epsilon_n = c\rho^2 a n^{\frac{1}{2}(b-1)}\,, \qquad (6.25)
$$

where $0 < b < 1$ is an arbitrarily small constant and $0 < c < 2$ is a constant. To identify whether message $j \in \mathcal{M}$ was sent, the decoder checks whether the channel output $\mathbf{y}$ belongs to the following decoding set:

$$\mathcal{D}_j = \left\{ \mathbf{y} \in \mathcal{Y}^n \ : \ |D(\mathbf{y}; \mathbf{u}_j)| \leq \delta_n \right\}, \tag{6.26}$$

where

$$D(\mathbf{y}; \mathbf{u}_j) = \frac{1}{n} \sum_{t=1}^{n} (y_t - (\rho u_{j,t} + \lambda))^2 - y_t \,, \tag{6.27}$$

is referred to as the decoding metric evaluated for observation vector $\mathbf{y}$ and codeword $\mathbf{u}_j$.

6.2.5.4 | Error Analysis

Consider the type I errors, i.e., the transmitter sends $\mathbf{u}_i$, yet $\mathbf{Y} \notin \mathcal{D}_i$. For every $i \in [\![L]\!]$, the type I error probability is bounded by

$$P_{e,1}(i) = \Pr(|D(\mathbf{Y}; \mathbf{u}_i)| > \delta_n \,|\, \mathbf{u}_i) \,, \tag{6.28}$$

where the condition means that $\mathbf{x} = \mathbf{u}_i$ was sent. In order to bound $P_{e,1}(i)$, we apply Chebyshev's inequality, namely

$$\Pr\left(\left| D(\mathbf{Y}; \mathbf{u}_i) - \mathbb{E}\left\{ D(\mathbf{Y}; \mathbf{u}_i) \,|\, \mathbf{u}_i \right\} \right| > \delta_n \middle| \mathbf{u}_i \right) \leq \frac{\operatorname{var}\left\{ D(\mathbf{Y}; \mathbf{u}_i) \,|\, \mathbf{u}_i \right\}}{\delta_n^2} \,. \tag{6.29}$$

First, we derive the expectation of the decoding metric as follows

$$\mathbb{E}\left\{ D(\mathbf{Y}; \mathbf{u}_i) \,|\, \mathbf{u}_i \right\} = \frac{1}{n} \sum_{t=1}^{n} \left[\operatorname{var}\left\{ Y_t \,|\, u_{i,t} \right\} - E\left[Y_t \,|\, u_{i,t} \right] \right] = 0 \,. \tag{6.30}$$

Now, since the channel is memoryless, we can compute the variance as follows

$$\begin{aligned} \operatorname{var}\left\{ D(\mathbf{Y}; \mathbf{u}_i) \,|\, \mathbf{u}_i \right\} &= \frac{1}{n^2} \sum_{t=1}^{n} \operatorname{var}\left\{ \left(Y_t - \left(\rho u_{i,t} + \lambda \right) \right)^2 \,\middle|\, u_{i,t} \right\} - \operatorname{var}\left\{ Y_t \,\middle|\, u_{i,t} \right\} \\ &\overset{(a)}{=} \frac{1}{n^2} \sum_{t=1}^{n} \operatorname{var}\left\{ \left(Y_t - \left(\rho u_{i,t} + \lambda \right) \right)^2 \,\middle|\, u_{i,t} \right\} - \frac{1}{n^2} \sum_{t=1}^{n} \rho u_{i,t} + \lambda \\ &\overset{(b)}{\leq} \frac{1}{n^2} \sum_{t=1}^{n} \operatorname{var}\left\{ \left(Y_t - \left(\rho u_{i,t} + \lambda \right) \right)^2 \,\middle|\, u_{i,t} \right\} - \frac{\lambda}{n} \,, \end{aligned} \tag{6.31}$$

where (a) holds since $\operatorname{var}\{Y_t \,|\, u_{i,t}\} = \rho u_{i,t} + \lambda$ and (b) follows since $u_{i,t} \geq 0 \,, \forall t \in [\![n]\!]$, $\forall i \in [\![L]\!]$. Next, to establish an upper bound on the first sum in (6.31), we present a useful lemma.

Lemma 6.2.1. *Let $Z \sim Pois(\lambda_Z)$ be a Poisson RV with mean λ_Z. The following inequality holds*

$$\mathbb{E}\{(Z-\lambda_Z)^4\} \leq 7\left(\lambda_Z^4+\lambda_Z^3+\lambda_Z^2+\lambda_Z\right).$$

Proof. The proof is provided in Appendix C. □

Using the above lemma, we bound the variance of the decoding metric as follows

$$\begin{aligned}
\mathbb{E}\{|D(\mathbf{Y};\mathbf{u}_i)|^2 \,|\, \mathbf{u}_i\} &\overset{(a)}{=} \mathrm{var}\{D(\mathbf{Y};\mathbf{u}_i) \,|\, \mathbf{u}_i\} \\
&\overset{(b)}{\leq} \mathbb{E}\{|D(\mathbf{Y};\mathbf{u}_i)|^4 \,|\, \mathbf{u}_i\} - \frac{\lambda}{n} \\
&\overset{(c)}{=} \frac{1}{n}\mathbb{E}\left[\left(Y_t-(\rho u_{i,t}+\lambda)\right)^4 \,\middle|\, u_{i,t}\right] - \frac{\lambda}{n} \\
&\leq \frac{7}{n}\left((\rho A+\lambda)^4+(\rho A+\lambda)^3+(\rho A+\lambda)^2+(\rho A+\lambda)\right) - \frac{\lambda}{n}. \quad (6.32)
\end{aligned}$$

where (a) follows since $\mathbb{E}\{D(\mathbf{Y};\mathbf{u}_i)\} = 0$, (b) follows since $\mathrm{var}\{Z\} \leq \mathbb{E}\{Z^2\}$, and (c) holds by letting $Z = (Y_t-(\rho u_{i,t}+\lambda))^2$ and exploiting an upper bound on the fourth non-central moment of a Poisson random variable (see Appendix C). Therefore, exploiting (6.29), (6.30) and (6.32), we can bound the type I error probability in (6.28) as follows

$$\begin{aligned}
P_{e,1}(i) &= \Pr\left(\left|D(\mathbf{Y};\mathbf{u}_i)\right| > \delta_n | \mathbf{u}_i\right) \\
&\leq \frac{7\left((\rho A+\lambda)^4+(\rho A+\lambda)^3+(\rho A+\lambda)^2+\lambda\right)}{n\delta_n^2} \\
&= \frac{7\left((\rho A+\lambda)^4+(\rho A+\lambda)^3+(\rho A+\lambda)^2+\lambda\right)}{c^2\rho^4 a^2 n^b} \\
&\leq \lambda_1, \quad (6.33)
\end{aligned}$$

for sufficiently large n and arbitrarily small $\lambda_1 > 0$.

Next, we address type II errors, i.e., when $\mathbf{Y} \in \mathcal{D}_j$ while the transmitter sent $\mathbf{u}_i$. Then, for every $i, j \in [\![L]\!]$, where $i \neq j$, the type II error probability is given by

$$P_{e,2}(i,j) = \Pr\left(\left|D(\mathbf{Y};\mathbf{u}_j)\right| \leq \delta_n \,|\, \mathbf{u}_i\right). \quad (6.34)$$

where

$$D(\mathbf{Y};\mathbf{u}_j) = \underbrace{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t-(\rho u_{i,t}+\lambda)+\rho\left(u_{i,t}-u_{j,t}\right)\right)^2}_{\triangleq \beta} - \underbrace{\frac{1}{n}\sum_{t=1}^{n} Y_t}_{\triangleq \alpha}. \quad (6.35)$$

Observe that term β can be expressed as follows

$$\beta = \underbrace{\frac{1}{n}\left[\left\|\mathbf{Y} - \left(\rho\mathbf{u}_i + \lambda\mathbf{1}_n\right)\right\|^2 + \left\|\rho\left(\mathbf{u}_i - \mathbf{u}_j\right)\right\|^2\right]}_{\triangleq \beta_1} + \underbrace{\frac{2\rho}{n}\sum_{t=1}^{n}\left(u_{i,t} - u_{j,t}\right)\left(Y_t - \left(\rho u_{i,t} + \lambda\right)\right)}_{\triangleq \beta_2}. \tag{6.36}$$

Then, define the following events

$$\mathcal{H}_i^j = \left\{\left|\beta - \alpha\right| \leq \delta_n \,\middle|\, \mathbf{u}_i\right\}, \quad \mathcal{E}_0 = \left\{\left|\beta_2\right| > \delta_n \,\middle|\, \mathbf{u}_i\right\}, \quad \mathcal{E}_1 = \left\{\beta_1 - \alpha \leq 2\delta_n \,\middle|\, \mathbf{u}_i\right\}. \tag{6.37}$$

Exploiting the reverse triangle inequality, i.e., $|\beta| - |\alpha| \leq |\beta - \alpha|$, we obtain the following upper bound on the type II error probability

$$\begin{aligned} P_{e,2}(i,j) &= \Pr\left(\mathcal{H}_i^j\right) \\ &= \Pr\left(\left|\beta - \alpha\right| \leq \delta_n \,\middle|\, \mathbf{u}_i\right) \\ &\leq \Pr\left(\left|\beta\right| - \left|\alpha\right| \leq \delta_n \,\middle|\, \mathbf{u}_i\right) \\ &\overset{(a)}{=} \Pr\left(\beta - \alpha \leq \delta_n \,\middle|\, \mathbf{u}_i\right), \end{aligned} \tag{6.38}$$

where (a) follows since $\alpha \geq 0$ and $\beta \geq 0$. Now, applying the law of total probability to event $\mathcal{B} = \left\{\beta - \alpha \leq \delta_n | \mathbf{u}_i\right\}$ over $\mathcal{E}_0$ and its complement $\mathcal{E}_0^c$, we obtain

$$\begin{aligned} P_{e,2}(i,j) &\leq \Pr\left(\mathcal{B} \cap \mathcal{E}_0\right) + \Pr\left(\mathcal{B} \cap \mathcal{E}_0^c\right) \\ &\overset{(a)}{\leq} \Pr\left(\mathcal{E}_0\right) + \Pr\left(\mathcal{B} \cap \mathcal{E}_0^c\right) \\ &\overset{(b)}{\leq} \Pr\left(\mathcal{E}_0\right) + \Pr\left(\mathcal{E}_1\right), \end{aligned} \tag{6.39}$$

where inequality (a) follows from $\mathcal{B} \cap \mathcal{E}_0 \subset \mathcal{E}_0$ and inequality (b) follows from $\Pr\left(\mathcal{B} \cap \mathcal{E}_0^c\right) \leq \Pr\left(\mathcal{E}_1\right)$, which is proved in the following. Observe,

$$\begin{aligned} \Pr\left(\mathcal{B} \cap \mathcal{E}_0^c\right) &= \Pr\left(\left\{\beta - \alpha \leq \delta_n\right\} \cap \left\{\left|\beta_2\right| \leq \delta_n\right\} \middle| \mathbf{u}_i\right) \\ &= \Pr\left(\left\{\beta_1 - \alpha \leq \delta_n - \beta_2\right\} \cap \left\{\left|\beta_2\right| \leq \delta_n\right\} \middle| \mathbf{u}_i\right) \end{aligned}$$

$$\overset{(a)}{\leq} \Pr\left(\left\{\beta_1 - \alpha \leq 2\delta_n\right\}\middle|\mathbf{u}_i\right)$$
$$= \Pr\left(\mathcal{E}_1\right), \tag{6.40}$$

where inequality (a) holds since $\delta_n - \beta_2 \leq 2\delta_n$ conditioned on $|\beta_2| \leq \delta_n$.

We now proceed with bounding $\Pr\left(\mathcal{E}_0\right)$. By Chebyshev's inequality, the probability of this event can be bounded as follows

$$\begin{aligned}
\Pr(\mathcal{E}_0) &\leq \frac{\operatorname{var}\left\{\sum_{t=1}^{n}\left(u_{i,t}-u_{j,t}\right)\left(Y_t-\left(\rho u_{i,t}+\lambda\right)\right)\,\middle|\,\mathbf{u}_i\right\}}{n^2\delta_n^2/(4\rho^2)} \\
&= \frac{4\rho^2\sum_{t=1}^{n}(u_{i,t}-u_{j,t})^2\cdot\operatorname{var}\{Y_t\,|\,u_{i,t}\}}{n^2\delta_n^2} \\
&= \frac{4\rho^2\sum_{t=1}^{n}(u_{i,t}-u_{j,t})^2\cdot(\rho u_{i,t}+\lambda)}{n^2\delta_n^2} \\
&\leq \frac{4\rho^2(\rho A+\lambda)\sum_{t=1}^{n}(u_{i,t}-u_{j,t})^2}{n^2\delta_n^2} \\
&= \frac{4\rho^2(\rho A+\lambda)\left\|\mathbf{u}_i-\mathbf{u}_j\right\|^2}{n^2\delta_n^2}.
\end{aligned} \tag{6.41}$$

Observe that

$$\begin{aligned}
\left\|\mathbf{u}_i-\mathbf{u}_j\right\|^2 &\overset{(a)}{\leq} \left(\|\mathbf{u}_i\|+\left\|\mathbf{u}_j\right\|\right)^2 \\
&\overset{(b)}{\leq} \left(\sqrt{n}\|\mathbf{u}_i\|_\infty+\sqrt{n}\left\|\mathbf{u}_j\right\|_\infty\right)^2 \\
&\overset{(c)}{\leq} \left(\sqrt{n}A+\sqrt{n}A\right)^2 \\
&= 4nA^2,
\end{aligned} \tag{6.42}$$

where (a) holds by the triangle inequality, (b) follows since $\|\cdot\| \leq \sqrt{n}\|\cdot\|_\infty$, and (c) is valid by (6.19). Hence, we obtain

$$\begin{aligned}
\Pr(\mathcal{E}_0) &\leq \frac{16n\rho^2(\rho A+\lambda)A^2}{n^2\delta_n^2} \\
&= \frac{16\rho^2(\rho A+\lambda)A^2}{n\delta_n^2} \\
&= \frac{16(\rho A+\lambda)A^2}{c^2\rho^2a^2n^b} \\
&= \frac{16(\rho A+\lambda)}{\rho n^b}
\end{aligned}$$

$$\leq \zeta_0 \,, \tag{6.43}$$

for sufficiently large n, where $\zeta_0 > 0$ is an arbitrarily small constant.

We now proceed with bounding $\Pr(\mathcal{E}_1)$ as follows. Based on the codebook construction, each codeword is surrounded by a sphere of radius $\sqrt{n\epsilon_n}$, that is

$$\left\|\mathbf{u}_i - \mathbf{u}_j\right\|^2 \geq 4n\epsilon_n \,. \tag{6.44}$$

Thus, we can establish the following upper bound for event $\mathcal{E}_1$:

$$\begin{aligned}
\Pr(\mathcal{E}_1) &= \Pr\left(\frac{1}{n}\left[\left\|\mathbf{Y} - (\rho\mathbf{u}_i + \lambda\mathbf{1}_n)\right\|^2 + \left\|\rho\left(\mathbf{u}_i - \mathbf{u}_j\right)\right\|^2 - \sum_{t=1}^{n} Y_t\right] \leq 2\delta_n \,\middle|\, \mathbf{u}_i\right) \\
&\overset{(a)}{\leq} \Pr\left(\frac{1}{n}\left[\left\|\mathbf{Y} - (\rho\mathbf{u}_i + \lambda\mathbf{1}_n)\right\|^2 - \sum_{t=1}^{n} Y_t\right] \leq 2(c-2)\rho^2\epsilon_n \,\middle|\, \mathbf{u}_i\right) \\
&= \Pr\left(\frac{1}{n}\left[\sum_{t=1}^{n}\left(Y_t - \left(\rho u_{i,t} + \lambda\right)\right)^2 - Y_t\right] \leq 2(c-2)\rho^2\epsilon_n \,\middle|\, \mathbf{u}_i\right) \\
&\overset{(b)}{\leq} \frac{\operatorname{var}\left\{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \left(\rho u_{i,t} + \lambda\right)\right)^2 - Y_t\right\}}{\left(2(c-2)\rho^2\epsilon_n\right)^2} \\
&\overset{(c)}{\leq} \frac{7\left(\left(\rho A + \lambda\right)^4 + \left(\rho A + \lambda\right)^3 + \left(\rho A + \lambda\right)^2 + \lambda\right)}{n\left(2(c-2)\rho^2\epsilon_n\right)^2} \\
&= \frac{7\left(\left(\rho A + \lambda\right)^4 + \left(\rho A + \lambda\right)^3 + \left(\rho A + \lambda\right)^2 + \lambda\right)}{4(c-2)^2\rho^4 a^2 n^b} \\
&\leq \zeta_1 \,,
\end{aligned} \tag{6.45}$$

for sufficiently large n, where $\zeta_1 > 0$ is an arbitrarily small constant. Here, (a) follows from (6.44) and (6.25), (b) holds by Chebyshev's inequality as given in (6.29), and (c) follows by Lemma 6.2.1. Therefore, $P_{e,2}(i,j) \leq \Pr(\mathcal{E}_0) + \Pr(\mathcal{E}_1) \leq \zeta_0 + \zeta_1 \leq \lambda_2$. We have thus shown that for every $\lambda_1, \lambda_2 > 0$ and sufficiently large n, there exists an $(L(n,R), n, \lambda_1, \lambda_2)$ code.

6.2.6 | Converse Proof

We show that the capacity is bounded by $\mathbb{C}_{DI}(\mathcal{W}, L) \leq \frac{3}{2}$. The derivation of this upper bound for the achievable rate of the DTPC is more involved than the derivation in the

Gaussian case [105]. In chapter 4 and 5 on the Gaussian channels with fading [105], the converse proof was based on establishing a minimum distance between each pair of codewords. Here, on the other hand, we use the stronger requirement that the ratio of the letters of every two different codewords is different from 1 for at least one index.

We begin with the following lemma on the ratio of the letters of every pair of codewords.

Lemma 6.2.2. *Suppose that R is an achievable rate for the DTPC. Consider a sequence of $(L(n,R),n,\lambda_1^{(n)},\lambda_2^{(n)})$ codes $(\mathcal{U}^{(n)},\mathcal{D}^{(n)})$ such that $\lambda_1^{(n)}$ and $\lambda_2^{(n)}$ tend to zero as $n\to\infty$. Then, given a sufficiently large n, the codebook $\mathcal{U}^{(n)}$ satisfies the following property. For every pair of codewords, $\mathbf{u}_{i_1}$ and $\mathbf{u}_{i_2}$, there exists at least one letter $t\in[\![n]\!]$ such that*

$$\left|1-\frac{\rho u_{i_2,t}+\lambda}{\rho u_{i_1,t}+\lambda}\right|>\epsilon_n', \tag{6.46}$$

for all $i_1,i_2\in[\![L]\!]$, such that $i_1\neq i_2$, with

$$\epsilon_n'=\frac{P_{max}}{n^{1+b}}, \tag{6.47}$$

where $b>0$ is an arbitrarily small constant.

Proof. In the following, we provide the proof of Lemma 6.2.2. The method of proof is by contradiction, namely, we assume that the condition given in (6.46) is violated and then we show that this leads to a contradiction (sum of the type I and type II error probabilities converge to one).

Fix $\lambda_1,\lambda_2>0$. Let $\kappa,\delta>0$ be arbitrarily small constants. Assume to the contrary that there exist two messages i_1 and i_2, where $i_1\neq i_2$, meeting the error constraints in (6.8), such that for all $t\in[\![n]\!]$, we have

$$\left|1-\frac{v_{i_2,t}}{v_{i_1,t}}\right|\leq\epsilon_n', \tag{6.48}$$

where $v_{i_k,t}=\rho u_{i_k,t}+\lambda,\ k=1,2$. In order to show contradiction, we will bound the sum of the two error probabilities, $P_{e,1}(i_1)+P_{e,2}(i_2,i_1)$, from below. To this end, define

$$\mathcal{B}_{i_1}=\left\{\mathbf{y}\in\mathcal{D}_{i_1}\ :\ \frac{1}{n}\sum_{t=1}^{n}y_t\leq\rho P_{\max}+\lambda+\delta\right\}. \tag{6.49}$$

Then, observe that

$$P_{e,1}(i_1)+P_{e,2}(i_2,i_1)=1-\sum_{\mathbf{y}\in\mathcal{D}_{i_1}}W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right)+\sum_{\mathbf{y}\in\mathcal{D}_{i_1}}W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_2}\right)$$

$$\geq 1 - \sum_{\mathbf{y}\in\mathcal{D}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right) + \sum_{\mathbf{y}\in\mathcal{D}_{i_1}\cap\mathcal{B}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_2}\right). \tag{6.50}$$

Now, consider the sum over $\mathcal{D}_{i_1}$ in (6.50),

$$\begin{aligned}\sum_{\mathbf{y}\in\mathcal{D}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right) &= \sum_{\mathbf{y}\in\mathcal{D}_{i_1}\cap\mathcal{B}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right) + \sum_{\mathbf{y}\in\mathcal{D}_{i_1}\cap\mathcal{B}_{i_1}^c} W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right)\\ &\leq \sum_{\mathbf{y}\in\mathcal{D}_{i_1}\cap\mathcal{B}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right) + \Pr\left(\frac{1}{n}\sum_{t=1}^{n} Y_t > \rho P_{\max} + \lambda + \delta \,\middle|\, \mathbf{u}_{i_1}\right).\end{aligned} \tag{6.51}$$

Next, we bound the probability on the right hand side of (6.51) as follows

$$\begin{aligned}&\Pr\left(\frac{1}{n}\sum_{t=1}^{n} Y_t - \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\{Y_t\} > \rho P_{\max} + \delta - \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\{Y_t\}\right)\\ &\overset{(a)}{\leq} \frac{\operatorname{var}\left\{\frac{1}{n}\sum_{t=1}^{n} Y_t \,\middle|\, \mathbf{u}_{i_1}\right\}}{\left(\rho P_{\max} + \delta - \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\{Y_t\}\right)^2}\\ &\overset{(b)}{=} \frac{\frac{1}{n^2}\sum_{t=1}^{n}(\rho u_{i_1,t} + \lambda)}{\left(\rho P_{\max} + \delta - \frac{1}{n}\sum_{t=1}^{n}\rho u_{i_1,t} + \lambda\right)^2}\\ &\overset{(c)}{\leq} \frac{\rho P_{\max} + \lambda}{n\delta^2}\\ &\leq \kappa\,,\end{aligned} \tag{6.52}$$

for sufficiently large n, where inequality (a) follows from Chebyshev's inequality, for equality (b), we exploited $\operatorname{var}\{Y_t|u_{i_1,t}\} = \mathbb{E}\{Y_t|u_{i_1,t}\} = \rho u_{i_1,t} + \lambda$, and for inequality (c), we used the fact that $u_{i_1,t} \leq P_{\max}\,, \forall\, t \in [\![n]\!]$.

Returning to the sum of error probabilities in (6.50), exploiting the bound (6.52) leads to

$$P_{e,1}(i_1) + P_{e,2}(i_2, i_1) \geq 1 - \sum_{\mathbf{y}\in\mathcal{D}_{i_1}\cap\mathcal{B}_{i_1}} \left[W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right) - W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_2}\right)\right] - \kappa\,. \tag{6.53}$$

Now, let us focus on the summand in the square brackets in (6.53). By (6.3), we have

$$W^n\left(\mathbf{y}\middle|\mathbf{u}_{i_1}\right) - W^n\left(\mathbf{y}\middle|\mathbf{u}_{i_2}\right) = W^n\left(\mathbf{y}\middle|\mathbf{u}_{i_1}\right)\cdot\left[1 - W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_2}\right) / W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right)\right]$$

$$= W^n\left(\mathbf{y}\middle|\mathbf{u}_{i_1}\right)\cdot\left[1-\prod_{t=1}^{n} e^{-(v_{i_2,t}-v_{i_1,t})}\left(\frac{v_{i_2,t}}{v_{i_1,t}}\right)^{y_t}\right]$$
$$= W^n\left(\mathbf{y}\middle|\mathbf{u}_{i_1}\right)\cdot\left[1-\prod_{t=1}^{n} e^{-\epsilon'_n v_{i_1,t}}\left(1-\epsilon'_n\right)^{y_t}\right], \tag{6.54}$$

where for the last inequality, we employed $v_{i_2,t}-v_{i_1,t} \leq \left|v_{i_2,t}-v_{i_1,t}\right| \leq \epsilon'_n v_{i_1,t}$ and $1-\frac{v_{i_2,t}}{v_{i_1,t}} \leq \left|1-\frac{v_{i_2,t}}{v_{i_1,t}}\right| \leq \epsilon'_n$, which follow from (6.48). Now, we bound the product term inside the bracket as follows:

$$\begin{aligned}
\prod_{t=1}^{n} e^{-\epsilon'_n v_{i_1,t}}\left(1-\epsilon'_n\right)^{y_t} &= e^{-\epsilon'_n\sum_{t=1}^{n} v_{i_1,t}}\cdot\left(1-\epsilon'_n\right)^{\sum_{t=1}^{n} y_t}\\
&\overset{(a)}{\geq} e^{-n\epsilon'_n\left(\rho P_{\max}+\lambda\right)}\cdot\left(1-\epsilon'_n\right)^{n\left(\rho P_{\max}+\lambda+\delta\right)}\\
&= e^{n\epsilon'_n\delta}\cdot e^{-n\epsilon'_n\left(\rho P_{\max}+\lambda+\delta\right)}\cdot\left(1-\epsilon'_n\right)^{n\left(\rho P_{\max}+\lambda+\delta\right)}\\
&\overset{(b)}{\geq} e^{n\epsilon'_n\delta}\cdot e^{-n\epsilon'_n\left(\rho P_{\max}+\lambda+\delta\right)}\cdot\left(1-n\epsilon'_n\right)^{\rho P_{\max}+\lambda+\delta}\\
&\geq e^{n\epsilon'_n\delta}\cdot f(n\epsilon'_n)\\
&\overset{(c)}{>} f(n\epsilon'_n)\\
&\overset{(d)}{\geq} 1-3\left(\rho P_{\max}+\lambda+\delta\right)n\epsilon'_n\\
&= 1-\frac{3\left(\rho P_{\max}+\lambda+\delta\right)P_{\max}}{n^b}\\
&\geq 1-\kappa\,.
\end{aligned} \tag{6.55}$$

for sufficiently large n. For inequality (a), we used $v_{i_1,t} \leq \rho P_{\max}+\lambda\,,\ \forall\, t \in [\![n]\!]$, and $\sum_{t=1}^{n} y_t \leq n\left(\rho P_{\max}+\lambda+\delta\right)$, where the latter inequality follows from $\mathbf{y} \in \mathcal{B}_{i_1}$, cf. (6.49). For (b), we used Bernoulli's inequality $(1-x)^r \geq 1-rx$ for all $x > -1$ and $r > 0$ [174, see Ch. 3]. For (c), we exploited $e^{n\epsilon'_n\delta} > 1$ and the following definition: $f(x) = e^{-cx}(1-x)^c$ with $c = \lambda+\rho P_{\max}+\delta$. Finally, for (d), we used the Taylor expansion $f(x) = 1-2cx+\mathcal{O}(x^2)$ to obtain the upper bound $f(x) \geq 1-3cx$ for sufficiently small values of x.

Equation (6.54) can then be written as follows

$$\begin{aligned}
W^n\left(\mathbf{y}\middle|\mathbf{u}_{i_1}\right)-W^n\left(\mathbf{y}\middle|\mathbf{u}_{i_2}\right) &\leq W^n\left(\mathbf{y}\middle|\mathbf{u}_{i_1}\right)\cdot\left[1-e^{-\epsilon'_n\sum_{t=1}^{n} v_{i_1,t}}\cdot\left(1-\epsilon'_n\right)^{\sum_{t=1}^{n} y_t}\right]\\
&\leq \kappa\cdot W^n\left(\mathbf{y}\,\middle|\,\mathbf{u}_{i_1}\right).
\end{aligned} \tag{6.56}$$

Combining, (6.53), (6.54), and (6.56) yields

$$\begin{aligned} P_{e,1}(i_1) + P_{e,2}(i_2, i_1) &\overset{(a)}{\geq} 1 - \sum_{\mathbf{y} \in \mathcal{B}_{i_1}} \left[W^n \left(\mathbf{y} \,\middle|\, \mathbf{u}_{i_1} \right) - W^n \left(\mathbf{y} \,\middle|\, \mathbf{u}_{i_2} \right) \right] - \kappa \\ &= 1 - \sum_{\mathbf{y} \in \mathcal{B}_{i_1}} \left[\kappa \cdot W^n \left(\mathbf{y} \,\middle|\, \mathbf{u}_{i_1} \right) \right] - \kappa \\ &\overset{(b)}{\geq} 1 - 2\kappa \,, \end{aligned} \tag{6.57}$$

where for (a), we replaced $\mathbf{y} \in \mathcal{B}_{i_1} \cap \mathcal{D}_{i_1}$ by $\mathbf{y} \in \mathcal{B}_{i_1}$ to enlarge the domain and for (b), we used $\sum_{\mathbf{y} \in \mathcal{B}_{i_1}} W^n(\mathbf{y} \,|\, \mathbf{u}_{i_1}) \leq 1$. Clearly, this is a contradiction since the error probabilities tend to zero as $n \to \infty$. Thus, the assumption in (6.48) is false. This completes the proof of Lemma 6.2.2. □

6.3 | Simulation Results

We emphasize that the main result of this chapter is the characterization of fundamental performance bounds in terms of the DI capacity for the DTPC (cf. Theorem 6.2.1), which by definition holds for asymptotically large codewords, i.e., as $n \to \infty$. Explicit code construction for the DTPC is not the focus of this chapter and hence the purpose of this section is not the evaluation/verification of our achievability proof in Section III[6]. Nonetheless, we are interested in studying whether our key finding, i.e., the possibility of reliable identification for super-exponentially large codebook sizes, also holds for a heuristically-designed (structure-less) finite-length code.

6.3.1 | Heuristic Codebook Construction

The codebook construction is briefly sketched in the following. At first, codewords are generated uniformly, that is, the value of each symbol is chosen uniformly distributed between 0 and A. Next, in order to realize the minimum distance property of the codebook, once a codeword is created, before adding it to the codebook, it is verified whether it has at least a minimum Euclidean distance of $2\sqrt{n\epsilon_n}$ from all previously generated codewords or not. In the course of codeword generation, if a codeword violates the

[6] In fact, our achievability proof in Section 6.2.5 shows only the existence of codes and does not provide any explicit construction of the codebook. That is, although we prove that a saturated arrangement of sphere exists, we do not know the coordinates of their centers explicitly for a given codeword length n.

Table 6.1: Parameters of The Simulations - Poisson Channel

Description	Notation	Value
Minimum of power constraints	$A = \min(P_{\text{ave}}, P_{\max})$	1000 molecules/s
Release time	T_{rls}	1 s
Prob. molecules reaching the receiver	p_{ch}	0.01
Expected number of interfering molecules	λ	0.2
Code rate	R	0.1
Codeword length	n	[19 - 28]
Codebook size	$L = 2^{(n \log n)R}$	[268 - 11273]
Codebook parameters	a, b, c	10^5, 0.99, $\frac{1}{3}$
Codebook precision	$\epsilon_n = an^{\frac{1}{2}(b-1)}$	[9.853 - 9.834]
Decoding threshold	$\delta_n = c\rho^2\epsilon_n$	[3.284 - 3.278]
Codebook minimum distance	$2\sqrt{n\epsilon_n}$	[27.36 - 33.18]
Number of iterations	-	7×10^5

minimum distance property, it is discarded and a new codeword is generated and the procedure is repeated until the desired codebook size is obtained. To simulate the receiver's task, the distance decoder in (6.12) is implemented and the empirical type I and type II error rates for finite codeword lengths are obtained via Monte Carlo simulation. We focus on a range of small codeword lengths, i.e., $19 \leq n \leq 28$, since the above simple look-up table code construction and full search decoding are not scalable for large n. Moreover, since rates $R \geq \frac{1}{4}$ are achievable by the proposed scheme only as $n \to \infty$, we choose a smaller rate, i.e., $R = 0.1$, for codebook generation for finite n. However, we study a codebook with super-exponential size in n, i.e., $L = 2^{(n \log n)R}$, which is the key element of Theorem 6.2.1. Without loss of generality, we assume that the transmitter sends message $i = 1$ and denote the empirical type I and type II error rates (average and maximum) by $\bar{e}_1(i)$, and $\bar{e}_2^{\text{ave}}$, $\bar{e}_2^{\max}$, respectively. The values of the parameters used in the proposed simulation setup and codebook construction are summarized in Table 6.1.

6.3.2 | Results and Discussions

Figures 6.5 and 6.6 show respectively the empirical type I and type II error rates versus the codeword length. The results in Figures 6.5 and 6.6 show that fast-decaying error rates are attainable for the considered codebook which has a super exponentially large size in codeword length n even though the code construction is sub-optimal and n is finite. This is an interesting observation given the fact that our theoretical results only prove that asymptotically as $n \to \infty$, reliable identification with super-exponentially large codebook size in n is achievable. Furthermore, the simulation results in Figures 6.5 and 6.6 show the general trend of the empirical error rates as functions of the codeword length is well captured by the analytical upper bounds. However, the achieved error rates for the constructed code with $R = 0.1$, decay faster than the theoretical upper bounds provided in (6.33), (6.43), and (6.45) evaluated for $b = 0.99$. We note that since our theoretical bounds are derived for $n \to \infty$, it is not contradictory if the slopes of the empirical error rates are slightly higher for the simulated curves at finite n.

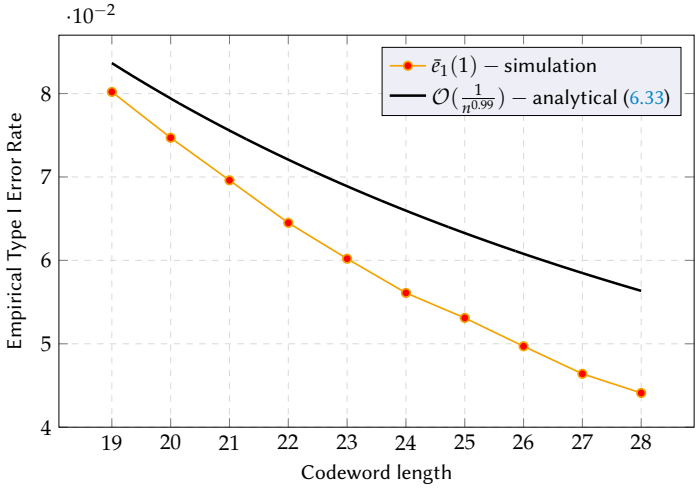


Figure 6.5: Impact of codeword length on the empirical type I error rate. Larger lengths decrease the empirical type I error rate.

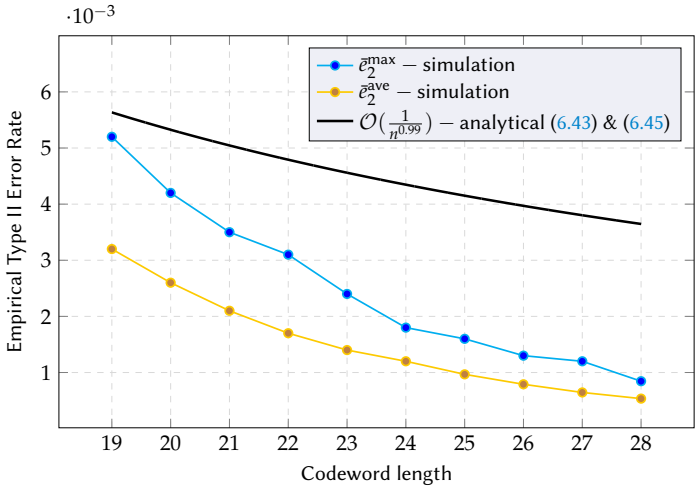


Figure 6.6: Impact of codeword length on the empirical type II error rate. Larger lengths decrease the empirical type II error rate.

6.4 | Summary

In this chapter, we studied the DI problem over the DTPC, which may serve as a model for event-triggered based tasks in the context of MC for applications such as targeted drug delivery, health condition monitoring, olfactory systems, etc. In particular, we derived lower and upper bounds on the DI capacity of the DTPC subject to average and peak power constraints in the codebook size of $L(n,R) = 2^{(n\log n)R} = n^{nR}$. Our results revealed that the super-exponential scale of n^{nR} is the appropriate scale for the DI capacity of the DTPC, which was proved by finding a suitable sphere packing arrangement embedded in a hyper cube. We emphasize that this scale is sharply different from the ordinary scales in transmission and RI settings, where the codebook size grows exponentially and double exponentially, respectively.

CHAPTER 7

DI FOR POISSON CHANNELS WITH MEMORY

> *The True Art of Memory is The Art of Attention.*
>
> Samuel Johnson,

7.1 | Introduction

One of the basic and widely-accepted abstract models for MC systems with molecule counting receivers is the discrete-time Poisson channel (DTPC) model with *inter-symbol interference* (ISI) model [9, 156, 157]. The DTPC model with memory has been used to study the performance limits of MC systems. Despite the recent theoretical and technological advancements in the field of MCs, the transmission capacity of most MC systems with DTPC with memory model are still unknown [9]. However, a number of approaches to examining the behavior of Poisson channel are being explored. For instance, an analytic expression for the transmission capacity of a DTPC with memory under an average power constraint alone, is still open [9, 175, 176]. However, several bounds and asymptotic behaviors for the DTPC with memory in different setups have been established. For instance, analytical lower and upper bounds on the transmission capacity of the DTPC with input constraints and memory are provided in [37]. Bounds on the transmission capacity of the DTPC with memory are developed in [32, 33]. The design of optimal code for DTPC with memory under a peak and average power constraint is studied in [177]. In [178], the impact of memory on the performance for a diffusive MC channel is characterized. Performance analysis of modulation schemes for diffusive MC with memory is considered in [179] and impact of degree of memory on the performance is shown. Design of the filter and detector parts in a receiver for Poisson channel with time-varying mean when transmitted symbols are exposed to the ISI is studied in [180]. The code design problem for diffusive MC channel under ISI

is considered in [181] where influence of the ISI is incorporated into the code design. The authors in [51, 108] studied the DTPC in absence of ISI, i.e., $K = 1$, and established lower and upper bounds on the DI capacity where the codebook size scales as $\sim 2^{(n \log n)R}$.

7.1.1 | Related Work on The Transmission Capacity of ISI-Poisson Channel

In the following we present previous result for DTPC with $\lfloor K \rfloor$ degree of ISI.

Capacity of the diffusive MC networks over linear time-invariant Poisson (LTI-Poisson) channel is considered in [33] where LTI-Poisson model as generalization of classical memoryless Poisson channel where they defined a class of memory limited networks that generalizes both the linear ISI channel and the LTI-Poisson model. They provide the following result for a memory-limited network: A network is called memory limited network (MLN) of order K if the input-output relation satisfies

$$p\left(\mathbf{y}_{K+1:n} \,|\, \mathbf{x}_{1:n}\right) = \prod_{t=K+1}^{n} p\left(y_t \,|\, \mathbf{x}_{t:t-K}\right) . \tag{7.1}$$

Given an MLN defined by (7.1) and $r \in \mathbb{N}$, a block memoryless channel is defined as follows

$$p\left(\mathbf{y}_{K+1:k+r} \,|\, \mathbf{x}_{1:K+r}\right) = \prod_{t=K+1}^{K+r} p\left(y_t \,|\, \mathbf{x}_{t:t-K}\right) . \tag{7.2}$$

For an arbitrary $r \in \mathbb{N}$, capacity of a MLN is bounded by

$$\frac{r}{K+r}\,\mathbb{C}_r \leq \mathbb{C} \leq \mathbb{C}_r\,, \tag{7.3}$$

where $\mathbb{C}_r = \frac{1}{r}$ fraction of the capacity region of a block memoryless system of size r.

In [182], transmission of a bit over DTPC with memory is studied and problem of code design is presented as follows: Consider a DTPC with memory of order K and interference signal level $\lambda > 0$. Assume that $n \geq K+1$. Further assume that codewords $\mathbf{c}_i$ and $\mathbf{c}_j$ associated to the two messages satisfy maximum power constraint of the form $\sum_{t=1}^{n} c_{i,t} \leq P_{\max}$. Then the following codeword pairs is optimal in the sense of minimizing maximum likelihood decoder error probability:

$$\begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} = \begin{pmatrix} \underbrace{\begin{matrix} P_{\max} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{matrix}}_{K} & \underbrace{\begin{matrix} 0 & 0 & \dots & 0 \\ P_{\max} & 0 & \dots & 0 \\ P_{\max} & 0 & \dots & 0 \end{matrix}}_{n-K} \end{pmatrix} . \tag{7.4}$$

In [178], impact of the amount of memory on the performance for a diffusion-based MC channel modelled by Poisson channel with memory is characterized and a simple memory-limited decoder is proposed.

7.1.2 | Contributions

In this chapter, we consider MC systems employing molecule counting receivers with a large number of released molecules at the transmitter, see [5, Sec. IV]. Further, we assume that the received signal experiences ISI and follows the Poisson distribution. We formulate the problem of DI over the DTPC with memory under average and peak molecule release rate constraints to account for the limited molecule production / release rates of the transmitter. As our main objective, we investigate the fundamental performance limits of DI over the DTPC with ISI. In particular, this chapter makes the following contributions:

♢ **Generalized ISI Model**: In MC systems, often the number of channel taps K can be large, particularly for non-degrading signalling molecules in bounded environments, which leads to a long channel impulse response (CIR). In addition, the value of K increases not only with the dispersiveness of the channel but also with the symbol rate. Therefore, it is of interest to investigate the asymptotic limits of the system for large symbol rates (leading to large K) and large codeword lengths n. To do so, we consider a generalized ISI model that captures the ISI-free channel (i.e., $K = 1$), ISI channels with constant $K > 1$, and ISI channels for which K increases with the codeword length n (e.g., due to increasing symbol rate). To the best of the authors' knowledge, such a generalized ISI model has not been studied in the literature, yet.

♢ **Codebook Scale**: We establish that the codebook size of the DTPC with ISI for deterministic encoding scales in n similar to the memoryless DTPC [51], namely super-exponentially in the codeword length ($\sim 2^{(n \log n)R}$), even when the number of ISI taps scale as $K = 2^{\kappa \log n}$ for any $\kappa \in [0, 1)$, which we refer to as the ISI rate. This observation suggests that memory does not change the scale of the codebook derived for memoryless DTPC [51] and Gaussian channels [105].

♢ **Capacity Bounds**: We derive DI capacity bounds for the DTPC with constant $K \geq 1$ and growing ISI $K = 2^{\kappa \log n}$, respectively. We show that for constant K, the proposed lower and upper bounds on R are independent of K, whereas for growing ISI, they are functions of the ISI rate κ. Moreover, we show that optimizing κ leads to an effective identification rate [bits/s] that scales linearly with n, which is in contrast to the typical transmission rate [bits/s] that is independent of n.

♢ **Technical Novelty in The Capacity Proof**: To obtain the proposed lower bound, the existence of an appropriate sphere packing within the input space, for which the distance between the centers of the spheres does not fall below a certain value, is guaranteed. This packing incorporates the effect of ISI as a function of κ. In particular, we consider the packing of hyper spheres inside a larger hyper cube, whose radius grows in both the codeword length n and the ISI rate κ, i.e., $\sim n^{\frac{1+\kappa}{4}}$. For derivation of the upper bound, we assume that for given sequences of codes with vanishing error probabilities, a certain minimum distance between the codewords is asserted, where this distance depends on the ISI rate and decreases as K grows.

7.1.3 | Organization

The remainder of this chapter is structured as follows. In Section 7.2, system model is explained and the required preliminaries regarding DI codes are established. Section 7.3 provides the main contributions and results on the message identification capacity of the DTPC with ISI. Finally, Section 7.4 of the paper concludes with a summary and directions for future research.

7.2 | System Model and Preliminaries

In this section, we present the adopted system model and establish some preliminaries regarding DI coding.

7.2.1 | System Model

We consider an identification-focused communication setup, where the decoder seeks to accomplish the following task: Determining whether or not an specific message was

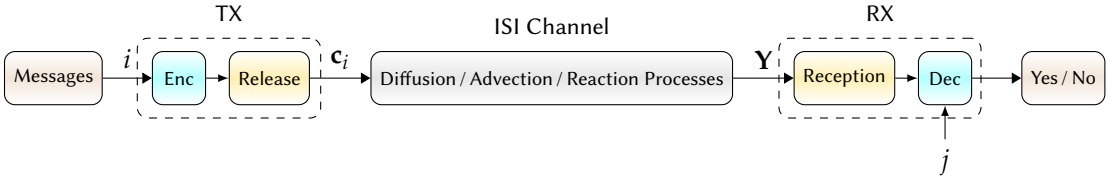


Figure 7.1: End-to-end transmission chain for DI communication in a generic MC system modelled as a DTPC. Relevant processes in the molecular channel include diffusion, advection, and chemical reactions. The transmitter maps message i onto a codeword $\mathbf{c}_i$. The receiver is provided with an arbitrary message j, and given the channel output vector $\mathbf{Y}$, it asks whether j is identical to i or not.

sent by the transmitter[1]; see Figure 7.1. To attain this objective, a coded communication between the transmitter and the receiver over n channel uses of an MC channel[2] is established. We consider the Poisson channel $\mathcal{P}$ which arises as a channel model in the context of MC for molecular counting receivers [9]. Let $X \in \mathbb{R}_{\geq 0}$ and $Y \in \mathbb{N}_0$ denote random variables (RVs) modeling the rate of molecule release by the transmitter and the number of molecules observed at the receiver, respectively. We consider a stochastic release model, where for the t-th channel use, the transmitter releases molecules with rate x_t (molecules/second) over a time slot of T_s seconds into the channel [9]. These molecules propagate through the channel via diffusion and/or advection, and may even be degraded in the channel via enzymatic reactions [5]. The receive is assumed to be equipped with a counting-type mechanism which is able to enumerate the number of received molecules observed in a determined volume.

The channel memory is modelled by a length K sequence of probability values, i.e., $\mathbf{p} = [p_0, p_1, \ldots, p_{K-1}]$. The value p_k in specifies the probability that a given molecule released by the transmitter at the beginning time slot t, is observed at the receiver during time slot $t + k$ and depends on the propagation environment (e.g., diffusion, advection, and reaction processes) and the reception mechanism (e.g., transparent, absorbing, or reactive receiver) as well as the distance between transmitter and receiver, see [5, Sec. III] for the characterization of $\mathbf{p}$ for various MC setups. Let $\rho_k \stackrel{\text{def}}{=} p_k T_s$ where the value $p_k \in (0, 1]$ denotes the probability that a given molecule released by the transmitter at the beginning time slot t, is observed at the receiver during time slot

[1] We assume that the transmitter does not know which message the decoder is interested in. This assumption is justified by the fact that otherwise, entire communication setting is specialized to transmission of only one indicator bit between Alice and Bob.

[2] The proposed performance bounds works regardless of whether or not an specific code is used for communication, although proper codes may be required to approach such performance limits.

$t+k$.

When the number of released molecules is large but only a small fraction of them arrives at the receiver, the relation of channel output Y and input X is characterized as follows [5, 9]:

$$Y_t = \text{Pois}\left(X_t^{\rho} + \lambda\right), \tag{7.5}$$

where

$$X_t^{\rho} \stackrel{\text{def}}{=} \sum_{k=0}^{K-1} \rho_k X_{t-k}, \tag{7.6}$$

is the mean number of observed molecules; see Figure 7.2, due to the release of the transmitter and the constant $\lambda \in \mathbb{R}_{>0}$ is the mean number of observed interfering molecules originating from external noise sources which employ the same type of molecule as the considered MC system. Let $\mathbf{x}_t^* \stackrel{\text{def}}{=} (x_{t-K+1}, \ldots, x_t)$ be the vector of the K most recently released symbols. Then, the letter-wise transition probability law is given by

$$V(y_t|\mathbf{x}_t^*) = \frac{e^{-\left(x_t^{\rho}+\lambda\right)}\left(x_t^{\rho}+\lambda\right)^{y_t}}{y_t!}. \tag{7.7}$$

We assume that different channel uses given any K previous input symbols are statistically independent, which is a valid assumption for, e.g., fully absorbing receivers [5]. Therefore, for n channel uses, the transition probability law is given by

$$V^{\bar{n}}(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{\bar{n}} V(y_t|\mathbf{x}_t^*) = \prod_{t=1}^{\bar{n}} \frac{e^{-\left(x_t^{\rho}+\lambda\right)}\left(x_t^{\rho}+\lambda\right)^{y_t}}{y_t!}, \tag{7.8}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_{\bar{n}})$ denote the transmitted codeword and the received signal, respectively, with $\bar{n} = n + K - 1$. We assume that $x_t = 0$ when $t > n$ or $t < 0$. The peak and average molecule release rate constraints on the codewords are

$$0 \le x_t \le P_{\max} \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n} x_t \le P_{\text{avg}}, \tag{7.9}$$

respectively, $\forall t \in [\![n]\!]$, where $P_{\max} > 0$ and $P_{\text{avg}} > 0$ constrain the rate of molecule release per channel use and over the entire n channel uses in each codeword, respectively.

Remark 7.2.1 (Input Constraint Interpretation). *We note that while the average power constraint for the Gaussian channel is a non-linear (square) function of the symbols (signifying the signal energy), here for the DTPC, it is a linear function (signifying the number of released molecules) [9].*

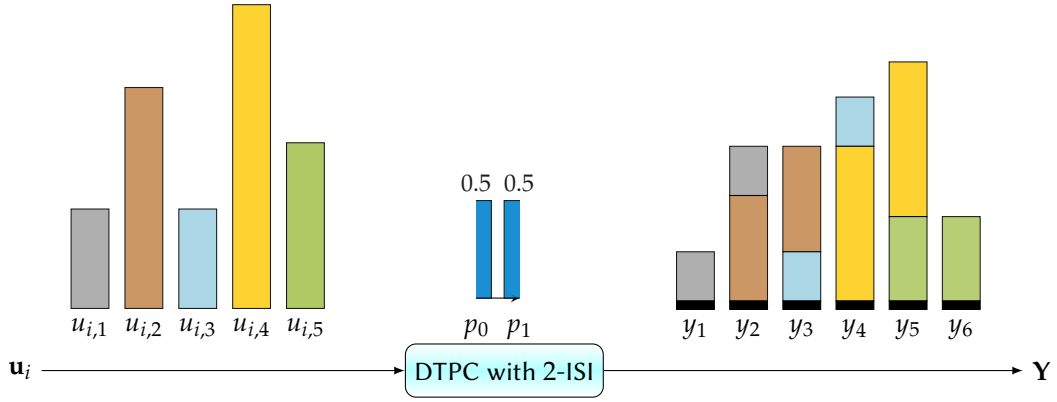


Figure 7.2: A DTPC with 2-ISI channel with $\mathbf{p} = (0.5, 0.5)$. Channel takes an input sequence of non-negative real numbers and outputs a sequence with length $n + K - 1 = 5 + 2 - 1 = 6$ of integer numbers where each integer is a Poisson distributed random variable whose mean is sum of previous marked as accumulation of different colors. The constant interference λ is depicted in black.

7.2.2 | DI Coding For The DTPC With Memory

The definition of a DI code for the DTPC $\mathcal{P}$ is given below.

Definition 7.2.1 (ISI-Poisson DI Code). *An $(n, M(n,R), K(n,\kappa), e_1, e_2)$ DI code for a DTPC $\mathcal{P}$ under average and peak molecule release rate constraints of P_{ave} and P_{max}, respectively, and for integers $M(n,R)$ and $K(n,\kappa)$, respectively, where n and R are the codeword length and coding rate, respectively, is defined as a system $(\mathcal{C}, \mathscr{T})$, which consists of a codebook $\mathcal{C} = \{\mathbf{c}_i\}_{i\in[\![M]\!]} \subset \mathbb{R}_+^n$, such that*

$$0 \leq c_{i,t} \leq P_{max} \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n} c_{i,t} \leq P_{avg}\,, \tag{7.10}$$

$\forall i \in [\![M]\!]$, $\forall t \in [\![n]\!]$, and a collection of decoding regions $\mathscr{T} = \{\mathbb{T}_i\}_{i\in[\![M]\!]}$ with

$$\bigcup_{i=1}^{M(n,R)} \mathbb{T}_i \subset \mathbb{N}_0^{\bar{n}}\,, \tag{7.11}$$

and $1 \leq K = K(n,R) < n$ being the number of ISI channel taps[3]. Given a message $i \in [\![M]\!]$, the encoder transmits $\mathbf{c}_i$, and the decoder's aim is to answer the following question: Was a

[3] While in the definition of a DI code, no specific restriction on the functional form of $K(n,R)$ is imposed, in our capacity results, it will be turned out that for at most a sub-linear form of $\sim n^{\kappa/4}$ for $0 \leq \kappa < 1/4$ being an arbitrary constant approaching $1/4$, non-trivial achievability results would be yielded.

desired message j sent or not? There are two types of errors that may occur: Rejection of the true message (type I) or acceptance of a false message (type II). The corresponding error probabilities of the DI code $(\mathcal{C}, \mathscr{T})$ are given by

$$P_{e,1}(i) = 1 - \sum_{\mathbf{y} \in \mathbb{T}_i} V^{\bar{n}}(\mathbf{y} \,|\, \mathbf{c}_i) \quad \text{and} \quad P_{e,2}(i,j) = \sum_{\mathbf{y} \in \mathbb{T}_j} V^{\bar{n}}(\mathbf{y} \,|\, \mathbf{c}_i)\,, \tag{7.12}$$

and satisfy the following bounds $P_{e,1}(i) \leq e_1$ and $P_{e,2}(i,j) \leq e_2$, $\forall\, i, j \underset{i \neq j}{\in} [\![M]\!]$ and every e_1, $e_2 > 0$. A rate $R > 0$ is called achievable if for every $e_1, e_2 > 0$ and sufficiently large n, there exists an $(n, M(n,R), K(n,\kappa), e_1, e_2)$ DI code. The DI capacity of the DTPC $\mathcal{P}$ is defined as the supremum of all achievable rates, and is denoted by $\mathbb{C}_{DI}(\mathcal{P}, M, K)$.

7.3 | DI Capacity of DTPC With Memory

In this section, we first present our main results, i.e., lower and upper bounds on the achievable identification rates for the DTPC with ISI. Subsequently, we provide the detailed proofs of these bounds.

7.3.1 | Main Results

The DI capacity theorem for DTPC with ISI $\mathcal{P}$ is stated below.

Theorem 7.3.1. *Consider the DTPC with ISI, $\mathcal{P}$, and assume that the number of ISI channel taps grow* sub-linearly *with the codeword length, i.e., $K(n,\kappa) = 2^{\kappa \log n}$ where $\kappa \in [0, 1/4)$. Then, the DKI capacity of $\mathcal{P}$ subject to average and peak molecule release rate constraints of the form $n^{-1}\sum_{t=1}^{n} c_{i,t} \leq P_{ave}$ and $0 \leq c_{i,t} \leq P_{max}$, respectively, with $i \in [\![M]\!]$ and a codebook of* super-exponential *scale, i.e., $M(n,R) = 2^{(n \log n)R}$, is bounded by*

$$\frac{1 - 4\kappa}{4} \leq \mathbb{C}_{DI}(\mathcal{P}, M, K) \leq \frac{3}{2} + \kappa\,. \tag{7.13}$$

Proof. The proof of Theorem 7.3.1 consists of two parts, namely the achievability and the converse proofs, which are provided in Sections 7.3.2 and 7.3.3, respectively. □

Remark 7.3.1. *The result in Theorem 7.3.1 comprises the following three special cases in terms of K:*

- ***Unit $K = 1$**: This cases accounts for an ISI-free setup ($\kappa = 0$), which is valid when the symbol duration is large ($T_s \geq T_{cir}$), and implies $K = 1$ and $\kappa = 0$. Thereby, $\bar{R}_{eff}$ scales*

logarithmically with the codeword length n. This is in contrast to the transmission setting in which $\bar{R}_{eff}$ is independent of n (e.g., the well-known Shannon formula for the Gaussian channel). This result is known in the identification literature [23, 51].

- ***Constant $K > 1$:*** *When T_s is constant and $T_s < T_{cir}$, we have constant $K > 1$ which implies $\kappa \to 0$ as $n \to \infty$. Surprisingly, our capacity result in Theorem 7.3.1 reveals that the bounds for the DTPC with memory are in fact identical to those for the memoryless DTPC given in [51].*

- ***Growing K:*** *Our capacity results reveal that reliable identification is possible even when K scales with the codeword length as $\sim 2^{\kappa \log n}$. Moreover, the impact of ISI rate κ is reflected in the capacity lower and upper bounds in (7.13), where the bounds respectively decrease and increase in κ. While the upper bound on R_{eff} increases in κ, too, the lower bound in (7.16) suggests a trade-off in terms of κ, which is investigated in the Corollary 7.3.1.2.*

Corollary 7.3.1.1 (Effective Identification Rate). *Let us assume that the physical length of the CIR interval is fixed and given by T_{cir}. Further, assume that the K ISI taps span the CIR interval, T_{cir}. Then, the following relation between the symbol duration, T_S, and the number of ISI taps, K holds:*

$$T_S = T_{cir}/K = T_{cir} 2^{-\kappa \log n}, \tag{7.14}$$

for some $\kappa \in [0, 1/4)$. Now, let the effective identification rate, $\bar{R}_{eff}$, be defined as follows

$$\bar{R}_{eff} \stackrel{def}{=} \frac{\log M(n, R)}{n T_S} \tag{7.15}$$

(in bits/s). Then, the effective identification rate subject to average and peak molecule release rate constraints is bounded by

$$\frac{(1 - 4\kappa)\, n^{\kappa} \log n}{4 T_{cir}} \leq \bar{R}_{eff} \leq \frac{(3 + 2\kappa)\, n^{\kappa} \log n}{2 T_{cir}}. \tag{7.16}$$

Proof. The proof follows directly by substituting the capacity results in Theorem 7.3.1 into the definition of the effective rate and making further mathematical simplifications. □

Corollary 7.3.1.2 (Optimum ISI Rate). *The lower bound given in Corollary 7.3.1.1 is maximized for the following ISI rate*

$$l_{max}(n) = \frac{1}{4}\left(1 - \frac{4}{\ln n}\right), \tag{7.17}$$

where $n \in \mathbb{N}$. Moreover, the maximum ISI rate, l_{max}, provided in (7.17) yields the following lower bound on the effective identification rate, $\bar{R}_{eff}(n)$:

$$\bar{R}_{eff}(n) \geq \frac{\log e}{eT_{cir}} \cdot n^{\frac{1}{4}(1-\kappa)}. \tag{7.18}$$

Thereby, the normalized effective identification rate is lower bounded as follows

$$\liminf_{n\to\infty} \frac{\bar{R}_{eff}(n)}{n^{\frac{1}{4}(1-\kappa)}} \geq \frac{\log e}{eT_{cir}}. \tag{7.19}$$

Proof. The proof follows from differentiating the lower bound in Corollary 7.3.1.1 with respect to κ and equating it to zero. □

The effective identification rate $\bar{R}_{\text{eff}}$ [bits/s] in (7.15) consists of two terms, namely the identification rate per symbol $\frac{\log M(n,R)}{n}$ [bits/symbol] (which decreases with κ for the lower bound in (7.13)) and the symbol rate $\frac{1}{T_s}$ [symbol/s] (which increases with κ). The above corollary reveals that in order to maximize $\bar{R}_{\text{eff}}$, it is optimal to set the trade-off for κ such that the identification rate, i.e.,

$$\frac{\log M(n,R)}{n} = \frac{(1-4\kappa_{\max})\log n}{4} = \log e, \tag{7.20}$$

becomes independent of n but the symbol rate scales polynomially with fractional exponent in n, i.e.,

$$1/T_S = n^{\frac{1}{4}(1-\kappa)}/T_{\text{cir}} = 2^{O(\log n)}. \tag{7.21}$$

As a result, in contrast to the typical transmission settings where the effective rate is independent of n, here, the effective identification rate $\bar{R}_{\text{eff}}$ for the optimal κ linearly grows in n. This completes the proof of Corollary 7.3.1.2.

Before we provide the proof, we highlight some insights obtained from Theorem 7.3.1 and corresponding proof.

Codebook Scale: Theorem 7.3.1 reveals a different behavior compared to the traditional scaling of the codebook size as a function of codeword length n. The bounds given in Theorem 7.3.1 are valid in the super-exponential scale of $M = 2^{(n\log n)R}$ which is in between the conventional exponential and double exponential codebook sizes (see Figure 7.3). Therefore, Theorem 7.3.1 induces that the relevant coding rate for DI capacity should be defined as

$$R = \frac{\log M}{n \log n}. \tag{7.22}$$

to guarantee that the capacity bounds represent informative non-zero finite values. The capacity values in the standard codebook sizes, i.e., exponential and double exponential, are infinite and zero, respectively [98, see Rem. 1].

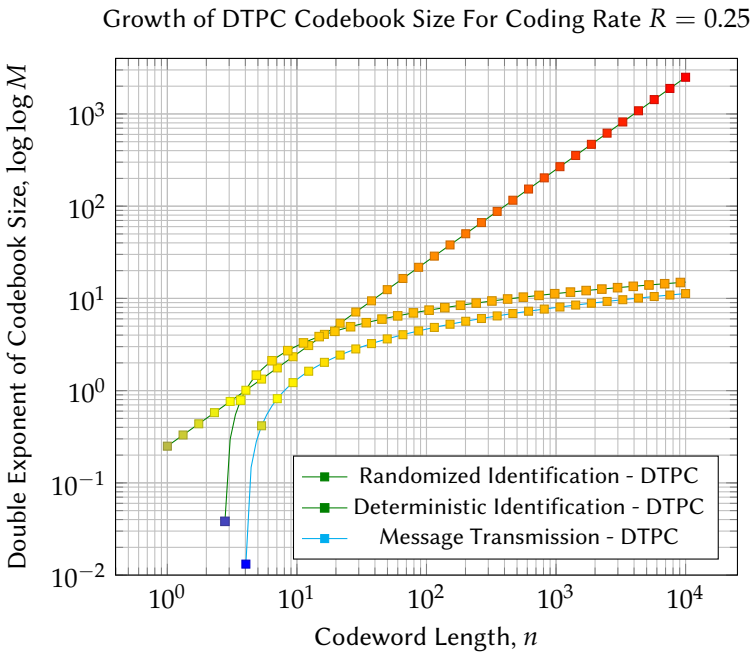


Figure 7.3: Double exponent of DI for DTPC with memory lies in between the conventional exponential and double exponential codebook sizes for transmission [28] and RI [23]. Even though the distance between double exponents of DI and TR decreases when in asymptotic regime of $n \to \infty$, still this distance is large enough that DI capacity with a codebook size of exponentially in the codeword length becomes infinite. The double exponent distance between DI and RI when $n \to \infty$ increases which means that RI in the asymptotic regime is much stronger than DI in terms of having a larger codebook size. Therefore, although the DI codebook scale lies in the middle of TR and RI, it is more inclined to the TR's exponential scale.

Molecule Budget: Our proposed capacity bounds with a codebook size of super-exponentially large in the codeword length n does not reflect the key values on the input constraints P_{ave} and $P_{\max}$ in the asymptotic regime of $n \to \infty$ and becomes independent of such values as n increases. However, for finite n, the codebook size is indeed a function of P_{ave} and $P_{\max}$. This observation can be drawn from our analysis in the achievability proof, where the codebook size before division by the dominant term (see (7.45)) is

$$M(n,R) = 2^{(n\log n)R + n(\log \frac{A}{e\sqrt{a}}) + o(n)}, \tag{7.23}$$

where $A = \min(P_{\text{ave}}, P_{\max})$ and $a > 0$ is a parameter of the codebook construction. In other words, the codebook size increases as A increases; however, since A appears in a term that is exponential in n, i.e., $\sim 2^{n(\log \frac{A}{e\sqrt{a}})}$, the influence of A becomes negligible compared to the dominant super-exponential term, i.e., $2^{(n \log n)R}$ as $n \to \infty$[4]. While the proof of Theorem 7.3.1 mainly concerns the asymptotic regime of $n \to \infty$, we are still able to get some insight for finite n, too. For instance, the error constraints in (7.12) can be met by the proposed achievable scheme even for finite n if A is sufficiently large and $a = \Omega(A^8)$ A comprehensive study of the achievable DI rates for finite n constitutes an interesting research topic for future work, but is beyond the scope of this chapter.

Memory: Our derived capacity bounds in the super-exponential scale incorporate the impact of memory degree K in terms of memory rate κ. Such a rate is different than the identification coding rate and characterize strength of the ISI effect by a scalar $\kappa \in [0, \frac{1}{2}]$ defined as

$$\kappa = \frac{\log K}{\log n} = \log \frac{K}{n}, \tag{7.24}$$

Roughly, the DI capacity of Poisson channels whose ISI is of order at strictly less than $\sqrt{n}$ have non-zero lower bounds. κ-dependent lower and upper bounds for the capacity reveal that when ISI is stronger, bounds are weaker and when ISI is weak, bounds are tight and optimum. See Remark7.3.1 for further discussion on different functional forms of K.

Adopted Decoder: Before going through the details of the achievability proof, we will present some insight into the proposed decoder. In particular, in the proposed achievable scheme, we adopt a distance decoder that decides in favour of a candidate codeword based on the distance between the received vector and expected value of the received vector if such a candidate codeword was really sent by the transmitter. More specifically, upon observing an output sequence $\mathbf{y}$ at the receiver, the decoder declares that message j was sent if the following condition is met

$$\left| \|\mathbf{y} - \mathbb{E}(\mathbf{Y}|\mathbf{c}_j)\|^2 - \|\mathbf{y}\|_1 \right| \leq \bar{n}\delta_n, \tag{7.25}$$

where δ_n is referred to as a decoding threshold and $\mathbf{c}_j = [c_{j,1}, \ldots, c_{j,n}]$ is the codeword associated with message j. Unlike the distance decoder used for Gaussian channels [105],

[4] Recall that the codebook size for the transmission capacities of both the DTPC [169, see Eq. (5)] and the Gaussian channel [170, 171] scale with $2^{n \log \sqrt{A}}$ in terms of A.

which includes only the distance term $\|\mathbf{y} - \mathbb{E}(\mathbf{Y}|\mathbf{c}_j)\|$, the proposed decoder provided in (7.25) requires subtraction of an additional correction term $\|\mathbf{y}\|_1$. This correction term stems from the fact that the noise in the DTPC with ISI is signal (input codeword) dependent [5]. Therefore, the variance of $\|\mathbf{y} - \mathbb{E}(\mathbf{Y}|\mathbf{c}_j)\|$ depends on the adopted codeword $\mathbf{c}_j$ which implies that, unlike for the Gaussian channel, here the radius of the decoding region is not constant for all the codewords. To account for this fact, we include the correction term $\|\mathbf{y}\|_1$.

7.3.2 | Achievability

The achievability proof consists of the following two steps.

- **Step 1:** We propose a codebook construction and derive an analytical lower bound on the corresponding codebook size using inequalities for the sphere packing density.
- **Step 2:** We prove that this codebook leads to an *achievable* rate by proposing a decoder and showing that the corresponding type I and type II error probabilities vanish as $n \to \infty$.

A DI code for the DTPC, $\mathcal{P}$, is constructed as follows.

Input constraint adaptation: We restrict ourselves to codewords that meet the condition $0 \leq c_{i,t} \leq P_{\text{ave}}, \forall\, i \in [\![M]\!], \forall\, t \in [\![n]\!]$, which ensures that both average and peak constraints in (7.10) are met for $P_{\text{ave}} > P_{\max}$ and $P_{\text{ave}} \leq P_{\max}$:

1. $P_{\text{ave}} > P_{\max}$: In this case, the condition $0 \leq c_{i,t} \leq P_{\max}, \forall\, i \in [\![M]\!], \forall\, t \in [\![n]\!]$, yields $n^{-1}\sum_{t=1}^{n} c_{i,t} \leq P_{\text{ave}}$. In this case, the average constraint trivially holds and we exclude this scenario from the analysis.
2. $P_{\text{ave}} \leq P_{\max}$: Then, the condition $0 \leq c_{i,t} \leq P_{\text{ave}}, \forall\, i \in [\![M]\!], \forall\, t \in [\![n]\!]$, implies both $0 \leq c_{i,t} \leq P_{\max}$ and $n^{-1}\sum_{t=1}^{n} c_{i,t} \leq P_{\text{ave}}$.

Thus, for the construction of the codebook in the next steps, we only require that $0 \leq c_{i,t} \leq P_{\text{ave}}, \forall\, i \in [\![M]\!], \forall\, t \in [\![n]\!]$.

Convoluted codebook construction: In the following, instead of directly constructing the original codebook $\mathcal{C} = \{\mathbf{c}_i\} \subset \mathbb{R}_+^n$, with $i \in [\![M]\!]$, we present a construction of a codebook called convoluted codebook and show that the original codebook can be uniquely reconstructed for a convoluted codebook. In particular, the convoluted codebook is denoted by $\mathcal{C}^{\rho} = \{\mathbf{c}_i^{\rho}\} \subset \mathbb{R}_+^n$, with $i \in [\![M]\!]$, where each $\mathbf{c}_i^{\rho} \triangleq (\mathbf{c}_{i,1}^{\rho}, \ldots, \mathbf{c}_{i,n}^{\rho})$ is

referred to as a convoluted codeword whose symbols are formed as a linear combination (convolution) of the L most recent symbols of codeword $\mathbf{c}_i \triangleq (\mathbf{c}_{i,1}, \ldots, \mathbf{c}_{i,n})$ and CIR vector $\boldsymbol{\rho}$, i.e.,

$$c^{\boldsymbol{\rho}}_{i,t} \triangleq \sum_{l=0}^{K-1} \rho_k c_{i,t-k}. \tag{7.26}$$

Observe that the convoluted symbol $c^{\boldsymbol{\rho}}_{i,t}$ represents the expected value of the signal observed at the receiver after the release of $c_{i,t}$ molecules by the transmitter. The proposed convoluted codebook construction is motivated by the structure of the ISI channel and the choice of the distance decoder given in (7.25). More specifically, the term $\mathbb{E}(\mathbf{Y} \,|\, \mathbf{c}_j)$ for $j \in [\![M]\!]$ given in (7.25) is the center of the distance decoder and includes the convoluted codeword, i.e., $\mathbf{c}^{\boldsymbol{\rho}}_j$.

In order to use the convoluted codebook, we have to show that the original codewords $\mathbf{c}_i$ can be uniquely derived from the convoluted codewords $\mathbf{c}^{\boldsymbol{\rho}}_i$, i.e., there is a one-to-one mapping between the convoluted and the original codebooks. To show this, let us first define the set of feasible original and convoluted codewords, respectively, as:

$$\mathbb{C}_0 = \mathbb{Q}_{\mathbf{0}(n,P_{\text{ave}})} \triangleq \Big\{ \mathbf{c}_i \in \mathbb{R}^n : \; 0 \leq c_{i,t} \leq P_{\text{ave}}, \forall\, i \in [\![M]\!], \forall\, t \in [\![n]\!] \Big\} \tag{7.27}$$

$$\mathbb{C}^{\boldsymbol{\rho}}_0 \triangleq \Big\{ \mathbf{c}^{\boldsymbol{\rho}}_i \in \mathbb{R}^n : \; c^{\boldsymbol{\rho}}_{i,t} \triangleq \sum_{l=0}^{K-1} \rho_k c_{i,t-k}, \; \mathbf{c}_i \in \mathbb{C}_0, \forall\, i \in [\![M]\!] \Big\}. \tag{7.28}$$

Unfortunately, unlike the feasible set of the original codewords $\mathbb{C}_0$, the feasible set of the convoluted codewords $\mathbb{C}^{\boldsymbol{\rho}}_0$ lacks the simple structure and geometry needed for the calculation of the volume and rate analysis. To cope with this issue, we target a *subset* of $\mathbb{C}^{\boldsymbol{\rho}}_0$ that enjoys a suitable structure with well-known geometry and analytic volume formula, namely the following hyper cube:

$$\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}}) = \{ \mathbf{c}^{\boldsymbol{\rho}}_i : \; 0 \leq c^{\boldsymbol{\rho}}_{i,t} \leq \bar{P}_{\text{ave}}, \forall\, i \in [\![M]\!], \forall\, t \in [\![n]\!] \}, \tag{7.29}$$

where

$$\bar{P}_{\text{ave}} \triangleq \min_{\substack{i \in [\![M]\!]; \\ \mathbf{c}^{\boldsymbol{\rho}}_i \in \mathbb{C}_1 \cap \mathbb{C}_2^c}} \; \min_{\substack{t \in [\![n]\!]; \\ t-K+1 \leq \bar{t} \leq t}} c^{\boldsymbol{\rho}}_{i,t}, \tag{7.30}$$

where $\bar{t}$ is a specific symbol index for which the corresponding input symbol yields a non-zero number of released molecules from the transmitter, i.e., $\left\lfloor T_R c_{i,\bar{t}} \right\rfloor \geq 1$. Moreover, sets $\mathbb{C}_1$ and $\mathbb{C}_2$ are given by

$$\mathbb{C}_1 = \mathbb{Q}_{\mathbf{0}}(n, P'_{\text{ave}}) \triangleq \Big\{ \mathbf{c}^{\boldsymbol{\rho}}_i \in \mathbb{R}^n \; : \; 0 \leq c^{\boldsymbol{\rho}}_{i,t} \leq P'_{\text{ave}}, \forall\, i \in [\![M]\!], \forall\, t \in [\![n]\!] \Big\},$$

$$\mathbb{C}_2 = \left\{ \mathbf{c}_i^{\boldsymbol{\rho}} \in \mathbb{R}^n \ : \ c_{i,t} \geq 0, \forall\, i \in [\![M]\!], \forall\, t \in [\![n]\!] \right\}, \tag{7.31}$$

where $P'_{\text{ave}} \triangleq \rho_0 P_{\text{ave}}$.

Next, we have to show that the volume of $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$ is non-zero (i.e., $\bar{P}_{\text{ave}}$ is bounded away from zero) and $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}}) \subseteq \mathbb{C}_0^{\boldsymbol{\rho}}$. The former follows from the fact that $\bar{P}_{\text{ave}}$ tends to zero only if all symbols of at least one of the original codewords are arbitrary close to zero. Such a single all-zero codeword can be excluded without affecting the rate analysis. To prove $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}}) \subseteq \mathbb{C}_0^{\boldsymbol{\rho}}$, we show that the original codeword $\mathbf{c}_i$ obtained from $\mathbf{c}_i^{\boldsymbol{\rho}} \in \mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$ belongs to $\mathbb{C}_0$, namely the extracted original symbols must meet $0 \leq c_{i,t} \leq P_{\text{ave}}$. We first show that $c_{i,t} \geq 0$ holds via contradiction. In other words, we assume $\mathbf{c}_i^{\boldsymbol{\rho}} \in \mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$ but the corresponding original codeword meets $\mathbf{c}_i \in \mathbb{C}_2^c$. This already contradicts the fact that $\bar{P}_{\text{ave}} > 0$, see (7.30). To show $c_{i,t} \leq P_{\text{ave}}$, we use the following chain of inequalities assuming $\mathbf{c}_i^{\boldsymbol{\rho}} \in \mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$:

$$\begin{aligned}
\rho_0 c_{i,1} &\leq \bar{P}_{\text{ave}} \leq P'_{\text{ave}} \\
\rho_0 c_{i,2} + \rho_1 c_{i,1} &\leq \bar{P}_{\text{ave}} \leq P'_{\text{ave}} \\
\rho_0 c_{i,3} + \rho_1 c_{i,2} + \rho_2 c_{i,1} &\leq \bar{P}_{\text{ave}} \leq P'_{\text{ave}} \\
&\vdots \\
\rho_0 c_{i,n} + \rho_1 c_{i,n-1} + \ldots + \rho_{K-1} c_{i,n-K+1} &\leq \bar{P}_{\text{ave}} \leq P'_{\text{ave}},
\end{aligned} \tag{7.32}$$

where $\bar{P}_{\text{ave}} \leq P'_{\text{ave}}$ holds since $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}}) \subset \mathbb{C}_1$, see (7.30). The above inequalities can be rewritten as follows

$$\begin{aligned}
c_{i,1} &\leq P'_{\text{ave}}/\rho_0 = P_{\text{ave}} \\
c_{i,2} &\leq \frac{P'_{\text{ave}} - \rho_0 c_{i,1}}{\rho_0} \leq P'_{\text{ave}}/\rho_0 = P_{\text{ave}} \\
&\vdots \\
c_{i,n} &\leq \frac{P'_{\text{ave}} - \sum_{t=1}^{K-1} \rho_k c_{i,t-k}}{\rho_0} \leq P'_{\text{ave}}/\rho_0 = P_{\text{ave}},
\end{aligned} \tag{7.33}$$

where we used the fact that $c_{i,t} \geq 0$. Hence, condition $\|\mathbf{c}_i\|_\infty \leq P_{\text{ave}}$ holds for the extracted original codewords. In summary, we showed that for convoluted codewords $\mathbf{c}_i^{\boldsymbol{\rho}} \in \mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$, there is a unique feasible original codeword $\mathbf{c}_i \in \mathbb{Q}_{\mathbf{0}}(n, P_{\text{ave}})$. Therefore, the rate analysis of the convoluted codebook is also valid for the original codebook.

Calculation of the codebook size/rate: We use a packing arrangement of non-overlapping hyper spheres of radius $r_0 = \sqrt{n\epsilon_n}$ in a hyper cube with edge length $\bar{P}_{\text{ave}}$,

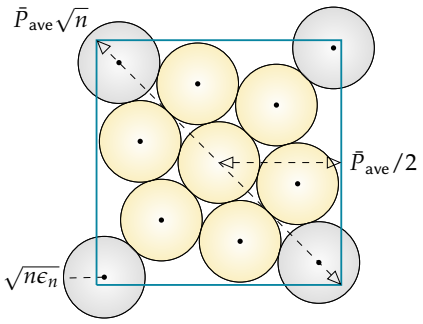


Figure 7.4: Illustration of a saturated sphere packing inside a cube, where small spheres of radius $r_0 = \sqrt{n\epsilon_n}$ cover a larger cube. Dark gray colored spheres are not entirely contained within the larger cube, and yet they contribute to the packing arrangement. As we assign a codeword to each sphere center, the 1-norm and arithmetic mean of a codeword are bounded by $\bar{P}_{\text{ave}}$ as required.

where

$$\epsilon_n = \frac{3a}{4n^{\frac{1}{2}(1-(b+4\kappa))}}, \tag{7.34}$$

and $a > 0$ is a non-vanishing fixed constant, $0 < b < 1$ is an arbitrarily small constant, and $0 \leq \kappa < 1/4$.

Let $\mathscr{S}$ denote a sphere packing, i.e., an arrangement of M non-overlapping spheres $\mathcal{S}_{\mathbf{c}_i^\rho}(n, r_0)$, $i \in [\![M]\!]$, that are packed inside the larger cube $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$ with edge length $\bar{P}_{\text{ave}}$, see Figure 7.4. As opposed to standard sphere packing coding techniques [138], the spheres are not necessarily entirely contained within the cube. That is, we only require that the centers of the spheres are inside $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$, the spheres are disjoint from each other, and they have a non-empty intersection with $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$. The packing density $\Delta_n(\mathscr{S})$ is defined as the ratio of the saturated packing volume to the cube volume $\text{Vol}\left(\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})\right)$, i.e.,

$$\Delta_n(\mathscr{S}) \triangleq \frac{\text{Vol}\left(\bigcup_{i=1}^{M} \mathcal{S}_{\mathbf{c}_i^\rho}(n, r_0)\right)}{\text{Vol}\left(\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})\right)}. \tag{7.35}$$

Sphere packing $\mathscr{S}$ is called *saturated* if no spheres can be added to the arrangement without overlap. In particular, we use a packing argument that has a similar flavor as that for the Minkowski–Hlawka theorem for saturated packings [138]. Specifically, consider the saturated packing arrangement of

$$\bigcup_{i=1}^{M(n,R)} \mathcal{S}_{\mathbf{c}_i^\rho}(n, \sqrt{n\epsilon_n}) \tag{7.36}$$

spheres with radius $r_0 = \sqrt{n\epsilon_n}$ embedded within cube $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$. Then, for such an arrangement, we have the following lower [118, Lem. 2.1] and upper bounds [138, Eq. 45] on the packing density

$$2^{-n} \leq \Delta_n(\mathscr{S}) \leq 2^{-0.599n}. \tag{7.37}$$

In particular, in our subsequent analysis, we employ the lower bound given in (7.37), which can be proved as follows: For the saturated packing arrangement given in (7.36), there cannot be a point in the larger cube $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$ with a distance of more than $2r_0$ from all sphere centers. Otherwise, a new sphere could be added which contradicts the assumption that the union of $M(n, R)$ spheres with radius $\sqrt{n\epsilon_n}$ is saturated. Now, if we double the radius of each sphere, the spheres with radius $2r_0$ cover thoroughly the entire volume of $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$, that is, each point inside the hyper cube $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$ belongs to at least one of the small spheres. In general, the volume of a hyper sphere of radius r is given by [138, Eq. (16)]

$$\text{Vol}\left(\mathcal{S}_{\mathbf{x}}(n, r)\right) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r^n. \tag{7.38}$$

Hence, if the radius of the small spheres is doubled, the volume of

$$\bigcup_{i=1}^{M(n,R)} \mathcal{S}_{\mathbf{c}_i^{\boldsymbol{\rho}}}(n, \sqrt{n\epsilon_n})$$

is increased by 2^n. Since the spheres with radius $2r_0$ cover $\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})$, it follows that the original r_0-radius packing[5] has a density of at least 2^{-n}. We assign a convoluted codeword to the center $\mathbf{c}_i^{\boldsymbol{\rho}}$ of each small hyper sphere. The convoluted codewords satisfy the input constraint as $0 \leq c_{i,t}^{\boldsymbol{\rho}} \leq P'_{\text{ave}}, \forall t \in [\![n]\!], \forall i \in [\![M]\!]$, which is equivalent to

$$\left\| \mathbf{c}_i^{\boldsymbol{\rho}} \right\|_\infty \leq \bar{P}_{\text{ave}}. \tag{7.39}$$

Since the volume of each sphere is equal to $\text{Vol}(\mathcal{S}_{\mathbf{c}_1^{\boldsymbol{\rho}}}(n, r_0))$ and the centers of all spheres lie inside the cube, the total number of spheres is bounded from below by

$$M = \frac{\text{Vol}\left(\bigcup_{i=1}^{M} \mathcal{S}_{\mathbf{c}_i^{\boldsymbol{\rho}}}(n, r_0)\right)}{\text{Vol}(\mathcal{S}_{\mathbf{c}_1^{\boldsymbol{\rho}}}(n, r_0))} = \frac{\Delta_n(\mathscr{S}) \cdot \text{Vol}\left(\mathbb{Q}_{\mathbf{0}}(n, \bar{P}_{\text{ave}})\right)}{\text{Vol}(\mathcal{S}_{\mathbf{c}_1^{\boldsymbol{\rho}}}(n, r_0))}$$

[5] We note that the proposed proof of the lower bound in (7.37) is non-constructive in the sense that, while the existence of the respective saturated packing is proved, no systematic construction method is provided.

$$\geq 2^{-n} \cdot \frac{P_{\text{ave}}^{n}}{\text{Vol}(\mathcal{S}_{\mathbf{c}_1^{\rho}}(n, r_0))}, \tag{7.40}$$

where the inequality holds by (7.37). The bound in (7.40) can be written as follows

$$\begin{aligned} \log M &\geq \log\left(\bar{P}_{\text{ave}}^{n} \,/\, \text{Vol}\left(\mathcal{S}_{\mathbf{c}_1^{\rho}}(n, r_0)\right)\right) - n \\ &\geq n \log\left(\bar{P}_{\text{ave}}/\sqrt{\pi} r_0\right) + \log\left(\Gamma\left(n/2+1\right)\right) - n, \end{aligned} \tag{7.41}$$

where the last inequality exploits (7.38). The above bound can be further simplified as follows

$$\log M \geq n \log\left(\bar{P}_{\text{ave}}/\sqrt{\pi} r_0\right) + \log\left(\lfloor n/2 \rfloor!\right) - n, \tag{7.42}$$

where the equality exploits the following relation:

$$\Gamma\left(n/2+1\right) \overset{(a)}{=} \frac{n}{2}\Gamma\left(n/2\right) \overset{(b)}{\geq} \lfloor n/2 \rfloor\, \Gamma\left(\lfloor n/2 \rfloor\right) \overset{(c)}{\triangleq} \lfloor n/2 \rfloor!. \tag{7.43}$$

In the above equation, (a) holds by the recurrence relation of the Gamma function [172] for real $n/2$, (b) follows from $\lfloor n/2 \rfloor \leq n/2$, the monotonicity of the Gamma function [172] for $\lfloor n/2 \rfloor \geq 1.46 \equiv n \geq 4$, and (c) holds since for positive integer $\lfloor n/2 \rfloor$, we have $\Gamma\left(\lfloor n/2 \rfloor\right) = (\lfloor n/2 \rfloor - 1)!$, cf. [172]. Next, we proceed to simplify the factorial term given in (7.42). To this end, we exploit *Stirling's approximation*, i.e., $\log n! = n \log n - n \log e + o(n)$ [173, p. 52] with the substitution of $n = \lfloor n/2 \rfloor$, where $\lfloor n/2 \rfloor \in \mathbb{Z}$. Thereby, we obtain

$$\begin{aligned} \log M \geq\; & n \log \bar{P}_{\text{ave}} - n \log r_0 + \lfloor n/2 \rfloor \log\left(\lfloor n/2 \rfloor\right) \\ & - \lfloor n/2 \rfloor \log e + o\left(\lfloor n/2 \rfloor\right) - n, \end{aligned} \tag{7.44}$$

Therefore, for $r_0 = \sqrt{n\epsilon_n} = \sqrt{a'} n^{\frac{1+b+4\kappa}{4}}$, where $a' \triangleq 3a/4$, we have

$$\begin{aligned} \log M &\overset{(a)}{\geq} n \log \frac{\bar{P}_{\text{ave}}}{\sqrt{\pi a'}} - \left(\frac{1+b+4\kappa}{4}\right) n \log \sqrt{a} n \\ &\quad + (n/2-1)\log\left(n/2-1\right) - \lfloor n/2 \rfloor \log e + o\left(n/2-1\right) - n \\ &\overset{(b)}{\geq} n \log \frac{\bar{P}_{\text{ave}}}{\sqrt{\pi a'}} - \left(\frac{1+b+4\kappa}{4}\right) n \log \sqrt{a} n \\ &\quad + \frac{1}{2} n \log n - 2n - \log n - \frac{n}{2} \log e + o\left(n/2\right) \\ &= \left(\frac{1-(b+4\kappa)}{4}\right) n \log n + n\left(\log \bar{P}_{\text{ave}}/\sqrt{\pi a' e}\right) + O(n), \end{aligned} \tag{7.45}$$

where (a) follows from $\left\lfloor \frac{n}{2} \right\rfloor > \frac{n}{2} - 1$ and (b) holds since $\log(t-1) \geq \log t - 1$ for $t \geq 2$ and $\left\lfloor \frac{n}{2} \right\rfloor \leq \frac{n}{2}$ for integer n. Observe that the dominant term in (7.45) is of order $n \log n$. Hence, to obtain a finite value for the lower bound on the rate, R, (7.45) reveals that the scaling law of M is $2^{(n \log n)R}$. Therefore, we obtain

$$R \geq \frac{1}{n \log n} \left[\left(\frac{1 - (b + 4\kappa)}{4} \right) n \log n + n \left(\log \frac{\bar{P}_{\text{ave}}}{\sqrt{\pi a' e}} \right) + O(n) \right], \tag{7.46}$$

which tends to $(1 - 4\kappa)/4$ when $n \to \infty$ and $b \to 0$.

Encoder: Given message $i \in [\![M]\!]$, transmit $\mathbf{x} = \mathbf{c}_i$.

Proposed decoder: In order to analyze the error performance of the proposed codebook, we need to adopt a decoder which is introduced next. Before we proceed, for the sake of a concise analysis, we introduce the following conventions. Let:

- $Y_t(i) \sim \text{Pois}(c_{i,t}^{\boldsymbol{\rho}} + \lambda)$ denote the channel output at time t given that $\mathbf{x} = \mathbf{c}_i$.
- The output vector is the vector of symbols, i.e., $\mathbf{Y}(i) = (Y_1(i), \ldots, Y_{\bar{n}}(i))$.
- $\bar{y}_t(i) \triangleq y_t(i) - (c_{i,t}^{\boldsymbol{\rho}} + \lambda)$, where $y_t(i)$ is a realization of $Y_t(i)$.

Furthermore, let

$$\delta_n \triangleq 4\epsilon_n/3 = 4a/(3n^{\frac{1}{2}(1-(b+4\kappa))}), \tag{7.47}$$

where $0 < b < 1$ is an arbitrarily small constant and $0 \leq \kappa < 1/4$ with κ being the ISI rate. To identify whether a message $j \in [\![M]\!]$ was sent, the decoder checks whether the channel output $\mathbf{y}$ belongs to the following decoding set,

$$\mathbb{T}_j = \left\{ \mathbf{y} \in \mathbb{N}_0^{\bar{n}} \,:\, \left| T(\mathbf{y}, \mathbf{c}_j) \right| \leq \delta_n \right\}, \tag{7.48}$$

where

$$T(\mathbf{y}; \mathbf{c}_j) = \frac{1}{\bar{n}} \sum_{t=1}^{\bar{n}} \left(y_t - \left(c_{j,t}^{\boldsymbol{\rho}} + \lambda \right) \right)^2 - y_t, \tag{7.49}$$

is referred to as the *decoding metric* evaluated for observation vector $\mathbf{y}$ and codeword $\mathbf{c}_j$. Finally, let $e_1, e_2 > 0$ and $\zeta_0, \zeta_0', \zeta_1, \zeta_1' > 0$ be arbitrarily small constants.

Error analysis:

In the following, we exploit Chebyshev's inequality in order to establish upper bounds for the type I and type II error probabilities.

Type I error analysis: Consider the type I errors, i.e., the transmitter sends $\mathbf{c}_i$, yet $\mathbf{Y} \notin \mathbb{T}_i$. For every $i \in [\![M]\!]$, the type I error probability is bounded as

$$P_{e,1}(i) = \Pr\left(\mathbf{Y}(i) \in \mathbb{T}_i^c\right) = \Pr\left(\left|T(\mathbf{Y}(i), \mathbf{c}_j)\right| > \delta_n\right), \tag{7.50}$$

In order to bound $P_{e,1}(i_1)$ in (7.50), we apply Chebyshev's inequality, namely

$$\Pr\left(\left|T(\mathbf{Y}(i), \mathbf{c}_i) - \mathbb{E}\left[T(\mathbf{Y}(i), \mathbf{c}_i)\right]\right| > \delta_n\right) \leq \frac{\mathrm{Var}\left[T(\mathbf{Y}(i), \mathbf{c}_i)\right]}{\delta_n^2}. \tag{7.51}$$

First, we calculate the expectation of the *decoding metric* as follows

$$\begin{aligned}
\mathbb{E}\left[T(\mathbf{Y}(i), \mathbf{c}_i)\right] &\overset{(a)}{=} \frac{1}{\bar{n}} \sum_{t=1}^{\bar{n}} \mathbb{E}\left[\left(Y_t(i) - \left(c_{i,t}^{\boldsymbol{\rho}} + \lambda\right)\right)^2\right] - \mathbb{E}\left[Y_t(i)\right] \\
&\overset{(b)}{=} \frac{1}{\bar{n}} \sum_{t=1}^{\bar{n}} \mathrm{Var}\left[Y_t(i)\right] - \left(c_{i,t}^{\boldsymbol{\rho}} + \lambda\right) \\
&\overset{(c)}{=} \frac{1}{\bar{n}} \sum_{t=1}^{\bar{n}} \left(c_{i,t}^{\boldsymbol{\rho}} + \lambda\right) - \left(c_{i,t}^{\boldsymbol{\rho}} + \lambda\right) = 0,
\end{aligned} \tag{7.52}$$

where (a) follows from the linearity of expectation, (b) holds since $\mathbb{E}[(Y_t(i) - \mathbb{E}[Y_t(i)])^2] = \mathrm{Var}[Y_t(i)]$ and $\mathbb{E}[Y_t(i)] = c_{i,t}^{\boldsymbol{\rho}} + \lambda$, and (c) follows since $\mathrm{Var}[Y_t(i)] = \mathbb{E}[Y_t(i)] = c_{i,t}^{\boldsymbol{\rho}} + \lambda$. Second, in order to compute the upper bound in (7.51), we proceed to compute the variance of the *decoding metric*. Let us define

$$\psi_{\mathrm{Var}} \triangleq \sum_{t=1}^{\bar{n}} \mathrm{Var}\left[\bar{Y}_t^2(i) - Y_t(i)\right]. \tag{7.53}$$

Since, conditioned on $\mathbf{c}_i$, the channel outputs conditioned on the L most recent input symbols are uncorrelated, we obtain

$$\mathrm{Var}\left[T(\mathbf{Y}(i); \mathbf{c}_i)\right] = \frac{\psi_{\mathrm{Var}}}{\bar{n}^2}. \tag{7.54}$$

Next, we proceed to establish an upper bound $\psi_{\mathrm{Var}}^{\mathrm{UB}}$ for ψ_{Var}. To this end, let us define

$$\begin{aligned}
\psi_{\mathrm{Var}} &\triangleq \mathrm{Var}\left[\bar{Y}_t^2(i) - Y_t(i)\right] \\
&\overset{(a)}{=} \mathrm{Var}\left[Y_t^2(i) - \left(2\left(c_{i,t}^{\boldsymbol{\rho}} + \lambda\right) + 1\right) Y_t(i)\right] \\
&\overset{(b)}{=} \mathrm{Var}\left[Y_t^2(i)\right] + \left(2\left(c_{i,t}^{\boldsymbol{\rho}} + \lambda\right) + 1\right)^2 \mathrm{Var}\left[Y_t(i)\right] \\
&\quad - \left(4\left(c_{i,t}^{\boldsymbol{\rho}} + \lambda\right) + 2\right) \mathrm{Cov}\left[Y_t^2(i), Y_t(i)\right],
\end{aligned} \tag{7.55}$$

where (a) holds since $\bar{Y}_t(i) \triangleq Y_t(i) - (c^{\boldsymbol{\rho}}_{i,t} + \lambda)$ and the decomposition in (b) follows from the following identity for constants a and b:

$$\mathrm{Var}\Big[aX - bY\Big] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y] - 2ab\,\mathrm{Cov}[X,Y]. \tag{7.56}$$

Next, let us define

$$\psi_{\mathrm{Cov}} \triangleq \Big(4\Big(\bar{K}P_{\mathrm{avg}} + \lambda\Big) + 2\Big)\sqrt{\exp(8/\lambda)\Big(\bar{K}P_{\mathrm{avg}} + \lambda\Big)} = O(K^{3/2}), \tag{7.57}$$

with $\bar{K} \triangleq KT_R$. Now, we proceed to establish an upper bound on (7.55) as follows

$$\begin{aligned}
\psi_{\mathrm{Var}} &\overset{(a)}{\leq} \mathbb{E}\left[Y_t^4(i)\right] + \Big(2\Big(c^{\boldsymbol{\rho}}_{i,t} + \lambda\Big) + 1\Big)^2\Big(c^{\boldsymbol{\rho}}_{i,t} + \lambda\Big) \\
&\quad + \Big(4\Big(c^{\boldsymbol{\rho}}_{i,t} + \lambda\Big) + 2\Big)\sqrt{\mathbb{E}\left[Y_t^4(i)\right]\mathrm{Var}\left[Y_t(i)\right]} \\
&\overset{(b)}{\leq} \Big(\bar{K}P_{\mathrm{avg}} + \lambda\Big)^4\exp(8/\lambda) + \Big(2\Big(\bar{K}P_{\mathrm{avg}} + \lambda\Big) + 1\Big)^2 + \psi_{\mathrm{Cov}},
\end{aligned} \tag{7.58}$$

where (a) follows from the triangle inequality, i.e., $\alpha - \beta \leq |\alpha - \beta| \leq |\alpha| + |\beta|$ for real a and b, $\mathrm{Var}[Y_t^2(i)] \leq \mathbb{E}[Y_t^4(i)]$, $\mathrm{Var}[Y_t(i)] = c^{\boldsymbol{\rho}}_{i,t} + \lambda$, and $\mathrm{Cov}[X,Y] \leq \sqrt{\mathrm{Var}[X]\cdot\mathrm{Var}[Y]}$ for RVs with finite variances, (b) follows from $c_{i,t} \leq P_{\mathrm{avg}}, \forall i \in [\![M]\!], \forall t \in [\![n]\!]$, for a Poisson RV $Y_t(i) \sim \mathrm{Pois}\Big(\lambda\Big)$, an upper bound on the non-centered moments:

$$\mathbb{E}[Y_t^k(i)] \leq \mathbb{E}^k[Y_t(i)]\cdot\exp(k^2/2\,\mathbb{E}[Y_t(i)]), \tag{7.59}$$

(see [183, Th. 1]), and (7.57). Thereby, exploiting (7.51)–(7.55) and (7.58), we can establish the following upper bound on the type I error probability given in (7.50):

$$\begin{aligned}
P_{e,1}(i_1) &= \Pr\Big(\Big|T(\mathbf{Y}(i), \mathbf{c}_j)\Big| > \delta_n\Big) \\
&\overset{(a)}{\leq} \frac{\Big(\bar{K}P_{\mathrm{avg}}+\lambda\Big)^4\exp(8/\lambda) + \Big(2\Big(\bar{K}P_{\mathrm{avg}}+\lambda\Big)+1\Big)^2 + \psi_{\mathrm{Cov}}\Big)}{n\delta_n^2} \\
&\overset{(b)}{=} \frac{9\Big(\Big(\bar{K}P_{\mathrm{avg}}+\lambda\Big)^4\exp(8/\lambda) + \Big(2\Big(\bar{K}P_{\mathrm{avg}}+\lambda\Big)+1\Big)^2 + \psi_{\mathrm{Cov}}\Big)\Big)}{16a^2n^{b+4\kappa}} \\
&= \frac{O(K^4)}{n^{b+4\kappa}} = \frac{O(1)}{n^{b+4\kappa}} \\
&\leq e_1,
\end{aligned} \tag{7.60}$$

for sufficiently large n and arbitrarily small e_1, where (a) follows from (7.51), (7.54) and (7.58), and (b) follows from (7.47).

Type II error analysis: Next, we address type II errors, i.e., when $\mathbf{Y}(i) \in \mathbb{T}_j$ while the transmitter sent $\mathbf{c}_i$. Then, for every $i, j \in [\![M]\!]$, where $i \neq j$, the type II error probability is given by

$$P_{e,2}(i,j) = \Pr\left(\mathbf{Y}(i) \in \mathbb{T}_j\right) \leq \max_{1\leq j\leq K} \Pr\left(\left|T(\mathbf{Y}(i);\mathbf{c}_j)\right| \leq \delta_n\right), \tag{7.61}$$

where $T(\mathbf{Y}(i);\mathbf{c}_j)$ is a random variable modeling the decoding metric in (7.49), i.e.,

$$T(\mathbf{Y}(i);\mathbf{c}_j) = \frac{1}{\bar{n}} \sum_{t=1}^{\bar{n}} \left(Y_t(i) - \left(c^{\rho}_{j,t} + \lambda\right)\right)^2 - Y_t(i). \tag{7.62}$$

Next, we establish an upper bound on the RHS of (7.61), while we assume that j can be an arbitrary value from set $[\![K]\!]$. Further, let

$$\tilde{j} \triangleq \arg\max_{1\leq j\leq M} \Pr\left(\left|T(\mathbf{Y}(i);\mathbf{c}_j)\right| \leq \delta_n\right). \tag{7.63}$$

We note that if our analysis gives an upper bound on $\Pr(|T(\mathbf{Y}(i);\mathbf{c}_j)| \leq \delta_n)$ for arbitrary $j \in [\![M]\!]$, then the same upper bound is valid for $\Pr(|T(\mathbf{Y}(i);\mathbf{c}_{\tilde{j}})| \leq \delta_n)$. That is, we immediately obtain an upper bound for $\max_{1\leq j\leq M} \Pr(|T(\mathbf{Y}(i);\mathbf{c}_j)| \leq \delta_n)$ in (7.61).

Observe that (7.62) for $j = \tilde{j}$ can be rewritten as follows

$$T(\mathbf{Y}(i);\mathbf{c}_{\tilde{j}}) = \frac{1}{\bar{n}} \sum_{t=1}^{\bar{n}} \underbrace{\left(Y_t(i) - \left(c^{\rho}_{i,t} + \lambda\right) + \left(c^{\rho}_{i,t} - c^{\rho}_{\tilde{j},t}\right)\right)^2}_{\triangleq \phi_{i,\tilde{j},t}} - Y_t(i). \tag{7.64}$$

Observe that $\phi_{i,\tilde{j},t}$ in (7.64) can be expressed as

$$\phi_{i,\tilde{j},t} = \bar{Y}_t(i)^2 + \psi^2_{i,\tilde{j},t} + 2\bar{Y}_t(i)\psi_{i,\tilde{j},t}, \tag{7.65}$$

where

$$\bar{Y}_t(i) = Y_t(i) - \left(c^{\rho}_{i,t} + \lambda\right) \quad \text{and} \quad \psi_{i,\tilde{j},t} = c^{\rho}_{i,t} - c^{\rho}_{j,t}. \tag{7.66}$$

Then, define the following events

$$\mathcal{E}_{i,\tilde{j}} = \left\{\left|\sum_{t=1}^{\bar{n}} \left(\bar{Y}_t(i) + \psi_{i,\tilde{j},t}\right)^2 - Y_t(i)\right| \leq \bar{n}\delta_n\right\},$$
$$\mathcal{E}'_{i,\tilde{j}} = \left\{\sum_{t=1}^{\bar{n}} \left(\bar{Y}_t(i) + \psi_{i,\tilde{j}},t\right)^2 - Y_t(i) \leq \bar{n}\delta_n\right\},$$
$$\mathcal{E}''_{i,\tilde{j}} = \left\{\left|\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)\psi_{i,\tilde{j},t}\right| > \bar{n}\delta_n/2\right\},$$

$$\mathcal{E}'''_{i,\tilde{j}} = \left\{ \sum_{t=1}^{\bar{n}} \bar{Y}_t(i)^2 + \psi^2_{i,\tilde{j},t} - Y_t(i) \leq 2\bar{n}\delta_n \right\}. \tag{7.67}$$

Hence,

$$\begin{aligned} P_{e,2}(i,j) &\leq \Pr\left(\mathcal{E}_{i,\tilde{j}}\right) \\ &= \Pr\left(\left| \sum_{t=1}^{\bar{n}} \left(\bar{Y}_t(i) + \psi_{i,\tilde{j},t}\right)^2 - Y_t(i) \right| \leq \bar{n}\delta_n \right) \\ &\overset{(a)}{\leq} \Pr\left(\sum_{t=1}^{\bar{n}} \left(\bar{Y}_t(i) + \psi_{i,\tilde{j},t}\right)^2 - Y_t(i) \leq \bar{n}\delta_n \right) \\ &= \Pr\left(\mathcal{E}'_{i,\tilde{j}}\right), \end{aligned} \tag{7.68}$$

where (a) holds since $\alpha - \beta \leq |\alpha - \beta|$ for real α, β. Now, we apply the law of total probability to event $\mathcal{E}'_{i,\tilde{j}}$ with respect to the pair of $(\mathcal{E}''_{i,\tilde{j}}, \mathcal{E}''^c_{i,\tilde{j}})$, and obtain the following upper bound on the type II error probability,

$$\begin{aligned} P_{e,2}(i,j) &\leq K \cdot \Pr\left(\mathcal{E}'_{i,\tilde{j}}\right) \\ &= K \cdot \left[\Pr\left(\mathcal{E}'_{i,\tilde{j}} \cap \mathcal{E}''_{i,\tilde{j}}\right) + \Pr\left(\mathcal{E}'_{i,\tilde{j}} \cap \mathcal{E}''^c_{i,\tilde{j}}\right) \right] \\ &\overset{(a)}{\leq} K \cdot \left[\Pr\left(\mathcal{E}''_{i,\tilde{j}}\right) + \Pr\left(\mathcal{E}'_{i,\tilde{j}} \cap \mathcal{E}''^c_{i,\tilde{j}}\right) \right] \\ &\overset{(b)}{=} K \cdot \left[\Pr\left(\mathcal{E}''_{i,\tilde{j}}\right) + \Pr\left(\mathcal{E}'''_{i,\tilde{j}}\right) \right], \end{aligned} \tag{7.69}$$

where (a) follows from $\mathcal{E}'_{i,\tilde{j}} \cap \mathcal{E}''_{i,\tilde{j}} \subset \mathcal{E}''_{i,\tilde{j}}$ and (b) holds since the event $\mathcal{E}'_{i,\tilde{j}} \cap \mathcal{E}''^c_{i,\tilde{j}}$ yields event $\mathcal{E}'''_{i,\tilde{j}}$, with the following argument. Observe that,

$$\begin{aligned} \Pr\left(\mathcal{E}'_{i,\tilde{j}} \cap \mathcal{E}''^c_{i,\tilde{j}}\right) &\overset{(a)}{\leq} \Pr\left(\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)^2 + \psi^2_{i,\tilde{j},t} - Y_t(i) \leq 2\bar{n}\delta_n \right) \\ &= \Pr\left(\mathcal{E}'''_{i,\tilde{j}}\right), \end{aligned} \tag{7.70}$$

where (a) holds since given the complementary event $\mathcal{E}''^c_{i,\tilde{j}}$, we obtain

$$-\bar{n}\delta_n/2 \leq \sum_{t=1}^{\bar{n}} \bar{Y}_t(i)\psi_{i,\tilde{j},t} \leq \bar{n}\delta_n/2,$$

which implies that $-2\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)\psi_{i,\tilde{j},t} \leq \bar{n}\delta_n$. That is, event $\mathcal{E}'_{i,\tilde{j}} \cap \mathcal{E}''^c_{i,\tilde{j}}$ *yields* the event

$$\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)^2 + \psi^2_{i,\tilde{j},t} - Y_t(i) \leq 2\bar{n}\delta_n.$$

Now, we establish an upper bound on $\Pr(\mathcal{E}''_{i,\tilde{j}})$ by exploiting Chebyshev's inequality:

$$\begin{aligned}\Pr(\mathcal{E}''_{i,\tilde{j}}) &= \Pr\left(\left|\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)\psi_{i,\tilde{j},t}\right| > \bar{n}\delta_n/2\right)\\ &\leq \frac{\mathrm{Var}\left[\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)\psi_{i,\tilde{j},t}\right]}{(\bar{n}\delta_n)^2}\\ &= \frac{\sum_{t=1}^{\bar{n}} \mathrm{Var}\left[\bar{Y}_t(i)\psi_{i,\tilde{j},t}\right]}{(\bar{n}\delta_n)^2},\end{aligned} \tag{7.71}$$

where the last equality holds since the variance of the sum of uncorrelated RVs is the sum of the respective variances. Thereby,

$$\begin{aligned}\Pr(\mathcal{E}''_{i,\tilde{j}}) &\leq \frac{\sum_{t=1}^{\bar{n}} \psi^2_{i,\tilde{j},t}\mathrm{Var}\left[\bar{Y}_t(i)\right]}{(\bar{n}\delta_n)^2}\\ &= \frac{\sum_{t=1}^{\bar{n}} \left(c^{\boldsymbol{\rho}}_{i,t} - c^{\boldsymbol{\rho}}_{\tilde{j},t}\right)^2\mathrm{Var}\left[\bar{Y}_t(i)\right]}{(\bar{n}\delta_n)^2}\\ &\overset{(a)}{\leq} \frac{\sum_{t=1}^{\bar{n}} \left(c^{\boldsymbol{\rho}}_{i,t} + c^{\boldsymbol{\rho}}_{\tilde{j},t}\right)^2\mathrm{Var}\left[\bar{Y}_t(i)\right]}{(\bar{n}\delta_n)^2}\\ &\overset{(b)}{=} \frac{\sum_{t=1}^{\bar{n}} \left(c^{\boldsymbol{\rho}}_{i,t} + c^{\boldsymbol{\rho}}_{\tilde{j},t}\right)^2(c^{\boldsymbol{\rho}}_{i,t} + \lambda)}{(\bar{n}\delta_n)^2}\\ &\overset{(c)}{\leq} \frac{\|\mathbf{c}^{\boldsymbol{\rho}}_i + \mathbf{c}^{\boldsymbol{\rho}}_{\tilde{j}}\|^2(\bar{K}P_{\mathrm{ave}} + \lambda)}{(\bar{n}\delta_n)^2},\end{aligned} \tag{7.72}$$

where (a) exploits the triangle inequality, i.e., $\left|c^{\boldsymbol{\rho}}_{i,t} - c^{\boldsymbol{\rho}}_{\tilde{j},t}\right| \leq \left|c^{\boldsymbol{\rho}}_{i,t} + c^{\boldsymbol{\rho}}_{\tilde{j},t}\right|$, (b) follows since $\mathrm{Var}[\bar{Y}_t(i)] = c^{\boldsymbol{\rho}}_{i,t} + \lambda, \forall t \in [\![\bar{n}]\!]$, and (c) follows since $c^{\boldsymbol{\rho}}_{i,t} \leq \bar{K}P_{\mathrm{ave}} + \lambda$. Now, observe that

$$\begin{aligned}\left\|\mathbf{c}^{\boldsymbol{\rho}}_i + \mathbf{c}^{\boldsymbol{\rho}}_{\tilde{j}}\right\|^2 &\overset{(a)}{\leq} \left(\left\|\mathbf{c}^{\boldsymbol{\rho}}_i\right\| + \left\|\mathbf{c}^{\boldsymbol{\rho}}_{\tilde{j}}\right\|\right)^2\\ &\overset{(b)}{\leq} \left(\sqrt{n}\left\|\mathbf{c}^{\boldsymbol{\rho}}_i\right\|_\infty + \sqrt{n}\left\|\mathbf{c}^{\boldsymbol{\rho}}_{\tilde{j}}\right\|_\infty\right)^2\\ &\overset{(c)}{\leq} \left(\sqrt{n}\bar{K}P_{\mathrm{avg}} + \sqrt{n}\bar{K}P_{\mathrm{avg}}\right)^2\\ &= 4\bar{K}^2nP^2_{\mathrm{avg}},\end{aligned} \tag{7.73}$$

where (a) holds by the triangle inequality, (b) follows since $\|\cdot\| \leq \sqrt{n}\|\cdot\|_\infty$, and (c) is

valid by the definition of $\mathbf{c}_i^{\boldsymbol{\rho}}$, i.e., $\mathbf{c}_i^{\boldsymbol{\rho}} = \sum_{l=0}^{K-1} \rho_k c_{i,t-k}$, and (7.39). Hence,

$$\begin{aligned}
\Pr(\mathcal{E}''_{i,\tilde{j}}) &\leq \frac{\|\mathbf{c}_i^{\boldsymbol{\rho}} + \mathbf{c}_{\tilde{j}}^{\boldsymbol{\rho}}\|^2(\bar{K}P_{\text{ave}} + \lambda)}{(\bar{n}\delta_n)^2} \\
&\leq \frac{4\bar{K}^2P_{\text{avg}}^2(\bar{K}P_{\text{ave}} + \lambda)}{n\delta_n^2} \\
&= \frac{9\bar{K}^3P_{\text{avg}}^2(P_{\text{avg}} + \lambda)}{4a^2n^{b+\kappa+4l}} \\
&= \frac{O(K^3)}{n^{b+\kappa+4l}} \\
&\triangleq \zeta_0.
\end{aligned} \tag{7.74}$$

We now proceed with bounding $\Pr\left(\mathcal{E}'''_{i,\tilde{j}}\right)$ as follows. Based on the convoluted codebook construction, each convoluted codeword is surrounded by a sphere of radius $\sqrt{n\epsilon_n}$, that is

$$\left\|\mathbf{c}_i^{\boldsymbol{\rho}} - \mathbf{c}_{\tilde{j}}^{\boldsymbol{\rho}}\right\|^2 \geq 4n\epsilon_n = 3\bar{n}\delta_n, \tag{7.75}$$

where the last equality exploits (7.47). Thus, we can establish the following upper bound for event $\mathcal{E}'''_{i,\tilde{j}}$:

$$\begin{aligned}
\Pr(\mathcal{E}'''_{i,\tilde{j}}) &= \Pr\left(\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)^2 + \psi_{i,\tilde{j},t}^2 - Y_t(i) \leq 2\bar{n}\delta_n\right) \\
&= \Pr\left(\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)^2 - Y_t(i) \leq 2\bar{n}\delta_n - \psi_{i,\tilde{j},t}^2\right) \\
&\overset{(a)}{\leq} \Pr\left(\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)^2 - Y_t(i) \leq 2\bar{n}\delta_n - 3\bar{n}\delta_n\right) \\
&\overset{(b)}{\leq} \frac{\text{Var}\left[\sum_{t=1}^{\bar{n}} \bar{Y}_t(i)^2 - Y_t(i)\right]}{\bar{n}^2\delta_n^2} \\
&\overset{(c)}{=} \frac{\text{Var}\left[T(\mathbf{Y}(i), \mathbf{c}_i)\right]}{\delta_n^2} \\
&\overset{(d)}{=} \frac{9\left(\left(\bar{K}P_{\text{avg}}+\lambda\right)^4 \exp(8/\lambda) + \left(2\left(\bar{K}P_{\text{avg}}+\lambda\right)+1\right)^2 + \psi_{\text{Cov}}\right)}{16a^2n^{b+4\kappa}} \\
&\triangleq \zeta_1,
\end{aligned} \tag{7.76}$$

where (a) follows from (7.75), (b) holds from applying Chebyshev's inequality, (c) follows from similar arguments as provided for the type I error probability, i.e., the calculations provided in (7.53) and (7.54), (d) holds by (7.58).

To sum up, recalling (7.74), we obtain

$$\zeta_0 = \frac{9\bar{K}^3 P_{\text{avg}}^2 (P_{\text{avg}} + \lambda)}{4a^2 n^{b+4\kappa}} \overset{(a)}{=} \frac{O(K^3)}{n^{b+4\kappa}} \overset{(b)}{=} \frac{O(1)}{n^{b+\kappa}} \triangleq \zeta_0', \tag{7.77}$$

where (a) exploits $K = n^{\kappa}$. On the other hand, recalling (7.76), we obtain

$$\begin{aligned}\zeta_1 &= \frac{9\Big(\Big(\Big(\bar{K}P_{\text{avg}}+\lambda\Big)^4 \exp(8/\lambda)+\Big(2\Big(\bar{K}P_{\text{avg}}+\lambda\Big)+1\Big)^2+\psi_{\text{Cov}}\Big)\Big)}{16a^2 n^{b+4\kappa}} \\ &\overset{(a)}{=} \frac{O(K^4)}{n^{b+4\kappa}} \overset{(b)}{=} \frac{O(1)}{n^{b}} \triangleq \zeta_1',\end{aligned} \tag{7.78}$$

where (a) exploits $K = n^{\kappa}$. Therefore, recalling (7.69) and (7.74), and (7.76) we obtain

$$\begin{aligned}P_{e,2}(i,j) &\leq \Pr\left(\mathcal{E}_{i,\tilde{j}}''\right) + \Pr\left(\mathcal{E}_{i,\tilde{j}}'''\right) \\ &\leq \zeta_0 + \zeta_1 \\ &= \zeta_0' + \zeta_1' \\ &\leq e_2,\end{aligned} \tag{7.79}$$

hence, $P_{e,2}(i,j) \leq e_2$ holds for sufficiently large n and arbitrarily small $e_2 > 0$.

We have thus shown that for every $e_1, e_2 > 0$ and sufficiently large n, there exists an $(n, M(n,R), K(n,\kappa), e_1, e_2)$-ISI-Poisson DI code.

Remark 7.3.2. *In the error analysis, we established upper bounds on the type I (cf. (7.60)) and type II error probabilities (cf. (7.77) and (7.78)). These results reveal that the fastest scales for the number of ISI taps $K(n,\kappa)$ which ensure the vanishing of the type I and type II error probabilities as $n \to \infty$, are allowed to be defined as follows: $K(n,\kappa) = 2^{n\kappa} = n^{\kappa}$.*

7.3.3 | Upper Bound (Converse Proof)

Before we start with the converse proof, for the sake of a concise presentation of the analysis, we introduce the following notations. Let:

- $I_t^{\mathbf{x}} \triangleq \lambda + \sum_{l=1}^{K-1} \rho_k x_{t-l}$.
- $d_{i,t} = \rho_0 c_{i,t} + I_t^{\mathbf{c}_i}, \forall t \in [\![n]\!]$.

The converse proof consists of the following two main steps.

- **Step 1:** First, we show in Lemma 7.3.1 that for any achievable DKI rate (for which the type I and type II error probabilities vanish as $n \to \infty$), the distance between any selected entry of one codeword and any entry of another codeword is at least larger than a threshold.

- **Step 2:** Employing Lemma 7.3.1, we then derive an upper bound on the codebook size of DI codes.

We start with the following lemma on the ratio of $d_{i_2,t}/d_{i_1,t}$ for two distinct messages i_1 and i_2, with $i_1, i_2 \in [\![M]\!]$.

Lemma 7.3.1 (Shifted Symbol Distance). *Suppose that $R > 0$ is an achievable DKI rate for the DTPC with ISI, $\mathcal{P}$. Consider a sequence of $(n, M(n,R), K(n,\kappa), e_1^{(n)}, e_2^{(n)})$-ISI-Poisson codes $(\mathcal{C}^{(n)}, \mathscr{T}^{(n)})$, where $K(n,\kappa) = 2^{\kappa \log n}$ with $\kappa \in [0, 1/4)$ such that $e_1^{(n)}$ and $e_2^{(n)}$ tend to zero as $n \to \infty$. Then, given a sufficiently large n, the codebook $\mathcal{C}^{(n)}$ satisfies the following property. For every pair of codewords, $\mathbf{c}_{i_1}$ and $\mathbf{c}_{i_2}$, there exists at least one letter $t \in [\![n]\!]$ such that*

$$\left|1 - \frac{d_{i_2,t}}{d_{i_1,t}}\right| > \theta_n, \tag{7.80}$$

for all $i_1, i_2 \in [\![M]\!]$, such that $i_1 \neq i_2$, with

$$\theta_n \triangleq \frac{P_{max}}{Kn^{1+b}} = \frac{P_{max}}{n^{1+b+\kappa}}, \tag{7.81}$$

where $b > 0$ is an arbitrarily small constant.

Proof. The method of proof is by contradiction, namely, we assume that the condition given in (7.80) is violated and then we show that this leads to a contradiction, namely the sum of the type I and type II error probabilities converges to one, i.e.,

$$\lim_{n\to\infty} \left[P_{e,1}(i_1) + P_{e,2}(i_1, i_2)\right] = 1.$$

Let $e_1, e_2 > 0$ and $\eta_0, \eta_1, \eta_2, \delta > 0$ be arbitrarily small constants. Assume to the contrary that there exist two messages i_1 and i_2, where $i_1 \neq i_2$, meeting the error constraints in (10.22) and (10.23), such that $\forall t \in [\![n]\!]$, we have

$$\left|1 - \frac{d_{i_2,t}}{d_{i_1,t}}\right| \leq \theta_n. \tag{7.82}$$

In order to show contradiction, we bound the sum of the two error probabilities, $P_{e,1}(i_1) + P_{e,2}(i_1, i_2)$, from below. Then, observe that

$$P_{e,1}(i_1) + P_{e,2}(i_1, i_2) = \left[1 - \sum_{\mathbf{y}\in\mathbb{T}_{i_1}} V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right)\right] + \sum_{\mathbf{y}\in\mathbb{T}_{i_1}} V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_2}\right). \tag{7.83}$$

To bound the error, let us define

$$\mathbb{F}_{i_1} = \left\{ \mathbf{y} \in \mathbb{T}_{i_1} \, : \, \bar{n}^{-1} \sum_{t=1}^{\bar{n}} Y_t - I_t^{\mathbf{c}_{i_1}} \leq \rho_0 P_{\max} + \delta \right\}, \tag{7.84}$$

where $\mathbb{T}_{i_1} \subseteq \mathbb{N}_0^{\bar{n}}$ is the decoding set adopted[6] for the set of target messages $\mathbb{K}$.

Now, consider the sum inside the bracket in (7.83),

$$\sum_{\mathbf{y} \in \mathbb{T}_{i_1}} V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) = \sum_{\mathbf{y} \in \mathbb{T}_{i_1} \cap \mathbb{F}_{i_1}} V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) + \sum_{\mathbf{y} \in \mathbb{T}_{i_1} \cap \mathbb{F}_{i_1}^c} V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right), \tag{7.85}$$

where the equality follows from applying the law of total probability on $\mathbb{T}_{i_1}$ with respect to $(\mathbb{F}_{i_1}, \mathbb{F}_{i_1}^c)$.

Now, we proceed to establish an upper bound on the RHS sum in (7.85) as follows

$$\sum_{\mathbf{y} \in \mathbb{T}_{i_1} \cap \mathbb{F}_{i_1}^c} V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) = \Pr\left(\mathbb{T}_{i_1} \cap \mathbb{F}_{i_1}^c\right) \leq \Pr\left(\bar{n}^{-1} \sum_{t=1}^{\bar{n}} Y_t(i_1) - I_t^{\mathbf{c}_{i_1}} > \rho_0 P_{\max} + \delta\right). \tag{7.86}$$

Next, we apply Chebyshev's inequality to the probability term in (7.86) and obtain

$$\begin{aligned}
\sum_{\mathbf{y} \in \mathbb{T}_{i_1} \cap \mathbb{F}_{i_1}^c} V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) &\overset{(a)}{\leq} \Pr\left(\bar{n}^{-1} \sum_{t=1}^{\bar{n}} Y_t(i_1) - \bar{n}^{-1} \sum_{t=1}^{\bar{n}} \mathbb{E}[Y_t(i_1)] > \rho_0 P_{\max} + \delta\right) \\
&\overset{(b)}{\leq} \frac{\mathrm{Var}\left[\bar{n}^{-1} \sum_{t=1}^{\bar{n}} Y_t(i_1)\right]}{\left(\rho_0 P_{\max} + \delta\right)^2} \\
&\overset{(c)}{=} \frac{\bar{n}^{-2} \sum_{t=1}^{\bar{n}} \rho_0 c_{i_1,t} + I_t^{\mathbf{c}_{i_1}}}{\left(\rho_0 P_{\max} + \delta\right)^2} \\
&\overset{(d)}{\leq} \frac{T_R P_{\max} + \lambda + (K-1) T_R P_{\max}}{n\delta^2} \\
&\leq \frac{K T_R P_{\max} + \lambda}{n\delta^2} = \frac{O(K)}{n\delta^2} \\
&\overset{(e)}{=} \frac{O(1)}{n^{1-\kappa}\delta^2} \triangleq \eta_0,
\end{aligned} \tag{7.87}$$

for sufficiently large n, where (a) holds since $\mathbb{E}[Y_t(i_1)] = I_t^{\mathbf{c}_{i_1}}$, for inequality (b), we exploited Chebyshev's inequality, and for equality (c), we used the fact that $\mathrm{Var}[Y_t(i_1)] = \mathbb{E}[Y_t(i_1)] =$

[6] We note that in the achievability proof given in Section 7.3.2 we impose a specific structure on the decoding set $\mathbb{T}_{i_1}$, namely, we defined $\mathbb{T}_{i_1}$ to be the union of the individual decoding set corresponding to messages that belong to set $\mathbb{K}$, i.e., $\mathbb{T}_{i_1} = \bigcup_{i_1 \in \mathbb{K}} \mathbb{T}_{i_1}$. In contrast, in the converse proof, we do not impose any structure on $\mathbb{T}_{i_1}$ and treat the decoding set $\mathbb{T}_{i_1}$ as a general choice $\mathbb{T}_{i_1} \subseteq \mathbb{N}_0^{\bar{n}}$.

$\rho_0 c_{i_1,t} + I_t^{\mathbf{c}_{i_1}}, \forall\, t \in [\![n]\!]$. Inequality (d) employs $c_{i_1,t} \leq P_{\max}, \forall\, i_1 \in [\![M]\!], \forall\, t \in [\![n]\!], \rho_0 \leq T_R, n \leq \bar{n}$ and (e) exploits $K = n^\kappa$. Thereby, recalling (7.85) and (7.87), we obtain

$$\begin{aligned}\sum_{\mathbf{y}\in\mathbb{T}_{i_1}} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) &\leq \sum_{\mathbf{y}\in\mathbb{T}_{i_1}\cap\mathbb{F}_{i_1}} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) + \sum_{\mathbf{y}\in\mathbb{T}_{i_1}\cap\mathbb{F}_{i_1}^c} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) \\ &\leq \sum_{\mathbf{y}\in\mathbb{T}_{i_1}\cap\mathbb{F}_{i_1}} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) + \eta_0. \end{aligned} \tag{7.88}$$

Next, recalling the sum of error probabilities in (7.83), where $i_1 \in \mathbb{K}$ and $i_2 \notin \mathbb{K}$, we obtain

$$\begin{aligned} P_{e,1}(i_1) + P_{e,2}(i_1,i_2) &= \left[1 - \sum_{\mathbf{y}\in\mathbb{T}_{i_1}} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right)\right] + \sum_{\mathbf{y}\in\mathbb{T}_{i_1}} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_2}\right) \\ &\overset{(a)}{\geq} 1 - \eta_0 - \sum_{\mathbb{F}_{i_1}} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) + \sum_{\mathbb{T}_{i_1}} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_2}\right) \\ &\geq 1 - \eta_0 - \sum_{\mathbb{F}_{i_1}} \left[V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) - V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_2}\right)\right], \end{aligned} \tag{7.89}$$

where (a) holds by (7.88) and (b) follows since $\mathbb{F}_{i_1} \subset \mathbb{T}_{i_1}$. Now, let us focus on the summand in the square brackets in (7.89). Employing (7.8), we have

$$\begin{aligned} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) - V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_2}\right) &= V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) \cdot \left[1 - V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_2}\right) / V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right)\right] \\ &= V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) \cdot \left[1 - \prod_{t=1}^{\bar{n}} e^{-(d_{i_2,t} - d_{i_1,t})} \left(\frac{d_{i_2,t}}{d_{i_1,t}}\right)^{Y_t}\right] \\ &= V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) \cdot \left[1 - \prod_{t=1}^{\bar{n}} e^{-\theta_n d_{i_1,t}} \left(1-\theta_n\right)^{Y_t}\right], \end{aligned} \tag{7.90}$$

where for the last inequality, we exploited

$$d_{i_2,t} - d_{i_1,t} \leq \left|d_{i_2,t} - d_{i_1,t}\right| \leq \theta_n d_{i_1,t} \tag{7.91}$$

and

$$1 - \frac{d_{i_2,t}}{d_{i_1,t}} \leq \left|1 - \frac{d_{i_2,t}}{d_{i_1,t}}\right| \leq \theta_n, \tag{7.92}$$

which holds by (7.82). Now, we bound the product term inside the bracket in (7.90) for space $\mathbf{y} \in \mathbb{F}_{i_1}$ as follows:

$$\prod_{t=1}^{\bar{n}} e^{-\theta_n d_{i_1,t}} \left(1-\theta_n\right)^{Y_t} = e^{-\theta_n \sum_{t=1}^{\bar{n}} d_{i_1,t}} \cdot \left(1-\theta_n\right)^{\sum_{t=1}^{\bar{n}} Y_t}$$

$$
\begin{aligned}
&\overset{(a)}{\geq} e^{-\bar{n}\theta_n\left(\rho_0 P_{\max}+\bar{n}^{-1}\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}}\right)}\,(1-\theta_n)^{\bar{n}\left(\rho_0 P_{\max}+\bar{n}^{-1}\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}}+\delta\right)}\\
&= e^{\bar{n}\theta_n\delta}\cdot e^{-\bar{n}\theta_n\left(\rho_0 P_{\max}+\bar{n}^{-1}\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}}+\delta\right)}\,(1-\theta_n)^{\bar{n}\left(\rho_0 P_{\max}+\bar{n}^{-1}\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}}+\delta\right)}\\
&\overset{(b)}{\geq} e^{\bar{n}\theta_n\delta}\cdot e^{-\bar{n}\theta_n\left(\rho_0 P_{\max}+\bar{n}^{-1}\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}}+\delta\right)}\,(1-\bar{n}\theta_n)^{\rho_0 P_{\max}+\bar{n}^{-1}\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}}+\delta}\\
&\overset{(c)}{=} e^{\bar{n}\theta_n\delta}\cdot f(\bar{n}\theta_n) \geq e^{n\theta_n\delta}\cdot f(\bar{n}\theta_n) \overset{(d)}{>} f(\bar{n}\theta_n)\\
&\overset{(e)}{\geq} 1-3\left(\rho_0 P_{\max}+\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}}+\delta\right)\bar{n}\theta_n\\
&\overset{(f)}{\geq} 1-\frac{3\left(T_R P_{\max}+\lambda+(K-1)T_R P_{\max}+\delta\right)P_{\max}}{n^{b+\kappa}}\cdot\frac{\bar{n}}{n}\\
&= 1-\frac{O(K)}{n^{b+\kappa}}\cdot\left(1+\frac{O(K)}{n}\right)\\
&= 1-\frac{O(1)}{n^{b+\kappa}}-\frac{O(K^2)}{n^{1+b+\kappa}}\\
&\overset{(g)}{=} 1-\left(\frac{O(1)}{n^{b+\kappa}}+\frac{O(1)}{n^{1+b-\kappa}}\right)\\
&\overset{(h)}{=} 1-\eta_1,
\end{aligned}
\tag{7.93}
$$

for sufficiently large n. We used the following facts for the above inequalities:

- Inequality (a) follows since

$$
d_{i_1,t} \leq \rho_0 P_{\max} + I_t^{\mathbf{c}_{i_1}},\ \forall\, t \in [\![n]\!], \tag{7.94}
$$

  and

$$
\sum_{t=1}^{\bar{n}} Y_t \leq \bar{n}\left(\rho_0 P_{\max} + \bar{n}^{-1}\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}} + \delta\right), \tag{7.95}
$$

  where the latter inequality follows from $\mathbf{y} \in \mathbb{F}_{i_1}$, cf. (7.84).

- For (b), we used Bernoulli's inequality [174, Ch. 3]:

$$
(1-x)^r \geq 1 - rx\ ,\ \ \forall\, x > -1\ ,\ \forall r > 0. \tag{7.96}
$$

- For (c), we used the following definition:

$$
f(x) = e^{-cx}(1-x)^c, \tag{7.97}
$$

  with $c = \rho_0 P_{\max} + \bar{n}^{-1}\sum_{t=1}^{\bar{n}} I_t^{\mathbf{c}_{i_1}} + \delta$.

□ For (d), we used the fact that

$$e^{n\theta_n\delta} = e^{P_{\max}\delta/n^{b+\kappa}} > 1. \tag{7.98}$$

□ For (e), we used the Taylor expansion

$$f(\bar{n}\theta_n) = 1 - 2c\bar{n}\theta_n + O((\bar{n}\theta_n)^2) \tag{7.99}$$

to obtain the upper bound $f(\bar{n}\theta_n) \geq 1 - 3c\bar{n}\theta_n$ for sufficiently small values of $\bar{n}\theta_n$, i.e.,

$$\begin{aligned}\bar{n}\theta_n &= \frac{P_{\max}}{n^{1+b+\kappa}} \cdot (n+K-1) = \frac{P_{\max}}{n^{b+\kappa}} \cdot \frac{\bar{n}}{n} \\ &= \frac{P_{\max}}{n^{b+\kappa}} \cdot \left(1 + \frac{O(K)}{n}\right) = \frac{P_{\max}}{n^{b+\kappa}} + \frac{O(1)}{n^{b+\kappa}}.\end{aligned} \tag{7.100}$$

□ Inequality (f) exploits (7.81).

□ Equality (g) employs $K = n^{\kappa}$, with $\kappa \in [0, 1/4)$.

□ Finally, (h) follows from

$$\frac{O(1)}{n^{b+\kappa}} + \frac{O(1)}{n^{1+b-\kappa}} \triangleq \eta_1.$$

Thereby, (7.90) can then be written as follows

$$\begin{aligned}V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) - V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_2}\right) &\leq V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) \cdot \left[1 - e^{-\theta_n \sum_{t=1}^{\bar{n}} d_{i_1,t}} \cdot (1-\theta_n)^{\sum_{t=1}^{\bar{n}} Y_t}\right] \\ &\leq \eta_1 \cdot V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right).\end{aligned} \tag{7.101}$$

Next, observe that for every pair of distinct messages (i_1, i_2) we have the following upper bounds on the type I and type II error probabilities

$$\begin{aligned}P_{e,1}(i_1) &= V^{\bar{n}}(\mathbb{T}_{i_1}^c \,|\, x^n = \mathbf{c}_{i_1}) \leq e_1^{(n)}, \\ P_{e,2}(i_1, i_2) &= V^{\bar{n}}(\mathbb{T}_{i_1} \,|\, x^n = \mathbf{c}_{i_2}) \leq e_2^{(n)}.\end{aligned} \tag{7.102}$$

Hence,

$$\begin{aligned}e_1^{(n)} + e_2^{(n)} &\geq P_{e,1}(i_1) + P_{e,2}(i_1, i_2) \\ &\overset{(a)}{\geq} 1 - \eta_0 - \sum_{\mathbb{F}_{i_1}} \left[V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) - V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_2}\right)\right] \\ &\overset{(b)}{\geq} 1 - \eta_0 - \eta_1 \sum_{\mathbb{F}_{i_1}} V^{\bar{n}}\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right)\end{aligned}$$

$$\overset{(c)}{\geq} 1-\eta_0-\eta_1 \overset{(d)}{=} 1-\eta_2, \tag{7.103}$$

where (a) follows from (7.89), and (b) holds by (7.101), (c) exploits $\sum_{\mathbb{F}_{i_1}} V^{\bar{n}}\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) = \Pr(\mathbb{F}_{i_1}) \leq 1$, (d) holds since $\eta_2 \triangleq \eta_0+\eta_1$.

Therefore, $e_1^{(n)}+e_2^{(n)} \geq 1-\eta_0-\eta_2$ which is a contradiction. In other words, Lemma 7.3.1 states that every given sequence of ISI-Poisson DI codes $(\mathcal{C}^{(n)},\mathcal{T}^{(n)})$ with the parameters $(n, M(n,R), K(n,\kappa)=2^{\kappa\log n}, e_1^{(n)}, e_2^{(n)})$ endows the following property: For an arbitrary pair of distinct messages (i_1,i_2) the upper bounds on the type I and type II error probabilities vanish, i.e., $e_1^{(n)}$ and $e_2^{(n)}$ tend to zero as $n\to\infty$. However, we show that if the condition given in (7.80) does not hold, then the sum of the corresponding upper bounds on the type I and type II errors is lower bounded by one, i.e., $e_1^{(n)}$ and $e_2^{(n)}$ do not vanish. This is clearly a contradiction and implies that the inequality given in (7.82) does not hold. This completes the proof of Lemma 7.3.1. □

Next, we use Lemma 7.3.1 to prove the upper bound on the DKI capacity. Observe that since

$$d_{i,t} = \rho_0 c_{i,t} + I_t^{\mathbf{c}_i} > \lambda, \tag{7.104}$$

Lemma 7.3.1 implies

$$\rho_0\left|c_{i_1,t}-c_{i_2,t}\right| = \left|d_{i_1,t}-d_{i_2,t}\right| \overset{(a)}{>} \theta_n d_{i_1,t} \overset{(b)}{>} \lambda\theta_n, \tag{7.105}$$

where (a) follows from (7.80) and (b) holds by (7.104). Now, since $\left\|\mathbf{c}_{i_1}-\mathbf{c}_{i_2}\right\| \geq \left|c_{i_1,t}-c_{i_2,t}\right|$, we deduce that the distance between every pair of codewords satisfies

$$\left\|\mathbf{c}_{i_1}-\mathbf{c}_{i_2}\right\| > \lambda\theta_n/\rho_0. \tag{7.106}$$

Thus, we can define an arrangement of non-overlapping spheres $\mathcal{S}_{\mathbf{c}_i}(n,\lambda\theta_n/2\rho_0)$, i.e., spheres of radius $r_0=\lambda\theta_n/2\rho_0$ that are centered at the codewords $\mathbf{c}_i$. Since all codewords belong to a hyper cube $\mathbb{Q}_{\mathbf{0}}(n,P_{\max})$ with edge length $P_{\max}$, it follows that the number of packed small spheres, i.e., the number of codewords M, is bounded by

$$\begin{aligned} M &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{M}\mathcal{S}_{\mathbf{c}_i}(n,r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n,r_0))} = \frac{\Delta_n(\mathscr{S})\cdot\mathrm{Vol}\left(\mathbb{Q}_{\mathbf{0}}(n,P_{\max})\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n,r_0))} \\ &\leq 2^{-0.599n}\cdot\frac{P_{\max}^n}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n,r_0))}, \end{aligned} \tag{7.107}$$

where the last inequality follows from (7.37). Thereby,

$$\begin{aligned}\log M &\leq \log\left(\frac{P_{\max}^{n}}{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{c}_1}(n,r_0)\right)}\right) - 0.599n\\ &= n\log(P_{\max}) - \log\Big(\mathrm{Vol}\left(\mathcal{S}_{\mathbf{c}_1}(n,r_0)\right)\Big) - 0.599n\\ &\overset{(a)}{=} n\log P_{\max} - n\log r_0 - n\log\sqrt{\pi} + \log\Big(\Gamma\left(n/2+1\right)\Big) \end{aligned} \tag{7.108}$$

where (a) exploits (7.38). Next, we proceed to establish an upper bound on the last term in (7.108). Observe that

$$\begin{aligned}\Gamma\left(n/2+1\right) &\overset{(a)}{=} (n/2)\Gamma\left(n/2\right)\\ &\overset{(b)}{<} \Big(\lfloor n/2\rfloor+1\Big)\Gamma\left(\lfloor n/2\rfloor+1\right)\\ &\overset{(c)}{=} \Big(\lfloor n/2\rfloor+1\Big)!, \end{aligned} \tag{7.109}$$

where (a) holds by the recurrence relation of the Gamma function [172] for real $n/2$, (b) follows since $n/2 < \lfloor n/2\rfloor + 1$ for real $n/2$, and (c) holds since for positive integer $\lfloor n/2\rfloor$, we have $\Gamma\Big(\lfloor n/2\rfloor+1\Big) = (\lfloor n/2\rfloor)!$, cf. [172]. Next, we proceed to simplify the factorial term given in (7.109). To this end, we exploit *Stirling's approximation*, i.e., $\log n! = n\log n - n\log e + o(n)$ [173, p. 52] with the substitution of $n = \lfloor n/2\rfloor + 1$, where $\lfloor n/2\rfloor \in \mathbb{Z}$. Thereby, we obtain

$$\begin{aligned}\log\Big(\Gamma\left(n/2+1\right)\Big) &< \Big(\lfloor n/2\rfloor+1\Big)\log\Big(\lfloor n/2\rfloor+1\Big) - \Big(\lfloor n/2\rfloor+1\Big)\log e + o\Big(\lfloor n/2\rfloor\Big)\\ &\overset{(a)}{\leq} \Big(n/2+1\Big)\log\Big(n/2+1\Big) - \Big(n/2\Big)\log e + o\Big(\lfloor n/2\rfloor\Big), \end{aligned} \tag{7.110}$$

where (a) follows from $\left\lfloor \frac{n}{2}\right\rfloor \leq \frac{n}{2}$ and $\left\lfloor \frac{n}{2}\right\rfloor > \frac{n}{2} - 1$, for integer n. Therefore, merging (7.108)–(7.110), we obtain

$$\begin{aligned}\log M &\leq n\log P_{\max} - n\log r_0 - n\log\sqrt{\pi}\\ &\quad + \Big(n/2+1\Big)\log\Big(n/2+1\Big) - \Big(n/2\Big)\log e + o\Big(\lfloor n/2\rfloor\Big)\\ &= n\log P_{\max} - n\log(\lambda P_{\max}/(2\rho_0)) + (1+b+l+\kappa)\,n\log n\\ &\quad - n\log\sqrt{\pi} + \Big(n/2+1\Big)\log\Big(n/2+1\Big)\\ &\quad - \Big(n/2\Big)\log e + o\Big(\lfloor n/2\rfloor\Big), \end{aligned} \tag{7.111}$$

where for the equality we used

$$r_0 = \frac{\lambda\theta_n}{2\rho_0} = \frac{\lambda P_{\max}}{2\rho_0 n^{1+b+\kappa}}. \tag{7.112}$$

The dominant term in (7.108) is again of order $n\log n$. Hence, to ensure a finite value for the upper bound of the rate, R, (7.108) induces the scaling law of M to be $2^{(n\log n)R}$. By setting $M(n,R)=2^{(n\log n)R}$, we obtain

$$R\leq\frac{1}{n\log n}\left[\left(\frac{1}{2}+(1+b+\kappa)\right)n\log n-n\left(\frac{1}{2}+\log(\lambda\sqrt{\pi e}/(2\rho_0))\right)+o(n)\right], \quad (7.113)$$

which tends to $\frac{3}{2}+\kappa$ as $n\to\infty$ and $b\to 0$. This completes the proof of Theorem 7.3.1.

Example 7.3.1 (logarithmic-ISI DTPC). *Let $\mathcal{P}_{\log n}$ denotes a DTPC with $K=\log n$ degree of memory with n being the codeword length. Find DI capacity bounds for such a channel.*

***Solution**: We set $\log n=n^{\kappa}$ and find κ as follows:*

$$\log\log n=\kappa\log n\Rightarrow\kappa=\frac{\log\log n}{\log n}, \quad (7.114)$$

Now since $n\to\infty$ we derive limit of κ as $n\to\infty$, i.e.,

$$\lim_{n\to\infty}\frac{\log\log n}{\log n}\overset{Hop}{=}\lim_{n\to\infty}\frac{1}{\log n}=0. \quad (7.115)$$

Hence, $\mathcal{P}_{\log n}$ is equivalent to $\mathcal{P}_K$ with $K=n^{\kappa=0}=1$ and share identical capacity bounds, i.e.,

$$\frac{1}{4}\leq\mathsf{C}_{DI}^{K}(\mathcal{P}_{\log n},M)\leq\frac{3}{2}. \quad (7.116)$$

7.4 | Summary

In this chapter, we studied the DI problem over the DTPC with K number of ISI channel taps. We assumed that $K=K(n,\kappa)=2^{\kappa\log n}=n^{\kappa}$ where $\kappa\in[0,1)$ scales sub-linearly with the codeword length n. In practice, the DTPC exhibits memory [9], therefore, our results in this chapter may serve as a model for event-triggered based tasks in the context of many practical MC applications. Especially, we obtained lower and upper bounds on the DI capacity of the DTPC with memory subject to average and peak power constraints with the codebook size of $M(n,R)=2^{(n\log n)R}=n^{nR}$. Our results for the DI capacity of the DTPC with memory revealed that the super-exponential scale of n^{nR} is the appropriate scale for codebook size. This scale coincides the scale that was observed in chapter 4 and 5 for codebook of memoryless DTPC and Gaussian channels; see [99, 105], and stands considerably different from the traditional scales as in transmission

and RI setups where corresponding codebooks size grows exponentially and double exponentially, respectively.

We show the achievability proof using a packing of hyper spheres and a distance decoder. In particular, we pack hyper spheres with radius $\sim n^{\frac{1+4\kappa}{4}}$ where

$$\kappa := \log_n K \in [0, 1/4)\,,$$

is the ISI rate, inside a larger hyper cube. While as observed in chapter 3, the radius of the spheres in a similar proof for Gaussian channels vanishes, as n increases [105], the radius here similar to the case for memoryless DTPC in chapter 6 diverges to infinity. Yet, likewise as in chapter 6 we can obtain a positive rate while packing a super-exponential number of spheres fulfilling the molecule release rates and error constraints.

For the converse proof, we follow a similar approach as in chapter 6 for the memoryless DTPC [51]. In chapter 6, we established a minimum distance between each pair of shifted codewords when the amount of shift was the constant interference signal $\lambda > 0$. Here, we let the value of shift vary according to the related codeword where it is lower bounded by $\lambda > 0$. In general, the derivation here is more involved than the derivation in the Gaussian case [105]. In chapters 4 and 5 on the Gaussian channels with fading [105], the converse proof was based on establishing a minimum distance between each pair of codewords (with no shift). Here, on the other hand, we use the stricter requirement that the ratio of the letters of every two different shifted codewords is different from 1 for at least one index.

CHAPTER 8

DI FOR BINOMIAL CHANNEL

> *Probability is Degree of Certainty and Differs From Absolute Certainty as The Part Differs From The Whole.*
>
> Jacques Bernoulli,

8.1 | Introduction

In the context of MC, information can be encoded in the concentration (rate) of molecules released by transmitter and can be decoded based on the number of molecules reaching the receiver. Assuming that the release, propagation, and reception of different molecules are independent from each other, the MC systems with molecule counting receivers are characterized by the Binomial channel. The transmission capacity of the Binomial channel is studied in [184, 185]. In [186] and [31], the Binomial channel law is used in modelling with imperfect particle intensity modulation and detection, is exploited for an MC channel. In the literature, the Binomial channel is often approximated by the Poisson channel when the number of released molecules, denoted by N, is large [5, Sec. IV], for which bounds on the DI capacity are studied in [51]. However, to the best of the authors' knowledge, the fundamental performance limits of DI for the original Binomial channel, which does not rely on very large N, has not been so far investigated in the literature.

A method is in [185] is developed to compute the capacity of Binomial channel. Transmission capacity of Binomial channel in any finite order n is computed in [184]. In [31,186] the Binomial channel law is used in modelling with imperfect particle intensity modulation and detection is exploited for an MC channel where biological transmitter may attempt to transmit $m_\tau = \lfloor \lambda \tau \rfloor$ particles for communication where λ is a

fixed rate for molecule generation and τ reads the symbol duration[1]. To the best of the authors' knowledge, the fundamental performance limits of DI for the Binomial channels has not been so far studied in the literature.

8.1.1 | Contributions

In this chapter, we consider identification systems employing deterministic encoder and receivers that are interested to accomplish the identification task, namely, finding an object in a set of size M where $M = 2^{(n \log n)R}$. We assume that the communication over n channel uses are independent of each other. Further, we assume that the CIR is available at the decoder. We formulate the problem of DI over the Binomial channel under average and peak power constraint which account for the restricted molecule numbers in the transmitter. As our main objective, we investigate the fundamental performance limits of DI over the Binomial channel. In particular, this chapter makes the following contributions:

♢ **Problem Formulation**: We formulate the problem of DI over the DTBC under average and peak power constraints to account for the limited molecule production / release rates of the transmitter. To the best of the authors' knowledge, the DI capacity of the DTBC has not been studied in the literature, yet.

♢ **Codebook Scale**: We establish that the codebook size of DI problem over the Binomial channels with average and peak power constraints for deterministic encoding scales super-exponentially in the codeword length ($\sim 2^{(n \log n)R}$). This result is in contrast with the scaling of the codebook size for conventional transmission (i.e., 2^{nR} [28]) and RI (i.e., $2^{2^{nR}}$ [23]). The enlarged codebook size of the identification problem compared to the transmission problem may have interesting implications for MC system design. For instance, it may help explain the extremely large identification capability of natural olfactory systems and guide the design of olfactory-inspired synthetic MC systems [81] by, e.g., determining the maximum number of identifiable molecule mixtures.

[1] In our setup, we assume that the rate at which transmitter generate the molecules for different time slots, vary according the deterministic letter x_t, that is, the number of released molecules intended for communication is $m_{T_s} = \lfloor x_t T_s \rfloor$. For the sake of analysis, we assume throughout the paper, that $x_t T_s$ is an integer number.

♢ **Capacity Bounds**: We derive lower and upper bounds on the DI capacity of the DTBC, which are the main results of this chapter. Such bounds does not reflect the impact of power constraints $P_{\text{ave}}, P_{\max}$ in the super-exponential scale, unless requiring them to be positive and finite values.

♢ **Technical Novelty**: To obtain the proposed lower bound, the existence of an appropriate sphere packing within the input space, for which the distance between the centers of the spheres does not fall below a certain value, is guaranteed. While the radius of the small spheres in the Gaussian case [105] tends to zero, here the radius grows in the codeword length, n. Yet, we show that we can pack a super-exponential number of spheres within the larger cube. In particular, we consider the packing of hyper spheres with radius $\sim n^{\frac{1}{4}}$ inside a large hyper cube, whose radius grows in both the codeword length n. For derivation of the upper bound, we assume that for given sequences of codes with vanishing error probabilities, a certain minimum distance between the codewords is asserted, where this distance decreases as n grows. Here, the derivation of the upper bound is less involved compared to that for the Gaussian [105] and Poisson channels [108, 109].

8.1.2 | Organization

The remainder of this chapter is structured as follows. In Section 8.2, system model is explained and the required preliminaries regarding DI codes are established. Section 8.3 provides the main contributions and results on the message K-identification capacity of the slow fading channel. Finally, Section 8.4 of the paper concludes with a summary and directions for future research.

8.2 | System Model and Preliminaries

In this section, we present the adopted system model and establish some preliminaries regarding DI coding.

8.2.1 | System Model

We address an identification-focused communication setup, where the decoder's purpose is accomplishing the following task: Determining whether or not a target message was sent by the transmitter; see Figure 8.1. To attain this objective, a coded communication between the transmitter and the receiver over n channel uses of an MC chan-

nel is established. We assume that for a given channel use, the transmitter generates $N = \lfloor T_s X \rfloor$ molecules where X is the rate for molecule generation and T_s denotes the symbol duration. The generated molecules are released instantaneously into the channel at the beginning of the next symbol interval. Let p denote the probability that a molecule released by the transmitter, is observed at the receiver, whose value depends on the parameters such as the diffusion coefficient of the molecules D, the distance between the transmitter and the receiver d, and the type of reception (e.g. absorbing or transparent receivers); see [5] for further details. For instance, assuming molecule propagation via diffusion in an unbounded three-dimensional environment and under the approximation of uniform concentration within the reception volume of a transparent receiver, p at sampling time τ after the release of molecules is obtained as [5]

$$p = \frac{V_{rx}}{(4\pi D\tau)^{3/2}} e^{-\frac{d^2}{4D\tau}}, \tag{8.1}$$

where V_{rx} is the reception volume size. Assuming that the release, propagation, and reception of molecules are independent from each other, probability of observing Y molecules at the receiver follows a Binomial distribution[2], i.e.,

$$\text{Binom}\left(\lfloor T_s X \rfloor, p\right) = \binom{\lfloor T_s X \rfloor}{Y} p^Y (1-p)^{\lfloor T_s X \rfloor - Y}. \tag{8.2}$$

While in principle, MC channels are dispersive, the contribution of ISI can be made negligible if the symbol intervals are chosen sufficiently large such that the channel impulse response (CIR) fully decays to zero within one symbol interval. Alternatively, enzymes [188] and reactive information molecules, such as acid / base molecules [189], [157], may be used to speed up the decay of the CIR as a function of time, which would increase the accuracy of this assumption. Therefore, in such scenarios, we can assume that the MC channel between transmitter and receiver is characterized by a memoryless Binomial channel $\mathcal{B}$, that is, the n channel uses are independent. Hence, the transition probability law for n channel uses is given by

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{n} \binom{\lfloor T_s x_t \rfloor}{y_t} p^{y_t} (1-p)^{\lfloor T_s x_t \rfloor - y_t}, \tag{8.3}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ denote the transmitted codeword and the received signal, respectively.

[2] The Binomial channel can be approximated by the Poisson channel (for large N and small Np) and the Gaussian channel (for large N and large Np); cf. [5, 9, 187]. However, here, we study the Binomial channel which does not rely on the assumption of large N.

The peak and average release rate constraints on the codewords $\mathbf{x} = (x_t)\Big|_{t=1}^{n}$ are $0 \leq x_t \leq P_{\max}$ and $n^{-1}\sum_{t=1}^{n} x_t \leq P_{\text{avg}}$, respectively, $\forall t \in [\![n]\!]$, where $P_{\max} > 0$ and $P_{\text{avg}} > 0$ constrain the maximum value of molecule release rate per channel use and average molecule release rate over the entire n channel uses in each codeword, respectively.

While, unlike the Poisson and Gaussian approximations, the Binomial model does not require that the number of released molecules to be *very* large [5], in practice, the number of released molecules cannot be too small either for reliable communication. This observation motivates us to adopt the approximation $\lfloor T_s x_t \rfloor \approx T_s x_t$ in the remaining part, since the relative error

$$\frac{T_s x_t - \lfloor T_s x_t \rfloor}{T_s x_t} \leq \frac{1}{T_s x_t}$$

become sufficiently small for *reasonably* large $T_s x_t$.

8.2.2 | DI Coding For The Binomial Channel

The definition of a DI code for the Binomial channel $\mathcal{B}$ is given below.

Definition 8.2.1 (Binomial DI Code). *An $(n, M(n,R), K(n,\kappa), e_1, e_2)$ DI code for a Binomial channel $\mathcal{B}$ under average and peak power constraints of P_{ave} and P_{max}, and for integers $M(n,R)$ and $K(n,\kappa)$, where n and R are the codeword length and coding rate, respectively, is defined as a system $(\mathcal{C}, \mathscr{T})$, which consists of a codebook $\mathcal{C} = \{\mathbf{c}_i\}_{i\in[\![M]\!]} \subset \mathbb{R}_+^n$, such that*

$$0 \leq c_{i,t} \leq P_{max} \quad \text{and} \quad \frac{1}{n}\sum_{t=1}^{n} c_{i,t} \leq P_{avg}\,, \tag{8.4}$$

$\forall i \in [\![M]\!]$, $\forall t \in [\![n]\!]$ and a collection of decoding regions $\mathscr{T} = \{\mathbb{T}_i\}_{i\in[\![M]\!]}$ where

$$\bigcup_{i=1}^{M(n,R)} \mathbb{T}_i \subset \mathbb{N}_0^n\,. \tag{8.5}$$

Given a message $i \in [\![M]\!]$, the encoder sends $\mathbf{c}_i$, and the decoder's task is to address a binary hypothesis: Was a target message $j \in \mathbb{K}$ sent or not? There exist two types of errors that may happen (see Figure 8.2):

- *Type I: Rejection of the actual message; $i \in [\![M]\!]$.*
- *Type II: Acceptance of a wrong message; $j \neq i$.*

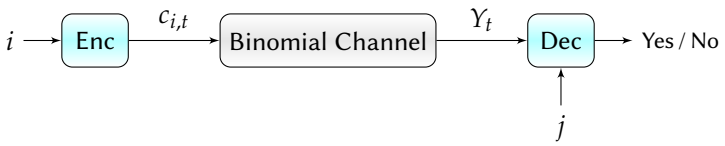


Figure 8.1: End-to-end transmission chain for DI communication in a generic molecular communication system modelled as a Binomial channel. The transmitter maps message i onto a codeword $\mathbf{c}_i = (c_{i,t})|_{t=1}^{n}$. The receiver is provided with an arbitrary message j, and given the channel output vector $\mathbf{Y} = (Y_t)|_{t=1}^{n}$, it asks whether j is identical to i or not.

The associated error probabilities of the DI code $(\mathcal{C}, \mathscr{T})$ *reads*

$$P_{e,1}(i) = 1 - \sum_{\mathbf{y} \in \mathbb{T}_i} W^n(\mathbf{y} \,|\, \mathbf{c}_i) \quad \textit{(miss-identification error)}\,, \tag{8.6}$$

$$P_{e,2}(i,j) = \sum_{\mathbf{y} \in \mathbb{T}_j} W^n(\mathbf{y} \,|\, \mathbf{c}_i) \quad \textit{(false identification error)}\,. \tag{8.7}$$

and satisfy the following bounds $P_{e,1}(i) \leq e_1\,, \forall i \in [\![M]\!]$ *and* $P_{e,2}(i,j) \leq e_2\,, \forall i \neq j$*, and every* $e_1, e_2 > 0$.

A rate $R > 0$ *is called achievable if for every* $e_1, e_2 > 0$ *and sufficiently large* n*, there exists an* $(n, M(n,R), K(n,\kappa), e_1, e_2)$ *DI code. The operational DI capacity of the Binomial channel* $\mathcal{B}$ *is defined as the supremum of all achievable rates, and is denoted by* $\mathbb{C}_{DI}(\mathcal{B}, M)$.

8.3 | DI Capacity of The Binomial Channel

In this section, we first present our main results, i.e., lower and upper bounds on the achievable identification rates for the Binomial channel. Subsequently, we provide the detailed proofs of these bounds.

8.3.1 | Main Results

The DI capacity theorem for the Binomial channel $\mathcal{B}$ is stated below.

Theorem 8.3.1. *Consider the Binomial channel* $\mathcal{B}$ *subject to average and peak power constraints of the form* $n^{-1}\sum_{t=1}^{n} c_{i,t} \leq P_{ave}$ *and* $0 \leq c_{i,t} \leq P_{max}$*, respectively. Then, the DI capacity in the super-exponential scale, i.e.,* $M(n,R) = 2^{(n\log n)R}$*, is bounded by*

$$\frac{1}{4} \leq \mathbb{C}_{DI}(\mathcal{B}, M) \leq \frac{3}{2}\,. \tag{8.8}$$

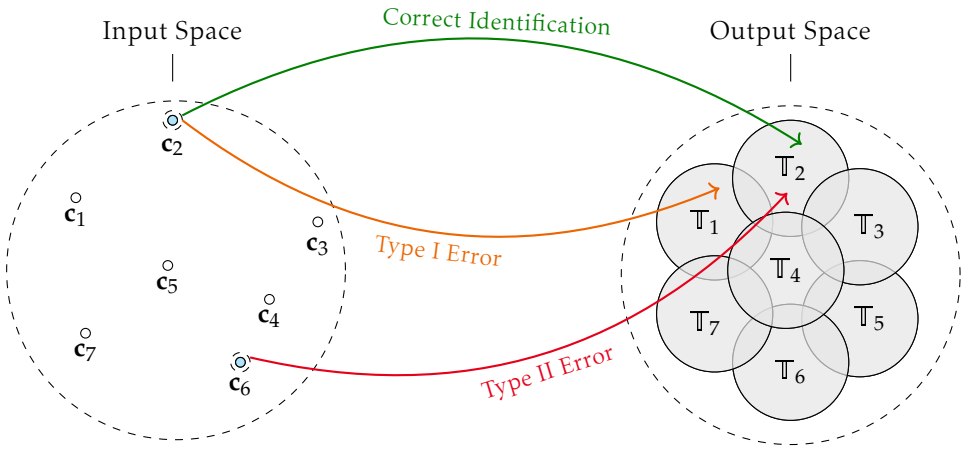


Figure 8.2: Illustration of a deterministic identification setting. Assuming that the decision maker is the decoder $\mathbb{T}_2$, in the correct identification scenario, channel output is observed in the decoder $\mathbb{T}_2$ whose index coincide the sent message. Type I (miss-identification) error occurs if the channel output is detected in the complement of the decoder whose index is identical to the sent message and type II error (false identification) happens where the index of decoder for which channel output belongs to, differs from the sent message.

Proof. The proof of Theorem 8.3.1 consists of two parts, namely the achievability and the converse proofs, which are provided in Sections 8.3.2 and 8.3.3, respectively. □

8.3.2 | Lower Bound (Achievability Proof)

The achievability proof consists of the following two main steps.

- **Step 1:** First, we propose a codebook construction and derive an analytical lower bound on the corresponding codebook size using inequalities for sphere packing density.

- **Step 2:** Then, to prove that this codebook leads to an achievable rate, we propose a decoder and show that the corresponding type I and type II error rates vanished as $n \to \infty$.

8.3.2.1 | Codebook Construction

Let $A = \min(P_{\max}, P_{\text{ave}})$. In the following, we confine ourselves to codewords that meet the condition $0 \leq c_t \leq A, \forall\, t \in [\![n]\!]$. We argue that this condition ensures both the average and the peak power constraints in (8.4). In particular,

- **Case 1:** $P_{\max} \leq P_{\text{ave}}$, then $A = P_{\max}$ and the constraint $0 \leq c_t \leq A \, \forall \, t \in [\![n]\!]$ yields $\frac{1}{n}\sum_{t=1}^{n} c_t \leq A = P_{\max}^2 \leq P_{\text{ave}}$, that is the average power constraint $\frac{1}{n}\sum_{t=1}^{n} c_t \leq P_{\text{ave}}$ is met. Furthermore, condition $0 \leq c_t \leq A \, \forall \, t \in [\![n]\!]$ implies $0 \leq c_t \leq P_{\max} \, \forall \, t \in [\![n]\!]$, i.e., the peak power constraint is attained.

- **Case 2:** $P_{\max} > P_{\text{ave}}$, then $A = P_{\text{ave}}$. Now by $0 \leq c_t \leq A \, \forall \, t \in [\![n]\!]$, we obtain $\frac{1}{n}\sum_{t=1}^{n} c_t \leq A = P_{\text{ave}}$, that is, the average power constraint is fulfilled. Furthermore, the condition $0 \leq c_t \leq A \, \forall \, t \in [\![n]\!]$ implies $0 \leq c_t \leq P_{\text{ave}} \leq P_{\max} \, \forall \, t \in [\![n]\!]$, that is, the peak power constraint is accomplished.

Hence, in the following, we restrict our considerations to a hyper cube with edge length A. We use a packing arrangement of non-overlapping hyper spheres of radius $r_0 = \sqrt{n\epsilon_n}$ in a hyper cube with edge length A, where

$$\epsilon_n = \frac{a}{n^{\frac{1}{2}(1-b)}} \, , \tag{8.9}$$

and $a > 0$ is a non-vanishing fixed constant and $0 < b < 1$ is an arbitrarily small constant[3].

Let $\mathscr{S}$ denote a sphere packing, i.e., an arrangement of L non-overlapping spheres $\mathcal{S}_{\mathbf{c}_i}(n, r_0)$, $i \in [\![L]\!]$, that are packed inside the larger cube $\mathcal{Q}_{\mathbf{0}}(n, A)$ with an edge length A, see Figure 8.3. As opposed to standard sphere packing coding techniques [138], the spheres are not necessarily entirely contained within the cube. That is, we only require that the centers of the spheres are inside $\mathcal{Q}_{\mathbf{0}}(n, A)$ and are disjoint from each other and have a non-empty intersection with $\mathcal{Q}_{\mathbf{0}}(n, A)$. The packing density $\Delta_n(\mathscr{S})$ is defined as the ratio of the saturated packing volume to the cube volume $\text{Vol}\left(\mathcal{Q}_{\mathbf{0}}(n, A)\right)$, i.e.,

$$\Delta_n(\mathscr{S}) \triangleq \frac{\text{Vol}\left(\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{c}_i}(n, r_0)\right)}{\text{Vol}\left(\mathcal{Q}_{\mathbf{0}}(n, A)\right)} \, . \tag{8.10}$$

Sphere packing $\mathscr{S}$ is called *saturated* if no spheres can be added to the arrangement without overlap. In particular, we use a packing argument that has a similar flavor as that observed in the Minkowski–Hlawka theorem for saturated packing [138].

Specifically, consider a saturated packing arrangement of

$$\bigcup_{i=1}^{M(n,R)} \mathcal{S}_{\mathbf{c}_i}(n, \sqrt{n\epsilon_n}) \tag{8.11}$$

[3] We note that our achievability proof is valid for any $b \in (0, 1)$; however, arbitrarily small values of b leads to the tightest lower bound and hence are of interest here.

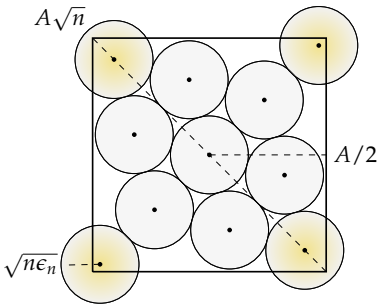


Figure 8.3: Illustration of a saturated sphere packing inside a cube, where small spheres of radius $r_0 = \sqrt{n\epsilon_n}$ cover a larger cube. Yellow colored spheres are not entirely contained within the larger cube, and yet they contribute to the packing arrangement. As we assign a codeword to each sphere center, the 1-norm and arithmetic mean of a codeword are bounded by A as required.

spheres with radius $r_0 = \sqrt{n\epsilon_n}$ embedded within cube $\mathcal{Q}_{\mathbf{0}}(n, A)$. Then, for such an arrangement, we have the following lower [118, Lem. 2.1] and upper bounds [138, Eq. 45] on the packing density

$$2^{-n} \leq \Delta_n(\mathscr{S}) \leq 2^{-0.599n} \,. \tag{8.12}$$

In our subsequent analysis, we use the above lower bound which can be proved as follows: For the saturated packing arrangement given in (8.11), there cannot be a point in the larger cube $\mathcal{Q}_{\mathbf{0}}(n, A)$ with a distance of more than $2r_0$ from all sphere centers. Otherwise, a new sphere could be added which contradicts the assumption that the union of $M(n, R)$ spheres with radius $\sqrt{n\epsilon_n}$ is saturated. Now, if we double the radius of each sphere, the spheres with radius $2r_0$ cover thoroughly the entire volume of $\mathcal{Q}_{\mathbf{0}}(n, A)$, that is, each point inside the hyper cube $\mathcal{Q}_{\mathbf{0}}(n, A)$ belongs to at least one of the small spheres. In general, the volume of a hyper sphere of radius r is given by [138, Eq. (16)]

$$\text{Vol}\left(\mathcal{S}_{\mathbf{x}}(n, r)\right) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r^n \,. \tag{8.13}$$

Hence, if the radius of the small spheres is doubled, the volume of

$$\bigcup_{i=1}^{M(n,R)} \mathcal{S}_{\mathbf{c}_i}(n, \sqrt{n\epsilon_n}) \,, \tag{8.14}$$

is increased by 2^n. Since the spheres with radius $2r_0$ cover $\mathcal{Q}_{\mathbf{0}}(n, A)$, it follows that the original r_0-radius packing has a density of at least 2^{-n} [4]. We assign a codeword

[4] We note that the proposed proof of the lower bound in (8.12) is non-constructive in the sense that,

to the center $\mathbf{c}_i$ of each small sphere. The codewords satisfy the input constraint as $0 \leq c_{i,t} \leq A$, $\forall t \in [\![n]\!]$, $\forall i \in [\![L]\!]$, which is equivalent to

$$\|\mathbf{c}_i\|_\infty \leq A \,. \tag{8.15}$$

Since the volume of each sphere is equal to $\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))$ and the centers of all spheres lie inside the cube, the total number of spheres is bounded from below by

$$\begin{aligned} M &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{c}_i}(n, r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))} \\ &= \frac{\Delta_n(\mathscr{S}) \cdot \mathrm{Vol}\left(\mathcal{Q}_{\mathbf{0}}(n, A)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))} \\ &\geq 2^{-n} \cdot \frac{A^n}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))} \,, \end{aligned} \tag{8.16}$$

where the first inequality holds by (8.10) and the second inequality holds by (8.12).

The above bound can be further simplified as follows

$$\begin{aligned} \log M &\geq \log\left(\frac{A^n}{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{c}_1}(n, r_0)\right)}\right) - n \\ &\overset{(a)}{\geq} n \log\left(\frac{A}{\sqrt{\pi} r_0}\right) + \log\left(\left\lfloor \frac{n}{2} \right\rfloor !\right) - n \\ &\overset{(b)}{=} n \log A - n \log r_0 + \left\lfloor \frac{n}{2} \right\rfloor \log\left(\left\lfloor \frac{n}{2} \right\rfloor\right) - \left\lfloor \frac{n}{2} \right\rfloor \log e + o\left(\left\lfloor \frac{n}{2} \right\rfloor\right) - n \,, \end{aligned} \tag{8.17}$$

where (a) exploits (8.13) and

$$\begin{aligned} \Gamma\left(\frac{n}{2} + 1\right) &\overset{(a)}{=} \frac{n}{2} \Gamma\left(\frac{n}{2}\right) \\ &\overset{(b)}{\geq} \left\lfloor \frac{n}{2} \right\rfloor \Gamma\left(\left\lfloor \frac{n}{2} \right\rfloor\right) \\ &\triangleq \left\lfloor \frac{n}{2} \right\rfloor ! \,, \end{aligned} \tag{8.18}$$

where (a) holds by the recurrence relation of the Gamma function for a positive real argument and (b) follows from $\left\lfloor \frac{n}{2} \right\rfloor \leq \frac{n}{2}$ for positive integer n and the monotonicity

while the existence of the respective saturated packing is proved, no systematic construction method is provided.

of the Gamma function for $n \geq 4 \equiv \left\lfloor \frac{n}{2} \right\rfloor \in [z_1, \infty)$ where $z_1 \approx 1.46$ is the first root of the Digamma function; and (b) follows from Stirling's approximation, that is, $\log n! = n \log n - n \log e + o(n)$ for integer n, [173, P. 52]. Now, for $r_0 = \sqrt{n\epsilon_n} = \sqrt{a} n^{\frac{1+b}{4}}$, we obtain

$$
\begin{aligned}
&\log M \\
&\geq n \log \frac{A}{\sqrt{a}} - \frac{1}{4}(1+b)\, n \log n + \left\lfloor \frac{n}{2} \right\rfloor \log \left(\left\lfloor \frac{n}{2} \right\rfloor \right) - \left\lfloor \frac{n}{2} \right\rfloor \log e + o\left(\left\lfloor \frac{n}{2} \right\rfloor \right) - n \\
&\overset{(a)}{>} n \log \frac{A}{\sqrt{a}} - \frac{1}{4}(1+b)\, n \log n + \left(\frac{n}{2} - 1 \right) \log \left(\frac{n}{2} - 1 \right) - \left\lfloor \frac{n}{2} \right\rfloor \log e + o\left(\frac{n}{2} - 1 \right) - n \\
&\overset{(b)}{\geq} n \log \frac{A}{\sqrt{a}} - \frac{1}{4}(1+b)\, n \log n + \left(\frac{n}{2} - 1 \right) \left(\log \left(\frac{n}{2} \right) - 1 \right) - \frac{n}{2} \log e + o\left(\frac{n}{2} - 1 \right) - n \\
&\overset{(c)}{=} n \log \frac{A}{\sqrt{a}} - \frac{1}{4}(1+b)\, n \log n + \frac{1}{2} n \log n - 2n - \log n - \frac{n}{2} \log e + o\left(\frac{n}{2} - 1 \right) \\
&= \left(\frac{1-b}{4} \right) n \log n + n \left(\log \frac{A}{e\sqrt{a}} \right) - 2n - \log n - \frac{n}{2} \log e + o\left(\frac{n}{2} - 1 \right), \qquad (8.19)
\end{aligned}
$$

where (a) holds by $\left\lfloor \frac{n}{2} \right\rfloor > \frac{n}{2} - 1$ for integer n, (b) holds since $\log(t-1) \geq \log t - 1$ for $t \geq 2$ and $\left\lfloor \frac{n}{2} \right\rfloor \leq \frac{n}{2}$ for integer n, and (c) follows since base of logarithm is 2. Observe that the dominant term (8.19) is of order $n \log n$. Hence, for obtaining a finite value for the lower bound of the rate, R, (8.19) induces the scaling law of M to be $2^{(n \log n)R}$. Therefore, we obtain

$$
R \geq \frac{1}{n \log n} \left[\left(\frac{1-b}{4} \right) n \log n + n \log \left(\frac{A}{e\sqrt{a}} \right) - 2n - \log n - \frac{n}{2} \log e + o\left(\frac{n}{2} - 1 \right) \right], \qquad (8.20)
$$

which tends to $\frac{1}{4}$ when $n \to \infty$ and $b \to 0$.

8.3.2.2 | Encoding

Given a message $i \in [\![M]\!]$, transmit $\mathbf{x} = \mathbf{c}_i$.

8.3.2.3 | Decoding

Fix $e_1, e_2 > 0$ and let $\zeta_0, \zeta_1 > 0$ be arbitrarily small constants. Before we proceed, for the sake of brevity of analysis, we introduce the following conventions. Let:

- $Y_t(i) \sim \text{Binom}(c_{i,t}T_s,\text{p})$ denote the channel output at time t *conditioned* that $\mathbf{x} = \mathbf{c}_i$ was sent.
- $\mathbf{Y}(i) = (Y_1(i), \ldots, Y_n(i))$

Let

$$\delta_n = \frac{A}{n^{\frac{1}{2}(1-b)}}\,, \tag{8.21}$$

where $0 < b < 1$ is an arbitrarily small constant. To identify whether a message $j \in [\![M]\!]$ was sent, given the sequence $\mathbf{g}$, the decoder checks whether the channel output $\mathbf{y}$ belongs to the following decoding set,

$$\mathbb{T}_j = \left\{ \mathbf{y} \in \mathbb{N}_0^n \,:\, \left| T(\mathbf{y}, \mathbf{c}_j) \right| \leq \delta_n \right\}. \tag{8.22}$$

where

$$T(\mathbf{y}, \mathbf{c}_j) = \frac{1}{n} \sum_{t=1}^{n} \left(y_t - pT_s c_{j,t} \right)^2 - (1-p) y_t\,, \tag{8.23}$$

is referred to as the *decoding metric* evaluated for observation vector $\mathbf{y}$ and codeword $\mathbf{c}_j$.

8.3.2.4 | Error Analysis

Consider the type I error, i.e., when the transmitter sends $\mathbf{c}_i$, yet $\mathbf{Y} \notin \mathbb{T}_i$. For every $i \in [\![M]\!]$, the type I error probability is bounded by

$$P_{e,1}(i) = \Pr\left(\left|T(\mathbf{Y}(i), \mathbf{c}_j)\right| > \delta_n\right), \tag{8.24}$$

In order to bound $P_{e,1}(i)$, we apply the Chebyshev's inequality, namely

$$\Pr\left(\left|T(\mathbf{Y}(i), \mathbf{c}_j) - \mathbb{E}\left[T(\mathbf{Y}(i), \mathbf{c}_j)\right]\right| > \delta_n\right) \leq \frac{\mathrm{Var}\left[T(\mathbf{Y}(i), \mathbf{c}_j)\right]}{\delta_n^2}. \tag{8.25}$$

First, we calculate the expectation of the decoding metric as follows

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{t=1}^{n}\left(Y_t(i) - pT_s c_{i,t}\right)^2 - (1-p)Y_t(i)\right] &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(Y_t(i) - pT_s c_{i,t}\right)^2 - (1-p)Y_t(i)\right] \\
&= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(Y_t(i) - pT_s c_{i,t}\right)^2\right] - \mathbb{E}\left[(1-p)Y_t(i)\right] \\
&= \frac{1}{n}\sum_{t=1}^{n}\mathrm{Var}\left[Y_t(i)\right] - \mathbb{E}\left[Y_t(i)\right](1-p) \\
&= \frac{1}{n}\sum_{t=1}^{n} pT_s c_{i,t}(1-p) - pT_s c_{i,t}(1-p) \\
&= 0.
\end{aligned} \tag{8.26}$$

Second, since the channel is memoryless, we can derive the variance of decoding metric as follows

$$\mathrm{Var}\left[\frac{1}{n}\sum_{t=1}^{n}\left(Y_t(i) - pT_s c_{i,t}\right)^2 - (1-p)Y_t(i)\right] = \frac{1}{n^2}\mathrm{Var}\left[\sum_{t=1}^{n}\left(Y_t(i) - pT_s c_{i,t}\right)^2 - (1-p)Y_t(i)\right]$$

$$\stackrel{(a)}{=} \frac{1}{n^2}\sum_{t=1}^{n}\mathrm{Var}\left[\left(Y_t(i)-pT_sc_{i,t}\right)^2-(1-p)Y_t(i)\right]$$
$$=\frac{1}{n^2}\sum_{t=1}^{n}\mathrm{Var}\left[Y_t^2(i)-\left(2pT_sc_{i,t}+1-p\right)Y_t(i)\right], \tag{8.27}$$

where (a) holds since the channel is memoryless. Now we apply the identity $\mathrm{Var}\left[aX-bY\right] = a^2\mathrm{Var}[X]+b^2\mathrm{Var}[Y]-2ab\mathrm{Cov}[X,Y]$ to summand in (8.27), namely,

$$\mathrm{Var}\left[Y_t^2(i)-\left(2pT_sc_{i,t}+1-p\right)Y_t(i)\right]$$
$$=\mathrm{Var}\left[Y_t^2(i)\right]+\left(2pT_sc_{i,t}+1-p\right)^2\mathrm{Var}\left[Y_t(i)\right]-\left(2pT_sc_{i,t}+1-p\right)\mathrm{Cov}\left[Y_t^2(i),Y_t(i)\right]. \tag{8.28}$$

Now, in order to bound (8.27), we use the following three knowledge. First, observe that for a Binomial variable $Y_t(i)\sim\mathrm{Binom}(c_{i,t}T_s,\mathrm{p})$, we have $\mathrm{Var}[Y_t^2(i)]\le\mathbb{E}[Y_t(i)^4]$. Second, the identity $\mathrm{Cov}[X,Y]\le\sqrt{\mathrm{Var}[X]\cdot\mathrm{Var}[Y]}$, provide upper bound for covariance of variables with finite variances. Third, triangle inequality provide upper bound for summation of terms with different sign, i.e., $a-b\le|a-b|\le|a|+|b|$. Therefore,

$$\frac{1}{n^2}\sum_{t=1}^{n}\mathrm{Var}\left[Y_t^2(i)-\left(2pT_sc_{i,t}+1-p\right)Y_t(i)\right]$$
$$\le\frac{1}{n^2}\sum_{t=1}^{n}\mathbb{E}\left[Y_t^4(i)\right]+\left(2pT_sc_{i,t}+1-p\right)^2pT_sc_{i,t}+\left(2pT_sc_{i,t}+1-p\right)\sqrt{\mathbb{E}\left[Y_t^4(i)\right]\cdot pT_sc_{i,t}(1-p)}$$
$$\stackrel{(a)}{=}\frac{n}{n^2}\left[(c_{i,t}T_s)^4\exp\left(8/pT_sc_{i,t}\right)+(2AT_s+1)^2AT_s+(2AT_s+1)\sqrt{(pT_sc_{i,t})^4\exp\left(8/pT_sc_{i,t}\right)AT_s}\right]$$
$$\le\frac{1}{n}\left[A^4T_s^4\exp\left(8/pT_sc_{i,t}\right)+(2AT_s+1)^2AT_s+(2AT_s+1)A^2T_s^2\sqrt{\exp\left(8/pT_sc_{i,t}\right)AT_s}\right]. \tag{8.29}$$

where (a) holds since since $0<p<1$; and for a Binomial variable $Y_t(i)\sim\mathrm{Binom}(c_{i,t}T_s,\mathrm{p})$, the non-centered moments are upper bounded as follows

$$\mathbb{E}\left[Y_t^k(i)\right]\le\mathbb{E}^k\left[Y_t(i)\right]\cdot\exp\left(k^2/2\mathbb{E}\left[Y_t(i)\right]\right). \tag{8.30}$$

Therefore, exploiting (8.25), (8.26) and (8.29), we can establish the following upper bound on the type I error probability given in (8.24), namely

$$\begin{aligned} P_{e,1}(i) &= \Pr\left(\left|T(\mathbf{Y}(i),\mathbf{c}_j)\right| > \delta_n\right) \\ &= \frac{A^4T_s^4\exp\left(8/pT_sc_{i,t}\right) + \left(2AT_s+1\right)^2 AT_s + \left(2AT_s+1\right)A^2T_s^2\sqrt{\exp\left(8/pT_sc_{i,t}\right)AT_s}}{n\delta_n^2} \\ &\overset{(a)}{=} \frac{A^4T_s^4\exp\left(8/pT_sc_{i,t}\right) + \left(2AT_s+1\right)^2 AT_s + \left(2AT_s+1\right)A^2T_s^2\sqrt{\exp\left(8/pT_sc_{i,t}\right)AT_s}}{n^b} \\ &\leq e_1\,, \end{aligned} \tag{8.31}$$

where (a) follows from (8.21).

Next, we address the type II error, i.e., when $\mathbf{Y} \in \mathbb{T}_j$ while the transmitter sent $\mathbf{c}_i$. Then, for every $i, j \in [\![M]\!]$, where $i \neq j$, the type II error probability is given by

$$P_{e,2}(i,j) = \Pr\left(\left|T(\mathbf{Y}(i);\mathbf{c}_j)\right| \leq \delta_n\right)\,. \tag{8.32}$$

where

$$T(\mathbf{Y}(i);\mathbf{c}_j) = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t(i) - pT_sc_{j,t}\right)^2 - (1-p)Y_t(i)\,. \tag{8.33}$$

Observe that (8.33) can be expressed as follows

$$T(\mathbf{Y}(i);\mathbf{c}_j) = \frac{1}{n}\sum_{t=1}^{n}\left(Y_t(i) - pT_sc_{i,t} + (c_{i,t}-c_{j,t})pT_s\right)^2 - (1-p)Y_t(i)\,. \tag{8.34}$$

Observe that the sum in (8.34) can be expressed as

$$\sum_{t=1}^{n}\left(Y_t(i) - pT_sc_{i,t} + (c_{i,t}-c_{j,t})pT_s\right)^2 = \sum_{t=1}^{n}\left(Y_t(i) - pT_sc_{i,t}\right)^2 + \sum_{t=1}^{n}\left((c_{i,t}-c_{j,t})pT_s\right)^2$$

$$+2\sum_{t=1}^{n}\left(Y_t(i)-pT_sc_{i,t}\right)\left((c_{i,t}-c_{j,t})pT_s\right). \tag{8.35}$$

Then, define the following events

$$\mathcal{E}_0=\left\{\mathbf{Y}\in\mathbb{N}_0^n\,:\,\left|\sum_{t=1}^{n}\left(Y_t(i)-pT_sc_{i,t}\right)\left((c_{i,t}-c_{j,t})pT_s\right)\right|>n\delta_n\right\}, \tag{8.36}$$

$$\mathcal{E}_1=\left\{\mathbf{Y}\in\mathbb{N}_0^n\,:\,\sum_{t=1}^{n}\left(Y_t(i)-pT_sc_{i,t}\right)^2+\sum_{t=1}^{n}\left((c_{i,t}-c_{j,t})pT_s\right)^2-(1-p)Y_t(i)\le 2n\delta_n\right\}, \tag{8.37}$$

$$\mathcal{E}_{i,j}=\left\{\mathbf{Y}\in\mathbb{N}_0^n\,:\,\left|\sum_{t=1}^{n}\left(Y_t(i)-pT_sc_{i,t}+(c_{i,t}-c_{j,t})pT_s\right)^2-(1-p)Y_t(i)\le n\delta_n\right|\right\} \tag{8.38}$$

$$\mathcal{E}'_{i,j}=\left\{\mathbf{Y}\in\mathbb{N}_0^n\,:\,\sum_{t=1}^{n}\left(Y_t(i)-pT_sc_{i,t}+(c_{i,t}-c_{j,t})pT_s\right)^2-(1-p)Y_t(i)\le n\delta_n\right\}, \tag{8.39}$$

Then,

$$\begin{aligned}
P_{e,2}(i,j)&=\Pr\left(\mathcal{E}_{i,j}\right)\\
&=\Pr\left(\left|\sum_{t=1}^{n}\left(Y_t(i)-pT_sc_{i,t}+(c_{i,t}-c_{j,t})pT_s\right)^2-(1-p)Y_t(i)\right|\le n\delta_n\right)\\
&\overset{(a)}{\le}\Pr\left(\left|\sum_{t=1}^{n}\left(Y_t(i)-pT_sc_{i,t}+(c_{i,t}-c_{j,t})pT_s\right)^2\right|-\left|\sum_{t=1}^{n}(1-p)Y_t(i)\right|\le n\delta_n\right)\\
&\overset{(b)}{\le}\Pr\left(\sum_{t=1}^{n}\left(Y_t(i)-pT_sc_{i,t}+(c_{i,t}-c_{j,t})pT_s\right)^2-\sum_{t=1}^{n}(1-p)Y_t(i)\le n\delta_n\right)\\
&=\Pr\left(\mathcal{E}'_{i,j}\right),
\end{aligned} \tag{8.40}$$

where (a) exploits the reverse triangle inequality, $|\alpha| - |\beta| \leq |\alpha - \beta|$, and (b) holds since $\alpha, \beta \geq 0$.

Now, we apply the law of total probability to event $\mathcal{E}'_{i,j}$ over $\mathcal{E}_0$ and its complement $\mathcal{E}_0^c$, and obtain the following upper bound on the type II error probability,

$$\begin{aligned} P_{e,2}(i,j) &\leq \Pr\left(\mathcal{E}'_{i,j}\right) \\ &= \Pr\left(\mathcal{E}'_{i,j} \cap \mathcal{E}_0\right) + \Pr\left(\mathcal{E}'_{i,j} \cap \mathcal{E}_0^c\right) \\ &\overset{(a)}{\leq} \Pr\left(\mathcal{E}_0\right) + \Pr\left(\mathcal{E}'_{i,j} \cap \mathcal{E}_0^c\right) \\ &\overset{(b)}{=} \Pr\left(\mathcal{E}_0\right) + \Pr\left(\mathcal{E}_1\right), \end{aligned} \tag{8.41}$$

where (a) follows from $\mathcal{E}'_{i,j} \cap \mathcal{E}_0 \subset \mathcal{E}_0$ and (b) holds since the event $\mathcal{E}'_{i,j} \cap \mathcal{E}_0^c$ yields the event $\mathcal{E}_1$, with the following argument. Observe that,

$$\begin{aligned} \Pr\left(\mathcal{E}'_{i,j} \cap \mathcal{E}_0^c\right) &= \Pr\left(\sum_{t=1}^{n} \left(Y_t(i) - pT_s c_{i,t}\right)^2 + \sum_{t=1}^{n} \left((c_{i,t} - c_{j,t}) pT_s\right)^2 - (1-p)Y_t(i) \leq n\delta_n - (-n\delta_n)\right) \\ &\overset{(a)}{=} \Pr\left(\sum_{t=1}^{n} \left(Y_t(i) - pT_s c_{i,t}\right)^2 + \sum_{t=1}^{n} \left((c_{i,t} - c_{j,t}) pT_s\right)^2 - (1-p)Y_t(i) \leq 2n\delta_n\right) \\ &\overset{(b)}{=} \Pr\left(\mathcal{E}_1\right), \end{aligned} \tag{8.42}$$

where (a) holds since given the complementary event $\mathcal{E}_0^c$, we obtain

$$\sum_{t=1}^{n} \left(Y_t(i) - pT_s c_{i,t}\right) \left((c_{i,t} - c_{j,t}) pT_s\right) \geq -n\delta_n \,, \tag{8.43}$$

and (b) follows from (8.37).

Now, we proceed to bound $\Pr(\mathcal{E}_0)$. By Chebyshev's inequality, we can establish the following upper bound on $\Pr(\mathcal{E}_0)$ as follows

$$\begin{aligned}
\Pr(\mathcal{E}_0) &= \Pr\left(\left|\sum_{t=1}^{n}\left(Y_t(i) - pT_s c_{i,t}\right)\left((c_{i,t}-c_{j,t})pT_s\right)\right| > n\delta_n\right) \\
&\leq \frac{\operatorname{Var}\left[\sum_{t=1}^{n}\left(Y_t(i) - pT_s c_{i,t}\right)\left((c_{i,t}-c_{j,t})pT_s\right)\right]}{(n\delta_n)^2} \\
&= \frac{\sum_{t=1}^{n}\operatorname{Var}\left[\left(Y_t(i) - pT_s c_{i,t}\right)\left((c_{i,t}-c_{j,t})pT_s\right)\right]}{(n\delta_n)^2} \\
&= \frac{\sum_{t=1}^{n}\left((c_{i,t}-c_{j,t})pT_s\right)^2\operatorname{Var}\left[\left(Y_t(i) - pT_s c_{i,t}\right)\right]}{(n\delta_n)^2} \\
&= \frac{T_s^2p^2\sum_{t=1}^{n}\left(c_{i,t}-c_{j,t}\right)^2\operatorname{Var}\left[Y_t(i)\right]}{(n\delta_n)^2} \\
&\leq \frac{T_s^2p^2\sum_{t=1}^{n}\left(c_{i,t}-c_{j,t}\right)^2\cdot ApT_s(1-p)}{(n\delta_n)^2} \\
&= \frac{AT_s^3p^3(1-p)\left\|\mathbf{c}_i-\mathbf{c}_j\right\|^2}{(n\delta_n)^2}, \qquad (8.44)
\end{aligned}$$

Observe that

$$\begin{aligned}
\left\|\mathbf{c}_i-\mathbf{c}_j\right\|^2 &\overset{(a)}{\leq} \left(\|\mathbf{c}_i\| + \left\|\mathbf{c}_j\right\|\right)^2 \\
&\overset{(b)}{\leq} \left(\sqrt{n}\|\mathbf{c}_i\|_\infty + \sqrt{n}\left\|\mathbf{c}_j\right\|_\infty\right)^2 \\
&\overset{(c)}{\leq} \left(\sqrt{n}A + \sqrt{n}A\right)^2 \\
&= 4nA^2, \qquad (8.45)
\end{aligned}$$

where (a) holds by the triangle inequality, (b) follows since $\|.\| \leq \sqrt{n}\|.\|_\infty$ and (c) is valid by (8.15). Hence,

$$\begin{aligned}
\Pr(\mathcal{E}_0) &\leq \frac{T_s^3p^3A(1-p)\left\|\mathbf{c}_i-\mathbf{c}_j\right\|^2}{(n\delta_n)^2} \\
&\leq \frac{4A^2T_s^3p^3A(1-p)n}{n^2\delta_n^2}
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{4A^3T_s^3p^3(1-p)}{n\delta_n^2} \\
&= \frac{4A^3T_s^3p^3(1-p)}{n^b} \\
&\triangleq \zeta_0\,,
\end{aligned} \tag{8.46}$$

We now proceed with bounding $\Pr(\mathcal{E}_1)$ as follows. Based on the codebook construction, each pair of codeword are distanced by at least $r_0 = \sqrt{n\epsilon_n}$, hence,

$$\begin{aligned}
\left\| pT_s(\mathbf{c}_i - \mathbf{c}_j) \right\|^2 &\geq T_s^2p^2n\epsilon_n \\
&= 3n\delta_n\,,
\end{aligned} \tag{8.47}$$

where the equality holds by (8.21). Thus, we can establish the following upper bound for event $\mathcal{E}_1$:

$$\begin{aligned}
&\Pr(\mathcal{E}_1) \\
&\leq \left(\sum_{t=1}^{n} \left(Y_t(i) - pT_sc_{i,t}\right)^2 - (1-p)Y_t(i) \leq 2n\delta_n - \left\| pT_s(\mathbf{c}_i - \mathbf{c}_j) \right\|^2 \right) \\
&\leq \left(\sum_{t=1}^{n} \left(Y_t(i) - pT_sc_{i,t}\right)^2 - (1-p)Y_t(i) \leq 2n\delta_n - 3n\delta_n \right) \\
&= \left(\sum_{t=1}^{n} \left(Y_t(i) - pT_sc_{i,t}\right)^2 - (1-p)Y_t(i) \leq -n\delta_n \right) \\
&\overset{(a)}{\leq} \frac{\mathrm{Var}\left[\sum_{t=1}^{n} \left(Y_t(i) - pT_sc_{i,t}\right)^2 - (1-p)Y_t(i)\right]}{n^2\delta_n^2} \\
&\overset{(b)}{=} \frac{A^4T_s^4 \exp\left(8/pT_sc_{i,t}\right) + \left(2AT_s+1\right)^2 AT_s + \left(2AT_s+1\right) A^2T_s^2 \sqrt{\exp\left(8/pT_sc_{i,t}\right) AT_s}}{n^b} \\
&\triangleq \zeta_1\,,
\end{aligned} \tag{8.48}$$

where (a) follows from applying Chebyshev's inequality, (b) holds by similar line of arguments as we made in type I error probability analysis, see (8.25) and the derivations afterward.

Therefore, recalling (8.41), we obtain

$$P_{e,2}(i,j) \leq \Pr(\mathcal{E}_0) + \Pr(\mathcal{E}_1) \leq \zeta_0 + \zeta_1 \leq e_2\,, \tag{8.49}$$

hence, $P_{e,2}(i,j) \leq e_2$ holds for sufficiently large n and arbitrarily small $e_2 > 0$. We have thus shown that for every $e_1, e_2 > 0$ and sufficiently large n, there exists an $(n, M(n,R), K(n,\kappa), e_1, e_2)$ code.

8.3.3 | Upper Bound (Converse Proof)

The converse proof consists of the following two main steps.

- **Step 1:** We show in Lemma 8.3.1 that for any achievable rate (for which the type I and type II error probabilities vanish as $n \to \infty$), the distance between every pair of codeword should be at least larger than a threshold.
- **Step 2:** Employing Lemma 8.3.1, we derive an upper bound on the codebook size of achievable DI codes.

We start with the following lemma which establish a lower bound on the letter-wise ratio for every pair of codewords.

Lemma 8.3.1. *Suppose that R is an achievable rate for the Binomial channel $\mathcal{B}$. Consider a sequence of $(n, M(n,R), K(n,\kappa), e_1^{(n)}, e_2^{(n)})$ codes $(\mathcal{C}^{(n)}, \mathcal{T}^{(n)})$ such that $e_1^{(n)}$ and $e_2^{(n)}$ tend to zero as $n \to \infty$. Then, given a sufficiently large n, the codebook $\mathcal{C}^{(n)}$ satisfies the following property. For every pair of codewords, $\mathbf{c}_{i_1}$ and $\mathbf{c}_{i_2}$, such that $i_1, i_2 \in [\![M]\!]$ and $i_1 \neq i_2$, there exist $t \in [\![n]\!]$, such that,*

$$\left| c_{i_1,t} - c_{i_2,t} \right| > \epsilon'_n \,, \tag{8.50}$$

where

$$\epsilon'_n = \frac{P_{max}}{n^{1+b}} \,, \tag{8.51}$$

with $b > 0$ being an arbitrarily small constant.

In the following, we provide the proof of Lemma 8.3.1. The method of proof is by contradiction, namely, we assume that the condition given in (8.50) is violated and then we show that this leads to a contradiction, namely, sum of the type I and type II error probabilities converge to one, i.e., $\lim_{n\to\infty} \left[P_{e,1}(i_1) + P_{e,2}(i_2, i_1) \right] = 1$.

Proof. Fix e_1 and e_2. Let μ, θ, η, ζ be arbitrarily small positive. Assume to the contrary that there exist two messages i_1 and i_2, where $i_1 \neq i_2$, such that,

$$\left| c_{i_1,t} - c_{i_2,t} \right| \leq \epsilon'_n \,, \tag{8.52}$$

which implies

$$c_{i_1,t} - c_{i_2,t} \geq -\epsilon'_n \,, \tag{8.53}$$

$$c_{i_2,t} - c_{i_1,t} \geq -\epsilon'_n \,, \tag{8.54}$$

$$c_{i_1,t} - c_{i_2,t} \leq \epsilon'_n \,, \tag{8.55}$$

$$c_{i_2,t} - c_{i_1,t} \leq \epsilon'_n \,. \tag{8.56}$$

Observe that

$$P_{e,1}(i_1) + P_{e,2}(i_2, i_1) = \left[1 - \sum_{\mathbf{y} \in \mathbb{T}_{i_1}} W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right)\right] + \sum_{\mathbf{y} \in \mathbb{T}_{i_1}} W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_2}\right). \tag{8.57}$$

$$\begin{aligned}
&W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) - W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_2}\right) \\
&= W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) \left[1 - \frac{W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_2}\right)}{W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right)}\right] \\
&= W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) \left[1 - \frac{\prod_{t=1}^n \binom{T_s c_{i_2,t}}{y_t} p^{y_t} (1-p)^{T_s c_{i_2,t} - y_t}}{\prod_{t=1}^n \binom{T_s c_{i_1,t}}{y_t} p^{y_t} (1-p)^{T_s c_{i_1,t} - y_t}}\right] \\
&= W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) \left[1 - \prod_{t=1}^n \frac{\binom{T_s c_{i_2,t}}{y_t}}{\binom{T_s c_{i_2,t}}{y_t}} \cdot (1-p)^{T_s\left(c_{i_2,t} - c_{i_1,t}\right)}\right] \\
&= W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_{i_1}\right) \left[1 - \prod_{t=1}^n \frac{T_s c_{i_2,t}!}{T_s c_{i_1,t}!} \cdot \frac{(T_s c_{i_1,t} - y_t)!}{(T_s c_{i_2,t} - y_t)!} \cdot (1-p)^{T_s\left(c_{i_2,t} - c_{i_1,t}\right)}\right] .
\end{aligned} \tag{8.58}$$

In order to bound (8.58), we exploit the following useful double-inequality regarding ratio of two Gamma functions [190, Eq. 4.15]. For $0 < a < b$, we have

$$\min\left\{a, \frac{a+b-1}{2}\right\} \leq \left(\frac{\Gamma(a)}{\Gamma(b)}\right)^{\frac{1}{a-b}} \leq \max\left\{a, \frac{a+b-1}{2}\right\} . \tag{8.59}$$

To analyze more accurately, we divide into three cases.

- □ **Case 1:** Where $c_{i_1,t} < c_{i_2,t}\,, \forall t \in [\![n]\!]$
- □ **Case 2:** Where $c_{i_2,t} < c_{i_1,t}\,, \forall t \in [\![n]\!]$
- □ **Case 3:** $\begin{cases} c_{i_1,t} < c_{i_2,t} & \text{for } n_{1 \geq 1} \text{ indices} \\ c_{i_2,t} < c_{i_1,t} & \text{for } n_{2 \geq 1} \text{ indices} \end{cases}, n_1 + n_2 = n$

8.3.4 | Case 1

Consider the case 1, i.e, where $c_{i_1,t} < c_{i_2,t}, \forall t \in [\![n]\!]$. Then, we set $a = T_s c_{i_1,t} + 1$ and $b = T_s c_{i_2,t} + 1$. Now, condition $0 < a < b$ is met and we obtain

$$\begin{aligned}
\frac{T_s c_{i_2,t}!}{T_s c_{i_1,t}!} &= \frac{\Gamma(T_s c_{i_2,t} + 1)}{\Gamma(T_s c_{i_1,t} + 1)} \\
&= \frac{\Gamma(b)}{\Gamma(a)} \\
&\geq \left(\frac{1}{\max\left\{a, \frac{a+b-1}{2}\right\}} \right)^{a-b} \\
&= \left(\frac{1}{\max\left\{T_s c_{i_1,t} - y_t + 1, \frac{T_s(c_{i_1,t}+c_{i_2,t})-2y_t+1}{2}\right\}} \right)^{T_s(c_{i_1,t}-c_{i_2,t})} \\
&\geq \left(\frac{1}{\max\left\{AT_s + 1, AT_s + \frac{1}{2}\right\}} \right)^{T_s(c_{i_1,t}-c_{i_2,t})} \\
&\geq \left(\frac{1}{AT_s + 1} \right)^{T_s(c_{i_1,t}-c_{i_2,t})} .
\end{aligned} \tag{8.60}$$

and

$$\begin{aligned}
\frac{T_s c_{i_2,t}!}{T_s c_{i_1,t}!} &= \frac{\Gamma(T_s c_{i_2,t} + 1)}{\Gamma(T_s c_{i_1,t} + 1)} \\
&= \frac{\Gamma(b)}{\Gamma(a)} \\
&\leq \left(\frac{1}{\min\left\{a, \frac{a+b-1}{2}\right\}} \right)^{a-b} \\
&= \left(\frac{1}{\min\left\{T_s c_{i_1,t} - y_t + 1, \frac{T_s(c_{i_1,t}+c_{i_2,t})-2y_t+1}{2}\right\}} \right)^{T_s(c_{i_1,t}-c_{i_2,t})} \\
&\leq \left(\frac{1}{\frac{1}{2}} \right)^{T_s(c_{i_1,t}-c_{i_2,t})} \\
&\leq 2^{T_s(c_{i_1,t}-c_{i_2,t})} .
\end{aligned} \tag{8.61}$$

Second, we set $a = T_s c_{i_1,t} - y_t + 1$ and $b = T_s c_{i_2,t} - y_t + 1$. Now, again condition $0 < a < b$ is met and we obtain

$$\begin{aligned}
\frac{(T_s c_{i_1,t} - y_t)!}{(T_s c_{i_2,t} - y_t)!} &= \frac{\Gamma(T_s c_{i_1,t} - y_t + 1)}{\Gamma(T_s c_{i_2,t} - y_t + 1)} \\
&= \frac{\Gamma(a)}{\Gamma(b)} \\
&\geq \left(\min\left\{ a, \frac{a+b-1}{2} \right\} \right)^{a-b} \\
&= \left(\min\left\{ T_s c_{i_2,t} - y_t + 1, \frac{T_s(c_{i_1,t} + c_{i_2,t}) - 2y_t + 1}{2} \right\} \right)^{T_s(c_{i_1,t} - c_{i_2,t})} \\
&\geq \left(\frac{1}{2} \right)^{T_s(c_{i_1,t} - c_{i_2,t})} . \qquad (8.62)
\end{aligned}$$

and

$$\begin{aligned}
\frac{(T_s c_{i_1,t} - y_t)!}{(T_s c_{i_2,t} - y_t)!} &= \frac{\Gamma(T_s c_{i_1,t} - y_t + 1)}{\Gamma(T_s c_{i_2,t} - y_t + 1)} \\
&= \frac{\Gamma(a)}{\Gamma(b)} \\
&\leq \left(\max\left\{ a, \frac{a+b-1}{2} \right\} \right)^{a-b} \\
&= \left(\max\left\{ T_s c_{i_1,t} - y_t + 1, \frac{T_s(c_{i_1,t} + c_{i_2,t}) - 2y_t + 1}{2} \right\} \right)^{T_s(c_{i_1,t} - c_{i_2,t})} \\
&= \left(\max\left\{ T_s c_{i_1,t} + 1, \frac{T_s(c_{i_1,t} + c_{i_2,t}) + 1}{2} \right\} \right)^{T_s(c_{i_1,t} - c_{i_2,t})} \\
&= \left(\max\left\{ AT_s + 1, \frac{2AT_s + 1}{2} \right\} \right)^{T_s(c_{i_1,t} - c_{i_2,t})} \\
&\leq (AT_s + 1)^{T_s(c_{i_1,t} - c_{i_2,t})} . \qquad (8.63)
\end{aligned}$$

Now, observe that

$$\begin{aligned}
&\prod_{t=1}^{n} \frac{T_s c_{i_2,t}!}{T_s c_{i_1,t}!} \cdot \frac{(T_s c_{i_1,t} - y_t)!}{(T_s c_{i_2,t} - y_t)!} \cdot (1-p)^{T_s\left(c_{i_2,t} - c_{i_1,t}\right)} \\
&\geq \prod_{t=1}^{n} \left(\frac{1}{AT_s + 1} \right)^{T_s(c_{i_1,t} - c_{i_2,t})} \cdot \left(\frac{1}{2} \right)^{T_s(c_{i_1,t} - c_{i_2,t})} \cdot (1-p)^{T_s\left(c_{i_2,t} - c_{i_1,t}\right)}
\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{1}{2(AT_s+1)}\right)^{\sum_{t=1}^{n} T_s(c_{i_1,t}-c_{i_2,t})} \cdot \left(\frac{1}{1-p}\right)^{\sum_{t=1}^{n} T_s(c_{i_1,t}-c_{i_2,t})} \\
&= \left(1-\frac{2AT_s+1}{2(AT_s+1)}\right)^{\sum_{t=1}^{n} T_s(c_{i_1,t}-c_{i_2,t})} \cdot \left(\frac{1}{1-p}\right)^{\sum_{t=1}^{n} T_s(c_{i_1,t}-c_{i_2,t})} \\
&\overset{(a)}{\geq} \left(1-\frac{2AT_s+1}{2(AT_s+1)}\right)^{-\sum_{t=1}^{n} T_s\epsilon'_n} \cdot \left(\frac{1}{1-p}\right)^{-\sum_{t=1}^{n} T_s\epsilon'_n} \\
&= \left(\frac{1}{\left(1-\frac{2AT_s+1}{2(AT_s+1)}\right)^{\sum_{t=1}^{n} T_s\epsilon'_n}}\right) \cdot \left((1-p)^{\sum_{t=1}^{n} T_s\epsilon'_n}\right) \\
&\overset{(b)}{\geq} \left(\frac{1}{\left(1-\frac{2AT_s+1}{2(AT_s+1)} \cdot \sum_{t=1}^{n} T_s\epsilon'_n\right)}\right) \cdot \left(\left(1-p\cdot\sum_{t=1}^{n} T_s\epsilon'_n\right)\right) \\
&\overset{(c)}{\geq} \left(\frac{1}{\left(1-\frac{2AT_s+1}{2(AT_s+1)} \cdot \sum_{t=1}^{n} T_s\epsilon'_n\right)}\right) \cdot \left(1-pT_s n\epsilon'_n\right) \\
&\overset{(d)}{\geq} \left(\frac{1}{\left(1-\frac{2AT_s+1}{2(AT_s+1)} \cdot \sum_{t=1}^{n} T_s\epsilon'_n\right)}\right) \cdot \left(1-\frac{pT_s P_{\max}}{n^b}\right) \\
&\overset{(e)}{\geq} \frac{1}{1-0} \cdot (1-\kappa) \\
&\geq 1-\kappa\,.
\end{aligned} \tag{8.64}
$$

where (a) holds by (8.53), (b) follows from the Bernoulli inequality, i.e., $(1-x)^r \leq 1+rx$ for $x \geq -1$ and $0 \leq \forall r_{\in\mathbb{R}} \leq 1$, (c) holds since $\sum_{t=1}^{n} T_s\epsilon'_n = T_s n\epsilon'_n$, (d) holds by (8.51), and (e) follows from $\sum_{t=1}^{n} T_s\epsilon'_n \geq 0$.

On the other hand, to provide an upper bound, we have

$$
\begin{aligned}
&\prod_{t=1}^{n} \frac{T_s c_{i_2,t}!}{T_s c_{i_1,t}!} \cdot \frac{(T_s c_{i_1,t}-y_t)!}{(T_s c_{i_2,t}-y_t)!} \cdot (1-p)^{T_s\left(c_{i_2,t}-c_{i_1,t}\right)} \\
&\overset{(a)}{\leq} \prod_{t=1}^{n} 2^{T_s(c_{i_1,t}-c_{i_2,t})} \cdot (AT_s+1)^{T_s(c_{i_1,t}-c_{i_2,t})} \\
&= 2^{\sum_{t=1}^{n} T_s(c_{i_1,t}-c_{i_2,t})} \cdot (AT_s+1)^{\sum_{t=1}^{n} T_s(c_{i_1,t}-c_{i_2,t})} \\
&\overset{(b)}{\leq} 2^{T_s n\epsilon'_n} \cdot (AT_s+1)^{T_s n\epsilon'_n} \\
&= \left(2(AT_s+1)\right)^{T_s n\epsilon'_n}
\end{aligned}
$$

$$
\begin{aligned}
&\leq (1+2AT_{\mathrm{s}})^{T_{\mathrm{s}}P_{\max}/n^b} \\
&\overset{(c)}{\leq} 1+2AT_{\mathrm{s}}\cdot T_{\mathrm{s}}P_{\max}/n^b \\
&\leq 1+\kappa\,. \qquad (8.65)
\end{aligned}
$$

where (a) holds since $1-p\leq 1$, (b) follows from (8.56), and (c) holds by the Bernoulli inequality, i.e., $(1-x)^r\leq 1+rx$ for $x\geq -1$ and $0\leq \forall r_{\in\mathbb{R}}\leq 1$.

8.3.5 | Case 2

Consider the case 2, i.e, where $c_{i_2,t} < c_{i_1,t}, \forall t \in [\![n]\!]$. Then, we set $a = T_{\mathrm{s}}c_{i_2,t}+1$ and $b = T_{\mathrm{s}}c_{i_1,t}+1$. Now, condition $0<a<b$ is met and we obtain

$$
\begin{aligned}
\frac{T_{\mathrm{s}}c_{i_2,t}!}{T_{\mathrm{s}}c_{i_1,t}!} &= \frac{\Gamma(T_{\mathrm{s}}c_{i_2,t}+1)}{\Gamma(T_{\mathrm{s}}c_{i_1,t}+1)} \\
&= \frac{\Gamma(a)}{\Gamma(b)} \\
&\geq \left(\min\left\{a, \frac{a+b-1}{2}\right\}\right)^{a-b} \\
&= \left(\min\left\{T_{\mathrm{s}}c_{i_2,t}+1, \frac{T_{\mathrm{s}}(c_{i_2,t}+c_{i_1,t})+1}{2}\right\}\right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})} \\
&\geq \left(\frac{1}{2}\right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}\,. \qquad (8.66)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{T_{\mathrm{s}}c_{i_2,t}!}{T_{\mathrm{s}}c_{i_1,t}!} &= \frac{\Gamma(T_{\mathrm{s}}c_{i_2,t}+1)}{\Gamma(T_{\mathrm{s}}c_{i_1,t}+1)} \\
&= \frac{\Gamma(a)}{\Gamma(b)} \\
&\leq \left(\max\left\{a, \frac{a+b-1}{2}\right\}\right)^{a-b} \\
&= \left(\max\left\{T_{\mathrm{s}}c_{i_2,t}+1, \frac{T_{\mathrm{s}}(c_{i_2,t}+c_{i_1,t})-2y_t+1}{2}\right\}\right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})} \\
&= \left(\max\left\{T_{\mathrm{s}}c_{i_2,t}+1, \frac{T_{\mathrm{s}}(c_{i_2,t}+c_{i_1,t})+1}{2}\right\}\right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}
\end{aligned}
$$

$$\leq \left(\max\left\{ AT_{\mathrm{s}} + 1, \frac{2AT_{\mathrm{s}}+1}{2} \right\} \right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}$$
$$\leq \left(AT_{\mathrm{s}}+1\right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})} . \tag{8.67}$$

Second, we set $a = T_{\mathrm{s}}c_{i_2,t} - y_t + 1$ and $b = T_{\mathrm{s}}c_{i_1,t} - y_t + 1$. Now, again condition $0 < a < b$ is met and we obtain

$$\begin{aligned}
\frac{(T_{\mathrm{s}}c_{i_1,t}-y_t)!}{(T_{\mathrm{s}}c_{i_2,t}-y_t)!} &= \frac{\Gamma(T_{\mathrm{s}}c_{i_1,t}-y_t+1)}{\Gamma(T_{\mathrm{s}}c_{i_2,t}-y_t+1)} \\
&= \frac{\Gamma(b)}{\Gamma(a)} \\
&\geq \left(\frac{1}{\max\left\{ a, \frac{a+b-1}{2} \right\}} \right)^{a-b} \\
&= \left(\frac{1}{\max\left\{ T_{\mathrm{s}}c_{i_2,t}-y_t+1, \frac{T_{\mathrm{s}}(c_{i_1,t}+c_{i_2,t})-2y_t+1}{2} \right\}} \right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})} \\
&\geq \left(\frac{1}{\max\left\{ AT_s+1, AT_s+\frac{1}{2} \right\}} \right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})} \\
&\geq \left(\frac{1}{AT_s+1} \right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})} .
\end{aligned} \tag{8.68}$$

and

$$\begin{aligned}
\frac{(T_{\mathrm{s}}c_{i_1,t}-y_t)!}{(T_{\mathrm{s}}c_{i_2,t}-y_t)!} &= \frac{\Gamma(T_{\mathrm{s}}c_{i_1,t}-y_t+1)}{\Gamma(T_{\mathrm{s}}c_{i_2,t}-y_t+1)} \\
&= \frac{\Gamma(b)}{\Gamma(a)} \\
&\leq \left(\frac{1}{\min\left\{ a, \frac{a+b-1}{2} \right\}} \right)^{a-b} \\
&= \left(\frac{1}{\min\left\{ T_{\mathrm{s}}c_{i_2,t}-y_t+1, \frac{T_{\mathrm{s}}(c_{i_1,t}+c_{i_2,t})-2y_t+1}{2} \right\}} \right)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}
\end{aligned}$$

$$\leq \left(\frac{1}{\frac{1}{2}}\right)^{T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}$$
$$\leq 2^{T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}. \tag{8.69}$$

Now, observe that

$$\begin{aligned}
&\prod_{t=1}^{n}\frac{T_{\mathrm{s}}c_{i_{2,t}}!}{T_{\mathrm{s}}c_{i_{1,t}}!}\cdot\frac{(T_{\mathrm{s}}c_{i_{1,t}}-y_t)!}{(T_{\mathrm{s}}c_{i_{2,t}}-y_t)!}\cdot(1-p)^{T_{\mathrm{s}}\left(c_{i_{2,t}}-c_{i_{1,t}}\right)}\\
&\geq\prod_{t=1}^{n}\left(\frac{1}{2}\right)^{T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\cdot\left(\frac{1}{AT_s+1}\right)^{T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\cdot(1-p)^{T_{\mathrm{s}}\left(c_{i_{2,t}}-c_{i_{1,t}}\right)}\\
&=\left(\frac{1}{2(AT_s+1)}\right)^{\sum_{t=1}^{n}T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\cdot(1-p)^{\sum_{t=1}^{n}T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\\
&=\left(1-\frac{2AT_s+1}{2(AT_s+1)}\right)^{\sum_{t=1}^{n}T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\cdot(1-p)^{\sum_{t=1}^{n}T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\\
&\overset{(a)}{\geq}\left(1-\frac{2AT_s+1}{2(AT_s+1)}\right)^{-\sum_{t=1}^{n}T_s\epsilon'_n}\cdot(1-p)^{-\sum_{t=1}^{n}T_s\epsilon'_n}\\
&=\left(\frac{1}{\left(1-\frac{2AT_s+1}{2(AT_s+1)}\right)^{\sum_{t=1}^{n}T_s\epsilon'_n}}\right)\cdot\left(\frac{1}{(1-p)^{\sum_{t=1}^{n}T_s\epsilon'_n}}\right)\\
&\overset{(b)}{\geq}\left(\frac{1}{\left(1-\frac{2AT_s+1}{2(AT_s+1)}\cdot\sum_{t=1}^{n}T_s\epsilon'_n\right)}\right)\cdot\left(\frac{1}{(1-p\cdot\sum_{t=1}^{n}T_s\epsilon'_n)}\right)\\
&\overset{(c)}{\geq}\frac{1}{1-0}\cdot\frac{1}{1-0}\\
&=1\\
&\geq 1-\kappa\,.
\end{aligned}\tag{8.70}$$

where (a) holds by (8.54), (b) follows from the Bernoulli inequality, i.e., $(1-x)^r\leq 1+rx$ for $x\geq -1$ and $0\leq\forall r_{\in\mathbb{R}}\leq 1$, and (c) holds since $\sum_{t=1}^{n}T_s\epsilon'_n\geq 0$.

On the other hand, to provide an upper bound, we have

$$\begin{aligned}
&\prod_{t=1}^{n}\frac{T_{\mathrm{s}}c_{i_{2,t}}!}{T_{\mathrm{s}}c_{i_{1,t}}!}\cdot\frac{(T_{\mathrm{s}}c_{i_{1,t}}-y_t)!}{(T_{\mathrm{s}}c_{i_{2,t}}-y_t)!}\cdot(1-p)^{T_{\mathrm{s}}\left(c_{i_{2,t}}-c_{i_{1,t}}\right)}\\
&\overset{(a)}{\leq}\prod_{t=1}^{n}(AT_{\mathrm{s}}+1)^{T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\cdot 2^{T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\\
&=(AT_{\mathrm{s}}+1)^{\sum_{t=1}^{n}T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}\cdot 2^{\sum_{t=1}^{n}T_{\mathrm{s}}(c_{i_{2,t}}-c_{i_{1,t}})}
\end{aligned}$$

$$
\begin{aligned}
&\overset{(b)}{\leq} \left(AT_{\mathrm{s}}+1\right)^{T_{\mathrm{s}} n \epsilon_n'} \cdot 2^{T_{\mathrm{s}} n \epsilon_n'} \\
&= \left(2(AT_{\mathrm{s}}+1)\right)^{T_{\mathrm{s}} n \epsilon_n'} \\
&\leq \left(1+2AT_{\mathrm{s}}\right)^{T_{\mathrm{s}} P_{\max}/n^b} \\
&\overset{(c)}{\leq} 1+2AT_{\mathrm{s}} \cdot T_{\mathrm{s}} P_{\max}/n^b \\
&\leq 1+\kappa\,. \qquad (8.71)
\end{aligned}
$$

where (a) holds since $1-p \leq 1$, (b) follows from (8.56), and (c) holds by the Bernoulli inequality, i.e., $(1-x)^r \leq 1+rx$ for $x \geq -1$ and $0 \leq \forall r_{\in \mathbb{R}} \leq 1$.

8.3.6 | Case 3

Consider the case 3, i.e., where for n_1 indices we have $c_{i_1,t} < c_{i_2,t}$ and for n_2 indices we have $c_{i_2,t} < c_{i_1,t}$ such that the total indices sum up to n, i.e., $n_1+n_2 = n$. Therefore, for n_1 indices the rule (corresponding inequalities) of case 1 applies and for n_2 indices the rule (corresponding inequalities) of case 2 applies. That is, we have (follow the continuation in the landscape page in below)

Therefore,

$$
\begin{aligned}
e_1+e_2 &\geq P_{e,1}(i_1)+P_{e,2}(i_2,i_1) \\
&\overset{(a)}{=} \left[1-\sum_{\mathbf{y}\in\mathbb{T}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right)\right] + \sum_{\mathbf{y}\in\mathbb{T}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_2}\right) \\
&\overset{(b)}{=} 1-\sum_{\mathbf{y}\in\mathbb{T}_{i_1}} \left[W^n\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) - W^n\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_2}\right)\right] \\
&\overset{(c)}{=} 1-\kappa \sum_{\mathbf{y}\in\mathbb{T}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) \\
&= 1-\kappa\,, \qquad (8.72)
\end{aligned}
$$

where (a) follows from, (b) holds by (8.58), and (c) follows since

$$
\begin{aligned}
\sum_{\mathbf{y}\in\mathbb{T}_{i_1}} W^n\left(\mathbf{y}\,\middle|\,\mathbf{c}_{i_1}\right) &= \Pr\left(\mathbf{y}\in\mathbb{T}_{i_1}\right) \\
&\leq \Pr\left(\mathbf{y}\in\mathbb{N}_0^n\right) \\
&= 1\,. \qquad (8.73)
\end{aligned}
$$

$$
\begin{aligned}
&\prod_{t=1}^{n}\frac{T_{\mathrm{s}}c_{i_2,t}!}{T_{\mathrm{s}}c_{i_1,t}!}\cdot\frac{(T_{\mathrm{s}}c_{i_1,t}-y_t)!}{(T_{\mathrm{s}}c_{i_2,t}-y_t)!}\cdot(1-p)^{T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)} \leq \left[\prod_{t=1}^{n_1}2^{T_{\mathrm{s}}(c_{i_1,t}-c_{i_2,t})}\cdot\prod_{t=1}^{n_2}(AT_{\mathrm{s}}+1)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}\right]\\
&\cdot\left[\prod_{t=1}^{n_1}(AT_{\mathrm{s}}+1)^{T_{\mathrm{s}}(c_{i_1,t}-c_{i_2,t})}\cdot\prod_{t=1}^{n_2}2^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}\right]\cdot\left[\prod_{t=1}^{n_1}(1-p)^{T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)}\cdot\prod_{t=1}^{n_2}(1-p)^{T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)}\right]\\
&\leq\left[\prod_{t=1}^{n_1}2^{T_{\mathrm{s}}(c_{i_1,t}-c_{i_2,t})}\cdot(AT_{\mathrm{s}}+1)^{T_{\mathrm{s}}(c_{i_1,t}-c_{i_2,t})}\cdot(1-p)^{T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)}\right]\cdot\left[\prod_{t=1}^{n_2}(AT_{\mathrm{s}}+1)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}\cdot2^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}\cdot(1-p)^{T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)}\right]\\
&=\left[\prod_{t=1}^{n_1}\big(2(AT_{\mathrm{s}}+1)\big)^{T_{\mathrm{s}}(c_{i_1,t}-c_{i_2,t})}\cdot(1-p)^{T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)}\right]\cdot\left[\prod_{t=1}^{n_2}\big(2(AT_{\mathrm{s}}+1)\big)^{T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}\cdot(1-p)^{T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)}\right]\\
&=\left[\big(2(AT_{\mathrm{s}}+1)\big)^{\sum_{t=1}^{n_1}T_{\mathrm{s}}(c_{i_1,t}-c_{i_2,t})}\cdot(1-p)^{\sum_{t=1}^{n_1}T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)}\right]\cdot\left[\big(2(AT_{\mathrm{s}}+1)\big)^{\sum_{t=1}^{n_2}T_{\mathrm{s}}(c_{i_2,t}-c_{i_1,t})}\cdot(1-p)^{\sum_{t=1}^{n_2}T_{\mathrm{s}}\left(c_{i_2,t}-c_{i_1,t}\right)}\right]\\
&=\left[\big(2(AT_{\mathrm{s}}+1)\big)^{T_{\mathrm{s}}n_1\epsilon'_n}\cdot(1-p)^{T_{\mathrm{s}}n_1\epsilon'_n}\right]\cdot\left[\big(2(AT_{\mathrm{s}}+1)\big)^{T_{\mathrm{s}}n_2\epsilon'_n}\cdot(1-p)^{T_{\mathrm{s}}n_2\epsilon'_n}\right]\\
&=\left[\big(2(AT_{\mathrm{s}}+1)\big)^{T_{\mathrm{s}}(n_1+n_2)\epsilon'_n}\cdot(1-p)^{T_{\mathrm{s}}(n_1+n_2)\epsilon'_n}\right]\\
&=\left[(1+2AT_{\mathrm{s}})^{T_{\mathrm{s}}P_{\max}/n^b}\cdot(1-p)^{T_{\mathrm{s}}P_{\max}/n^b}\right]\\
&\overset{(b)}{\leq}(1+2AT_{\mathrm{s}})^{T_{\mathrm{s}}P_{\max}/n^b}\overset{(c)}{\leq}1+2AT_{\mathrm{s}}\cdot T_{\mathrm{s}}P_{\max}/n^b\\
&\leq 1+\kappa\,.
\end{aligned}
\tag{8.75}
$$

where (b) follows since $1-p\leq 1$, and (c) follows from the Bernoulli inequality, i.e., $(1-x)^r\leq 1+rx$ for $x\geq -1$ and $0\leq \forall r_{\in\mathbb{R}}\leq 1$,

Therefore,

$$e_1 + e_2 \geq 1 - \kappa \,, \tag{8.74}$$

which leads to a contradiction since for sufficiently small μ and vanishing error probabilities we obtain $\kappa < 1 - e_1 - e_2$. This completes the proof of Lemma 8.3.1. □

Next, we employ Lemma 8.3.1 to determine the upper bound on scale of codebook size for $\mathcal{B}$. Observe that Lemma 8.3.1 implies that the distance between every pair of codewords fulfill

$$\begin{aligned} \left\| \mathbf{c}_{i_1} - \mathbf{c}_{i_2} \right\| &\geq \left| c_{i_1,t} - c_{i_1,t} \right| \\ &\geq \epsilon'_n \\ &= \frac{P_{\max}}{n^{1+b}} \,, \end{aligned} \tag{8.76}$$

Thus, we can define an arrangement of non-overlapping spheres $\mathcal{S}_{\mathbf{c}_i}(n, \epsilon'_n)$, i.e., spheres of radius ϵ'_n that are centered at the codewords $\mathbf{c}_i$. Since the codewords all belong to a hyper cube $\mathcal{Q}_{\mathbf{0}}(n, P_{\max})$ with edge length $P_{\max}$, it follows that the number of packed small spheres, i.e., the number of codewords M, is bounded by

$$\begin{aligned} M &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{L} \mathcal{S}_{\mathbf{u}_i}(n, r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))} \\ &= \frac{\Delta_n(\mathscr{S}) \cdot \mathrm{Vol}\left(\mathcal{Q}_{\mathbf{0}}(n, P_{\max})\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))} \\ &\leq 2^{-0.599n} \cdot \frac{P_{\max}^n}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))} \,, \end{aligned} \tag{8.77}$$

where the last inequality follows from inequality (8.12). Thereby,

$$\begin{aligned} \log M &\leq \log\left(\frac{P_{\max}^n}{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{c}_1}(n, r_0)\right)}\right) - 0.599n \\ &= n \log P_{\max} - n \log r_0 - n \log \sqrt{\pi} + \frac{1}{2} n \log \frac{n}{2} - \frac{n}{2} \log e + o(n) - 0.599n \,, \end{aligned} \tag{8.78}$$

where the dominant term is again of order $n \log n$. Hence, for obtaining a finite value for the upper bound of the rate, R, (8.78) induces the scaling law of M to be $2^{(n \log n)R}$. Hence, by setting $M(n, R) = 2^{(n \log n)R}$ and $r_0 = \epsilon'_n = P_{\max}/n^{1+b}$, we obtain

$R \leq$

$$\frac{1}{n\log n}\left[n\log P_{\max}-n\log r_0-n\log\sqrt{\pi}+\frac{1}{2}n\log\frac{n}{2}-\frac{n}{2}\log e+o(n)-0.599n\right]=$$
$$\frac{1}{n\log n}\left[\left(\frac{1}{2}+(1+b)\right)n\log n-n\left(\log P_{\max}\sqrt{\pi e}+1.099\right)+o(n)\right], \tag{8.79}$$

which tends to $\frac{3}{2}$ as $n\to\infty$ and $b\to 0$. This completes the proof of Theorem 8.3.1.

8.4 | Summary

In this chapter, we studied the DI problem over the Binomial channel. We assume that the transmitter is subject to both the average and peak molecule release rate constraints. Our results in this chapter may serve as a model for event-triggered based tasks in the context of future XG applications. In particular, we obtained lower and upper bounds on the DI capacity of the Binomial channel with the codebook size of $M(n,R)=2^{(n\log n)R}=n^{nR}$. Our results for the DI capacity of the Binomial channel revealed that the super-exponential scale of $n^{nR}=2^{(n\log n)R}$ is again the appropriate scale for codebook size. This scale coincides as of the codebook for the memoryless Gaussian channels [99,105] and Poisson channels [51,108] and stands considerably different from the traditional scales in transmission and RI setups where corresponding codebooks size grows exponentially and double exponentially, respectively.

We show the achievability proof using a sphere packing arrangement of hyper spheres and a distance decoder. In particular, we pack hyper spheres with radius $\sqrt{n\epsilon_n}\sim n^{\frac{1}{4}}$, inside a larger hyper sphere, which results in $\sim 2^{\frac{1}{4}n\log n}$ codewords. For the converse proof, we follow a similar approach as in chapter 4 and 5 for the DI over the Gaussian channels [99,106]. That is, given a certain condition imposed on the codebook, using the continuity of the channel law, we reach to a contradiction on sum of the error probabilities. In general, the derivation here is more involved than the derivation in the DI case [105] and entails employing of new analysis and inequalities. Here, proving the continuity of Binomial law requires dealing with binomial coefficients and factorial terms. We used inequalities on the ratio of two Gamma function depending on the relation of two codeword's symbols with each other in all possible cases. In chapter 4 and 5 on Gaussian channels with fading [105], the converse proof was based on establishing a minimum distance between Euclidean norm of each pair of codewords. Here, we consider a distance (absolute value norm) between symbols of two different codeword in the relevant Lemma; cf. (8.50).

CHAPTER 9

DKI FOR SLOW FADING CHANNELS

“ If The Stone Does Not Want to Shake, If it is Stuck,
First Move The Stones Around it. ”

Ludwig Witgenstein,

9.1 | Introduction

Modern communications within the scope of future generation wireless networks (XG) [26, 27] require the transfer of extensive amount of data in wireless communication, including smart applications for internet of things [127], cellular communication, sensor networks, etc. One of the basic and abstract models for wireless communication systems is the fading channel [128, 144]. Unlike the fast fading setting, where the coherence time of the channel is small relative to the latency requirement of the application [144, 191], in the slow fading regime, the latency is short compared to the coherence time [144, 191]. In some appliances, the receiver may acquire channel side information (CSI) by instantaneous estimation of the channel parameters [149, 192].

In the (standard) identification problem [23], the receiver is interested in a *single* message which we refer to as the target message in the rest of the paper. However, for the K-identification problem [193], the receiver aims to determine the presence of a single message within a set of messages referred to as the target message set[1]. The K-identification scenario may be understood as the generalization of the original identification problem within this interpretation: the target message (singleton) is substituted with a set of more than one element with size K. The first result for K-identification is

[1] For instance, the K-identification scenario may be used whenever a person aims to determine whether a winner is among their favourite teams or within the context of lottery prize; when people seek to know if a lottery number is among their collection of numbers.

derived by Ahlswede for a DMC $\mathcal{W}$ with randomized encoder setting as follows: Assume that $K = 2^{\kappa n}$, then the set of all achievable coding and target identification rate pairs, i.e., (R, κ) with a codebook of double exponentially large, i.e., $M = 2^{2^{nR}}$, contains $\{(R, \kappa) : 0 \leq R, \kappa; R + 2\kappa \leq \mathbb{C}_{RI}(\mathcal{W}, M, K)\}$; see [193, Th. 1]. DI for block fading channels *without* CSI is studied in [194]. To the best of the authors' knowledge, the fundamental performance limits of DKI for the Gaussian channels has not been studied in the literature, yet.

9.1.1 | Contributions

In this chapter, we consider identification systems employing deterministic encoder and receivers that are interested to accomplish the K-identification task, namely, finding an object in a target message set of size K where $K = 2^{\kappa \log n}$ for $\kappa \in [0, 1)$ scales sub-linearly in the codeword length n. We assume that the noise is additive Gaussian and the signal experiences slow fading process. Further, we assume that the channel side information (CSI) is available at the decoder. We formulate the problem of DKI over the GSF under average power constraint which account for the restricted signal energy in the transmitter. As our main objective, we investigate the fundamental performance limits of DKI over the slow fading channel. In particular, this paper makes the following contributions:

- ♢ **Generalized Identification Model**: In several identification systems, often the size of target message set K can be large, particularly when one by one comparison is not demanded due to the delay constraint. In addition, the value of K may increases with the codeword lengths n. To do so, we consider a generalized identification model that captures the standard channel (i.e., $K = 1$), identification channels with constant $K > 1$, and identification channels for which K increases with the codeword length n. To the best of the authors' knowledge, such a generalized deterministic identification model has not been studied in the literature, yet.

- ♢ **Codebook Scale**: We establish that the codebook size of DKI problem over the Gaussian channels with slow fading for deterministic encoding scales in n similar to the DI problem ($K = 1$) [99, 105], namely super-exponentially in the codeword length ($\sim 2^{(n \log n)R}$), even when the size of target message set scale as $K = 2^{\kappa \log n}$ for any $\kappa \in [0, 1)$, which we refer to as the target identification rate. This observa-

tion suggests that increasing the number of target messages does not change the scale of the codebook derived for DI over the Gaussian channels [105].

♢ **Capacity Bounds**: We derive DKI capacity bounds for the slow fading channel with constant $K \geq 1$ and growing size of the target message set $K = 2^{\kappa \log n}$, respectively. We show that for constant K, the proposed lower and upper bounds on R are independent of K, whereas for growing number of target messages, they are functions of the target identification rate κ.

♢ **Technical Novelty**: To obtain the proposed lower bound, the existence of an appropriate sphere packing within the input space, for which the distance between the centers of the spheres does not fall below a certain value, is guaranteed. This packing incorporates the effect of number of target messages as a function of κ. In particular, we consider the packing of hyper spheres inside a larger large hyper sphere, whose radius grows in both the codeword length n and the target identification rate κ, i.e., $\sim n^{\frac{1+\kappa}{4}}$. For derivation of the upper bound, we assume that for given sequences of codes with vanishing error probabilities, a certain minimum distance between the codewords is asserted, where this distance depends on the target identification rate and decreases as K grows.

9.1.2 | Organization

The remainder of this paper is structured as follows. In Section 9.2, system model is explained and the required preliminaries regarding DKI codes are established. Section 9.3 provides the main contributions and results on the K-identification capacity of the slow fading channel. Finally, Section 9.4 of the paper concludes with a summary and directions for future research.

9.2 | System Model and Preliminaries

In this section, we present the adopted system model and establish some preliminaries regarding DKI coding.

9.2.1 | System Model

We consider an identification-focused communication setup, where the decoder seeks to accomplish the following task: Determining whether or not an specific message be-

longs[2] to a set of messages called target message set; see Figure 9.1. We assume that the signal experiences an additive Gaussian noise and slow fading process.

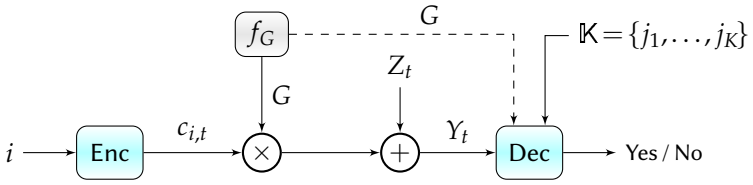


Figure 9.1: End-to-end transmission chain for DKI communication in a wireless communication system modelled as a GSF. The transmitter maps message i onto a codeword $\mathbf{c}_i = (c_{i,1}, \ldots, c_{i,n})$. The receiver is provided with an arbitrary target message set $\mathbb{K} = \{j_1, \ldots, j_K\}$, and given the channel output vector $\mathbf{Y}$, it asks whether the sent message i belong to set of K messages $\{j_1, \ldots, j_K\}$ or not.

To attain this objective, a coded communication between the transmitter and the receiver over n channel uses of a Gaussian channel with slow fading is established[3]. We consider the slow fading channel $\mathcal{G}_{\text{slow}}$ which arises as a channel model in the context of wireless communication [144] where the input-output relation is given by

$$Y_t = Gx_t + Z_t\,, \tag{9.1}$$

where $G_t = G \sim f_G$ is a continuous random variable $\sim f_G(g)$, and the noise sequence $\bar{\mathbf{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{\sigma_Z^2}{n}\right)$ where $\sigma_Z^2 > 0$ is bounded away from zero. We assume that G has finite expectation and variance $\text{var}(G) > 0$. Further, assume that the values of G belong to a set $\mathcal{G}$ where $\gamma \overset{\text{def}}{=} \inf_{G\in\mathcal{G}} |G|$, that is, the set $\mathcal{G}$ has a constant infimum or equivalently, the fading coefficients are bounded away from zero, i.e., $|G_t| > \gamma\,, \forall t \in [\![n]\!]$ with probability 1.

The average power constraint on the codewords is

$$\frac{1}{n}\|\mathbf{x}\|^2 \leq P_{\text{avg}}\,, \tag{9.2}$$

where $P_{\text{avg}} > 0$ constrain the energy of codeword over the entire n channel uses.

9.2.2 | DI Coding For The GSF

The definition of a DKI code for the GSF $\mathcal{G}_{\text{slow}}$ is given below.

[2] We assume that the transmitter does not know which specific K messages the decoder is interested in. This assumption is justified by the fact that otherwise, entire communication setting is specialized to transmission of only one indicator bit between Alice and Bob.

[3] The proposed performance bounds works regardless of whether or not an specific code is used for communication, although proper codes may be required to approach such performance limits.

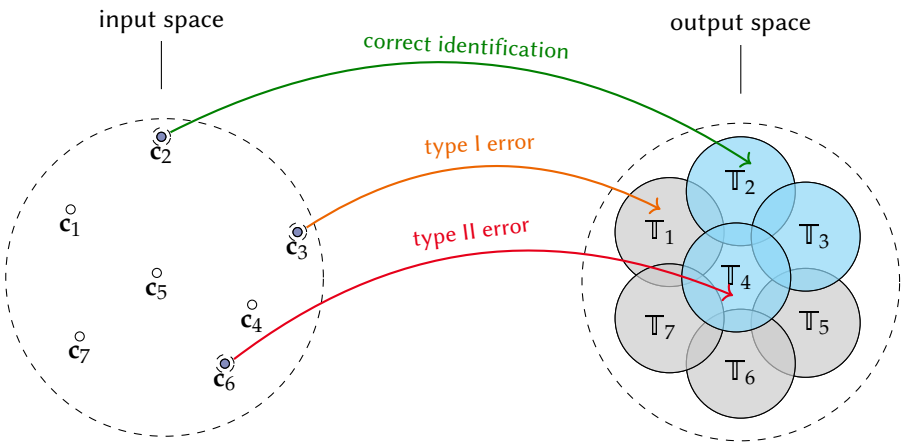


Figure 9.2: Illustration of a deterministic 3-identification setting with target message set $\mathbb{K} = \{2,3,4\}$. In the correct identification scenario, channel output is observed in the union of individual decoder $\mathbb{T}_{j,g}$ where j belongs to the target message set. Type I error occurs if the channel output is detected in the complement of union of individual decoders for which the index of codeword at the lest belongs to. The case where the index of codeword at the left does not match to any of the individual decoders for which the channel output belongs to the their union, is referred to as the type II error.

Definition 9.2.1 (Slow Fading DKI code). *An $(n, M(n,R), K(n,\kappa), e_1, e_2)$ DKI code for a GSF $\mathcal{G}_{\text{slow}}$ under average power constraint of P_{ave}, and for integers $M(n,R)$ and $K(n,\kappa)$, where n and R are the codeword length and coding rate, respectively, with CSI at the decoder is defined as a system $(\mathcal{C}, \mathscr{T}_{\mathbb{K}})$, which consists of a codebook $\mathcal{C} = \{\mathbf{c}_i\}_{i\in[\![M]\!]} \subset \mathbb{R}^n$, such that*

$$\frac{1}{n}\|\mathbf{c}_i\|^2 \leq P_{avg}\,, \tag{9.3}$$

$\forall i \in [\![M]\!]$ and a decoder

$$\mathscr{T}_{\mathbb{K}} = \bigcup_{j\in\mathbb{K}} \mathbb{T}_{j,g}\,, \tag{9.4}$$

where $\mathbb{T}_{j,g} \subset \mathbb{R}^n$, for $j \in [\![M]\!]$, $g \in \mathcal{G}$, and $\mathbb{K} \in \binom{M}{K}$[4]. Given a message $i \in [\![M]\!]$, the encoder transmits $\mathbf{c}_i$, and the decoder's aim is to answer the following question: Was a desired message $j \in \mathbb{K}$ sent or not? There are two types of errors that may occur (see Figure 9.2): Rejection of the true message for $i \in \mathbb{K}$ (type I), or acceptance of a false message for $i \notin \mathbb{K}$ (type II). The corresponding error probabilities of the DKI code $(\mathcal{C}, \mathscr{T}_{\mathbb{K}})$ are given by

$$P_{e,1}(i) = \sup_{g\in\mathcal{G}} \left[\Pr\left(\mathbf{Y} \in \mathscr{T}_{\mathbb{K}}^c \,\middle|\, \mathbf{x} = \mathbf{c}_i\right)\right]_{i\in\mathbb{K}} = \sup_{g\in\mathcal{G}} \left[1 - \int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_i)d\mathbf{y}\right]_{i\in\mathbb{K}} \tag{9.5}$$

[4] We recall that $\binom{M}{K}$ is the family of all subsets of $[\![M]\!]$ with size K and DKI code definition applies to every possible choice of set $\mathbb{K}$ with K arbitrary messages from the original message set $[\![M]\!]$.

$$P_{e,2}(i,\mathbb{K}) = \sup_{g\in\mathcal{G}} \left[\Pr\left(\mathbf{Y} \in \mathscr{T}_{\mathbb{K}} \,\middle|\, \mathbf{x} = \mathbf{c}_i \right) \right]_{i\notin\mathbb{K}} = \sup_{g\in\mathcal{G}} \left[\int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_i) d\mathbf{y} \right]_{i\notin\mathbb{K}} \tag{9.6}$$

where

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{z}) &= f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_i) \\ &= \prod_{t=1}^{n} f_{\mathbf{Z}}(y_t - g c_{i,t}) \\ &= \prod_{t=1}^{n} \frac{1}{(2\pi\sigma_Z^2)^{1/2}} e^{-z_t^2/2\sigma_Z^2} \\ &= \frac{1}{(2\pi\sigma_Z^2)^{n/2}} e^{-\|\mathbf{z}\|^2/2\sigma_Z^2}, \end{aligned} \tag{9.7}$$

(see Figure 9.1) and satisfy the following bounds $P_{e,1}(i) \leq e_1$, $\forall i \in \mathbb{K}$ *and* $P_{e,2}(i,\mathbb{K}) \leq e_2$, $\forall i \notin \mathbb{K}$, *where* $\mathbb{K} \in \binom{M}{K}$ *and every* $e_1, e_2 > 0$.

A rate $R > 0$ *is called achievable if for every* $e_1, e_2 > 0$ *and sufficiently large* n, *there exists an* $(n, M(n,R), K(n,\kappa), e_1, e_2)$ *DKI code. The DKI capacity of the GSF* $\mathcal{G}_{\text{slow}}$ *is defined as the supremum of all achievable rates, and is denoted by* $\mathbb{C}_{DI}(\mathcal{G}_{\text{slow}}, M, K)$.

Remark 9.2.1. *If the fading coefficients can be zero or arbitrarily close to zero, i.e.,* $0 \in cl(\mathcal{G})$, *then it immediately follows that the DKI capacity is zero. To see this, observe that if* $0 \in cl(\mathcal{G})$, *then*

$$\begin{aligned} P_{e,1}(i) + P_{e,2}(i,\mathbb{K}) &= \sup_{g\in\mathcal{G}} \left[1 - \int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_i)\, d\mathbf{y} \right] + \sup_{g\in\mathcal{G}} \left[\int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_i)\, d\mathbf{y} \right] \\ &\geq \left[1 - \int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_i)\, d\mathbf{y} \right]_{\substack{g=0,\\ i\in\mathbb{K}}} + \left[\int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_i)\, d\mathbf{y} \right]_{\substack{g=0,\\ i\notin\mathbb{K}}} \\ &= 1 . \end{aligned} \tag{9.8}$$

9.3 | DKI Capacity of The GSF

In this section, we first present our main results, i.e., lower and upper bounds on the achievable identification rates for the GSF. Subsequently, we provide the detailed proofs of these bounds.

9.3.1 | Main Results

The DKI capacity theorem for GSF $\mathcal{G}_{\text{slow}}$ is stated below.

Theorem 9.3.1. *Consider the GSF $\mathcal{G}_{\text{slow}}$ and assume that the fading coefficients are bounded away from zero, i.e., $0 \notin cl(\mathcal{G})$. Further, assume that the number of target messages scales sub-linearly with codeword length n, i.e., $K(n,\kappa) = 2^{\kappa \log n}$, where $\kappa \in [0,1)$. Then the DKI capacity of $\mathcal{G}_{\text{slow}}$ subject to average power constraint of the form $\|\mathbf{c}_i\|^2 \leq nP_{ave}$ and a codebook of super-exponential scale, i.e., $M(n,R) = 2^{(n\log n)R}$, is bounded by*

$$\frac{1-\kappa}{4} \leq \mathsf{C}_{DI}(\mathcal{G}_{\text{slow}}, M, K) \leq 1+\kappa\,. \tag{9.9}$$

Proof. The proof of Theorem 9.3.1 consists of two parts, namely the achievability and the converse proofs, which are provided in Sections 9.3.2 and 9.3.3, respectively. □

Remark 9.3.1. *The result in Theorem 9.3.1 comprises the following three special cases in terms of K:*

- ***Unit $K = 1$**: This cases accounts for a standard identification setup ($\kappa = 0$), that is, when the target message set is a degenerate case $\mathbb{K} = \{i\}_{i\in[\![M]\!]}$, i.e., $K = |\mathbb{K}| = 1$. Therefore, the identification setup as studied in [23] can be regarded as a special case of K-identification. This result is known in the identification literature [23, 99, 105, 106].*
- ***Constant $K > 1$**: Constant $K > 1$ implies $\kappa \to 0$ as $n \to \infty$. Surprisingly, our capacity result in Theorem 9.3.1 reveals that the bounds for the GSF with constant finite $K > 1$ are in fact identical to those for the memoryless GSF given in [99, 105, 106].*
- ***Growing K**: Our capacity results reveal that reliable identification is possible even when K scales with the codeword length as $\sim 2^{\kappa \log n}$ for $\kappa \in [0,1)$. Moreover, the impact of target identificaiton rate κ is reflected in the capacity lower and upper bounds in (9.9), where the bounds respectively decrease and increase in κ.*

9.3.2 | Achievability

The achievability proof consists of the following two main steps.

- **Step 1:** We propose a codebook construction and derive an analytical lower bound on the corresponding codebook size using inequalities for sphere packing density.
- **Step 2:** To prove that this codebook leads to an achievable rate, we propose a decoder and show that the corresponding type I and type II error rates vanished as $n \to \infty$.

9.3.2.1 | Normalization

Since the decoder can normalize the output symbols by $\sqrt{n}$, we have an equivalent input-output relation,

$$\bar{Y}_t = G\bar{x}_t + \bar{Z}_t \,, \tag{9.10}$$

where $G_t = G \sim f_G$, and the noise sequence $\bar{\mathbf{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{\sigma_Z^2}{n}\right)$, with an input power constraint

$$\|\bar{\mathbf{x}}\| \leq \sqrt{A} \,, \tag{9.11}$$

where $A \overset{\text{def}}{=} P_{\text{ave}}$ and

$$\bar{\mathbf{x}} = \frac{1}{\sqrt{n}}\mathbf{x} \quad , \quad \bar{\mathbf{Z}} = \frac{1}{\sqrt{n}}\mathbf{Z} \quad , \quad \bar{\mathbf{Y}} = \frac{1}{\sqrt{n}}\mathbf{Y} \,. \tag{9.12}$$

9.3.2.2 | Codebook Construction

We use a packing arrangement of non-overlapping hyper spheres of radius $r_0 = \sqrt{\theta_n}$ in a large hyper sphere with radius $\sqrt{A} - \sqrt{\theta_n}$, with

$$\theta_n = \frac{A\sqrt{K}}{n^{\frac{1}{2}(1-b)}} = \frac{A}{n^{\frac{1}{2}(1-(b+\kappa))}} \,, \tag{9.13}$$

where $0 < b < 1$ is an arbitrarily small constant[5], and $\kappa \in [0, 1)$.

Let $\mathscr{S}$ denote a sphere packing, i.e., an arrangement of M non-overlapping spheres $\mathcal{S}_{\bar{\mathbf{c}}_i}(n, r_0)$, $i \in [\![M]\!]$, that are packed inside the larger sphere $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})$ with radius $\sqrt{A} - \sqrt{\theta_n}$. As opposed to standard sphere packing in coding techniques [138], the spheres are not necessarily entirely contained within the larger sphere. That is, we only require that the centers of the spheres are inside $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})$ and are disjoint from each other and have a non-empty intersection with $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})$. The packing density $\Delta_n(\mathscr{S})$ is defined as the ratio of the saturated packing volume to the larger sphere's volume $\text{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})\right)$, i.e.,

$$\Delta_n(\mathscr{S}) \triangleq \frac{\text{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n}) \cap \bigcup_{i=1}^{M} \mathcal{S}_{\bar{\mathbf{c}}_i}(n, r_0)\right)}{\text{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})\right)} \,. \tag{9.14}$$

[5] we recall that our achievability proof works for any $b \in (0, 1)$; however, arbitrarily small values of b are of interest since they result in the tightest lower bound.

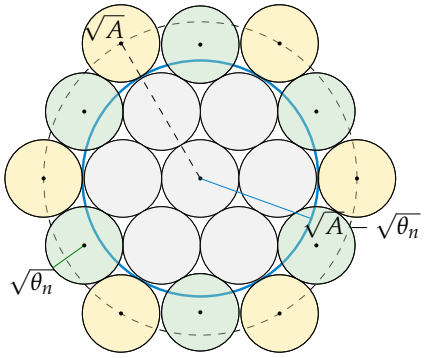


Figure 9.3: Illustration of a saturated sphere packing inside a hyper sphere, where small spheres of radius $r_0 = \sqrt{\theta_n}$ cover a larger hyper sphere. The small spheres are disjoint from each other and have a non-empty intersection with the large sphere. Some of the small spheres, colored in green, are not entirely contained within the larger sphere, and yet they are considered to be a part of the packing arrangement, since their centers fulfill the power constraint in 9.11. Yellow colored spheres whose centers exactly lies on the circle with radius A do not contribute to the packing. As we assign a codeword to each sphere center (white and green), the 2-norm of a codeword is bounded by $\sqrt{A}$ as required.

Sphere packing $\mathscr{S}$ is called *saturated* if no spheres can be added to the arrangement without overlap.

In particular, we use a packing argument that has a similar flavor as that observed in the Minkowski–Hlawka theorem for saturated packing [138]. Specifically, consider a saturated packing arrangement of

$$\bigcup_{i=1}^{M(n,R)} \mathcal{S}_{\mathbf{c}_i}(n, \sqrt{\theta_n}) \tag{9.15}$$

spheres with radius $r_0 = \sqrt{\theta_n}$ embedded within sphere $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})$. Then, for such an arrangement, we have the following lower [118, Lem. 2.1] and upper bounds [138, Eq. 45] on the packing density

$$2^{-n} \leq \Delta_n(\mathscr{S}) \leq 2^{-0.599n} \,. \tag{9.16}$$

In our subsequent analysis, we use the above lower bound which can be proved as follows: For the saturated packing arrangement given in (9.15), there cannot be a point in the larger sphere $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})$ with a distance of more than $2r_0$ from all sphere centers. Otherwise, a new sphere could be added which contradicts the assumption that the union of $M(n, R)$ spheres with radius $\sqrt{\theta_n}$ is saturated. Now, if we double the radius of each sphere, the spheres with radius $2r_0$ cover thoroughly the entire volume of $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})$, that is, each point inside the large hyper sphere $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})$

belongs to at least one of the small spheres. In general, the volume of a hyper sphere of radius r is given by [138, Eq. (16)]

$$\mathrm{Vol}\left(\mathcal{S}_{\mathbf{x}}(n,r)\right) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r^n \,. \tag{9.17}$$

Hence, if the radius of the small spheres is doubled, the volume of

$$\bigcup_{i=1}^{M(n,R)} \mathcal{S}_{\mathbf{c}_i}(n, \sqrt{\theta_n})$$

is increased by 2^n. Since the spheres with radius $2r_0$ cover $\mathcal{S}_{\mathbf{0}}(n, \sqrt{A} - \sqrt{\theta_n})$, it follows that the original r_0-radius packing has a density of at least 2^{-n} [6]. We assign a codeword to the center $\mathbf{c}_i$ of each small sphere. The codewords satisfy the input constraint as

$$\|\bar{\mathbf{c}}_i\| \leq \sqrt{A} \,. \tag{9.18}$$

Since the volume of each sphere is equal to $\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n,r_0))$ and the centers of all spheres lie inside the sphere, the total number of spheres is bounded from below by

$$\begin{aligned} M &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{M} \mathcal{S}_{\bar{\mathbf{c}}_i}(n,r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n,r_0))} \\ &\geq \frac{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n,\sqrt{A}-\sqrt{\theta_n}) \cap \bigcup_{i=1}^{M} \mathcal{S}_{\bar{\mathbf{c}}_i}(n,r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\bar{\mathbf{c}}_1}(n,r_0))} \\ &= \frac{\Delta_n(\mathscr{S}) \cdot \mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n,\sqrt{A}-\sqrt{\theta_n})\right)}{\mathrm{Vol}(\mathcal{S}_{\bar{\mathbf{c}}_1}(n,r_0))} \\ &\geq 2^{-n} \cdot \frac{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n,\sqrt{A}-\sqrt{\theta_n})\right)}{\mathrm{Vol}(\mathcal{S}_{\bar{\mathbf{c}}_1}(n,r_0))} \,, \end{aligned} \tag{9.19}$$

where the first inequality holds by (9.14) and the second inequality holds by (9.16). The above bound can be further simplified as follows

$$\begin{aligned} \log M &\overset{(a)}{\geq} \log\left(\frac{\sqrt{A}-\sqrt{\theta_n}}{r_0}\right)^n - n \\ &\overset{(b)}{=} n \log\left(\frac{\sqrt{A}-\sqrt{\theta_n}}{\sqrt{\theta_n}}\right) - n \end{aligned}$$

[6] We note that the proposed proof of the lower bound in (9.16) is non-constructive in the sense that, while the existence of the respective saturated packing is proved, no systematic construction method is provided.

$$= n \log \left(\sqrt{\frac{A}{\theta_n}} - 1 \right) - n$$
$$\overset{(c)}{\geq} \frac{1}{2} n \log \left(\frac{A}{\theta_n} \right) - 2n \,, \tag{9.20}$$

where (a) exploits (9.17), (b) follows from $r_0 = \sqrt{\theta_n}$, and (c) holds by $\log(t-1) \geq \log t - 1 \,, \forall t \geq 2$. Therefore, for $\theta_n = A/n^{\frac{1}{2}(1-(b+\kappa))}$, we obtain

$$\log M \geq \frac{1}{2} n \log n^{\frac{1}{2}(1-(b+\kappa))} - 2n$$
$$= \left(\frac{1-(b+\kappa)}{4} \right) n \log n - 2n \,, \tag{9.21}$$

where the dominant term is of order $n \log n$. Hence, for obtaining a finite value for the lower bound of the rate, R, (9.21) induces the scaling law of M to be $2^{(n \log n)R}$. Therefore, we obtain

$$R \geq \frac{1}{\log n} \left[\left(\frac{1-(b+\kappa)}{4} \right) \log n - 2 \right] \,, \tag{9.22}$$

which tends to $\frac{1-\kappa}{4}$ when $n \to \infty$ and $b \to 0$.

9.3.2.3 | Encoding

Given message $i \in [\![M]\!]$, transmit $\bar{\mathbf{x}} = \bar{\mathbf{c}}_i$.

9.3.2.4 | Decoding

Let

$$\tau_n = \frac{\gamma^2 \theta_n}{3} = \frac{A\gamma^2}{3n^{\frac{1}{2}(1-(b+\kappa))}} \,, \tag{9.23}$$

where $0 < b < 1$ is an arbitrarily small constant, $0 < c < 2$ is a constant, $\kappa \in [0,1)$, and γ is the infimum value of all fading coefficients g.

To identify whether message $j \in \mathcal{M}$ was sent, given the fading coefficient g, the decoder checks whether the channel output $\bar{\mathbf{y}}$ belongs to the following decoding set:

$$\mathscr{T}_{\mathbb{K}} = \bigcup_{j \in \mathbb{K}} \mathbb{T}_{j,g} \,, \tag{9.24}$$

where

$$\mathbb{T}_{j,g} = \left\{ \bar{\mathbf{y}} \in \mathbb{R}^n \,:\, \sum_{t=1}^{n} (\bar{y}_t - g\bar{c}_{j,t})^2 \leq \sigma_Z^2 + \tau_n \right\} \,. \tag{9.25}$$

is referred to as the individual decoding territory evaluated for observation vector $\mathbf{y}$ and codeword $\mathbf{c}_j$.

9.3.2.5 | Error Analysis

Fix $e_1, e_2 > 0$ and let $\zeta_0, \zeta_1 > 0$ be arbitrarily small constants. Before we proceed, for the sake of brevity of analysis, we introduce the following conventions:

- Let $Y_t(.|i,g)$ denote the channel output at time t given that $\bar{\mathbf{x}} = \bar{\mathbf{c}}_i$ and $G = g$.
- $\mathbf{Y}(.|i,g) = (Y_1(.|i,g), \ldots, Y_n(.|i,g))$.

Consider the type I errors, i.e., the transmitter sends $\bar{\mathbf{c}}_i$, yet $\mathbf{Y}(.|i,g) \notin \mathbb{T}_{\mathbb{K},g}$. For every $i \in [\![M]\!]$, the type I error probability is given by

$$P_{e,1}(i) = \sup_{g \in \mathcal{G}} \left[P_{e,1}\left(i \,|g\right) \right] , \tag{9.26}$$

where

$$\begin{aligned}
P_{e,1}\left(i \,|g\right) &= \Pr\left(\bar{\mathbf{Y}}(.|i,g) \in \mathbb{T}^c_{\mathbb{K},g}\right) \\
&= \Pr\left(\bar{\mathbf{Y}}(.|i,g) \in \left(\bigcup_{i \in \mathbb{K}} \mathbb{T}_{i,g}\right)^c\right) \\
&\overset{(a)}{=} \Pr\left(\bar{\mathbf{Y}}(.|i,g) \in \bigcap_{i \in \mathbb{K}} \mathbb{T}^c_{i,g}\right) \\
&\overset{(b)}{\leq} \Pr\left(\bar{\mathbf{Y}}(.|i,g) \in \mathbb{T}^c_{i,g}\right) \\
&\overset{(c)}{=} \Pr\left(\sum_{t=1}^{n} (\bar{Y}_t(.|i,g) - G\bar{c}_{i,t})^2 > \sigma_Z^2 + \tau_n\right) \\
&\overset{(d)}{=} \Pr\left(\sum_{t=1}^{n} \bar{Z}_t{}^2 > \sigma_Z^2 + \tau_n\right) ,
\end{aligned} \tag{9.27}$$

where (a) follows by *De Morgan*'s law for finite number of unions, i.e., $\left(\bigcup_{i \in \mathbb{K}} \mathbb{T}_{i,g}\right)^c = \bigcap_{i \in \mathbb{K}} \mathbb{T}^c_{i,g}$, (b) holds since $\bigcap_{i \in \mathbb{K}} \mathbb{T}^c_{i,g} \subset \mathbb{T}_{i,g}$, (c) follows by definition of the individual decoding territory in (9.25), and (d) holds since the fading coefficient G and the noise vector $\bar{\mathbf{Z}}$ are statistically independent.

Now, in order to bound $P_{e,1}\left(i \,|g\right)$, we apply Chebyshev's inequality, namely

$$P_{e,1}\left(i \,|g\right) \leq \Pr\left(\sum_{t=1}^{n} \bar{Z}_t{}^2 - \sigma_Z^2 > \tau_n\right)$$

$$\begin{aligned}
&\overset{(a)}{\leq} \frac{3\sigma_Z^4}{n\tau_n^2} \\
&\overset{(b)}{=} \frac{27\sigma_Z^4}{A^2\gamma^4 n^{\kappa+b}} \\
&\leq e_1 \ , \qquad (9.28)
\end{aligned}$$

where (a) holds since the fourth moment of a Gaussian variable $V \sim \mathcal{N}(0,\sigma_V^2)$ is $\mathbb{E}[V^4] = 3\sigma_V^4$ and (b) follows from (9.23). Hence, $P_{e,1}\left(i\,|g\right) \leq e_1$, $\forall g \in \mathcal{G}$ holds for sufficiently large n and arbitrarily small $e_1 > 0$. Thereby, the type I error probability satisfies $P_{e,1}\left(i\right) \leq e_1$; see (9.26).

Next, we address type II errors, i.e., when $\bar{\mathbf{Y}}(.|i,g) \in \mathbb{T}_{\mathbb{K},g}$ while the transmitter sent $\bar{\mathbf{c}}_i$ with $i \notin \mathbb{K}$. Then, for every $\mathbb{K} \in \binom{M}{K}$, where $i \notin \mathbb{K}$, the type II error probability is given by

$$P_{e,2}(i,\mathbb{K}) = \sup_{g\in\mathcal{G}} \left[P_{e,2}\left(i,\mathbb{K}\,|g\,\right)\right] \ , \qquad (9.29)$$

where

$$\begin{aligned}
P_{e,2}\left(i,\mathbb{K}\,|g\right) &= \Pr\left(\bar{\mathbf{Y}}(.|i,g) \in \mathbb{T}_{\mathbb{K},g}\right) \\
&= \Pr\left(\bar{\mathbf{Y}}(.|i,g) \in \left(\bigcup_{j\in\mathbb{K}} \mathbb{T}_{j,g}\right)\right) \\
&\equiv \Pr\left(\bigcup_{j\in\mathbb{K}} \left\{\sum_{t=1}^{n}\left(\bar{Y}_t(.|i,g) - G\bar{c}_{j,t}\right)^2 \leq \sigma_Z^2 + \tau_n\right\}\right) \\
&\overset{(a)}{=} \Pr\left(\bigcup_{j\in\mathbb{K}} \left\{\sum_{t=1}^{n}\left(g\left(\bar{c}_{i,t} - \bar{c}_{j,t}\right) + \bar{Z}_t\right)^2 \leq \sigma_Z^2 + \tau_n\right\}\right) \\
&\overset{(b)}{\leq} \sum_{j\in\mathbb{K}} \Pr\left(\sum_{t=1}^{n}\left(g\left(\bar{c}_{i,t} - \bar{c}_{j,t}\right) + \bar{Z}_t\right)^2 \leq \sigma_Z^2 + \tau_n\right) , \qquad (9.30)
\end{aligned}$$

where (a) hold since the fading coefficient G and the noise vector $\bar{\mathbf{Z}}$ are statistically independent and (b) follows by the union bound, i.e., the probability of union of events is upper bounded by sum of probability of the individual events.

In order to bound (9.30), we divide into two cases. First, consider $g \in \mathcal{G}$ such that $\|g(\bar{\mathbf{c}}_i - \bar{\mathbf{c}}_j)\| > 2\sqrt{\sigma_Z^2 + \tau_n}$. Therefore, by the reverse triangle inequality, $\|\mathbf{a} - \mathbf{b}\| \geq$

$\big|\|\mathbf{a}\| - \|\mathbf{b}\|\big|$, we have

$$\sqrt{\sum_{t=1}^{n}\left(\left(g\left(\bar{c}_{i,t}-\bar{c}_{j,t}\right)\right)+\bar{Z}_t\right)^2} \geq \left\|g\left(\bar{\mathbf{c}}_i-\bar{\mathbf{c}}_j\right)\right\| - \|\bar{\mathbf{Z}}\| \geq 2\sqrt{\sigma_Z^2+\tau_n} - \|\bar{\mathbf{Z}}\| \ . \tag{9.31}$$

Hence, for every g such that $\|g\left(\bar{\mathbf{c}}_\mathbf{i}-\bar{\mathbf{c}}_\mathbf{j}\right)\| > 2\sqrt{\sigma_Z^2+\tau_n}$, we can bound the type II error probability by

$$\begin{aligned} P_{e,2}\left(i,\mathbb{K}\,\middle|\,g\right) &\leq \sum_{j\in\mathbb{K}} \Pr\left(\|\bar{\mathbf{Z}}\| \geq \sqrt{\sigma_Z^2+\tau_n}\right) \\ &= \sum_{j\in\mathbb{K}} \Pr\left(\sum_{t=1}^{n} \bar{Z}_t{}^2 > \sigma_Z^2+\tau_n\right) \\ &\leq \frac{3K\sigma_Z^4}{n\tau_n^2} \\ &= \frac{27\sigma_Z^4}{A^2\gamma^4 n^b} \\ &\leq e_2 \ , \end{aligned} \tag{9.32}$$

where (a) follows from applying Chebyshev's inequality and since the fourth moment of a Gaussian variable $V\sim\mathcal{N}(0,\sigma_V^2)$ is $\mathbb{E}[V^4]=3\sigma_V^4$ and (b) follows from (9.23). Hence, $P_{e,1}\left(i\,|g\right) \leq e_1$, $\forall g \in \mathcal{G}$ holds for sufficiently large n and arbitrarily small $e_1>0$. Thereby, the type I error probability satisfies $P_{e,2}\left(i,\mathbb{K}\right)\leq e_2$; see (9.26).

Now, we focus on the second case, i.e., when

$$\left\|g\left(\bar{\mathbf{c}}_i-\bar{\mathbf{c}}_j\right)\right\| \leq 2\sqrt{\sigma_Z^2+\tau_n} \ . \tag{9.33}$$

Observe that for every given $g\in\mathcal{G}$,

$$\sum_{t=1}^{n}(g(\bar{c}_{i,t}-\bar{c}_{j,t})+\bar{Z}_t)^2 = \sum_{t=1}^{n} g^2(\bar{c}_{i,t}-\bar{c}_{j,t})^2 + \sum_{t=1}^{n}\bar{Z}_t^2 + 2\sum_{t=1}^{n} g(\bar{c}_{i,t}-\bar{c}_{j,t})Z_t \ . \tag{9.34}$$

Then define the event

$$\mathcal{E}_0(\mathbf{Z}|g) = \left\{\mathbf{Z}\in\mathbb{R}^n \ : \ \left|\sum_{t=1}^{n} g(\bar{c}_{i,t}-\bar{c}_{j,t})\bar{Z}_t\right| > \frac{\tau_n}{2}\right\} \ , \tag{9.35}$$

Now, in order to bound $\Pr(\mathcal{E}_0(\mathbf{Z}|g))$, we apply Chebyshev's inequality, namely

$$\Pr\left(\mathcal{E}_0(\mathbf{Z}|g)\right) \leq \frac{\mathrm{Var}\left[\sum_{t=1}^{n} g(\bar{c}_{i,t}-\bar{c}_{j,t})\bar{Z}_t\right]}{\left(\tau_n/2\right)^2}$$

$$\begin{aligned}
&\overset{(a)}{=} \frac{4\sum_{t=1}^{n} g^2(\bar{c}_{i,t}-\bar{c}_{j,t})^2\mathbb{E}[\bar{Z}_t^2]}{\tau_n^2} \\
&\overset{(b)}{=} \frac{4\sigma_Z^2\|g(\bar{\mathbf{c}}_i-\bar{\mathbf{c}}_j)\|^2}{n\tau_n^2} \\
&\overset{(c)}{=} \frac{16\sigma_Z^2\left(\sigma_Z^2+\tau_n\right)}{n\tau_n^2} \\
&= \frac{144\sigma_Z^2\left(\sigma_Z^2+\tau_n\right)}{A^2\gamma^4 n^{\kappa+b}} \\
&\overset{\text{def}}{=} \zeta_0 \,, \qquad (9.36)
\end{aligned}$$

where (a) and (b) holds since the noise sequence $\bar{\mathbf{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{\sigma_Z^2}{n}\right)$, that is, $\text{Var}[\bar{Z}_t] = \mathbb{E}[\bar{Z}_t^2] - \mathbb{E}^2[\bar{Z}_t] = \frac{\sigma_Z^2}{n}$, and (c) follows from (9.33). Observe that given the complementary event $\mathcal{E}_0^c(\mathbf{Z}|g)$, we have

$$2\sum_{t=1}^{n} g\left(\bar{c}_{i,t}-\bar{c}_{j,t}\right)\bar{Z}_t \geq -\tau_n \,. \qquad (9.37)$$

Therefore, the event $\mathcal{E}_0^c(\mathbf{Z}|g)$, the type II error event in (9.30), and the identity in (9.33) together imply that the following event occurs,

$$\mathcal{E}_1(\mathbf{Z}|g) = \left\{\mathbf{Z}\in\mathbb{R}^n \;:\; \sum_{t=1}^{n} g^2(\bar{c}_{i,t}-\bar{c}_{j,t})^2 + \sum_{t=1}^{n}\bar{Z}_t^2 \leq \sigma_Z^2 + 2\tau_n\right\} \,. \qquad (9.38)$$

Now lets define

$$\mathcal{H}_{i,j}(\mathbf{Z}|g) = \left\{\mathbf{Z}\in\mathbb{R}^n \;:\; \sum_{t=1}^{n}(g(\bar{c}_{i,t}-\bar{c}_{j,t})+\bar{Z}_t)^2 \leq \sigma_Z^2+\tau_n\right\} \,. \qquad (9.39)$$

Therefore, applying the law of total probability to (9.40), we have

$$\begin{aligned}
P_{e,2}(i,\mathbb{K}|\,g) &= \sum_{j\in\mathbb{K}}\left[\Pr\left(\mathcal{H}_{i,j}(\mathbf{Z}|g)\cap\mathcal{E}_0(\mathbf{Z}|g)\right)+\Pr\left(\mathcal{H}_{i,j}(\mathbf{Z}|g)\cap\mathcal{E}_0^c(\mathbf{Z}|g)\right)\right] \\
&\leq \sum_{j\in\mathbb{K}}\left[\Pr(\mathcal{E}_0(\mathbf{Z}|g))+\Pr\left(\mathcal{E}_1(\mathbf{Z}|g)\right)\right] \\
&\leq K\left[\zeta_0+\Pr\left(\mathcal{E}_1(\mathbf{Z}|g)\right)\right] \,, \qquad (9.40)
\end{aligned}$$

where the last inequality holds by (9.36).

We now proceed with bounding $\Pr(\mathcal{E}_1(\mathbf{Z}|g))$ as follows. Based on the codebook construction, each codeword is surrounded by a sphere of radius $\sqrt{\theta_n}$, that is

$$\left\|\bar{\mathbf{c}}_i-\bar{\mathbf{c}}_j\right\| \geq \sqrt{\theta_n} \,. \qquad (9.41)$$

which implies

$$g^2 \left\| \bar{\mathbf{c}}_i - \bar{\mathbf{c}}_j \right\|^2 \geq \gamma^2 \theta_n \,, \tag{9.42}$$

where γ is the infimum value in $\mathcal{G}$. Thus, we can establish the following upper bound for event $\mathcal{E}_1(\mathbf{Z}|g)$:

$$\begin{aligned}
\Pr(\mathcal{E}_1(\mathbf{Z}|g)) &\leq \Pr\left(\left\| \bar{\mathbf{Z}} \right\|^2 \leq \sigma_Z^2 + 2\tau_n - \gamma^2 \theta_n \right) \\
&= \Pr\left(\left\| \bar{\mathbf{Z}} \right\|^2 - \sigma_Z^2 \leq -\tau_n \right) \\
&= \Pr\left(\sum_{t=1}^{n} \bar{Z}_t^2 - \sigma_Z^2 \leq -\tau_n \right) \\
&\overset{(a)}{\leq} \frac{\sum_{t=1}^{n} \mathrm{Var}[\bar{Z}_t^2]}{\tau_n^2} \\
&\overset{(b)}{\leq} \frac{\sum_{t=1}^{n} \mathbb{E}[\bar{Z}_t^4]}{\tau_n^2} \\
&= \frac{3n \left(\frac{\sigma_Z^2}{n} \right)^2}{\tau_n^2} \\
&= \frac{3\sigma_Z^4}{n\tau_n^2} \\
&\overset{(c)}{=} \frac{27\sigma_Z^4}{A^2 \gamma^4 n^{\kappa+b}} \\
&\overset{\text{def}}{=} \zeta_1 \,,
\end{aligned} \tag{9.43}$$

where (a) follows from applying Chebyshev's inequality, (b) holds since the fourth moment of a Gaussian variable $V \sim \mathcal{N}(0, \sigma_V^2)$ is $\mathbb{E}[V^4] = 3\sigma_V^4$ and (c) follows from (9.23) and (9.36). Therefore, we can proceed to bound the rightmost in (9.40) as follows

$$\begin{aligned}
P_{e,2}(i, \mathbb{K}) &\leq K \left[\Pr(\mathcal{E}_0(\mathbf{Z}|g)) + \Pr(\mathcal{E}_1(\mathbf{Z}|g)) \right] \\
&\leq K \left[\zeta_0 + \zeta_1 \right] \\
&= \frac{144 K \sigma_Z^2 \left(\sigma_Z^2 + \tau_n \right)}{A^2 \gamma^4 n^{\kappa+b}} + \frac{27 K \sigma_Z^4}{A^2 \gamma^4 n^{\kappa+b}} \\
&= \frac{144 \sigma_Z^2 \left(\sigma_Z^2 + \tau_n \right) + 27 \sigma_Z^4}{A^2 \gamma^4 n^b} \\
&\leq e_2 \,,
\end{aligned} \tag{9.44}$$

hence, $P_{e,2}\left(i,\mathbb{K}\,|g\right) \leq e_2$, $\forall g \in \mathcal{G}$ holds for sufficiently large n and arbitrarily small $e_2 > 0$. Thereby, the type II error probability satisfies $P_{e,2}\left(i,\mathbb{K}\right) \leq e_2$; see (9.29).

We have thus shown that for every $e_1, e_2 > 0$ and sufficiently large n, there exists an $(n, M(n,R), K(n,\kappa), e_1, e_2)$ code.

9.3.3 | Converse Proof

The converse proof consists of the following two main steps.

- **Step 1:** We show in Lemma 9.3.1 that for any achievable rate (for which the type I and type II error probabilities vanish as $n \to \infty$), the distance between every pair of codeword should be at least larger than a threshold.
- **Step 2:** Employing the Lemma 9.3.1, we derive an upper bound on the codebook size of achievable DKI codes.

We start with the following lemma which establish a lower bound on the Euclidean norm of two different codewords' difference.

Lemma 9.3.1. *Suppose that R is an achievable rate for the GSF $\mathcal{G}_{\text{slow}}$ and let $b > 0$ be an arbitrarily small constant that does not depend on codeword length n. Consider a sequence of $(n, M(n,R), K(n,\kappa), e_1^{(n)}, e_2^{(n)})$ codes $(\mathcal{C}^{(n)}, \mathcal{T}^{(n)})$ such that $e_1^{(n)}$ and $e_2^{(n)}$ tend to zero as $n \to \infty$. Then there exists $n_0(b)$, such that for all $n > n_0(b)$, every pair of codewords in the codebook $\mathcal{C}^{(n)}$ satisfies the following property.*

For every pair of codewords, $\mathbf{c}_{i_1}$ and $\mathbf{c}_{i_2}$,

$$\left\|\mathbf{c}_{i_1} - \mathbf{c}_{i_2}\right\| \geq 2\sqrt{n\epsilon'_n}\,, \tag{9.45a}$$

for all $i_1, i_2 \in [\![M]\!]$, such that $i_1 \neq i_2$, with

$$\epsilon'_n = \frac{A}{n^{2(1+\kappa+b)}}\,, \tag{9.45b}$$

The proof is given in the following.

Proof. In the following, we provide the proof of Lemma 9.3.1. The method of proof is by contradiction, namely, we assume that the condition given in (9.45a) is violated and then we show that this leads to a contradiction, namely, sum of the type I and type II error probabilities converge to one, i.e., $\lim_{n\to\infty}\left[P_{e,1}(i_1) + P_{e,2}(i_2,\mathbb{K})\right] = 1$. Fix e_1 and

e_2. Let $\zeta, \eta, \mu, \pi > 0$ be arbitrarily small constants. Assume to the contrary that there exist two messages i_1 and i_2, where $i_1 \neq i_2$, such that

$$\left\| \mathbf{c}_{i_1} - \mathbf{c}_{i_2} \right\| < 2\sqrt{n\epsilon'_n} = \alpha_n \,, \tag{9.46}$$

where

$$\alpha_n \equiv \frac{2\sqrt{A}}{n^{\frac{1}{2}(1+2(\kappa+b)}} \,. \tag{9.47}$$

Now let us define the following subsets

$$\mathbb{P}_{i_1,i_2} = \left\{ \mathbf{y} \in \mathbb{T}_{i_1,\mathbf{g}} \,:\, \|\mathbf{y} - g\mathbf{c}_{i_2}\| \leq \sqrt{n\left(\sigma_Z^2 + \zeta\right)} \right\} \,, \tag{9.48}$$

$$\mathbb{Q}_{i_1,i_2} = \left\{ \mathbf{y} \in \mathbb{Y}^n \,:\, \|\mathbf{y} - g\mathbf{c}_{i_2}\| \leq \sqrt{n\left(\sigma_Z^2 + \zeta\right)} \right\} \,. \tag{9.49}$$

Then, observe that

$$\begin{aligned} &P_{e,1}(i_1) + P_{e,2}(i_2, \mathbb{K}) \\ &= \sup_{g\in\mathcal{G}} \left[1 - \int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\, d\mathbf{y} \right]_{i_1 \in \mathbb{K}} + \sup_{g\in\mathcal{G}} \left[\int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_2})\, d\mathbf{y} \right]_{i_2 \notin \mathbb{K}} \,. \end{aligned} \tag{9.50}$$

Now consider the first integral in (9.50) where for every $g \in \mathcal{G}$ we have,

$$\begin{aligned} &\int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\, d\mathbf{y} \\ &\overset{(a)}{\leq} \int_{\bigcup_{i_1\in\mathbb{K}} \mathbb{T}_{i_1,g}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\, d\mathbf{y} \\ &\overset{(a)}{=} \int_{\left(\bigcup_{i_1\in\mathbb{K}} \mathbb{T}_{i_1,g}\right)\cap\mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\; d\mathbf{y} + \int_{\left(\bigcup_{i_1\in\mathbb{K}} \mathbb{T}_{i_1,g}\right)\cap\mathbb{P}^c_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\; d\mathbf{y} \\ &\overset{(b)}{\leq} \int_{\bigcup_{i_1\in\mathbb{K}} \left(\mathbb{T}_{i_1,g}\cap\mathbb{P}_{i_1,i_2}\right)} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\; d\mathbf{y} + \int_{\bigcup_{i_1\in\mathbb{K}} \left(\mathbb{T}_{i_1,g}\cap\mathbb{P}^c_{i_1,i_2}\right)} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\; d\mathbf{y} \\ &\overset{(c)}{\leq} \int_{\bigcup_{i_1\in\mathbb{K}} \mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\; d\mathbf{y} + \int_{\bigcup_{i_1\in\mathbb{K}} \mathbb{Q}^c_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1}) d\mathbf{y} \,, \end{aligned} \tag{9.51}$$

where (a) holds by the union bound, (b) follows by the following

$$\left(\bigcup_{i_1\in\mathbb{K}} \mathbb{T}_{i_1,g}\right) \cap \mathbb{P}_{i_1,i_2} \subset \bigcup_{i_1\in\mathbb{K}} \left(\mathbb{T}_{i_1,g} \cap \mathbb{P}_{i_1,i_2}\right) \,, \tag{9.52a}$$

and

$$\left(\bigcup_{i_1\in\mathbb{K}}\mathbb{T}_{i_1,g}\right)\cap\mathbb{P}^c_{i_1,i_2}\subset\bigcup_{i_1\in\mathbb{K}}\left(\mathbb{T}_{i_1,g}\cap\mathbb{P}^c_{i_1,i_2}\right)\,, \tag{9.52b}$$

and (c) holds since

$$\mathbb{Q}^c_{i_1,i_2}\supset\mathbb{T}_{i_1,g}\cap\mathbb{P}^c_{i_1,g}\,. \tag{9.53}$$

Consider the second integral in (9.51). Then, by the triangle inequality,

$$\begin{aligned}\|\mathbf{y}-g\mathbf{c}_{i,1}\|&\geq\|\mathbf{y}-g\mathbf{c}_{i,2}\|-\|g(\mathbf{c}_{i,1}-\mathbf{c}_{i,2})\|\\&=\|\mathbf{y}-g\mathbf{c}_{i,2}\|-g\|\mathbf{c}_{i,1}-\mathbf{c}_{i,2}\|\\&>\sqrt{n(\sigma_Z^2+\zeta)}-g\|\mathbf{c}_{i,1}-\mathbf{c}_{i,2}\|\\&\geq\sqrt{n(\sigma_Z^2+\zeta)}-g\alpha_n\,.\end{aligned} \tag{9.54}$$

For sufficiently large n, this implies the following subset

$$\mathbb{R}^c_{i_1,i_2}=\left\{y^n\in\mathbb{Y}^n\,:\|\mathbf{y}-g\mathbf{c}_{i,1}\|>\sqrt{n\left(\sigma_Z^2+\eta\right)}\right\}\,, \tag{9.55}$$

for $\eta<\frac{\zeta}{2}$. That is,

$$\left\{\mathbf{y}\,:\|\mathbf{y}-g\mathbf{c}_{i,2}\|\geq\sqrt{n\left(\sigma_Z^2+\zeta\right)}\right\}\quad\overset{\text{implies}}{\longrightarrow}\quad\left\{\mathbf{y}\,:\|\mathbf{y}-g\mathbf{c}_{i,1}\|\geq\sqrt{n\left(\sigma_Z^2+\eta\right)}\right\}\,. \tag{9.56}$$

Thus we deduce that

$$\mathbb{R}^c_{i_1,i_2}\supset\mathbb{Q}^c_{i_1,i_2}\,, \tag{9.57}$$

Hence, the second integral in the right hand side of (9.51) is bounded by

$$\begin{aligned}\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{Q}^c_{i_1,i_2}}f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})d\mathbf{y}&\leq\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{R}^c_{i_1,i_2}}f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})d\mathbf{y}\\&=\sum_{i_1\in\mathbb{K}}\Pr\left(\|\mathbf{y}-g\mathbf{c}_{i,1}\|\geq\sqrt{n\left(\sigma_Z^2+\eta\right)}\right)\\&=K\cdot\Pr(\|\mathbf{Z}\|^2-n\sigma_Z^2>n\eta)\\&\overset{(a)}{\leq}\frac{3\sigma_Z^4}{n^{1-\kappa}\eta^2}\end{aligned}$$

$$\leq \mu \,, \tag{9.58}$$

for sufficiently large n with $\kappa \in [0,1)$, where (a) holds by Chebyshev's inequality, followed by the substitution of $\mathbf{z} \equiv \mathbf{y} - g\mathbf{c}_{i_1}$. Thus, by (9.51),

$$\begin{aligned}\int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\, d\mathbf{y} &\leq \int_{\bigcup_{i_1 \in \mathbb{K}} \mathbb{T}_{i_1,g}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\, d\mathbf{y} \\ &\leq \int_{\bigcup_{i_1 \in \mathbb{K}} \mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\, d\mathbf{y} + \mu \,. \end{aligned} \tag{9.59}$$

Now, let us focus on the first integral in (9.51) with domain of $\mathbb{P}_{i_1,i_2}$, i.e., where

$$\|\mathbf{y} - g\mathbf{c}_{i,2}\| \leq \sqrt{n(\sigma_Z^2 + \zeta)} \,. \tag{9.60}$$

Observe that

$$f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1}) - f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_2}) = f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1}) \left[1 - e^{-\frac{1}{2\sigma_Z^2}\left(\left\|\mathbf{y} - g\mathbf{c}_{i_2}\right\|^2 - \left\|\mathbf{y} - g\mathbf{c}_{i_1}\right\|^2\right)}\right] \,. \tag{9.61}$$

By the triangle inequality,

$$\left\|\mathbf{y} - g\mathbf{c}_{i_1}\right\| \leq \|\mathbf{y} - g\mathbf{c}_{i_2}\| + g\left\|\mathbf{c}_{i_1} - \mathbf{c}_{i_2}\right\| \,. \tag{9.62}$$

Taking the square of both sides, we have

$$\begin{aligned}\left\|\mathbf{y} - g\mathbf{c}_{i_1}\right\|^2 &\leq \|\mathbf{y} - g\mathbf{c}_{i_2}\|^2 + g^2\left\|\mathbf{c}_{i_2} - \mathbf{c}_{i_1}\right\|^2 + 2\|\mathbf{y} - g\mathbf{c}_{i_2}\| \cdot g\left\|\mathbf{c}_{i_2} - \mathbf{c}_{i_1}\right\| \\ &\overset{(a)}{\leq} \|\mathbf{y} - g\mathbf{c}_{i_2}\|^2 + g^2\alpha_n^2 + 2g\alpha_n\sqrt{n(\sigma_Z^2 + \zeta)} \\ &\overset{(b)}{=} \|\mathbf{y} - g\mathbf{c}_{i_2}\|^2 + \frac{4Ag^2}{n^{1+2(\kappa+b)}} + \frac{4g\sqrt{A(\sigma_Z^2 + \zeta)}}{n^{\kappa+b}} \,, \end{aligned} \tag{9.63}$$

where (a) follows from (9.46) and (9.60), and (b) holds by (9.47). Now, in order to bound (9.63), let us define,

$$N_{\max} \overset{\text{def}}{=} 2\sigma_Z^2 \cdot \max\left(4Ag^2, 8g\sqrt{A(\sigma_Z^2 + \zeta)}\right) \,. \tag{9.64}$$

Therefore, (9.63) is bounded as follows

$$\begin{aligned}\left\|\mathbf{y} - g\mathbf{c}_{i_1}\right\|^2 - \|\mathbf{y} - g\mathbf{c}_{i_2}\|^2 &\leq \frac{4Ag^2}{n^{1+2(\kappa+b)}} + \frac{4g\sqrt{A(\sigma_Z^2 + \zeta)}}{n^{\kappa+b}} \\ &\leq \frac{2\sigma_Z^2 N_{\max}}{n^{\kappa+b}} \,, \end{aligned} \tag{9.65}$$

where the last inequality holds since $n^{1+2(\kappa+b)} \geq n^{\kappa+b}$ for a given κ and b, and every n. Now let us define

$$\omega_n \stackrel{\text{def}}{=} \frac{N_{\max}}{n^{\kappa+b}}. \tag{9.66}$$

Then we employ inequality $1-\frac{1}{x} \leq \ln x$, $\forall x > 0$ ([195, Eq. 1]) by setting $x = \frac{1}{1-\omega_n}$ and provide an upper bound on ω_n as follows

$$\begin{aligned} \omega_n &\leq \ln\left(\frac{1}{1-\omega_n}\right) \\ &= \ln\left(\frac{n^{\kappa+b}}{n^{\kappa+b}-N_{\max}}\right), \end{aligned} \tag{9.67}$$

where conditions $x > 0$ and $\omega_n < 1$ are fulfilled for sufficiently large n. Therefore by (9.65) we obtain

$$\left\|\mathbf{y} - g\mathbf{c}_{i_1}\right\|^2 - \left\|\mathbf{y} - g\mathbf{c}_{i_2}\right\|^2 \leq 2\sigma_Z^2 \cdot \ln\left(\frac{n^{\kappa+b}}{n^{\kappa+b}-N_{\max}}\right), \tag{9.68}$$

Hence,

$$\begin{aligned} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1}) - f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_2}) &\leq f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\left(1 - e^{-\frac{\omega_n}{2\sigma_Z^2}}\right) \\ &\leq f_{\mathbf{Z}}\left(\mathbf{y} - g\mathbf{c}_{i_1}\right)\left(1 - e^{-\ln\left(\frac{n^{\kappa+b}}{n^{\kappa+b}-N_{\max}}\right)}\right) \\ &\leq f_{\mathbf{Z}}\left(\mathbf{y} - g\mathbf{c}_{i_1}\right)\left(1 - \frac{n^{\kappa+b}-N_{\max}}{n^{\kappa+b}}\right) \\ &\leq f_{\mathbf{Z}}\left(\mathbf{y} - g\mathbf{c}_{i_1}\right) \cdot \frac{N_{\max}}{n^{\kappa+b}} \\ &= f_{\mathbf{Z}}\left(\mathbf{y} - g\mathbf{c}_{i_1}\right) \cdot \omega_n, \end{aligned} \tag{9.69}$$

Now we obtain,

$$\begin{aligned} e_1 + e_2 &\geq P_{e,1}(i_1) + P_{e,2}(i_2, \mathbb{K}) \\ &\stackrel{(a)}{\geq} \sup_{g\in\mathcal{G}}\left[P_{e,1}(i_1|g)\right] + \sup_{g\in\mathcal{G}}\left[P_{e,2}(i_2, \mathbb{K}|g)\right] \\ &\stackrel{(b)}{\geq} \sup_{g\in\mathcal{G}}\left[P_{e,1}(i_1|g) + P_{e,2}(i_2, \mathbb{K}|g)\right] \\ &\stackrel{(c)}{=} \sup_{g\in\mathcal{G}}\left[1 - \int_{\mathcal{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\, d\mathbf{y} + \int_{\mathcal{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_2})\, d\mathbf{y}\right] \end{aligned} \tag{9.70}$$

where (a) follows by (9.26) and (9.29), (b) holds since supremum is sub-additive and (c) is due to definitions of error in (9.5) and (9.6). Now we proceed to bound (9.70) as follows

$$\begin{aligned}
&\sup_{g\in\mathcal{G}}\left[1-\int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})\,d\mathbf{y}+\int_{\mathscr{T}_{\mathbb{K}}} f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_2})\,d\mathbf{y}\right]\\
&\overset{(a)}{\geq}\sup_{g\in\mathcal{G}}\left[1-\mu-\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})\,d\mathbf{y}+\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{T}_{i_1,g}} f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_2})\,d\mathbf{y}\right]\\
&\overset{(b)}{\geq}\sup_{g\in\mathcal{G}}\left[1-\mu-\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})\,d\mathbf{y}+\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_2})\,d\mathbf{y}\right]\\
&\overset{(c)}{=}\sup_{g\in\mathcal{G}}\left[1-\mu-\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{P}_{i_1,i_2}}\left[f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})-f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_2})\right]\,d\mathbf{y}\right]
\end{aligned}\tag{9.71}$$

where (a) holds by (9.59) and (b) follows from $\mathbb{P}_{i_1,i_2}\subset\mathbb{T}_{i_1,g}$. Now we proceed to bound (9.71) as follows

$$\begin{aligned}
&\sup_{g\in\mathcal{G}}\left[1-\mu-\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{P}_{i_1,i_2}}\left[f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})-f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_2})\right]\,d\mathbf{y}\right]\\
&\overset{(a)}{\geq}\sup_{g\in\mathcal{G}}\left[1-\mu-\omega_n\int_{\bigcup_{i_1\in\mathbb{K}}\mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})\,d\mathbf{y}\right]\\
&\overset{(b)}{\geq}\sup_{g\in\mathcal{G}}\left[1-\mu-\omega_n\sum_{i_1\in\mathbb{K}}\int_{\mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y}-g\mathbf{c}_{i_1})\,d\mathbf{y}\right]\\
&\overset{(c)}{\geq}\sup_{g\in\mathcal{G}}\left[1-\mu-\omega_n\cdot|\mathbb{K}|\right]\\
&\overset{(d)}{=}\sup_{g\in\mathcal{G}}\left[1-\mu-\frac{KN_{\max}}{n^{b+\kappa}}\right]\\
&\overset{(e)}{\geq}\sup_{g\in\mathcal{G}}\left[1-\mu-\pi\right]\\
&=1-2\mu-\pi\,,
\end{aligned}\tag{9.72}$$

where (a) follows by (9.69), (b) holds by the union bound, (c) follows from

$$\int_{\mathbb{P}_{i_1,i_2}} f_{\mathbf{Z}}(\mathbf{y} - g\mathbf{c}_{i_1})\, d\mathbf{y} = \Pr\left(\left\|\mathbf{y} - g\mathbf{c}_{i_1}\right\| \leq \sqrt{n\left(\sigma_Z^2 + \zeta\right)}\right) \leq 1\,, \tag{9.73}$$

and (c) follows since $|\mathbb{K}| = K = n^{\kappa}$, (d) follows from (9.66), and (e) holds since $\frac{KN_{\max}}{n^{b+\kappa}} = \frac{1}{n^b} \leq \pi$ for sufficiently large n. Thereby, recalling (9.70),(9.71),(9.72) we obtain

$$e_1 + e_2 \geq 1 - 2\mu - \pi\,. \tag{9.74}$$

Clearly, this is a contradiction since the error probabilities tend to zero as $n \to \infty$. Thus, the assumption in (9.46) is false. This completes the proof of Lemma 9.3.1. □

Next, we use Lemma 9.3.1 to prove the upper bound on the DKI capacity. Observe that Lemma 9.3.1 implies that the distance between every pair of codewords satisfies

$$\left\|\mathbf{c}_{i_1} - \mathbf{c}_{i_2}\right\| \geq 2\sqrt{n\epsilon_n'}\,. \tag{9.75}$$

Thus, we can define an arrangement of non-overlapping spheres $\mathcal{S}_{\mathbf{c}_i}(n, \sqrt{n\epsilon_n'})$, i.e., spheres of radius $\sqrt{n\epsilon_n'}$ that are centered at the codewords $\mathbf{c}_i$. Since the codewords all belong to a large hyper sphere $\mathcal{S}_{\mathbf{0}}(n, \sqrt{nA})$ of radius $\sqrt{nA}$, it follows that the number of packed small spheres, i.e., the number of codewords M, is bounded by

$$\begin{aligned} M &= \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{M} \mathcal{S}_{\mathbf{c}_i}(n, r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, \sqrt{nA} + r_0))} \\ &\overset{(a)}{=} \Delta_n(\mathscr{S}) \cdot \frac{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, \sqrt{nA} + r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))} \\ &\overset{(b)}{\leq} 2^{-0.599n} \cdot \frac{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{0}}(n, \sqrt{nA} + r_0)\right)}{\mathrm{Vol}(\mathcal{S}_{\mathbf{c}_1}(n, r_0))}\,, \end{aligned} \tag{9.76}$$

where (a) holds by definition of packing density, (b) follows from inequality (9.16). The above bound can be further simplified as follows

$$\begin{aligned} \log M &\overset{(a)}{\leq} \log\left(\frac{\sqrt{nA} + r_0}{r_0}\right)^n - 0.599n \\ &\leq n \log\left(\frac{\sqrt{nA} + r_0}{r_0}\right) - 0.599n \\ &\overset{(b)}{=} \frac{1}{2} n \log\left(\frac{A}{\epsilon_n'} + 1\right) - 0.599n\,, \end{aligned} \tag{9.77}$$

where (a) exploits (9.17) and (b) follows from $r_0 = \frac{1}{2}(2\sqrt{n\epsilon'_n})$. Therefore, for $\epsilon'_n = A/n^{2(1+\kappa+b)}$, we obtain

$$\begin{aligned}
\log M &\leq \frac{1}{2} n \log\left(n^{2(1+\kappa+b)} + 1\right) - 0.599n \\
&= \frac{1}{2} n \log\left(n^{2(1+\kappa+b)}\left(1 + 1/n^{2(1+\kappa+b)}\right)\right) - 0.599n \\
&= \frac{1}{2} n \log\left(n^{2(1+\kappa+b)}\right) + \frac{1}{2} n \log\left(1 + 1/n^{2(1+\kappa+b)}\right) - 0.599n \\
&= (1+\kappa+b)\, n \log n + \frac{1}{2} n \log\left(1 + 1/n^{2(1+\kappa+b)}\right) - 0.599n\,, \qquad (9.78)
\end{aligned}$$

where the dominant term is again of order $n \log n$. Hence, for obtaining a finite value for the upper bound of the rate, R, (9.77) induces the scaling law of M to be $2^{(n \log n)R}$. Hence, we obtain

$$\begin{aligned}
R &\leq \frac{1}{n \log n}\left[(1+\kappa+b)\, n \log n + \frac{1}{2} n \log\left(1 + 1/n^{2(1+\kappa+b)}\right) - 0.599n\right] \\
&= 1 + \kappa + b + \log\left(1 + 1/n^{2(1+\kappa+b)}\right) / \log n - 0.599/\log n\,, \qquad (9.79)
\end{aligned}$$

which tends to $1+\kappa$ as $n \to \infty$ and $b \to 0$. This completes the proof of Theorem 9.3.1.

9.4 | Summary

In this chapter, we studied the DKI problem over the GSF with K number target messages. We assumed that $K = K(n,\kappa) = 2^{\kappa \log n} = n^{\kappa}$ where $\kappa \in [0,1)$ scales sub-linearly with the codeword length n. In practice, the receiver sometimes suspend the exact matching task as is considered for the standard identification [99, 99] and requires only to spot an object among a group, therefore, our results in this chapter may serve as a model for event-triggered based tasks in the context of many practical XG applications where population of the target group scales sub-linearly in the codeword length. Especially, we obtained lower and upper bounds on the DKI capacity of the GSF with $K = 2^{\kappa \log n}$ many target messages subject to average power constraint with the codebook size of $M(n,R) = 2^{(n \log n)R} = n^{nR}$. Our results for the DKI capacity of the GSF revealed that the super-exponential scale of $n^{nR} = 2^{(n \log n)R}$ is again the appropriate scale for codebook size. This scale coincides as of the codebook for the memoryless GSF and Gaussian channels [99, 105] and stands considerably different from the traditional scales in transmission and RI setups where corresponding codebooks size grows exponentially and double exponentially, respectively.

We show the achievability proof using a packing of hyper spheres and a distance decoder. In particular, we pack hyper spheres with radius $\sqrt{n\theta_n} \sim n^{\frac{1+\kappa}{4}}$ where $\kappa \in [0,1)$ is the target identification rate, inside a larger hyper sphere, which results in $\sim 2^{((1-\kappa)/4)n\log n}$ codewords. For the converse proof, we follow a similar approach as in the chapter 3 for the standard identification over the slow fading channel [99, 106]. In general, the derivation here is more involved than the derivation in the standard identification case [105] and entails employing of new analysis and inequalities. In chapter 4 on Gaussian channels with slow fading [105], the converse proof was based on establishing a minimum distance between each pair of codewords. Here, we incorporate effect of the number of target messages into the minimum distance in the relevant Lemma; see Eq. 1 9.3.1.

CHAPTER 10

DKI FOR BINARY SYMMETRIC CHANNEL

> *Few, But Ripe.*
>
> Carl Friedrich Gauss,

> *The Mathematical Theory of Information Had Come Into Being When It Was Realized That The Flow of Information Can be Expressed Numerically in The Same Way as Distance, Mass, Temperature, etc.*
>
> Alfréd Rényi,

10.1 | Introduction

The binary symmetric channel (BSC) is deemed as a basic mathematical model through which one bit per unit of time can be transmitted. The capacity of such a channel is attained by Bernoulli input with $1/2$ success probability, i.e., $X \sim \text{Bern}(1/2)$. In [196] consider the BSC and used a random linear code for the achievability proof which result in a exponential search in the decoding.

10.1.1 | Previous Results

In the (standard) identification problem [23], the receiver aims to identify the occurrence of a *single* message. However, there exist a generalized variation of the DI problem, called the K-Identification problem [193], in which the receiver may seek to determine the presence of a single message *within* a set of messages (subset of the message

set) referred to as the target message set. The K-Identification problem can be understood as the generalization of the original identification problem in the following fashion: The target message (singleton set) is enlarged to be a general set of $K_{>1}$ messages. Ahlswede in [114, 193, Th. 1,Propos. 1] showed that the number of target messages for RI problem over a DMC scales exponentially in the codeword length n, i.e., $K = 2^{\kappa n}$ and proved that the set of all deterministic K-identification (DKI) achievable pairs contains

$$\left\{(R,\kappa) : 0 \leq R,\kappa \; ; \; R + 2\kappa \leq \mathbb{C}_{\mathrm{TR}}\right\} \; , \tag{10.1}$$

where $\mathbb{C}_{\mathrm{TR}}$ is the message transmission capacity of the DMC.

The DKI problem for the slow fading channels $\mathcal{G}_{\mathrm{slow}}$, assuming that the number of target messages scales sub-linearly with codeword length n, i.e., $K(n,\kappa) = 2^{\kappa \log n}$, subject to an average power constraint and a codebook size of super-exponential scale, i.e., $M(n,R) = 2^{(n \log n)R}$, is studied in [197] where the following K-depending bounds on the DKI capacity are derived:

$$\frac{1-\kappa}{4} \leq \mathbb{C}_{\mathrm{DKI}}(\mathcal{G}_{\mathrm{slow}}, M, K) \leq 1 + \kappa \, . \tag{10.2}$$

10.1.2 | Contributions

In this chapter, we consider identification systems employing deterministic encoder and receivers that are interested to accomplish the K-Identification task, namely, finding an object in a target message set of size $K = 2^{\kappa n}$ for $\kappa \in [0,1)$. We assume that the communication over n channel uses are independent of each other. We assume that the noise is additive Bernoulli process and formulate the problem of DKI over the DTBC under Hamming weight input constraint.

To the best of the authors' knowledge, the fundamental performance limits of DKI for the BSC model has not been so far studied in the literature. As our main objective, we investigate the DKI capacity of the BSC. In particular, this chapter makes the following contributions:

♢ **Generalized Identification Model**: In several identification systems, often the size of target message set K can be large, particularly when one by one comparison is not demanded due to the delay constraint. In addition, the value of K may increases as a function of the codeword lengths n. To address these cases, we consider a generalized

identification model[1] that captures the standard (i.e., $K = 1$), identification channels with constant $K > 1$, and identification channels for which K increases with the codeword length n.

♢ **Codebook Scale**: We establish that the codebook size of the DKI problem over the BSC for deterministic encoding scales *exponentially* in the codeword length n, i.e., $\sim 2^{nR}$, even when the size of target message set scales as $K = 2^{\kappa n}$ for some $\kappa \in [0, 1)$ (see Theorem 10.3.1 for exact upper bound on the κ), which we refer to as the *target identification rate*. Such an exponential scale for the codebook size coincide with that of the message transmission problem [28] and the standard identification problem (DI) in which $K = 2^0 = 1$ [52, 98, 99] and is lower than the *super-exponential* scale for that of the DKI problem over the slow fading channels [197]. This observation suggests that increasing the number of target messages does not change the scale of the codebook derived for DI scheme over the BSC [52, 98, 99].

♢ **Capacity Formula**: We derived a closed form analytical function for the DKI capacity for the BSC, which are the main results of this chapter. Such formula does *reflect* the impact of the input constraint P_{ave} in the *optimal* scale of the codebook size, i.e., 2^{nR}. This observation is in contrast to the result obtained for the DKI problem for the slow fading channel [197] or the DI problem for Gaussian and Poisson channels [99, 105, 108, 110]. We derive the DKI capacity formula for the BSC with constant $K \geq 1$ and growing size of the target message set $K = 2^{\kappa n}$, respectively. We show that for both cases of the constant K and the growing number of target messages, the proposed capacity expression is not a function of the target identification rate κ and remains only as a function of the Hamming weight constraint.

♢ **Technical Novelty**: To obtain the proposed lower bound, the existence of an appropriate ball packing within the input space, for which the Hamming distance between the centers of the balls does not fall below a certain value, is established. In particular, we consider the packing of hyper balls inside a larger n-dimensional Hamming hyper ball, whose radius grows in the codeword length n, i.e., nA. For the achievability proof, we exploit a greedy construction similar to the *Gilbert bound* method. While

[1] The proposed *generalized identification* setting may be used in a more advanced scheme called *generalized identification with decoding* [198, Ch. 1] where first the K-Identification and second, the standard identification are accomplished to find an object. Furthermore, *generalized identification* should be distinguished from *multiple object identification* [199] where K objects whose corresponding target message sets are *unknown* to the receiver, are identified at once.

the radius of the small balls in the DKI problem for the slow fading channel [197], grows in the codeword length n as $n \to \infty$, here, the radius similar to the DI problem for the Gaussian channel with slow and fast fading [105] tends to zero. In general, the derivation of lower bound for the BSC is more involved compared to that for the Gaussian [105] and Poisson channels with/out memory [108] and entails exploiting of new analysis and inequalities. Here, the error analysis in the achievability proof requires dealing with several combinatorial arguments and using of bounds on the tail of the cumulative distribution function of the Binomial distribution.

10.1.3 | Organization

The remainder of this chapter is structured as follows. In Section 10.2, system model is explained and the required preliminaries regarding DKI codes are established. Section 10.3 provides the main contributions and results on the message DKI capacity of the BSC. Finally, Section 10.4 of the paper concludes with a summary and directions for future research.

10.2 | System Model and Preliminaries

In this section, we present the adopted system model and establish some preliminaries regarding TR and DKI coding.

10.2.1 | System Model

We address an identification-focused communication setup, for which the objective of the decoder is defined as follows: Determining whether or not a desired message was sent by the transmitter; see Figure 10.1. To accomplish this purpose, a coded communication between the transmitter and the receiver over n channel uses of a binary symmetric channel is established[2]. Let $X \in \{0,1\}$ and $Y \in \{0,1\}$ indicate random variables (RVs) which model the input and output of the channel. Each binary input symbol is flipped with probability $0 < \varepsilon < \frac{1}{2}$[3]. The stochastic flipping of the input symbol is modelled via an additive Binary Bernoulli noise, i.e., $Z \in \{0,1\}$; see Figure 10.1.

[2] The proposed capacity formula works regardless of whether or not a specific code is used for communication, although proper explicit constructed codes may be required to approach the capacity limits.

[3] The extreme cases of $\varepsilon = 0$ or $\varepsilon = \frac{1}{2}$ result in $\mathbb{C}_{\mathrm{TR}} = 1$ and $\mathbb{C}_{\mathrm{TR}} = 0$, respectively, hence these cases are commonly excluded from the analysis.

Therefore, the input-output relation of channel reads: $Y = X \oplus Z$, where $\oplus$ indicate the modulo two addition. Throughout the paper, the considered binary symmetric channel with crossover probability $0 < \varepsilon < \frac{1}{2}$ is denoted by $\mathcal{W}_\varepsilon$. We consider the BSC channel $\mathcal{W}_\varepsilon$ which arises as a basic channel model in the context of information theory where the noise distribution, i.e., the probability of observing channel output Y at the receiver given that channel input X was sent at the transmitter, is characterized as follows:

$$W(Y|X) = \begin{cases} 1-\varepsilon & Y = X \\ \varepsilon & Y \neq X \end{cases}. \tag{10.3}$$

for all $x, y \in \{0,1\}$ and $0 < \varepsilon < \frac{1}{2}$.

We assume that $\mathcal{W}_\varepsilon$ is memoryless, that is, the different channel uses are independent. Hence, the transition probability law for n channel uses is given by

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{n} W(y_t|x_t) = \varepsilon^{d_{\mathrm{H}}(\mathbf{x},\mathbf{y})}(1-\varepsilon)^{n-d_{\mathrm{H}}(\mathbf{x},\mathbf{y})}, \tag{10.4}$$

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ denote the transmitted codeword and the received signal, respectively. Observe that $d_{\mathrm{H}}(\mathbf{x}, \mathbf{y})$ is a random variable and follows a Binomial distribution; see Remark10.2.1.

10.2.2 | Message Transmission Coding For BSC

The definition of a TR code for the BSC $\mathcal{W}_\varepsilon$ is given below.

Definition 10.2.1 (BSC-TR Code). *An $(n, M(n,R), e_1)$-BSC-TR code for a BSC $\mathcal{W}_\varepsilon$ for integer $M(n,R)$, where n and R are the codeword length[4] and coding rate, respectively, is defined as a system $(\mathcal{C}, \mathscr{T})$, which consists of a codebook $\mathcal{C} = \{\mathbf{c}_i\}_{i\in[\![M]\!]}$, with $\mathbf{c}_i = (c_{i,t})|_{t=1}^{n} \subset \{0,1\}^n$, such that*

$$\frac{1}{n}\sum_{t=1}^{n} c_{i,t} \leq A, \tag{10.5}$$

$\forall\, i \in [\![M]\!]$, and a collection of decoders $\mathscr{T} = \{\mathbb{T}_i\}_{i\in[\![M]\!]}$, where $\mathbb{T}_i \subset \{0,1\}^n$, such that the decoders are mutually disjoint, *i.e.,*

$$\mathbb{T}_i \cap \mathbb{T}_j = \varnothing, \tag{10.6}$$

[4] This code definition restricts itself to codewords of the same length for different messages, which sometimes in the literature is called as the *block* codes. Throughout this chapter, we always assume that different TR or DKI codes are *block* codes.

for every $i, j \in [\![M]\!]$ such that $i \neq j$. Given a message $i \in [\![M]\!]$, the encoder transmits codeword $\mathbf{c}_i$, and the decoder's task is to address a multiple hypothesis *as follows: Which message $\hat{i} \in [\![M]\!]$ was sent? There exist one type of error that may happen:*

- *Error Event: Rejection of the actual message; $i \in [\![M]\!]$.*

The associated error probability of the BSC-TR code $(\mathcal{C}, \mathscr{T})$ reads

$$P_{e,1}(i) = 1 - \sum_{\mathbf{y} \in \mathbb{T}_i} W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_i\right), \tag{10.7}$$

see Figure 10.1; and satisfy the following bounds $P_{e,1}(i) \leq e_1$, $\forall i \in [\![M]\!]$, $\forall e_1 > 0$.

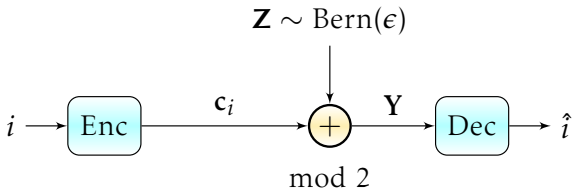


Figure 10.1: System model of message transmission for the binary symmetric channel. Message m is mapped to the codeword $\mathbf{c}_i$ where it is flipped with a probability of ε. Decoder employs the output vector $\mathbf{y}$ to declare a reconstructed version of the original sent message m, denoted by $\hat{i}$ in a *reliably* form.

Next we define the achievable TR rate and the TR capacity.

Definition 10.2.2 (Achievable Rate). *A TR rate $R > 0$ is called achievable if $\forall e_1 > 0$ and sufficiently large n, there exists an $(n, M(n,R), e_1)$-BSC-TR code. The operational TR capacity of the BSC $\mathcal{W}_\varepsilon$ is defined as the supremum of all achievable rates, and is denoted by $\mathbb{C}_{TR}(\mathcal{W}_\varepsilon, M, K)$.*

10.2.3 | Message Transmission Capacity of BSC

In this Subsection we introduce the message transmission problem for the BSC which was introduced originally by Shannon [28] and introduce some fundamental results on the achievability and the converse proofs.

Theorem 10.2.1 (see [200, Ch. 5 - Corrol. 2]). *Assume the binary symmetric channel $\mathcal{B}$ with cross-over probability ε and message transmission capacity of $\mathbb{C}_{TR}(\mathcal{W}_\varepsilon) = 1 - H(\varepsilon)$ and consider an $(n, M(n,R), e_1)$-BSC-TR code for which the codebook size scales exponentially in the codeword length n, i.e., $M(n,R) = 2^{nR}$ where R is the message transmission coding rate. Then $0 \leq \forall R < \mathbb{C}_{TR}(\mathcal{W}_\varepsilon)$ is achievable, namely, there exists a coding and decoding*

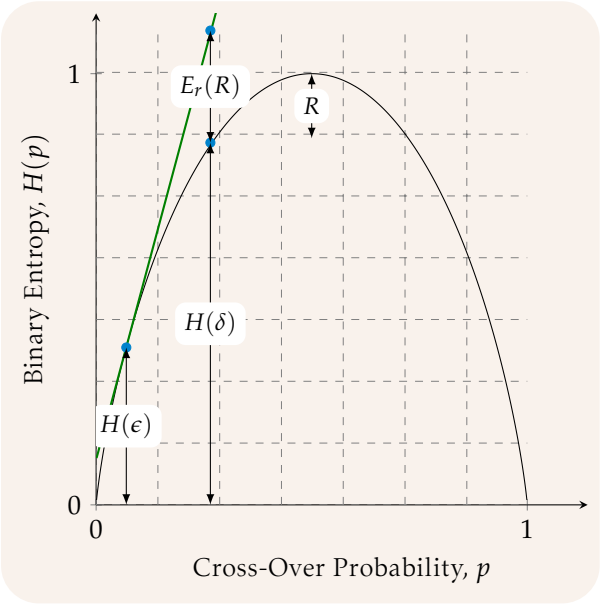


Figure 10.2: Depiction of the error exponent for a BSC. For a given crossover probability $0 < \varepsilon < \frac{1}{2}$, and The difference between the tangent line to the binary entropy function and the binary entropy function itself is referred to as the error exponent.

scheme, that is, the existence of a codebook with size $M(n,R) = 2^{nR}$ *is guaranteed and the maximum type I error probability converge to zero as* $n \to \infty$*, i.e.,*

$$P_{e,1}(i) \leq 2^{-nE_r(R)+2}, \tag{10.8}$$

for every $i \in [\![M]\!]$*, where* $E_r(R) > 0$ *is a positive, decreasing and convex function*[5] *of* R*.*

Here, we restrict ourselves to the following settings: Let the probability assignment on the channel input symbols 0 and 1 be the uniform probability mass function. Further, for the sake of accurate analysis of the $E_r(R)$, depending on the range of δ, we divide in two cases as follows:

$$E_r(R) = \begin{cases} T_\varepsilon(\delta) - H(\delta) & \varepsilon \leq \delta \leq \sqrt{\varepsilon}/(\sqrt{\varepsilon} + \sqrt{1-\varepsilon}) \quad (10.9) \\ 1 - R - \log(\sqrt{\varepsilon}/(\sqrt{\varepsilon} + \sqrt{1-\varepsilon})) & \sqrt{\varepsilon}/(\sqrt{\varepsilon} + \sqrt{1-\varepsilon}) < \delta < \frac{1}{2} \quad (10.10) \end{cases}$$

where the corresponding values for rate R are given by

[5] The function $E_r(R)$ in the literature is referred to as the *random coding exponent*. Even for the simple BSC, there is no simple way to express the $E_r(R)$ in an analytic functional way for all the values of $0 \leq R < C$ except than in *parametric* form. Cf. [200] for further properties of such function. Further discussion are explained in 10.10

1. Case 10.9 $\Rightarrow$ $1 - H(\sqrt{\varepsilon}/(\sqrt{\varepsilon} + \sqrt{1-\varepsilon})) \leq R = 1 - H(\delta) \leq \mathbb{C}_{\mathrm{TR}}(\mathcal{W}_\varepsilon)$
2. Case 10.10 $\Rightarrow$ $R < 1 - H(\sqrt{\varepsilon}/(\sqrt{\varepsilon} + \sqrt{1-\varepsilon}))$

Corollary 10.2.1.1. *For the BSC consider a message set consisting of 2^m messages[6] and let the length of the associated codewords be $n = m(1 + H(\varepsilon)/(1 - H(\varepsilon)) + r)$, where $r > 0$ is an arbitrarily small constant. Then a codebook consisting of codewords with length n. Further, the maximum error probability over the entire message set is upper bounded as follows*

$$\max_{i \in [\![M]\!]} \left[P_{e,1}(i)\right] \leq 2^{-n\alpha(r,\varepsilon) + \log q(r,\varepsilon)}\,. \tag{10.11}$$

Proof. Let $\delta > \varepsilon$ be such that

$$H(\delta) = \frac{r + (1-r)H(\varepsilon)}{r + 1 - rH(\varepsilon)}\,, \tag{10.12}$$

assuming $r > 0$ is sufficiently small such that the condition $\delta \leq \sqrt{\varepsilon}/(\sqrt{\varepsilon} + \sqrt{1-\varepsilon})$ as required by Theorem 10.2.1 is fulfilled. Then the exponent $\alpha(R)$ provided in Theorem 10.2.1 can be taken to be

$$\alpha = T_\varepsilon(\delta) - H(\delta)\,. \tag{10.13}$$

Observe that the exponent given in 10.13 can not be improved by the following theorem. □

Theorem 10.2.2 (see [200, Th. 5.8.5]). *Consider the BSC with crossover probability ε; $\mathcal{W}_\varepsilon$ with the message transmission capacity of $\mathbb{C}_{TR}(\mathcal{W}_\varepsilon) = 1 - H(\varepsilon)$. Further assume an $(n, M(n, R), e_1)$-BSC-TR code where the codebook size scales **exponentially** in the codeword length n, i.e., $2^m = M(m) = M(n, R) = 2^{nR}$ where R is the message transmission coding rate. Now if*

$$\frac{\log M(n,R)}{n} = R > \mathbb{C}_{TR}(\mathcal{W}_\varepsilon)\,, \tag{10.14}$$

then the average error probability of BSC-TR code is lower bounded as follows

$$\bar{P}_{e,1} \geq 1 - \frac{4L}{n(R - \mathbb{C}_{TR}(\mathcal{W}_\varepsilon))^2} - 2^{-\frac{n(R - \mathbb{C}_{TR}(\mathcal{W}_\varepsilon))}{2}}\,, \tag{10.15}$$

where $L > 0$ is a finite positive constant depending on the channel statistics ε and does not depend on the codeword length[7].

[6] Each message is a binary sequence of length m, which yields a total of 2^m message sequences, called the *message set*.

[7] The lower bound given in 10.15 converges to 1 from left as $n \to \infty$.

Corollary 10.2.2.1. *If codeword length n satisfies*

$$n < m\left(1 + \frac{H(\varepsilon)}{1 - H(\varepsilon)}\right), \tag{10.16}$$

then there exists ***at least one message*** *for which the error probability can not be upper bounded by any constant $q < 1$.*

Proof. Observe that (10.16) implies a chain of equations as follows

$$\begin{aligned}
n < m\left(1 + \frac{H(\varepsilon)}{1 - H(\varepsilon)}\right) &\Rightarrow \\
&= m\left(\frac{1 - H(\varepsilon) + H(\varepsilon)}{1 - H(\varepsilon)}\right) \\
&= m\left(\frac{1}{1 - H(\varepsilon)}\right) \\
&= \frac{m}{1 - H(\varepsilon)} \\
&\overset{(a)}{=} \frac{m}{\mathbb{C}_{\mathrm{TR}}} \\
&= \frac{\log 2^m}{\mathbb{C}_{\mathrm{TR}}} \\
&\overset{(b)}{=} \frac{\log M(m)}{\mathbb{C}_{\mathrm{TR}}},
\end{aligned} \tag{10.17}$$

where (a) employs $\mathbb{C}_{\mathrm{TR}}(\mathcal{W}_\varepsilon) = 1 - H(\varepsilon)$ and (b) holds by $M(m) = 2^m = M(n, R)$. Thereby exploiting $M(m) = M(n, R)$ into (10.17) yields,

$$\frac{\log M(n, R)}{n} = R > \mathbb{C}_{\mathrm{TR}}(\mathcal{W}_\varepsilon). \tag{10.18}$$

Therefore, by Theorem 10.2.2 we conclude that the average error probability converges to 1 which implies that there exist at least one message whose maximum error probability converges to 1 or can not be upper bounded by any constant $q < 1$. This completes the proof of Corollary 10.2.2.1. □

Corollary 10.2.2.2. *Let $\lambda > 0$ be an arbitrarily finite large constant. Then, $0 < \exists\, \varepsilon < \frac{1}{2}$, such that no code of length $n < \lambda m$ can guarantee any bound $q < 1$ on the maximum error probability of message transmission.*

Proof. Choose an ε where $0 < \varepsilon < \frac{1}{2}$, so that the following condition is fulfilled

$$H(\varepsilon) = 1 - \frac{1}{\lambda}. \tag{10.19}$$

Now, apply the Corollary 10.2.2.1. □

10.2.4 | DKI Coding For BSC

The definition of a DKI code for the BSC $\mathcal{W}_\varepsilon$ is given below.

Definition 10.2.3 (BSC-DKI Code). *An $(n, M(n,R), K(n,\kappa), e_1, e_2)$-BSC-DKI code for a BSC $\mathcal{W}_\varepsilon$ for integers $M(n,R)$ and $K(n,\kappa)$, where n and R are the codeword length and coding rate, respectively, is defined as a system $(\mathcal{C}, \mathscr{T}_{\mathbb{K}})$, which consists of a codebook $\mathcal{C} = \{\mathbf{c}_i\}_{i\in[\![M]\!]} \subset \{0,1\}^n$, with $\mathbf{c}_i = (c_{i,t})|_{t=1}^n \subset \{0,1\}^n$, such that*

$$\frac{1}{n}\sum_{t=1}^{n} c_{i,t} \leq P_{avg}\,, \tag{10.20}$$

$\forall i \in [\![M]\!]$ and a decoder[8]

$$\mathscr{T}_{\mathbb{K}} = \bigcup_{j\in\mathbb{K}} \mathbb{T}_j\,, \tag{10.21}$$

where $\mathbb{T}_j \subset \{0,1\}^n$ is the decoding set corresponding to the single message $\mathbf{c}_j$, $\forall \mathbb{K} \in \{X \subseteq [\![M]\!]\,; |X| = K\}$ where $\mathbb{K}$ is an arbitrary subset[9] *with size K. Given a message $i \in [\![M]\!]$, the encoder transmits codeword $\mathbf{c}_i$, and the decoder's task is to address a* binary hypothesis*: Was a target message $j \in \mathbb{K}$ sent or not? There exist two types of errors that may happen:*

- *Type I Error Event: Rejection of the actual message; $i \in \mathbb{K}$*
- *Type II Error Event: Acceptance of a wrong message; $i \notin \mathbb{K}$.*

The associated error probabilities of the DKI code $(\mathcal{C}, \mathscr{T})$ reads

$$P_{e,1}(i) = \Pr\left(\mathbf{Y} \in \mathscr{T}_{\mathbb{K}}^c \,\middle|\, \mathbf{x} = \mathbf{c}_i\right)_{i\in\mathbb{K}} = 1 - \sum_{\mathbf{y}\in\mathscr{T}_{\mathbb{K}}} W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_i\right)_{i\in\mathbb{K}} \text{ (Miss-Identification)}\,, \tag{10.22}$$

[8] We recall that the decoding sets for the DKI problem may have in general intersection, a behaviour similar to that of the RI problem. However to guarantee a vanishing type II error probability we will observe that size of such intersection becomes negligible asymptotically, i.e., as $n \to \infty$

[9] We recall that $\{X \subseteq [\![M]\!]\,; |X| = K\}$ is the system (family) of all subsets of the set $[\![M]\!]$, with size K. Observe that in general we have $\left|\{X \subseteq [\![M]\!]\,; |X| = K\}\right| = \binom{M}{K}$ and the error requirement as imposed by the DKI code definition applies to *each* possible choice of the set $\mathbb{K}$ with K arbitrary messages among all $\binom{M}{K}$ cases.

$$P_{e,2}(i,\mathbb{K}) = \Pr\left(\mathbf{Y} \in \mathscr{T}_{\mathbb{K}} \,\middle|\, \mathbf{x} = \mathbf{c}_i\right)_{i\notin\mathbb{K}} = \sum_{\mathbf{y}\in\mathscr{T}_{\mathbb{K}}} W^n\left(\mathbf{y} \,\middle|\, \mathbf{c}_i\right)_{i\notin\mathbb{K}} \textit{(False Identification)}\,. \tag{10.23}$$

(see Figure 10.4) and satisfy the following bounds

$$P_{e,1}(i) \leq e_1\,,\ \forall i \in \mathbb{K}\,, \tag{10.24}$$

$$P_{e,2}(i,\mathbb{K}) \leq e_2\,,\ \forall i \notin \mathbb{K}\,. \tag{10.25}$$

where $\mathbb{K} \in \{X \subseteq [\![M]\!]\,; |X| = K\}$ *is an arbitrary K-size subset of* $[\![M]\!]$ *and* $\forall e_1, e_2 > 0$.

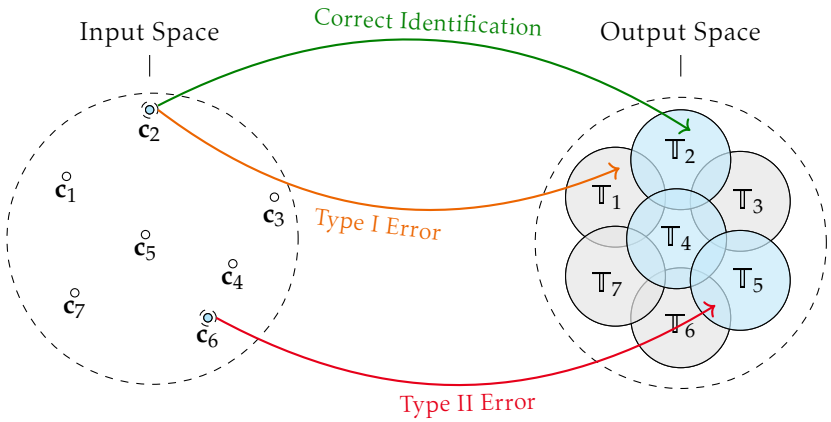


Figure 10.3: Depiction of a deterministic 3-identification setting with target message set $\mathbb{K} = \{2,3,5\}$. In the correct identification event, channel output is detected in the union of the individual decoder $\mathbb{T}_j$ where j belongs to the target message set. Type I error event occurs if the channel output is observed in the complement of the union of individual decoders for which the index of the codeword belongs to. The case where the index of codeword does not coincide to any of the individual decoders for which the channel output belongs to the their union, is referred to as the type II error.

Next we define the achievable DKI rate and the DKI capacity.

Definition 10.2.4 (Achievable Rate). *A DKI rate* $R > 0$ *is called achievable if* $\forall e_1, e_2 > 0$ *and sufficiently large* n*, there exists an* $(n, M(n,R), K(n,\kappa), e_1, e_2)$ *DKI code. The operational DKI capacity of the BSC* $\mathcal{W}_\varepsilon$ *is defined as the supremum of all achievable rates, and is denoted by* $\mathbb{C}_{DKI}(\mathcal{W}_\varepsilon, M, K)$.

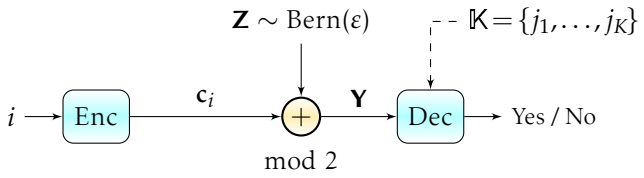


Figure 10.4: System model for DKI communication setting in a BSC. Employing a deterministic encoder in the transmitter, the message i is mapped onto the codeword $\mathbf{c}_i = (c_{i,t})|_{t=1}^{n}$ using a deterministic known function. The decoder at the receiver is provided with an arbitrary target message set $\mathbb{K}$, and given the channel output vector $\mathbf{Y} = (Y_t)|_{t=1}^{n}$, it asks whether i belong to $\mathbb{K}$ or not.

Remark 10.2.1. *Assuming that codeword $\mathbf{c}_i$ is sent and channel output $\mathbf{y}$ is observed at receiver, the number of crossovers (flips) that occur in the channel is $d_H(\mathbf{y}, \mathbf{c}_i)$. Therefore, the probability that k crossovers among the n channel uses occur, follows a Binomial distribution with parameters n and ε, as follows*

$$\Pr\left(d_H(\mathbf{Y}, \mathbf{c}_i) = k\right) = \binom{n}{k} \varepsilon^k (1-\varepsilon)^{n-k}, \tag{10.26}$$

10.3 | DKI Capacity of The BSC

In this section, we present our main results, i.e., achievability and converse proofs for the BSC. Subsequently, we provide the detailed proofs.

10.3.1 | Main Results

The DKI capacity theorem for the BSC channel $\mathcal{W}_\varepsilon$ is stated below.

Theorem 10.3.1. *Let $\mathcal{W}_\varepsilon$ indicate a BSC with cross-over probability $0 < \varepsilon < \frac{1}{2}$ and let $\beta > 0$ be an arbitrary small positive. Further let $H(p)$ indicate the binary entropy function and $T_\varepsilon(p) = H(\varepsilon) + (p-\varepsilon)\dfrac{\mathrm{d}H(p)}{\mathrm{d}p}|_{p=\varepsilon}$ specify the tangent to $H(p)$ at point ε. Then the deterministic K-Identification capacity of $\mathcal{W}_\varepsilon$ subject to the Hamming weight constraint of the form $n^{-1}\sum_{t=1}^{n} x_t \leq A$ with **exponentially** large target message set, i.e., $K(n,\kappa) = 2^{\kappa n}$ where the target identificaiton rate κ satisfy*

$$0 \leq \kappa < T_\varepsilon\Big(\big(1-\beta\big)\,\varepsilon + \beta/2\Big) - H\Big(\big(1-\beta\big)\,\varepsilon + \beta/2\Big)\,, \tag{10.27}$$

*and in the **exponential** codebook size, i.e., $M(n,R) = 2^{nR}$, is given by*

$$\mathbb{C}_{DKI}\left(\mathcal{W}_\varepsilon, K\right) = \begin{cases} H(A) & \text{if } A < \frac{1}{2} \\ 1 & \text{if } A \geq \frac{1}{2}\,. \end{cases} \tag{10.28}$$

Proof. The proof of Theorem 10.3.1 consists of two parts, namely the achievability and the converse proofs, which are provided in Sections 10.3.2 and 10.3.3, respectively. □

10.3.2 | Lower Bound (Achievability Proof)

The achievability proof consists of the following two main steps.

- □ **Step 1:** First, we propose a greedy-wise codebook construction and derive an analytical lower bound on the corresponding codebook size using similar argument as provided in the Gilbert-Varshamov (GV) bound[10] for packing of non-overlapping balls embedded in the input space.
- □ **Step 2:** Then, to prove that this codebook leads to an achievable rate, we propose a decoder and show that the corresponding type I and type II error rates vanished as $n \to \infty$.

10.3.2.1 | Codebook Construction

Let $A = P_{\text{ave}}$. In the following, we confine ourselves to codewords that meet the condition $n^{-1}\sum_{t=1}^{n} c_{i,t} \leq A, \forall\, i \in [\![M]\!]$.

- □ **Case 1 - With Hamming Weight Constraint:** $A \leq 1$, then the condition $n^{-1}\sum_{t=1}^{n} c_{i,t} \leq 1, i \in [\![M]\!]$ is non-trivial in the sense that it induces a strict subset of the entire input space $\mathbf{H}^n$. We denote such subset by $\mathcal{B}_{\mathbf{0}}(n, nA)$ and is equivalent to $\|\mathbf{c}_i\|_1 \leq A$.
- □ **Case 2 - Without Hamming Weight Constraint:** $A \geq 1$, then each codeword belonging to the n-dimensional Hamming cube $\mathbf{H}^n$ fulfilled the Hamming weight constraint, since $\frac{1}{n}\sum_{t=1}^{n} c_{i,t} \leq 1 \leq A, i \in [\![M]\!]$. Therefore, we address the entire input space $\mathbf{H}^n = \{0,1\}^n$ as the possible set of codewords and attempt to exhaust it in a brute-force manner in the asymptotic, i.e., as $n \to \infty$.

▷ **Analysis For Case 1:** Observe that within this case, we again divide into two cases:

1. $0 < A < \frac{1}{2}$
2. $A \geq \frac{1}{2}$

[10] The early introduction of such bound in the literature is accomplished by Gilbert in [201].

The argument for the need of such division is that the binary entropy function is monotonic increasing only for $0 \leq A \leq \frac{1}{2}$ and for $A \geq \frac{1}{2}$ is decreasing. That is, in the latter case, we can introduce an alternative Bernoulli process which result in a larger volume space, and at the same time, it guarantees the Hamming weight constraint.

For the sub-case 1, i.e., where $0 < A < \frac{1}{2}$, we restrict our considerations to an n-dimensional Hamming hyper ball with edge length A. We use a packing arrangement of overlapping hyper balls of radius $r_0 = \lfloor n\beta \rfloor$ in an n-dimensional Hamming hyper ball $\mathcal{B}_{\mathbf{0}}(n, nA)$, where

Lemma 10.3.1. *Let $R < H(A)$ and let $\beta > 0$ be an arbitrary small constant. Then for sufficiently large codeword length n, there exist a codebook $\mathcal{C} = \{\mathbf{c}_i\}_{i \in [\![M]\!]} \subset \{0,1\}^n$, with $\mathbf{c}_i = (c_{i,t})|_{t=1}^n \subset \{0,1\}^n$, which consists of M sequences in the n-dimensional Hamming hyper ball $\mathcal{B}_{\mathbf{0}}(n, nA)$, such that the following holds:*

- *Hamming Distance Property:* $d_H(\mathbf{c}_i, \mathbf{c}_j) \geq \lfloor n\beta \rfloor + 1 \qquad \forall i, j \in [\![M]\!]$ *where* $i \neq j$.
- *Codebook Size: The codebook size is at least* 2^{nR-1}, *that is,* $M \geq 2^{n(R-\frac{1}{n})}$.

Proof. Recall that the minimum Hamming distance of a code $\mathcal{C}$ is given by

$$d_{\min} \triangleq \min_{(i,j) \in [\![M]\!] \times [\![M]\!]} d_{\mathrm{H}}(\mathbf{c}_i, \mathbf{c}_j) \,. \tag{10.29}$$

We begin to obtain some codeword that fulfill the Hamming weight constraint, namely,

$$\frac{1}{n} \sum_{t=1}^{n} c_t \leq A \,. \tag{10.30}$$

First, we generate a codeword $\mathbf{C} \overset{i.i.d}{\sim} \text{Bernoulli}(A)$[11]. Since $\mathbb{E}[C_t] = A$, by the *weak law of large numbers*, we obtain

$$\lim_{n \to \infty} \Pr\left(\left| \frac{1}{n} \sum_{t=1}^{n} C_t - A \right| \leq \tau \right) = 1 \,, \tag{10.31}$$

[11] Such a random generation should not be confused with a similar procedure as is accomplished in the encoding stage of the RI problem. While therein, each message is mapped to a codeword through a random distribution, here for the DI problem, we first solely restrict ourselves to generation of codewords through the Bernoulli distribution to guarantee the Hamming weight constraint, and employ them in the next procedure called the greedy construction up to an exhaustion. Then, after the exhaustion, we establish a deterministic mapping between the message set and the codebook, that is, each message is associated to a codeword. Further, in the RI problem, it is in general possible that two different message are mapped to a common codeword, however, considering the DI problem in here, there exist a one-to-one mapping between the set of messages and the set of codewords.

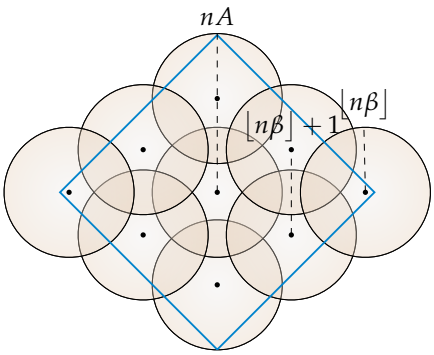


Figure 10.5: Illustration of an exhausted greedy-wise ball packing inside a hyper ball in 1-norm, where union of the small balls of radius $r_0 = \lfloor n\beta \rfloor$ cover a larger cube. As the codewords are assigned to the center of each ball lying inside a larger hyper ball according to the greedy construction, the 1-norm of a codeword is bounded by nA as required.

where $\tau > 0$ is an arbitrary small positive. Therefore, for sufficiently large codeword length n, the event $\left| n^{-1} \sum_{t=1}^{n} C_t - A \right| \leq \tau$ occurs with probability 1, which implies

$$\frac{1}{n} \sum_{t=1}^{n} C_t \leq A + \tau \,. \tag{10.32}$$

Now, observe that since (10.32) holds for arbitrary values of τ, it implies that the following condition for sufficiently large n, is fulfilled

$$\frac{1}{n} \sum_{t=1}^{n} C_t \leq A \,, \tag{10.33}$$

which is the Hamming weight constraint as required.

Next, we begin with the greedy procedure as follows: Let denote the first codeword determined by the Bernoulli distribution by $\mathbf{c}_1$ and assign it to message with index 1. Then, we remove all the sequences that have a Hamming distance of less or equal than $\lfloor n\beta \rfloor$ from $\mathbf{c}_1$. That is, we delete all the codewords that lies inside the Hamming ball with center $\mathbf{c}_1$ and radius $r = \lfloor n\beta \rfloor$. Then, we generate a second codeword by the Bernoulli distribution and repeat this procedure until all the sequences belonging to the legit subspace, i.e., the Hamming hyper ball in 1-norm; $\mathcal{B}_{\mathbf{0}}(n, nA)$, are exhausted. Therefore, such a construction fulfill the property provided in Lemma 10.3.1 regarding the minimum Hamming distance of the code, i.e.,

$$d_{\mathrm{H}}(\mathbf{c}_i, \mathbf{c}_j) \geq \lfloor n\beta \rfloor + 1 \,. \tag{10.34}$$

In general, the volume of a Hamming ball of radius r, assuming that the alphabet size is q, is the number of codewords that it encompass and is given by [202, see Ch. 1]

$$\mathrm{Vol}\left(\mathcal{B}_{\mathbf{x}}(n,r)\right) = \sum_{i=0}^{r} \binom{n}{i} (q-1)^i . \tag{10.35}$$

Let $\mathscr{B}$ denote the obtained ball packing after the exhaustion of the entire Hamming hyper ball $\mathcal{B}_{\mathbf{0}}$, i.e., an arrangement of M overlapping small hyper balls $\mathcal{B}_{\mathbf{c}_i}(n,r_0)$, with radius $r_0 = \lfloor n\beta \rfloor$ where $i \in [\![M]\!]$, that cover the entire Hamming hyper ball in 1-norm; $\mathcal{B}_{\mathbf{0}}(n,nA)$, where their centers are coordinated inside the $\mathcal{B}_{\mathbf{0}}(n,nA)$, and the distance between the closest centers is $\lfloor n\beta \rfloor + 1$; see Figure 10.5. As opposed to the standard ball packing observed in coding techniques [138], the balls here are neither necessarily entirely contained within the Hamming hyper ball, nor disjoint. That is, we only require that the centers of the balls are inside $\mathcal{B}_{\mathbf{0}}(n,nA)$ and have a non-empty intersection with $\mathcal{B}_{\mathbf{0}}(n,nA)$, which is rather a *ball covering problem*.

Th ball packing $\mathscr{B}$ is called *exhausted* if no point within the input space; $\mathcal{B}_{\mathbf{0}}(n,nA)$, is remained as an *isolated point*, that is, with the property that it does not belong to *at least* one of the small Hamming hyper balls.

In particular, we use a covering argument that has a similar flavor as that observed in the GV bound [203, Th. 5.1.7]. Specifically, consider an exhausted packing arrangement of

$$\bigcup_{i=1}^{M(n,R)} \mathcal{B}_{\mathbf{c}_i}(n, \lfloor n\beta \rfloor), \tag{10.36}$$

balls with radius $r_0 = \lfloor n\beta \rfloor$ embedded within the space $\mathcal{B}_{\mathbf{0}}(n,nA)$. According to the greedy construction, the center $\mathbf{c}_i$ of each small Hamming hyper ball, corresponds to a codeword. Since the volume of each hyper ball is equal to $\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n,r_0))$, the centers of all balls lie inside the space $\mathcal{B}_{\mathbf{0}}(n,nA)$, and the Hamming hyper balls *overlap* with each other, the total number of balls is bounded from below by

$$\begin{aligned} M &\geq \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{M} \mathcal{B}_{\mathbf{c}_i}(n,r_0)\right)}{\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n,r_0))} \\ &\overset{(a)}{\geq} \frac{\mathrm{Vol}\left(\mathcal{B}_{\mathbf{0}}(n,nA)\right)}{\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n,r_0))} \\ &\overset{(b)}{\geq} \frac{\sum_{j=0}^{\lfloor nA \rfloor} \binom{n}{j}}{\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n,r_0))}, \end{aligned} \tag{10.37}$$

where (a) holds since the Hamming hyper balls may have in general *intersection* and (b) follows by (10.35) with setting $q = 2$ and since $\lfloor nA \rfloor \leq nA$. Now, the bound in (10.37) can be further simplified as follows

$$\begin{aligned}\log M &\geq \log \left(\frac{\sum_{j=0}^{\lfloor nA \rfloor} \binom{n}{j}}{\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n, r_0))} \right) \\ &\overset{(a)}{\geq} nH(A) + o\left(\log n\right) - nH(\beta)\,. \end{aligned} \tag{10.38}$$

where (a) exploits (I.1) for setting radius $r = \lfloor n\varepsilon \rfloor = \lfloor nA \rfloor$ and $q = 2$, and (J.1) with $r_0 = \lfloor n\varepsilon \rfloor = \lfloor n\beta \rfloor$. Now, we obtain

$$\log M \geq nH(A) + o\left(\log n\right) - nH\left(\beta\right)\,, \tag{10.39}$$

where the dominant term has an order of n. Therefore, in order to obtain finite value for the lower bound on the DKI rate, R, (10.39) induces the scaling law of codebook size, M, to be 2^{nR}. Hence, we obtain

$$\begin{aligned} R &\geq \frac{1}{n}\left[nH(A) + o\left(\log n\right) - nH\left(\beta\right)\right]\,, \\ &= H(A) + \frac{o\left(\log n\right)}{n} - H\left(\beta\right)\,, \end{aligned} \tag{10.40}$$

which tends to $H(A)$ as $n \to \infty$ and $\beta \to 0$.

Now, we proceed to the sub-case 2, i.e., where $A \geq \frac{1}{2}$. In this case, instead of sticking to generation of codewords $\sim$ Bernoulli(A), we generate the codewords according to Bernoulli process with success probability of $\frac{1}{2}$, that is, $\mathbf{C} \overset{i.i.d}{\sim}$ Bernoulli$(\frac{1}{2})$. Observe that now the required Hamming weight constraint given in (10.30) is met, since for $\mathbb{E}\left[C_t\right] = \frac{1}{2}$ we have

$$\frac{1}{n}\sum_{t=1}^{n} c_t \leq \frac{1}{2} \leq A\,. \tag{10.41}$$

Therefore, following similar line of arguments as provided for the sub-case 1, we obtain the following lower bound on the DKI rate, R,

$$\begin{aligned} R &\geq \frac{1}{n}\left[nH\left(\frac{1}{2}\right) + o\left(\log n\right) - nH\left(\beta\right)\right]\,, \\ &= H\left(\frac{1}{2}\right) + \frac{o\left(\log n\right)}{n} - H\left(\beta\right)\,, \end{aligned} \tag{10.42}$$

which tends to $H\left(\frac{1}{2}\right) = 1$ as $n \to \infty$ and $\beta \to 0$. □

▷ **Analysis For Case 2:**

Lemma 10.3.2 (see [52, Claim 1]). *The entire Hamming cube* $\mathbf{H}^n$ *can be exhausted asymptotically as the codebook, that is, all the message sequences, i.e., the indices between* 1 *and* 2^m *can be coded with binary sequences of length* n*, subject to the Hamming distance property, i.e.,*

$$d_H(\mathbf{c}_i, \mathbf{c}_j) \geq \lfloor n\beta \rfloor + 1 \,. \tag{10.43}$$

for every $i, j \in [\![M]\!]$*, where* $i \neq j$ *and with* $\beta > 0$ *being an arbitrary small positive.*

Proof. Recall that the minimum Hamming distance of a code $\mathcal{C}$ is given by

$$d_{\min} \triangleq \min_{(i,j)\in[\![M]\!]\times[\![M]\!]} d_{\mathrm{H}}(\mathbf{c}_i, \mathbf{c}_j) \,. \tag{10.44}$$

Next, we begin with the greedy procedure as follows: Let denote the first codeword determined by the Bernoulli distribution by $\mathbf{c}_1$ and assign it to message with index 1. Then, we remove all the sequences that have a Hamming distance of less or equal than $\lfloor n\beta \rfloor$ from $\mathbf{c}_1$. That is, we delete all the codewords that lies inside the Hamming ball with center $\mathbf{c}_1$ and radius $r = \lfloor n\beta \rfloor$. Then, we generate a second codeword by the Bernoulli distribution and repeat this procedure until all the sequences are exhausted.

Let $\mathscr{B}$ denote the obtained ball packing after the exhaustion of the entire input space $\mathbf{H}^n = \{0,1\}^n$, i.e., an arrangement of M overlapping small hyper balls $\mathcal{B}_{\mathbf{c}_i}(n, r_0)$, with radius $r_0 = \lfloor n\beta \rfloor$ where $i \in [\![M]\!]$, that cover n-dimensional Hamming cube $\mathbf{H}^n = \{0,1\}^n$, where their centers are coordinated inside $\mathbf{H}^n$, and the distance between the closest centers is $\lfloor n\beta \rfloor + 1$. As opposed to the standard ball packing observed in coding techniques [138], the balls here are neither necessarily entirely contained within the Hamming hyper ball, nor disjoint. That is, we only require that the centers of the balls are inside $\mathbf{H}^n$ and have a non-empty intersection with $\mathbf{H}^n$, which is rather a *ball covering problem*. The ball packing $\mathscr{B}$ is called *exhausted* if no point within the input space; $\mathbf{H}^n$, is remained as an *isolated point*, that is, with the property that it does not belong to *at least* one of the small Hamming hyper balls.

In particular, we use a covering argument that has a similar flavor as that observed in the GV bound [203, Th. 5.1.7]. Specifically, consider an exhausted packing arrangement of

$$\bigcup_{i=1}^{M(n,R)} \mathcal{B}_{\mathbf{c}_i}(n, \lfloor n\beta \rfloor) \,, \tag{10.45}$$

balls with radius $r_0 = \lfloor n\beta \rfloor$ embedded within the space $\mathbf{H}^n$. According to the greedy construction, the center $\mathbf{c}_i$ of each small Hamming hyper ball, corresponds to a codeword. Since the volume of each hyper ball is equal to $\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n, r_0))$, the centers of all balls lie inside the space $\mathbf{H}^n$, and the Hamming hyper balls *overlap* with each other, the total number of balls is bounded from below by

$$\begin{aligned} M &\geq \frac{\mathrm{Vol}\left(\bigcup_{i=1}^{M} \mathcal{B}_{\mathbf{c}_i}(n, r_0)\right)}{\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n, r_0))} \\ &\overset{(a)}{\geq} \frac{\mathrm{Vol}\left(\mathbf{H}^n\right)}{\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n, r_0))} \\ &\overset{(b)}{\geq} \frac{|\mathcal{X}|^n}{\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n, r_0))}, \end{aligned} \tag{10.46}$$

where (a) holds since the Hamming hyper balls may have in general *intersection* and (b) follows since $\mathrm{Vol}\left(\mathbf{H}^n\right) = |\mathcal{X}^n| = |\mathcal{X}|^n$. Now, the bound in (10.46) can be further simplified as follows

$$\begin{aligned} \log M &\geq \log\left(\frac{|\mathcal{X}|^n}{\mathrm{Vol}(\mathcal{B}_{\mathbf{c}_1}(n, r_0))}\right) \\ &\overset{(a)}{\geq} n \log|\mathcal{X}| + o\left(\log n\right) - nH(\beta) \\ &\overset{(b)}{\geq} n + o\left(\log n\right) - nH(\beta). \end{aligned} \tag{10.47}$$

where (a) exploits (J.1) with $\varepsilon = \beta$. Now for $\beta > 0$ being an arbitrary small positive, we obtain

$$\begin{aligned} \log M &\geq n + o\left(\log n\right) - nH\left(\beta\right) \\ &= n\left(1 - H\left(\beta\right)\right) + o\left(\log n\right), \end{aligned} \tag{10.48}$$

where the dominant term has an order of n. Therefore, in order to obtain finite value for the lower bound on the DKI rate, R, (10.39) induces the scaling law of codebook size, M, to be 2^{nR}. Hence, we obtain

$$\begin{aligned} R &\geq \frac{1}{n}\left[n\left(1 - H\left(\beta\right)\right) + o\left(\log n\right)\right], \\ &= 1 - H\left(\beta\right) + \frac{o\left(\log n\right)}{n}, \end{aligned} \tag{10.49}$$

which tends to 1 as $n \to \infty$ and $\beta \to 0$. □

10.3.2.2 | Encoding

Given a message $i \in [\![M]\!]$, transmit $\mathbf{x} = \mathbf{c}_i$.

10.3.2.3 | Decoding

Let define $\delta_\beta \neq \frac{1}{2}$ as follows

$$\delta_\beta = \left(1 - \beta/2\right) \varepsilon + \beta/4\,, \tag{10.50}$$

which is referred to as the *decoding threshold* with $\beta > 0$ being an arbitrary small. Observe that given $0 < \varepsilon < 1/2$ and (10.50), we obtain the following double bound on the δ_β

$$\varepsilon < \delta_\beta < (1 - \beta)\,\varepsilon + \beta/2\,. \tag{10.51}$$

To identify whether message $j \in [\![M]\!]$ was sent, the decoder checks whether the channel output $\mathbf{y}$ belongs to the decoding set $\mathscr{T}_{\mathbb{K}} = \bigcup_{j\in\mathbb{K}} \mathbb{T}_j$, or not, with

$$\mathbb{T}_j = \left\{ \mathbf{y} \in \mathbf{H}^n \,;\, T(\mathbf{y}, c_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right\}, \tag{10.52}$$

where

$$T(\mathbf{y}, c_j) = d_{\mathrm{H}}(\mathbf{y}, \mathbf{c}_j) \triangleq \sum_{t=1}^{n} \delta_\beta(y_t, c_{j,t})\,, \tag{10.53}$$

is referred to as the *decoding metric* evaluated for observation vector $\mathbf{y}$ and the individual codeword $\mathbf{c}_j$, with $\delta_\beta(.,.)$ being the *Kronecker delta*. That is, given the channel output vector $\mathbf{y} \in \mathbf{H}^n$, if there exist at least one $j \in \mathbb{K}$ such that $d_{\mathrm{H}}(\mathbf{y}, \mathbf{c}_j) \leq \left\lfloor n\delta_\beta \right\rfloor$, then the decoder declares that the message j was sent. And for the other case, i.e., where for each index $j \in \mathbb{K}$, the inequality $d_{\mathrm{H}}(\mathbf{y}, \mathbf{c}_j) > \left\lfloor n\delta_\beta \right\rfloor$ holds, then the decoder decides that j was not sent.

Remark 10.3.1. ***Adopted Decoder****: For the achievability proof, we adopt a decoder which upon observing an output sequence* $\mathbf{y}$*, it declares that the message* $j \in \mathbb{K}$ *was sent if the output vector* $\mathbf{y}$ *belongs to the following set*

$$\bigcup_{j\in\mathbb{K}} \left\{ \mathbf{y} \in \mathbf{H}^n \,;\, d_H(\mathbf{y}, \mathbf{c}_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right\}, \tag{10.54}$$

where $\mathbf{c}_j = [c_{j,1}, \ldots, c_{j,n}]$ *is the codeword associated with message* j *and* δ_β *is a decoding threshold. We notice that the decoder in* (10.54) *combine the elements of set* $\mathbb{K}$ *through a*

fundamental union operator. Such a simple operator may feature a penalty with respect to the error exponents for the type I/II error probabilities or the obtained achievable rates. Therefore, we recall that in principle a more optimum decoder for the K-Identification scheme which guarantee vanishing type I/II error probabilities, might demand a more complicated algebraic operators between the realization of members for each specific set $\mathbb{K}$ and entails advanced dependencies on the elements of set $\mathbb{K}$.

10.3.2.4 | Error Analysis

Fix $e_1, e_2 > 0$ and let $\zeta_0, \zeta_1 > 0$ be arbitrarily small constants. Further, let introduce the following conventions:

- $Y_t(i)$ denote the channel output at time t *conditioned* that the sent codewords was $\mathbf{x} = \mathbf{c}_i$, that is, $Y_t(i) = \mathbf{c}_{i,t} \oplus Z_t$
- The output vector is defined as the vector of symbols, i.e., $\mathbf{Y}(i) \triangleq (Y_1(i), \ldots, Y_n(i))$

Consider the type I error, i.e., the transmitter sends $\mathbf{c}_i$, yet $\mathbf{y} \notin \mathscr{T}_{\mathbb{K}}$ for every $i \in \mathbb{K}$. The type I error probability is given by

$$\begin{aligned} P_{e,1}(i) &= \Pr\left(\mathbf{Y}(i) \in \mathscr{T}_{\mathbb{K}}^c\right) \\ &= \Pr\left(\mathbf{Y}(i) \in \left(\bigcup_{j\in\mathbb{K}} \mathbb{T}_j\right)^c\right) \\ &\overset{(a)}{=} \Pr\left(\mathbf{Y}(i) \in \bigcap_{j\in\mathbb{K}} \mathbb{T}_j^c\right) \\ &\overset{(b)}{\leq} \Pr\left(\mathbf{Y}(i) \in \mathbb{T}_i^c\right) \\ &= \Pr\left(T(\mathbf{Y}(i), c_i) > \left\lfloor n\delta_\beta \right\rfloor\right), \end{aligned} \tag{10.55}$$

where (a) holds by *De Morgan*'s law for a finite number of union of set, i.e., $\left(\bigcup_{i\in\mathbb{K}} \mathbb{T}_i\right)^c = \bigcap_{i\in\mathbb{K}} \mathbb{T}_i^c$ and (b) follows since $\bigcap_{j\in\mathbb{K}} \mathbb{T}_j^c \subset \mathbb{T}_i$. Now, observe that

$$\begin{aligned} \Pr\left(T(\mathbf{Y}(i), c_i) > \left\lfloor n\delta_\beta \right\rfloor\right) &\overset{(a)}{=} \Pr\left(d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_i) > \left\lfloor n\delta_\beta \right\rfloor\right) \\ &\overset{(b)}{=} \sum_{l=\left\lfloor n\delta_\beta \right\rfloor+1}^{n} \binom{n}{l} \varepsilon^l (1-\varepsilon)^{n-l}. \end{aligned} \tag{10.56}$$

where (a) follows by (10.53) and (b) holds by (10.26). In order to bound (10.56), we proceed to apply the bound provided in (L.1) given in Lemma L.0.1: Observe that

$$\begin{aligned}\frac{l}{n} &= \frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \\ &\overset{(a)}{>} \frac{n\delta_\beta}{n} \\ &= \delta_\beta \\ &\overset{(b)}{>} \varepsilon\,,\end{aligned} \tag{10.57}$$

where (a) follows since $x < \lfloor x \rfloor + 1$ for real x and (b) holds by (10.51). On the other hand,

$$\begin{aligned}\frac{l}{n} &= \frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \\ &\leq \frac{\max \left\lfloor n\delta_\beta \right\rfloor + 1}{n} \\ &\overset{(a)}{<} \frac{\left\lfloor n \max\left(\varepsilon + \beta\left(\frac{1}{2} - \varepsilon\right)\right) \right\rfloor + 1}{n} \\ &\overset{(b)}{<} \frac{\lfloor n/2 \rfloor + 1}{n} \\ &\overset{n\geq 3}{<} 1\,,\end{aligned} \tag{10.58}$$

where (a) follows by (10.51) and (b) holds since $\varepsilon + \beta\left(\frac{1}{2} - \varepsilon\right)$ is upper bounded by the boundary value of ε, i.e., where $\varepsilon = \frac{1}{2}$. Observe that the last inequality in (10.58) holds for sufficiently large n. Now, since the inequalities provided in (10.57) and (10.58) fulfill the conditions in Lemma L.0.1, we employ Lemma L.0.1 to establish the following lower bound on (10.56) as follows

$$\begin{aligned}&\Pr\left(T(\mathbf{Y}(i), c_i) > \left\lfloor n\delta_\beta \right\rfloor\right) \\ &= \sum_{l=\left\lfloor n\delta_\beta \right\rfloor + 1}^{n} \binom{n}{l} \varepsilon^l (1-\varepsilon)^{n-l}\end{aligned}$$

$$\leq \left[\frac{\left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right)(1-\varepsilon)}{\left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right)(1-\varepsilon) - \left[n - \left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right) \right] \varepsilon} \right] \cdot 2^{-n\left[T_\varepsilon \left(\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) - H\left(\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) \right]}. \tag{10.59}$$

Observe that the denominator in (10.59) is always a strict positive term, since assuming we arrive to a trivial inequality as follows

$$\left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right)(1-\varepsilon) > \left[n - \left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right) \right] \varepsilon \iff \tag{10.60}$$

$$\left\lfloor n\delta_\beta \right\rfloor + 1 - \varepsilon \left\lfloor n\delta_\beta \right\rfloor - \varepsilon > n\varepsilon - \varepsilon \left\lfloor n\delta_\beta \right\rfloor - \varepsilon \iff \tag{10.61}$$

$$\left\lfloor n\delta_\beta \right\rfloor + 1 > n\varepsilon \iff \tag{10.62}$$

$$\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} > \varepsilon, \tag{10.63}$$

which is already verified in (10.57). Now, we proceed to find a simplified upper bound on the left hand side coefficient in the bracket given in (10.59) as follow

$$\begin{aligned}
&\frac{\left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right)(1-\varepsilon)}{\left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right)(1-\varepsilon) - \left[n - \left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right) \right] \varepsilon} \\
&\overset{(a)}{=} \frac{\left(n\delta_\beta + 1 \right)(1-\varepsilon)}{\left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right) - \varepsilon \left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right) - n\varepsilon + \varepsilon \left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right)} \\
&\leq \frac{\left(n\delta_\beta + 1 \right)(1-\varepsilon)}{\left(\left\lfloor n\delta_\beta \right\rfloor + 1 \right) - n\varepsilon} \\
&\overset{(b)}{\leq} \frac{\left(n\delta_\beta + 1 \right)(1-\varepsilon)}{n\delta_\beta - n\varepsilon},
\end{aligned} \tag{10.64}$$

where (a) holds by exploiting $x \leq \lfloor x \rfloor$ for real x and simplifying the denominator by distributing ε over the bracket, and (b) follows since

$$n\delta_\beta < \left\lfloor n\delta_\beta \right\rfloor + 1 \iff \tag{10.65}$$

$$n\delta_\beta - n\varepsilon < \left\lfloor n\delta_\beta \right\rfloor + 1 - n\varepsilon \iff \tag{10.66}$$

$$\frac{1}{n\delta_\beta - n\varepsilon} > \frac{1}{\lfloor n\delta_\beta \rfloor + 1 - n\varepsilon}. \tag{10.67}$$

where the first inequality follows since $x < \lfloor x \rfloor + 1$ for real x. Thereby, employing (10.64) unto (10.59), we obtain

$$\begin{aligned}
\Pr\left(|T(\mathbf{Y}(i), c_i)| > \lfloor n\delta_\beta \rfloor\right) &= \sum_{l=\lfloor n\delta_\beta \rfloor + 1}^{n} \binom{n}{l} \varepsilon^l (1-\varepsilon)^{n-l} \\
&\leq \frac{\left(n\delta_\beta + 1\right)(1-\varepsilon)}{n\delta_\beta - n\varepsilon} \cdot 2^{-n\left[T_\varepsilon\left(\frac{\lfloor n\delta_\beta \rfloor + 1}{n}\right) - H\left(\frac{\lfloor n\delta_\beta \rfloor + 1}{n}\right)\right]} \\
&= \frac{\left(\delta_\beta + \frac{1}{n}\right)(1-\varepsilon)}{\delta_\beta - \varepsilon} \cdot 2^{-n\left[T_\varepsilon\left(\frac{\lfloor n\delta_\beta \rfloor + 1}{n}\right) - H\left(\frac{\lfloor n\delta_\beta \rfloor + 1}{n}\right)\right]} \\
&\triangleq \zeta_{1,n}\,.
\end{aligned} \tag{10.68}$$

Observe that the exponent of exponential term is always *strictly* positive, since for $\varepsilon \in (0, \frac{1}{2})$, the arguments of $T_\varepsilon(.)$ and $H(.)$ are strictly less than $\frac{1}{2}$. That is, we have the following

$$T_\varepsilon\left(\frac{\lfloor n\delta_\beta \rfloor + 1}{n}\right) > H\left(\frac{\lfloor n\delta_\beta \rfloor + 1}{n}\right), \tag{10.69}$$

The argument is as follows

$$\begin{aligned}
\frac{l}{n} &= \frac{\lfloor n\delta_\beta \rfloor + 1}{n} \\
&\leq \frac{\max \lfloor n\delta_\beta \rfloor + 1}{n} \\
&\overset{(a)}{<} \frac{\left\lfloor n \max\left(\varepsilon + \beta\left(\frac{1}{2} - \varepsilon\right)\right)\right\rfloor + 1}{n} \\
&\overset{(b)}{<} \frac{\lfloor n/2 \rfloor + 1}{n} \\
&\overset{(c)}{\leq} \frac{n/2 + 1}{n} \\
&\overset{(c)}{\leq} \frac{n\left(\frac{1}{2} + 1/n\right)}{n},
\end{aligned} \tag{10.70}$$

which is strictly less than $\frac{1}{2}$ in the asymptotic, i.e., as $n \to \infty$, where (a) and (b) follows by the same arguments given for (10.58), and (c) follows since $\lfloor x \rfloor \leq x$ for real x.

Therefore, the difference for the evaluation of $T_\varepsilon(.)$ and $H(.)$ is always a *strict* positive value; see Figure 10.1. Hence, $P_{e,1}(i) \leq e_1$, $\forall i \in \mathbb{T}_{\mathbb{K}}$ holds for sufficiently large n and arbitrarily small $e_1 > 0$. Thereby, the type I error probability satisfies $P_{e,1}(i) \leq \zeta_{1,n} \leq e_1$. This complete the analysis for the type I error probability.

Next, we address type II errors, i.e., when $\mathbf{Y}(i) \in \mathbb{T}_{\mathbb{K}}$ while the transmitter sent $\mathbf{c}_i$ with $i \notin \mathbb{K}$. Then, for each possible $\binom{M}{K}$ cases of $\mathbb{K}$, where $i \notin \mathbb{K}$, the type II error probability is given by

$$\begin{aligned} P_{e,2}\left(i, \mathbb{K}\right) &= \Pr\left(\mathbf{Y}(i) \in \mathbb{T}_{\mathbb{K}}\right) \\ &= \Pr\left(\mathbf{Y}(i) \in \bigcup_{j \in \mathbb{K}} \mathbb{T}_j\right) \\ &\overset{(a)}{=} \Pr\left(\bigcup_{j \in \mathbb{K}} \left\{ T(\mathbf{Y}(i), c_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right\}\right) \\ &\overset{(b)}{=} \Pr\left(\bigcup_{j \in \mathbb{K}} \left\{ d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right\}\right) \\ &\overset{(c)}{\leq} \sum_{j \in \mathbb{K}} \Pr\left(d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right) \\ &\leq K \cdot \Pr\left(d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right) \end{aligned} \tag{10.71}$$

where (a) follows by (10.52), (b) holds by (10.53) and (c) follows by the *union bound*, i.e., the probability of union of events is upper bounded by the sum of probability of the individual events. Let define the following events

$$\mathcal{F}_{\delta_\beta}(i) \triangleq \left\{ \mathbf{Y} \in \mathbf{H}^n \, ; \, d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_i) \leq \left\lfloor n\delta_\beta \right\rfloor \right\} , \tag{10.72}$$

$$\mathcal{F}_{\delta_\beta}(i,j) \triangleq \left\{ \mathbf{Y} \in \mathbf{H}^n \, ; \, d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right\} . \tag{10.73}$$

Next, employing the *law of total probability* with respect to the event $\left\{ d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_i) \leq \left\lfloor n\delta_\beta \right\rfloor \right\}$, we establish an upper bound on $\Pr\left(d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right)$ given in (10.71) as follows

$$\Pr\left(d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_j) \leq \left\lfloor n\delta_\beta \right\rfloor \right) \overset{(a)}{=} \Pr\left(\mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}_{\delta_\beta}(i) \right) + \Pr\left(\mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}^c_{\delta_\beta}(i) \right)$$

$$\begin{aligned}
&\overset{(b)}{\leq} \Pr\left(\mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}_{\delta_\beta}(i)\right) + \Pr\left(\mathcal{F}^c_{\delta_\beta}(i)\right) \\
&\overset{(c)}{=} \Pr\left(\mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}_i(\delta_\beta)\right) + \Pr\left(d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_i) > \left\lfloor n\delta_\beta \right\rfloor\right) \\
&\overset{(d)}{\leq} \Pr\left(\mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}_{\delta_\beta}(i)\right) + \zeta_{1,n}\,, \qquad (10.74)
\end{aligned}$$

where (a) holds by the *law of total probability*, (b) follows since $\mathcal{F}^c_i(\delta_\beta) \supset \mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}^c_i(\delta_\beta)$, (c) holds by (10.72), and (d) exploits (10.68).

Now we focus on the event $\mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}_{\delta_\beta}(i)$. Let

$$\begin{aligned}
d &\triangleq d_{\mathrm{H}}(\mathbf{c}_i, \mathbf{c}_j) \\
&\overset{(a)}{\geq} \left\lfloor n\beta \right\rfloor + 1\,. \qquad (10.75)
\end{aligned}$$

where (a) follows by the assumption made in the code construction regarding the minimum Hamming distance; see Lemma 10.3.1 and (10.43). Now, without loss of generality, we may assume that the two sequence $\mathbf{c}_i$ and $\mathbf{c}_j$ differ in the first d symbols, i.e.,

$$\begin{aligned}
\mathbf{c}_i &= \left(c_{i_1}, c_{i_2}, \ldots, c_{i_d}, c_{i_{d+1}}, \ldots, c_{i_n}\right) \\
\mathbf{c}_j &= \left(c_{j_1}, c_{j_2}, \ldots, c_{j_d}, c_{j_{d+1}}, \ldots, c_{j_n}\right) \\
\mathbf{y} &= \left(y_1, y_2, \ldots, y_d, y_{d+1}, \ldots, y_n\right)\,, \qquad (10.76)
\end{aligned}$$

where $\mathbf{y}$ is the realization of vector $\mathbf{Y}(i)$. Therefore, the $n-d$ last symbols (bits) of $\mathbf{c}_i$ and $\mathbf{c}_j$ are identical. Observe that the event $\left\{d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_i) \leq \left\lfloor n\delta_\beta \right\rfloor\right\}$ implies that the received vector $\mathbf{y}$ and $\mathbf{c}_i$ differ in p bits, where $p \leq \left\lfloor n\delta_\beta \right\rfloor$, i.e.,

$$d_{\mathrm{H}}(\mathbf{y}, \mathbf{c}_i) = p \leq \left\lfloor n\delta_\beta \right\rfloor\,. \qquad (10.77)$$

Now, we assume that p_1 bits out of the p bits happens in the first d bits, i.e.,

$$d_{\mathrm{H}}(\mathbf{y}|_1^d, \mathbf{c}_i|_1^d) = p_1\,, \qquad (10.78)$$

where

$$\begin{aligned}
\mathbf{c}_i|_1^d &\triangleq \left(c_{i_1}, c_{i_2}, \ldots, c_{i_d}\right) \\
\mathbf{y}|_1^d &\triangleq \left(y_1, y_2, \ldots, y_d\right)\,, \qquad (10.79)
\end{aligned}$$

and p_2 bits with $p_2 = p - p_1$ happens in last $n-d$ bits, i.e.,

$$d_{\mathrm{H}}(\mathbf{y}|_{d+1}^n, \mathbf{c}_i|_{d+1}^n) = p_2\,, \qquad (10.80)$$

where

$$\begin{aligned}\mathbf{c}_i|_{d+1}^{n} &\triangleq \left(c_{i_{d+1}}, \ldots, c_{i_n}\right)\\ \mathbf{y}|_{d+1}^{n} &\triangleq \left(y_{d+1}, \ldots, y_n\right),\end{aligned} \tag{10.81}$$

Observe that since the symbols of sequences are bits, i.e., either 0 or 1, therefore, $d = d_{\mathrm{H}}(\mathbf{c}_i, \mathbf{c}_j)$ implies that the two sequences $\mathbf{c}_i$ and $\mathbf{c}_j$ are complementary for the first d bits. Now, we infer that if the two sequences $\mathbf{y}|_1^d$ and $\mathbf{c}_i|_1^d$ differ in p_1, then $\mathbf{y}|_1^d$ and $\mathbf{c}_i|_1^d$ are identical in those p_1 bits. Hence,

$$d_{\mathrm{H}}(\mathbf{y}|_1^d, \mathbf{c}_j|_1^d) = d - p_1\,, \tag{10.82}$$

Now, if we collect all the positions for which $\mathbf{y}|_1^n$ and $\mathbf{c}_j|_1^n$ differ, we obtain

$$\begin{aligned}d_{\mathrm{H}}(\mathbf{y}, \mathbf{c}_j) &= d_{\mathrm{H}}(\mathbf{y}|_1^n, \mathbf{c}_j|_1^n)\\ &= d_{\mathrm{H}}(\mathbf{y}|_1^d, \mathbf{c}_j|_1^d) + d_{\mathrm{H}}(\mathbf{y}|_{d+1}^n, \mathbf{c}_j|_{d+1}^n)\\ &= d - p_1 + p_2\,.\end{aligned} \tag{10.83}$$

Observe that since we restrict ourselves to the event

$$\mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}_i^c(\delta_\beta) \triangleq \left\{d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_j) \le \left\lfloor n\delta_\beta \right\rfloor\right\} \cap \left\{d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_i) \le \left\lfloor n\delta_\beta \right\rfloor\right\}, \tag{10.84}$$

we deduce that $d_{\mathrm{H}}(\mathbf{y}|_1^n, \mathbf{c}_j|_1^n)$, therefore by (10.83), we obtain

$$\begin{aligned}d - p_1 + p_2 &\le \left\lfloor n\delta_\beta \right\rfloor \Rightarrow\\ p_2 &\le \left\lfloor n\delta_\beta \right\rfloor - d + p_1\,.\end{aligned} \tag{10.85}$$

On the other hand, since $d_{\mathrm{H}}(\mathbf{y}, \mathbf{c}_j) \le \left\lfloor n\delta_\beta \right\rfloor$, we obtain

$$p \le \left\lfloor n\delta_\beta \right\rfloor \Rightarrow \tag{10.86}$$

$$p_1 + p_2 \le \left\lfloor n\delta_\beta \right\rfloor \Rightarrow \tag{10.87}$$

$$p_2 \le \left\lfloor n\delta_\beta \right\rfloor - p_1\,. \tag{10.88}$$

Now in order to calculate $\Pr(d_{\mathrm{H}}(\mathbf{Y}(i), \mathbf{c}_j) \le \left\lfloor n\delta_\beta \right\rfloor)$ in (10.71), we first fix p_1 and then sum up over all possible cases for the p_2, then we would have a second sum which runs for values of p_1 from 0 to d. Observe that the p_2 has two upper bounds given in (10.85) and (10.86), therefore, in the calculation, we restrict ourselves to the minimum of those two upper bounds. Let define

$$p_2^{UB} \triangleq \min\left\{\left\lfloor n\delta_\beta \right\rfloor - p_1, \left\lfloor n\delta_\beta \right\rfloor - d + p_1\right\} \tag{10.89}$$

Thereby,

$$\begin{aligned}
&\Pr\left(\mathcal{F}_{\delta_\beta}(i,j)\cap\mathcal{F}_{\delta_\beta}(i)\right)\\
&\overset{(a)}{\leq}\sum_{p_1=0}^{d}\binom{d}{p_1}\cdot\sum_{p_2=0}^{p_2^{UB}}\binom{n-d}{p_2}\varepsilon^{p_1+p_2}(1-\varepsilon)^{n-(p_1+p_2)+d-d}\\
&\overset{(b)}{=}\left[\sum_{p_1=0}^{d}\binom{d}{p_1}\varepsilon^{p_1}(1-\varepsilon)^{d-p_1}\right]\cdot\left[\sum_{p_2=0}^{p_2^{UB}}\binom{n-d}{p_2}\varepsilon^{p_2}(1-\varepsilon)^{n-d-p_2}\right],
\end{aligned}\tag{10.90}$$

where (a) holds since $p=p_1+p_2$, and (b) follows since every expression that is independent of the sum's variable can be shifted left behind the inner sum. In (b), we have added $0=d-d$, to obtain the correct form for the two binomial distribution expressions. Now, observe that the first sum is the Binomial cumulative distribution function at point $x=d$ and can be upper bounded by 1, i.e.,

$$\begin{aligned}
\sum_{p_1=0}^{d}\binom{d}{p_1}\varepsilon^{p_1}(1-\varepsilon)^{d-p_1}&=\Pr\left(p_1\leq d\right)\\
&=B_X(x)|_{x=d}\\
&=B_X(d)\\
&=1.
\end{aligned}\tag{10.91}$$

Now, let focus on the second sum in (10.90) for which we establish an upper bound by maximizing p_2^{UB} through setting $p_1=\left\lfloor\frac{d}{2}\right\rfloor$, i.e.,

$$\arg\max_{p_1} p_2^{UB}=\left\lfloor\frac{d}{2}\right\rfloor\tag{10.92}$$

$$\begin{aligned}
\max p_2^{UB}&\triangleq\max\min\left\{\left\lfloor n\delta_\beta\right\rfloor-p_1,\left\lfloor n\delta_\beta\right\rfloor-d+p_1\right\}\\
&=\min\left\{\left\lfloor n\delta_\beta\right\rfloor-p_1,\left\lfloor n\delta_\beta\right\rfloor-d+p_1\right\}\Big|_{p_1=\left\lfloor\frac{d}{2}\right\rfloor}\\
&=\left\{\left\lfloor n\delta_\beta\right\rfloor-\left\lfloor\frac{d}{2}\right\rfloor,\left\lfloor n\delta_\beta\right\rfloor-d+\left\lfloor\frac{d}{2}\right\rfloor\right\}\\
&=\left\{\left\lfloor n\delta_\beta\right\rfloor-\left\lfloor\frac{d}{2}\right\rfloor,\left\lfloor n\delta_\beta\right\rfloor-\left(d-\left\lfloor\frac{d}{2}\right\rfloor\right)\right\}
\end{aligned}$$

$$= \left\lfloor n\delta_\beta \right\rfloor - d + \left\lfloor \frac{d}{2} \right\rfloor , \tag{10.93}$$

where the last equality holds since by $\left\lfloor \frac{d}{2} \right\rfloor \leq \frac{d}{2}$ for real $\frac{d}{2}$, we obtain $\frac{d}{2} \leq d - \left\lfloor \frac{d}{2} \right\rfloor$.

Now, we exploit the inequality (M.1) given in Lemma M.0.1 to obtain an upper bound for the second sum in (10.90) as follows: First we check whether the required condition in Lemma M.0.1 are satisfied or not. Namely, we set $k = \left\lfloor n\delta_\beta \right\rfloor - d + \left\lfloor \frac{d}{2} \right\rfloor$ and $n = n - d$. Now we calculate their ratio as follow

$$\begin{aligned}
\frac{k}{n-d} &= \frac{\left\lfloor n\delta_\beta \right\rfloor - d + \left\lfloor \frac{d}{2} \right\rfloor}{n-d} \\
&\overset{(a)}{\leq} \frac{n\delta_\beta - d + \frac{d}{2}}{n-d} \\
&= \frac{n\delta_\beta - \frac{d}{2}}{n-d} \\
&= \frac{\delta_\beta - \frac{d}{2n}}{1 - \frac{d}{n}} \\
&\overset{(b)}{<} \frac{\delta_\beta - \frac{\beta}{2}}{1-\beta} \\
&\triangleq \tau ,
\end{aligned} \tag{10.94}$$

where (a) holds since $\lfloor x \rfloor \leq x$ for real x and (b) holds by the following argument: We assume that (b) holds and assuming that $\delta_\beta \neq \frac{1}{2}$, we resulted in a trivial inequality, namely, $d > n\beta$, i.e,

$$\frac{\delta_\beta - \frac{d}{2n}}{1 - \frac{d}{n}} < \frac{\delta_\beta - \frac{\beta}{2}}{1-\beta} \Rightarrow \tag{10.95}$$

$$\left(\delta_\beta - \frac{d}{2n}\right)(1-\beta) < \left(\delta_\beta - \frac{\beta}{2}\right)\left(1 - \frac{d}{n}\right) \Rightarrow \tag{10.96}$$

$$\delta_\beta - \beta\delta_\beta - \frac{d}{2n} + \frac{\beta d}{2n} < \delta_\beta - \frac{\delta_\beta d}{n} - \frac{\beta}{2} + \frac{\beta d}{2n} \Rightarrow \tag{10.97}$$

$$\beta\left(\frac{1}{2} - \delta_\beta\right) < \frac{d}{2n} - \frac{\delta_\beta d}{n} \Rightarrow \tag{10.98}$$

$$\beta\left(\frac{1}{2} - \delta_\beta\right) < \frac{d}{n}\left(\frac{1}{2} - \delta_\beta\right) \Rightarrow \tag{10.99}$$

$$n\beta < d , \tag{10.100}$$

which can be deduced from the assumption made in the code construction given in

(10.43) and (10.43), i.e.,

$$\begin{aligned} d_{\mathrm{H}}(\mathbf{c}_i, \mathbf{c}_j) &\geq \lfloor n\beta \rfloor + 1 \\ &\overset{(a)}{>} n\beta - 1 + 1 \\ &= n\beta \,. \end{aligned} \tag{10.101}$$

where (a) holds since $\lfloor n\beta \rfloor > n\beta - 1$ for real $n\beta$. Now, we exploit (10.51), to show that (10.94) is upper bounded by ε as follows

$$\begin{aligned} \delta_\beta &< \varepsilon + \beta\left(\frac{1}{2} - \varepsilon\right) \Rightarrow \\ \delta_\beta &< \varepsilon + \frac{\beta}{2} - \beta\varepsilon \Rightarrow \\ \delta_\beta - \frac{\beta}{2} &< \varepsilon(1-\beta) \Rightarrow \\ \frac{\delta_\beta - \frac{\beta}{2}}{1-\beta} &< \varepsilon \,. \end{aligned} \tag{10.102}$$

Thereby, we apply safely the Lemma M.0.1 with parameters $j = p_2$, $k = p_2^{\mathrm{UB}} \triangleq \left\lfloor n\delta_\beta \right\rfloor - d + \left\lfloor \frac{d}{2} \right\rfloor$ and $n = n - d$, and obtain

$$\begin{aligned} \sum_{p_2=0}^{\left\lfloor n\delta_\beta \right\rfloor - d + \left\lfloor \frac{d}{2} \right\rfloor} \binom{n-d}{p_2} \varepsilon^{p_2}(1-\varepsilon)^{n-d-p_2} &\leq \frac{\varepsilon((n-d)-k)}{\varepsilon(n-d)-k} \cdot 2^{n\left[H(\frac{k}{n-d}) - T_\varepsilon(\frac{k}{n-d})\right]} \\ &\leq \frac{\varepsilon\left(1 - \frac{k}{n-d}\right)}{\varepsilon - \frac{k}{n-d}} \cdot 2^{n\left[H(\frac{k}{n-d}) - T_\varepsilon(\frac{k}{n-d})\right]} \,. \end{aligned} \tag{10.103}$$

Let focus on the coefficient in (10.103). In the following, assuming an upper bound for it, we arrive to a trivial inequality, therefore, the upper bound is valid.

$$\frac{\varepsilon\left(1 - \frac{k}{n-d}\right)}{\varepsilon - \frac{k}{n-d}} < \frac{\varepsilon(1-\tau)}{\varepsilon - \tau} \,. \tag{10.104}$$

Observe that (10.104) yield the following chain of expressions:

$$\frac{1 - \frac{k}{n-d}}{\varepsilon - \frac{k}{n-d}} < \frac{1-\tau}{\varepsilon - \tau} \Rightarrow \tag{10.105}$$

$$\varepsilon - \tau - \frac{k\varepsilon}{n-d} + \frac{k\tau}{n-d} < \varepsilon - \frac{k}{n-d} - \varepsilon\tau + \frac{k\tau}{n-d} \Rightarrow \tag{10.106}$$

$$-\tau - \frac{k\varepsilon}{n-d} < -\frac{k}{n-d} - \varepsilon\tau \Rightarrow \tag{10.107}$$

$$\frac{k}{n-d}(1-\varepsilon) < \tau(1-\varepsilon) \Rightarrow \tag{10.108}$$

$$\frac{k}{n-d} < \varepsilon\,, \tag{10.109}$$

which is trivial since it is already proved in 10.94. Now, observe that for $0 < \frac{k}{n-d} < \tau < \varepsilon$, the following holds

$$H\left(\frac{k}{n-d}\right) - T_\varepsilon\left(\frac{k}{n-d}\right) < H(\tau) - T_\varepsilon(\tau)\,, \tag{10.110}$$

see Figure 10.2. Therefore, since τ always yield a smaller exponent, we obtain an upper bound on the sum in (10.103) as follows

$$\begin{aligned}
\sum_{p_2=0}^{\lfloor n\delta_\beta\rfloor - d + \lfloor \frac{d}{2}\rfloor} \binom{n-d}{p_2} \varepsilon^{p_2}(1-\varepsilon)^{n-d-p_2} &\leq \frac{\varepsilon((n-d)-k)}{\varepsilon(n-d)-k} \cdot 2^{n\left[H(\frac{k}{n-d}) - T_\varepsilon(\frac{k}{n-d})\right]} \\
&\overset{(a)}{<} \frac{\varepsilon(1-\tau)}{\varepsilon-\tau} \cdot 2^{n\left[H(\frac{k}{n-d}) - T_\varepsilon(\frac{k}{n-d})\right]} \\
&\overset{(b)}{<} \frac{\varepsilon\left(1-\frac{k}{n-d}\right)}{\varepsilon-\frac{k}{n-d}} \cdot 2^{n\left[H(\tau) - T_\varepsilon(\tau)\right]} \\
&\triangleq \zeta_{0,n}\,,
\end{aligned} \tag{10.111}$$

where (a) exploits (10.104) and (b) follows by (10.110). Thereby, recalling (10.90) and employing (10.91), we obtain

$$\begin{aligned}
\Pr\left(\mathcal{F}_{\delta_\beta}(i,j) \cap \mathcal{F}_{\delta_\beta}(i)\right) &\leq 1 \cdot \sum_{j=0}^{k} \binom{n-d}{j} \varepsilon^j (1-\varepsilon)^{n-d-j} \\
&< \frac{\varepsilon(1-\tau)}{\varepsilon-\tau} \cdot 2^{n\left[H(\tau) - T_\varepsilon(\tau)\right]} \\
&\triangleq \zeta_{0,n}\,.
\end{aligned} \tag{10.112}$$

Hence, recalling (10.71) and (10.74) we obtain

$$
\begin{aligned}
P_{e,2}\left(i,\mathbb{K}\right) &\leq K\cdot\left[\Pr\left(d_{\mathrm{H}}(\mathbf{Y}(i),\mathbf{c}_j)\leq\left\lfloor n\delta_\beta\right\rfloor\right)\right]\\
&\leq K\cdot\left[\Pr\left(\mathcal{F}_{\delta_\beta}(i,j)\cap\mathcal{F}_{\delta_{\bar\beta}}(i)\right)+\zeta_{1,n}\right]\\
&= K\cdot\left[\frac{\varepsilon(1-\tau)}{\varepsilon-\tau}\cdot 2^{n\left[H(\tau)-T_\varepsilon(\tau)\right]}+\frac{\left(\delta_\beta+\frac{1}{n}\right)(1-\varepsilon)}{\delta_\beta-\varepsilon}\cdot 2^{-n\left[T_\varepsilon\left(\frac{\left\lfloor n\delta_\beta\right\rfloor+1}{n}\right)-H\left(\frac{\left\lfloor n\delta_\beta\right\rfloor+1}{n}\right)\right]}\right]\\
&\overset{(a)}{=} 2^{\kappa n}\cdot\left[\frac{\varepsilon(1-\tau)}{\varepsilon-\tau}\cdot 2^{-n\left[T_\varepsilon(\tau)-H(\tau)\right]}+\frac{\left(\delta_\beta+\frac{1}{n}\right)(1-\varepsilon)}{\delta_\beta-\varepsilon}\cdot 2^{-n\left[T_\varepsilon\left(\frac{\left\lfloor n\delta_\beta\right\rfloor+1}{n}\right)-H\left(\frac{\left\lfloor n\delta_\beta\right\rfloor+1}{n}\right)\right]}\right]\\
&= \frac{\varepsilon(1-\tau)}{\varepsilon-\tau}\cdot 2^{-n\left[T_\varepsilon(\tau)-H(\tau)-\kappa\right]}+\frac{\left(\delta_\beta+\frac{1}{n}\right)(1-\varepsilon)}{\delta_\beta-\varepsilon}\cdot 2^{-n\left[T_\varepsilon\left(\frac{\left\lfloor n\delta_\beta\right\rfloor+1}{n}\right)-H\left(\frac{\left\lfloor n\delta_\beta\right\rfloor+1}{n}\right)-\kappa\right]},
\end{aligned}
\tag{10.113}
$$

which implies that both the exponential factors given in (10.113) should yields strict positive exponents, that is, we obtain two separate upper bounds on the κ as follows

$$
\kappa<T_\varepsilon(\tau)-H(\tau)\qquad\text{and}\qquad\kappa<T_\varepsilon\left(\frac{\left\lfloor n\delta_\beta\right\rfloor+1}{n}\right)-H\left(\frac{\left\lfloor n\delta_\beta\right\rfloor+1}{n}\right),
\tag{10.114}
$$

Therefore,

$$\kappa < \min \left\{ T_\varepsilon(\tau) - H(\tau),\, T_\varepsilon \left(\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) - H \left(\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) \right\}, \tag{10.115}$$

Now we focus on the first argument in (10.115). In order to find an asymptotic value for the argument τ, we calculate the following

$$\begin{aligned} \lim_{\beta \to 0} \tau &\overset{(a)}{=} \lim_{\beta \to 0} \frac{\delta_\beta - \frac{\beta}{2}}{1 - \beta} \\ &= \delta_\beta \,, \end{aligned} \tag{10.116}$$

where (a) holds by (10.94). Thereby,

$$\lim_{n \to \infty} T_\varepsilon(\tau) - H(\tau) = T_\varepsilon \left(\delta_\beta \right) - H \left(\delta_\beta \right) , \tag{10.117}$$

Now we focus on the second argument in (10.115) and provide the following asymptotic behavior:

$$\begin{aligned} &\lim_{n \to \infty} T_\varepsilon \left(\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) - H \left(\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) \\ &= T_\varepsilon \left(\lim_{n \to \infty} \frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) - H \left(\lim_{n \to \infty} \frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) , \end{aligned} \tag{10.118}$$

where the equality holds since $T_\varepsilon(.)$ and $H(.)$ are continuous functions of δ_β. Now, observe that since $\left\lfloor n\delta_\beta \right\rfloor - 1 < \left\lfloor n\delta_\beta \right\rfloor \leq n\delta_\beta$ for real $n\delta_\beta$, we obtain

$$\begin{aligned} \lim_{n \to \infty} \frac{n\delta_\beta - 1 + 1}{n} \leq \lim_{n \to \infty} \frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \leq \lim_{n \to \infty} \frac{n\delta_\beta + 1}{n} \Rightarrow \\ \delta_\beta \leq \lim_{n \to \infty} \frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \leq \lim_{n \to \infty} \delta_\beta + \frac{1}{n} \overset{(a)}{\Rightarrow} \\ \lim_{n \to \infty} \frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} = \delta_\beta \,. \end{aligned} \tag{10.119}$$

where (a) holds by the *squeeze theorem*. Thereby,

$$\lim_{n \to \infty} T_\varepsilon \left(\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) - H \left(\frac{\left\lfloor n\delta_\beta \right\rfloor + 1}{n} \right) = T_\varepsilon \left(\delta_\beta \right) - H \left(\delta_\beta \right) . \tag{10.120}$$

Therefore, recalling (10.115), we obtain the following upper bound on the target identification rate κ:

$$\begin{aligned}\kappa &< \min\left\{T_\varepsilon(\tau) - H(\tau),\, T_\varepsilon\left(\frac{\lfloor n\delta_\beta\rfloor + 1}{n}\right) - H\left(\frac{\lfloor n\delta_\beta\rfloor + 1}{n}\right)\right\}\\ &= \min\left\{T_\varepsilon\left(\delta_\beta\right) - H\left(\delta_\beta\right),\, T_\varepsilon\left(\delta_\beta\right) - H\left(\delta_\beta\right)\right\}\\ &= T_\varepsilon\left(\delta_\beta\right) - H\left(\delta_\beta\right),\end{aligned} \tag{10.121}$$

where the equality holds since $T_\varepsilon(.)$ and $H(.)$ are continuous functions of δ_β. Therefore, recalling (10.113), we obtain

$$\begin{aligned}P_{e,2}(i,j) &\le \Pr\left(\mathcal{F}_{\delta_\beta}(i,j)\cap\mathcal{F}_{\delta_\beta}(i)\right) + \Pr\left(d_{\mathrm{H}}(\mathbf{Y}(i),\mathbf{c}_i) > \lfloor n\delta_\beta\rfloor\right)\\ &\le \zeta_{0,n} + \zeta_{1,n}\\ &\le \zeta_0 + \zeta_1\\ &\le e_2\,,\end{aligned} \tag{10.122}$$

hence, $P_{e,2}(i,j) \le e_2$ holds for sufficiently large n and arbitrarily small $e_2 > 0$. We have thus shown that for every $e_1, e_2 > 0$ and sufficiently large n, there exists an $(n, M(n,R), K(n,\kappa), e_1, e_2)$ code.

10.3.3 | Upper Bound (Converse Proof)

The converse proof consists of employing the following lemma on the size of a DKI code. In particular, depending on whether or not a Hamming weight constraint is present, we divide in two cases and address them separately. More specifically, we use the following observation. Let $R > 0$ be a DKI achievable rate. We assume to the contrary that there exist two distinct messages i_1 and i_2 that are represented by a common codeword, i.e., $\mathbf{c}_{i_1} = \mathbf{c}_{i_2} = x^n$, and show that this assumption result in a contradiction, namely, the sum of type I and type II error probabilities converges to one from left, i.e.,

$$\lim_{n\to\infty} P_{e,1}(i_1) + P_{e,2}(i_2, \mathbb{K}) = 1\,. \tag{10.123}$$

Hence our assumption is false and the number of messages 2^{nR} is bounded by either the size of the subset of the input sequences that satisfy the input constraint or the entire input space.

Lemma 10.3.3. *Consider a sequence of* $(n, M(n,R), K(n,\kappa), e_1^{(n)}, e_2^{(n)})$ *codes* $(\mathcal{C}^{(n)}, \mathscr{T}^{(n)})$ *such that* $e_1^{(n)}$ *and* $e_2^{(n)}$ *tend to zero as* $n \to \infty$. *Then, given a sufficiently large* n, *the codebook* $\mathcal{C}^{(n)}$ *satisfies the following property: There cannot be two distinct messages* $i_1, i_2 \in [\![M]\!]$ *that are represented by the same codeword, i.e.,*

$$i_1 \neq i_2 \quad \Rightarrow \quad \mathbf{c}_{i_1} \neq \mathbf{c}_{i_2} \,. \tag{10.124}$$

Proof. Assume to the contrary that there exist two messages i_1 and i_2, where $i_1 \neq i_2$, such that

$$\mathbf{c}_{i_1} = \mathbf{c}_{i_1} = x^n \,, \tag{10.125}$$

for some $x^n \in \mathcal{X}^n$. Since $(\mathcal{C}^{(n)}, \mathscr{T}^{(n)})$ forms a $(n, M(n,R), K(n,\kappa), e_1^{(n)}, e_2^{(n)})$ code, it implies that for every possible arrangement of $\left\{\mathbb{K}, \mathbb{K}^c\right\}$, e_1 and e_2 tends to zero. Therefore, the existence of a *desired* arrangement of $\left\{\mathbb{K}, \mathbb{K}^c\right\}$ where $\mathbb{K} \subseteq [\![M]\!]$ with the property that $i_1 \in \mathbb{K}$ and $i_2 \in \mathbb{K}^c$, is guaranteed. Thereby, we obtain

$$\begin{aligned} P_{e,1}(i_1) &= W^n(\mathscr{T}_{\mathbb{K}}^c \,|\, x^n = \mathbf{c}_{i_1})_{i_1 \in \mathbb{K}} \leq e_1^{(n)} \,, \\ P_{e,2}(i_2, \mathbb{K}) &= W^n(\mathscr{T}_{\mathbb{K}} \,|\, x^n = \mathbf{c}_{i_2})_{i_2 \notin \mathbb{K}} \leq e_2^{(n)} \,. \end{aligned} \tag{10.126}$$

This leads to a *contradiction* since

$$\begin{aligned} 1 &= W^n(\mathscr{T}_{\mathbb{K}}^c \,|\, x^n) + W^n(\mathscr{T}_{\mathbb{K}} \,|\, x^n) \\ &= P_{e,1}(i_1) + P_{e,2}(i_2, \mathbb{K}) \\ &\leq e_1^{(n)} + e_2^{(n)} \,, \end{aligned} \tag{10.127}$$

where the last inequality exploits the definition of type I/II error probabilities given in (10.22) and (10.23). Hence, the assumption is false, and distinct messages i_1 and i_2 cannot share the same codeword. □

▷ **Case 1 - With Hamming Weight Constraint** $(0 < A < 1)$**:** By Lemma 10.3.3, each message has a distinct codeword. Hence, the number of messages is bounded by the number of input sequences that satisfy the input constraint. We divide in two cases, namely, where $0 < A < \frac{1}{2}$ and $\frac{1}{2} \leq A < 1$. For the first case, we obtain the following upper bound on the size of the DKI codebook:

$$\begin{aligned} 2^{nR} &\leq \left|\mathcal{B}_{\mathbf{0}}(n, nA)\right| \\ &= \left|\left\{\mathbf{x} \in \mathbf{H}^n : 0 \leq \sum_{t=1}^{n} x_t \leq nA\right\}\right| \end{aligned}$$

$$\overset{(a)}{\leq} 2^{nH(A)}\,, \tag{10.128}$$

where (a) exploits the upper bound on the volume of the Hamming ball provided in Lemma J.0.1 for $0 < A < 1/2$. Thereby, (10.128) implies

$$R \leq H(A)\,. \tag{10.129}$$

Now, we proceed to calculate the upper bound on the size of the DKI codebook where $1/2 \leq A < 1$. We argue that this case is equivalent to having a Hamming weight constraint of the form $A^* = 1/2$. That is, the codewords with constraint $\sum_{t=1}^{n} x_t \leq nA^*$ where $A^* = 1/2$ fulfilled the same constraint with $\frac{1}{2} \leq A < 1$. The new Bernoulli input process has $1/2$ success probability, i.e., $X \sim \text{Bern}(1/2)$. Therefore, again employing Lemma J.0.1 for the critical point $\varepsilon = 1/2$, we obtain

$$\begin{aligned} 2^{nR} &\leq \left|\mathcal{B}_{\mathbf{0}}(n, nA^*)\right| \\ &= \left|\left\{\mathbf{x} \in \mathbf{H}^n : 0 \leq \sum_{t=1}^{n} x_t \leq nA^*\right\}\right| \\ &\leq 2^{nH(A^*=1/2)}\,, \end{aligned} \tag{10.130}$$

which implies

$$R \leq H\left(A^* = 1/2\right) = 1\,. \tag{10.131}$$

▷ **Case 2 - Without Hamming Weight Constraint ($A \geq 1$):** In this case, the number of messages is bounded by the number of the input sequences, that is, the size of entire input space, i.e., $|\mathcal{X}|^n$. Therefore, we can establish the following upper bound on the size of the DKI codebook $2^{nR} \leq |\mathcal{X}|^n$ which for $|\mathcal{X}| = 2$ implies

$$\begin{aligned} R &\leq \frac{1}{n}\log|\mathcal{X}|^n \\ &= 1\,. \end{aligned} \tag{10.132}$$

This completes the proof of Lemma 10.3.1.

Thus, by (10.129), (10.131) and (10.132), and exploiting the fact that DKI capacity is supremum of all achievable rate, the DKI coding rate is upper bounded by

$$\mathbb{C}_{DKI}\left(\mathcal{W}_\varepsilon, K\right) \leq \begin{cases} H(A) & \text{if } 0 < A < 1/2 \\ 1 & \text{if } A \geq 1/2\,, \end{cases} \tag{10.133}$$

which completes the proof of Theorem 10.3.1.

10.4 | Summary

In this chapter, we studied the DKI problem over the binary symmetric channel. We assume that the transmitter is subject to a Hamming weight constraint A. Our results in this chapter may serve as a model for event-triggered based tasks in the context of future XG applications. In particular, we obtained the DKI capacity of the BSC with the codebook size of $M(n,R) = 2^{nR}$ equals to the entropy of the Hamming constraint value, i.e., $H(A)$. Our results for the DKI capacity of the BSC revealed that the conventional exponential scale of 2^{nR} which is used in the standard message transmission setting, is again the appropriate scale for codebook size. Further, we find out that the BSC features an *exponentially* large set of target messages set, in the codeword length, n, i.e., $2^{\kappa n}$; and characterize all the possible valid range on the DKI target rate κ which depending on the value of channels statistic may varies.

We show the achievability proof using a Hamming distance decoder and employing packing arrangement of hyper balls in the same line of arguments as is conducted for the basic *Gilbert bound* method. In particular, in the presence of a Hamming weight constraint A, we pack hyper balls with radius $\lfloor n\beta \rfloor$, inside a larger Hamming hyper ball, which results in $\sim 2^{nH(A)}$ codewords.

For the converse part, a similar approach as chapter 2 for the DMC [98,99,137] is followed. That is, for the case where a non-trivial Hamming weight constraint is present ($0 < A < 1$), we establish an injection (one-to-one mapping) between the message set and the subset that is induced by the Hamming weight constraint. However, here we exploit the impact of generalized type I and type II error probability definitions with respect to the set of the target messages $\mathbb{K}$ in the course of the proof, namely, for any two distinct messages, there exist an arrangement of $\{\mathbb{K}, \mathbb{K}^c\}$ into which the two messages are categorized. In particular, we exploit the method of *proof by the contradiction*. Namely, we first assume that two generic different messages i_1 and i_2 share the common codewords, and then show that such an assumption leads to a contradiction regarding the sum of the error probabilities, i.e., we derive that the sum of type I and type II error probabilities converges to one from left. Hence the falsehood of the early assumption is guaranteed. Therefore, the total number of messages $M = 2^{nR}$ is bounded by the size of the induced subset, i.e., $M \leq 2^{nH(A)}$. For the case where $A \geq 1$, that is, in the absent of a Hamming weight constraint, a similar line of argument can be applied in order to establish the injective function.

CHAPTER 11

CONCLUSIONS

> *I Shall Know That The Conclusions Dwell in The Introduction. The Hegelian Dialectic Between The Two, Illuminates On The Next Step in Research.*
>
> M. J. Salariseddigh,

11.1 | Achieved Aims and Objectives

We develop the DI and DKI capacity for several channel models, including the basic scenarios such as the BSC, DMC or standard Gaussian or practical models such as the Poisson with/out memory, Binomial and fading channels. We determine the full characterization for DI or DKI capacity for the BSC [112] and DMC models, and derive lower and upper bounds for the Poisson, Binomial and the fading models. As our main objective, we obtained the fundamental performance limits of DI and DKI capacity for several settings which might be useful for the practical synthetic MC designs and provide insight for the code construction, performance evaluation, etc.

11.2 | Future Works

The results presented in this chapter can be extended in several directions, some of which are listed in the following as potential topics for future research works:

◊ **Continuous Alphabet Conjecture**: Our observations for the codebook size of following studies

- DI for the standard Gaussian channels without memory [105, 106],
- DI for the Poisson channels without memory [51, 108, 109],

- DI for the Poisson channels with memory [110],
- DKI for the slow fading channel without memory [197],
- DKI for Binomial channel [111],

lead us to conjecture that the codebook size for every *continuous* input alphabet channel either for DI or DKI and with / out memory is a super-exponential function, i.e., $2^{(n\log n)R}$. However, a formal proof of this conjecture remains unknown.

◊ **Memory Impact**: We assumed that the channel uses are orthogonal, which implies a memoryless channel and independent molecule reception for temporal and spatial channel use schemes, respectively. In practice, however, the DTBC may exhibit memory [6, 9] and non-orthogonal molecule reception [79], the investigation of which constitutes an interesting research problem.

◊ **DKI For Fast Fading Gaussian Channel**: The results in this chapter can be extended to the Gaussian channels with fast fading model.

◊ **Maximum Power Constraint**: Our achievability proof for the Gaussian channels with fading [99, 105] consider only the average power constraint, however, an interesting future research may include both the average and maximum power constraints at the same time which seems more practical.

◊ **Multi User**: This study has focused on a point-to-point system and may be extended to multi-user scenarios (e.g., broadcast and multiple access channels) or multiple-input multiple-output channels may seems more relevant in applications of complex MC nano-networks within the future generation wireless networks (XG). Recently, DI for multiple access channel has been studied in [204].

◊ **Fekete's Lemma**: Investigation of the behavior of the DI capacity in the sense of Fekete's Lemma [205]: To verify whether the pessimistic capacity,

$$\underline{C} = \liminf_{n\to\infty} \frac{\log M(n,R)}{n\log n},$$

and the optimistic capacity

$$\overline{C} = \limsup_{n\to\infty} \frac{\log M(n,R)}{n\log n},$$

[206] coincide or not; see [205] for more details.

◊ **Channel Reliability Function**: We note that to fully characterize the asymptotic behavior of the decoding errors as a function of the codeword length for every value of the rate $0 < R < C$, knowledge of the corresponding channel reliability function (CRF) is required [207]. To the best of the authors' knowledge, the CRF for DI has not been studied in the literature so far, neither for the Gaussian channel [99] nor the Poisson channel [51, 108, 109]. We note that even for the conventional message transmission problem, the characterization of the CRF is difficult, as the corresponding channel reliability function is not Turing computable [207].

◊ **Explicit Code Construction**: Our main focus in this dissertation was the establishment of fundamental performance limits of DI or DKI for various channel models, where an explicit code construction was not considered. Therefore, interesting directions for future research include the systematic design and explicit construction of DI and DKI codes and the development of low-complexity encoding and decoding schemes for practical applications. The efficiency of these designs can be evaluated against the performance bounds derived in this chapter.

◊ **Memory Gain**: We have not exploited the ISI knowledge in the decoding procedure. For instance, for the DTPC model with constant degree of ISI, capacity bounds coincide the bounds as of the memoryless DTPC. This observation suggest that testing a different decoding method which takes effect of ISI into account by conducting a symbol by symbol detection and exploits the previous K input symbols might probably yields different and more accurate capacity bounds.

Appendix A

NOTATIONS

We use the following notation conventions throughout all the chapters: Calligraphic letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \ldots$ and blackboard bold letters $\mathbb{K}, \mathbb{X}, \mathbb{Y}, \mathbb{Z} \ldots$ are used for finite alphabet sets. Lower case letters $x, y, z, \ldots$ stand for constants and values (realization) of random variables, and uppercase letters $X, Y, Z, \ldots$ stand for random variables. Lower case bold symbol $\mathbf{x}$ and $\mathbf{y}$ stand for row vectors of size n, that is, $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$. We use symbol $\triangleq$ to specify a definition convention. The probability distribution function (PMF) of a discrete random variable X is specified by $p_X(x)$ over a finite set $\mathcal{X}$. The distribution of a real random variable X is specified by a cumulative distribution function (CDF) $F_X(x) = \Pr(X \leq x)$ for $x \in \mathbb{R}$, or alternatively, by a probability density function (pdf) $f_X(x)$, when it exists. A random sequence $\mathbf{X}$ and its distribution $F_{\mathbf{X}}(\mathbf{x})$ are defined accordingly. The cumulative distribution function (CDF) of a Binomial random variable is indicated by $B_X(x) \triangleq \Pr(X \leq x)$. We use $x^j = (x_1, x_2, \ldots, x_j)$ to denote a sequence of letters from $\mathcal{X}$. A random sequence X^n and its distribution $p_{X^n}(x^n)$ are defined accordingly. Bold symbol $\mathbf{1}_n$ indicates the all-one row vector of size n. The set of whole numbers is denoted by $\mathbb{N}_0 \triangleq \{0, 1, 2, \ldots\}$. The set of real and non-negative numbers are denoted by $\mathbb{R}$ and $\mathbb{R}_+$, respectively.

The Gamma function for non-positive integer x is denoted by $\Gamma(x)$ and is defined as $\Gamma(x) = (x-1)!$, where $(x-1)! \triangleq (x-1) \times (x-2) \times \cdots \times 1$. The set of all PMFs over $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. $I(X;Y)$ indicate the mutual information. All logarithms and information quantities are taken to the base 2. The set of consecutive natural numbers from 1 through M is denoted by $[\![M]\!]$. The Hamming distance between two sequences a^n and b^n is defined as the number of positions for which the sequences have different symbols, i.e., $d_H(a^n, b^n) = |\{t \in [\![n]\!] \; ; \; a_t \neq b_t\}|$. The Hamming metric (distance) between two sequences $\mathbf{x}_1$ and $\mathbf{x}_2$ is defined as the number of positions for which the corresponding symbols are not identical, i.e.,

$$d_{\mathrm{H}}(\mathbf{x}_1, \mathbf{x}_2) \triangleq \sum_{t=1}^{n} \delta(x_{i_1,t}, x_{i_2,t}) \,, \tag{A.1}$$

where $\delta(.,.)$ the *Kronecker delta* and is defined as follows

$$\delta(x_i, x_j) = \begin{cases} 1 & x_i \neq x_j \\ 0 & x_i = x_j \end{cases} \tag{A.2}$$

The n-dimensional Hamming sphere of radius $n\epsilon$ centered at a^n is defined as

$$\mathcal{S}_\epsilon(a^n) = \{x^n \in \mathcal{X}^n \,:\, d_H(x^n, a^n) < n\epsilon\} \,. \tag{A.3}$$

We denote the hyper-sphere of radius r around $\mathbf{x}_0$ with respect to the ℓ_2-norm (used for Gaussian channel in chapters 3, 4, 5) by

$$\mathcal{S}_{\mathbf{x}_0}(n, r) = \left\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| \leq r\right\}, \tag{A.4}$$

and its volume by $\mathrm{Vol}(\mathcal{S})$. The closure of a set $\mathcal{A}$ is denoted by $\mathrm{cl}(\mathcal{A})$. The element-wise product of vectors is denoted by $\mathbf{x} \circ \mathbf{y} = (x_t y_t)_{t=1}^n$. We use the small O notation, $f(n) = o(g(n))$, to indicate that $f(n)$ is dominated by $g(n)$ asymptotically, that is, $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$. The big O notation, $f(n) = \mathcal{O}(g(n))$, is used to indicate that $|f(n)|$ is bounded above by $g(n)$ (up to constant factor) asymptotically, that is, $\limsup_{n\to\infty} \frac{|f(n)|}{g(n)} < \infty$. We use the big Omega notation, $f(n) = \Omega(g(n))$, to indicate that $f(n)$ is bounded below by $g(n)$ asymptotically, that is, $g(n) = \mathcal{O}(f(n))$. The the ℓ_1-norm, ℓ_2-norm and ℓ_∞-norm of vector $\mathbf{x}$ are denoted by $\|\mathbf{x}\|_1, \|\mathbf{x}\|$, and $\|\mathbf{x}\|_\infty$, respectively. We use $\mathbf{0} \triangleq (0, \ldots, 0)$ to represent coordination of the origin. An n-dimensional cube with center $(A/2, \ldots, A/2)$ and a corner at the origin, i.e., $\mathbf{0}$, whose edges have length A is denoted by $\mathcal{Q}_{\mathbf{0}}(n, A) = \left\{\mathbf{x} \in \mathbb{R}_+^n : 0 \leq x_t \leq A, \forall\, t \in [\![n]\!]\right\}$.

The n-dimensional Hamming hyper ball of radius r for integers n, r such that $n \geq r \geq 1$, in the binary alphabet, that is centered at $\mathbf{x}_0 = (x_{0,t})|_{t=1}^n$ is defined as

$$\mathcal{B}_{\mathbf{x}_0}(n, r) = \left\{x^n \in \mathcal{X}^n \,:\, d_{\mathrm{H}}(\mathbf{x}, \mathbf{x}_0) \leq r\right\} \,. \tag{A.5}$$

Volume of the Hamming hyper ball $\mathcal{B}_{\mathbf{x}_0}(n, r)$ in the q-ary alphabet is defined as the number of points that lies inside the ball and is denoted by $\mathrm{Vol}\left(\mathcal{B}_{\mathbf{x}_0}(n, r)\right)$. The Hamming cube is defined as the set of sequences with length n and is denoted by $\mathbf{H}^n = \{0, 1\}^n$. The n-dimensional Hamming hyper ball $\mathcal{B}_{\mathbf{x}_0}(n, r)$ for the choice of $\mathbf{x}_0 = \mathbf{0}$ and $r = nA$, is denoted by $\mathcal{B}_{\mathbf{0}}(n, nA)$ and is referred to as the n-dimensional Hamming hyper ball in 1-norm with a corner at the origin, i.e., $\mathbf{0} = (0, \ldots, 0)$, and radius equal to nA. The definition of $\mathcal{B}_{\mathbf{0}}(n, nA)$ is given as follows

$$\mathcal{B}_{\mathbf{0}}(n, nA) = \left\{\mathbf{x} \in \mathbf{H}^n : 0 \leq \sum_{t=1}^n x_t \leq nA\right\} \,. \tag{A.6}$$

The q-ary entropy function $H_q : [0,1] \to \mathbb{R}$ for $q \geq 2$; a positive integer, is defined as $H_q(\varepsilon) \triangleq x \log_q(q-1) - x \log_q x - (1-x) \log_q(1-x)$. The binary entropy function as a special case of the q-ary entropy function $H_q(.)$ is denoted by $H(.)$ or $H_2(.)$ and is defined as $H(\varepsilon) \triangleq -\varepsilon \log(\varepsilon) - (1-\varepsilon) \log(1-\varepsilon)$. We denote the DTPC with K ISI channel taps, Binomial channel, and GSF with K number of target messages by $\mathcal{P}$, $\mathcal{B}$, and $\mathcal{G}_{\text{slow}}$.

Appendix B

VOLUME OF A HYPER SPHERE WITH GROWING RADIUS

To solidify the idea of packing spheres within a hyper cube, we reveal and explain a counter-intuitive phenomenon regarding the packing of hyper spheres with growing radius in the codeword length inside a hyper cube. We observe that despite the fact that the hyper sphere's radius tends to infinity as the codeword length goes to infinity $\sim n^{\frac{1}{4}}$ its volume tends to zero super-exponentially inverse, i.e., $\sim n^{-\frac{n}{4}}$. This allows us to accommodate super-exponential number of such hyper spheres inside the hyper cube. The ratio of the spheres in our construction grows with n, as $\sim n^{\frac{1}{4}}$. It is well-known that the volume of an n-dimensional *unit*-hyper sphere, i.e., with a radius of $r_0 = 1$, tends to zero, as $n \to \infty$ [138, Ch. 1, Eq. (18)]. Nonetheless, we observe that the volume still tends to zero for a radius of $r_0 = n^c$, where $0 < c < \frac{1}{2}$. More precisely,

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Vol}\left(\mathcal{S}_{\mathbf{u}_1}(n, r_0)\right) &= \lim_{n\to\infty} \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r_0^n \\ &= \lim_{n\to\infty} \frac{\pi^{\frac{n}{2}}}{\frac{n}{2}!} \cdot r_0^n \\ &= \lim_{n\to\infty} \left(\sqrt{\frac{2\pi}{n}} r_0\right)^n , \end{aligned} \tag{B.1}$$

where the last equality follows by Stirling's approximation [173, P. 52], that is, $\log n! = n\log n - n\log e + o(n)$. The last expression in (B.1) tends to zero for all $r_0 = n^c$ with $c \in (0, \frac{1}{2})$. Observe that when $n \to \infty$, the volume of a hyper cube $\mathcal{Q}_{\mathbf{0}}(n, A)$ with edge length A is given by

$$\lim_{n\to\infty} \mathrm{Vol}\left[\mathcal{Q}_{\mathbf{0}}(n, A)\right] = \lim_{n\to\infty} A^n = \begin{cases} 0 & A < 1\,, \\ 1 & A = 1\,, \\ \infty & A > 1\,. \end{cases} \tag{B.2}$$

Now, to derive how many spheres can be packed inside the hyper cube $\mathcal{Q}_{\mathbf{0}}(n, A)$ we

derive the log-ratio of the volumes as follows

$$\begin{aligned}
\log\left(\frac{\mathrm{Vol}\left[\mathcal{Q}_{\mathbf{0}}(n,A)\right]}{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{u}_1}(n,r_0)\right)}\right) &= \log\left(\frac{A^n}{\pi^{\frac{n}{2}} r_0^n}\cdot\frac{n}{2}!\right) \\
&= n\log\left(\frac{A}{\sqrt{\pi}r_0}\right)+\log\left(\frac{n}{2}!\right) \\
&= n\log A - n\log r_0 - n\log\sqrt{\pi}+\frac{1}{2}n\log\frac{n}{2}-\frac{n}{2}\log e+o(n) \\
&= \left(\frac{1}{2}-c\right)n\log n + n\left(\log\frac{A}{\sqrt{\pi e}}-\frac{3}{2}\right)+o(n)\,, \qquad \text{(B.3)}
\end{aligned}$$

where the last equality follows from $r_0 = n^c$. Now, since the dominant term in (B.3) involves $n\log n$, we deduce that codebook size should be $L(n,R) = 2^{(n\log n)R}$, thereby by (9.19) we obtain

$$\begin{aligned}
R &\geq \frac{1}{n\log n}\left[\log\left(\frac{\mathrm{Vol}\left[\mathcal{Q}_{\mathbf{0}}(n,A)\right]}{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{u}_1}(n,r_0)\right)}\right)-n\right] \\
&= \frac{1}{n\log n}\left[\left(\frac{1}{2}-c\right)n\log n + n\left(\log\frac{A}{\sqrt{\pi e}}-\frac{3}{2}\right)+o(n)\right]\,, \qquad \text{(B.4)}
\end{aligned}$$

which tends to $\frac{1}{2}-c$ when $n\to\infty$.

Appendix C

MOMENT GENERATING FUNCTION OF POISSON RANDOM VARIABLE

The moment-generating function (MGF) of a Poisson variable $Z \sim \text{Pois}(\lambda_Z)$ is $G_Z(\alpha) = e^{\lambda_Z(e^\alpha - 1)}$. Hence, for $X = Z - \lambda_Z$, the MGF is given by $G_X(\alpha) = e^{\lambda_Z(e^\alpha - 1 - \alpha)}$. Since the fourth non-central moment equals the fourth order derivative of the MFG at $\alpha = 0$, we have

$$\begin{aligned}\mathbb{E}\{X^4\} = \frac{d^4}{d\alpha^4} G_X(\alpha)\bigg|_{\alpha=0} &= \lambda_Z \left(\lambda_Z^3 e^{3\alpha} + 6\lambda_Z^2 e^{2\alpha} + 7\lambda_Z e^\alpha + 1\right) e^{\alpha + \lambda_Z e^\alpha - \lambda_Z}\bigg|_{\alpha=0} \\ &= \lambda_Z^4 + 6\lambda_Z^3 + 7\lambda_Z^2 + \lambda_Z \leq 7\left(\lambda_Z^4 + \lambda_Z^3 + \lambda_Z^2 + \lambda_Z\right).\end{aligned}$$

Appendix D

SURVEY ON TRANSMISSION CAPACITY OF THE POISSON CHANNEL

In this Appendix, we review known results for the capacity of a DTPC in the asymptotic regimes of $P_{\text{avg}} \to 0$ and $P_{\text{avg}} \to \infty$. DTPC is an important channel model in optical communication and is addressed several times within the last decades. All of such investigation consider the well-known model of Shannon transmission. None of these models so far considered the identification problem for it. Therefore, in order to obtain a ground knowledge and basic insight about DTPC from the Shannon scheme perspective and observe the behavior of this model from a number of aspects such as presence of power constraints, working in extreme/asymptotic regions of key parameter (λ, P_{avg} and $P_{\max}$), and capacity achieving distributions we survey some of the best existing results in the literature. A second reason for studying such existing known results for the DTPC might be a technical issue regarding knowing the known and working techniques and analysis methods for a code construction, decoding rules and any smart methods aiming at tightening the bounds.

The asymptotic capacity with an average-power constraint P_{avg}, when $P_{\text{avg}} \to \infty$ and $\frac{P_{\text{avg}}}{\lambda}$ is fixed was studied in [208]. Furthermore, the same problem for a constant λ, with and without an additional peak power constraint was studied in [209]. The first-order asymptotic capacity for $P_{\text{avg}} \to 0$, both when $\frac{P_{\text{avg}}}{\lambda}$ is kept constant and when λ is fixed, with and without a peak power constraint, was determined in [210]. Later, the previously obtained first-order capacity approximation when $P_{\text{avg}} = \lambda$ is constant were improved in [211] where a refined approximation was determined, including an exact characterization of the second-order term, as well as an asymptotic characterization of the third-order term with respect to the dark current, λ. Asymptotic upper bounds for the DTPC capacity with an average-power constraint were given in [33, 209–211]. Explicit asymptotic and non-asymptotic capacity lower bounds for several settings were given in [209–214].

Previously, the best known non-asymptotic upper bound, which was in fact the best

capacity upper bound outside the limiting case $P_{\text{ave}} \to 0$, was derived in [212]. However, the proof suffers a small gap, as mentioned in [209], and is not considered completely rigorous. Later, in [215, Th. 8] strictly tighter upper bounds than the bound in [212] for all $P_{\text{avg}} > 0$ which are considered the best current capacity upper bounds for the DTPC with zero dark current term, i.e., $\lambda = 0$ subject to an average power constraint P_{avg} for all values of P_{avg} outside the limiting case $P_{\text{avg}} \to 0$. In the same paper, the result of [212] was recovered as an special (sub-optimal) case, thus yielding a rigorous proof for the bound proposed in [212]. As well, the same authors in [216, Th. 1] showed derived a significantly improved non-asymptotic upper bound on the capacity of the DTPC with constant positive dark current $\lambda \geq 0$ and an average power constraint P_{avg} in non-asymptotic regimes of P_{avg}.

D.0.1 | Capacity-Achieving Distributions

There has been an enormous literature focusing on the properties of capacity achieving distributions for different channels. This problem is well-understood for quite general classes of additive noise channels under several input constraints (see, e.g., the early works [160, 217, 218] and the recent works [219–221]). For the DTPC, in the absence of input constraints, capacity is infinite. The DTPC under a peak power constraint alone was addressed in [222], and was shown that the support size is of an order between $\sqrt{P_{\max}}$ and $\sim P_{\max} \log^2 P_{\max}$. In particular, they characterized the capacity in terms of the output optimal distribution where capacity equals $-\log P_{Y^*}(0)$ for $P_{Y^*}(0)$ to be the optimal output distribution. An analytic expression for the transmission capacity of a the DTPC with an average power constraint alone, is still open. However, several bounds and asymptotic behaviors for the DTPC in different setups have been established. For instance, it was shown that a capacity-achieving input distribution for the DTPC under an average power constraint must have a finite support. The number of mass points depends on the average and the peak power constraints, and increases to infinity as the constraints are relaxed [160]. It was conjectured by Shamai [160] that the support of such a distribution must be countably infinite. The result was extended in [175, 176] and it was stated that the support of such a distribution for the DTPC under an average-power constraint must have an *unbounded support*. Moreover, they also proved that such a distribution has a non-zero mass at $x = 0$, and, if a peak power constraint $P_{\max}$ is present, at $x = P_{\max}$ as well. Unlike additive noise channels, less is known about the capacity-achieving distributions of a DTPC when there is only an

average-power constraint. In [215] it was shown that such a distribution for the DTPC with an arbitrary $\lambda \geq 0$, under an average-power constraint and/or a peak power constraint, is discrete (see the conjecture by Shamai [160]). It was further shown that the support of such a capacity-achieving distribution for the DTPC under an average-power constraint and/or a peak power constraint has a finite intersection with every bounded interval [215, see Th. 14]. Further discussions on the capacity-achieving distributions are provided in [223]. Here, we will consider the identification setting, where the receiver is not required to determined the message, but rather identifies a specific task.

D.0.2 | Asymptotic Characterizations

In the sequel, we denote the capacity under an average power constraint P_{ave}, a peak power constraint $P_{\max}$ and the dark current λ by $\mathbf{C}(\lambda, P_{\text{ave}}, P_{\max})$. For small values of P_{ave}, in [210] it was shown that the asymptotic capacity of the DTPC, i.e., when the average input power tends to zero while the peak-power, if finite, is fixed, scales as $-P_{\text{ave}} \log P_{\text{ave}}$, i.e.,

$$\lim_{P_{\text{ave}} \to 0} \frac{\mathbf{C}(\lambda = cP_{\text{ave}}, P_{\text{ave}}, P_{\max})}{P_{\text{ave}} \log P_{\text{ave}}} = -1 \,, \tag{D.1}$$

for any $c \in [0, \infty)$ and $P_{\max} \in (0, \infty]$. Furthermore, they provided the following upper bound

$$\begin{aligned} \mathbf{C}(0, P_{\text{ave}}, \infty) \leq & -P_{\text{ave}} \log p - \log(1-p) + \frac{P_{\text{ave}}}{\beta} \\ & + P_{\text{ave}} \cdot \max\left(0, \left(\frac{1}{2}\log\beta + \log\left(\frac{\overline{\Gamma}\left(\frac{1}{2}, 1/\beta\right)}{\sqrt{\pi}} + \frac{1}{2\beta}\right)\right)\right) \,, \end{aligned} \tag{D.2}$$

where $p \in (0,1)$ and $\beta > 0$ are arbitrary constants and $\overline{\Gamma}(.)$ is the upper incomplete Gamma function. Later, in [211], for a small value of P_{ave} the higher order asymptotic behavior for $\mathbf{C}(P_{\text{ave}})$ was characterized and given by

$$\mathbf{C}(\lambda = cP_{\text{ave}}, P_{\text{ave}}, P_{\max}) = -P_{\text{ave}} \log P_{\text{ave}} - P_{\text{ave}} \log\left(-\log P_{\text{ave}}\right) + \mathcal{O}(P_{\text{ave}}) \,, \tag{D.3}$$

where $c \in [0, \infty)$. This upper bound holds irrespective of whether a peak power constraint is imposed or not as long as $P_{\max}$ is positive and does not approach zero together with ϵ. Also the following upper bound was stated.

$$\mathbf{C}(\lambda = cP_{\text{ave}}, P_{\text{ave}}, P_{\max}) \leq P_{\text{ave}} - P_{\text{ave}} \log\log P_{\text{ave}} - \log(1 - P_{\text{ave}})$$

$$- P_{\text{ave}} \log \left(1 - \frac{1}{\log P_{\text{ave}}}\right) + P_{\text{ave}} \cdot \sup_{x \geq 0} \phi_\mu(x)\,,$$

which matches the asymptotic behavior given in (D.3) where

$$\phi_\mu(x) := \frac{1 - e^{-x}}{x} \log \left(\frac{-x}{P_{\text{ave}} \log P_{\text{ave}}}\right)\,. \tag{D.4}$$

For a large value of P_{ave}, and in the absence of a peak power constraint, it was shown in [169] that

$$\lim_{P_{\text{ave}} \to \infty} \left[\mathbb{C}(\lambda, P_{\text{ave}}, \infty) - \frac{1}{2} \log P_{\text{ave}}\right] = 0\,, \tag{D.5}$$

where the dark current is a non-negative constant, i.e., $\lambda \geq 0$. Expressions for the capacity, based on the ratio of the average and the peak power constraint, i.e., $\alpha = \frac{P_{\text{ave}}}{P_{\text{max}}}$ was provided in [224] as follows

$$\mathbb{C}(\lambda, P_{\text{ave}}, P_{\text{max}}) = \begin{cases} \frac{1}{2} \log P_{\text{max}} + (\alpha - 1)u - \log\left(\frac{1}{2} - \alpha u\right) - \frac{1}{2} \log(2\pi e) + \mathcal{O}(1) & 0 < \alpha < \frac{1}{3}\,, \\ \frac{1}{2} \log P_{\text{max}} - \frac{1}{2} \log \frac{\pi e}{2} + \mathcal{O}(1) & \frac{1}{3} \leq \alpha \leq 1\,, \end{cases} \tag{D.6}$$

for u is the non-zero solution to

$$\sqrt{\pi}\, \text{erf}\left(\sqrt{u}\right) \left(\frac{1}{2} - \alpha u\right) - \sqrt{u} e^{-u} = 0\,, \tag{D.7}$$

with erf(.) being the Gauss error function. Observe that $\alpha \ll 1$ represents the regimes of very weak peak power constraints, whereas $\alpha = 1$ corresponds to the absence of an average power constraint. Also, the term $\mathcal{O}(1)$ vanishes as $P_{\text{ave}}, P_{\text{max}} \to \infty$ where α is considered to be constant.

The previously best upper bound in the presence of only average power constraint P_{ave} in any regime outside the regime $P_{\text{ave}} \to 0$ was derived [212, see Eq. (10)] and is given by

$$\mathbb{C}\left(\lambda, P_{\text{ave}}, \infty\right) \leq \left(P_{\text{ave}} + \frac{1}{2}\right) \log \left(P_{\text{ave}} + \frac{1}{2}\right) - P_{\text{ave}} \log P_{\text{ave}} - \frac{1}{2} + \log \left(1 + \frac{\sqrt{2e} - 1}{\sqrt{1 + 2P_{\text{ave}}}}\right)\,, \tag{D.8}$$

which is strictly less than the upper bound given in (D.2) for all $P_{\text{ave}} > 0$ and tends to $\frac{1}{2} \log(1 + P_{\text{ave}})$ for $P_{\text{ave}} \to \infty$. However, recently, authors in [215] yielded the best

known upper bound on the capacity $\mathbb{C}(0, P_{\text{ave}}, \infty)$ in the absence of dark current, i.e., $\lambda = 0$ for any P_{ave} outside the asymptotic regime $P_{\text{ave}} \to 0$, as follows

$$\mathbb{C}(0, P_{\text{ave}}, \infty) \leq P_{\text{ave}} \ln \left(\frac{1 + \left(1 + e^{1+\gamma}\right) P_{\text{ave}} + 2P_{\text{ave}}^2}{e^{1+\gamma} P_{\text{ave}} + 2P_{\text{ave}}^2} \right) + \ln \left(1 + \frac{1}{\sqrt{2e}} \left(\sqrt{\frac{1 + \left(1 + e^{1+\gamma}\right) P_{\text{ave}} + 2P_{\text{ave}}^2}{1 + P_{\text{ave}}}} - 1 \right) \right) \tag{D.9}$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant. This bound outperforms the previously best upper bound given in (D.8) and recovers that result with a rigorous proof. Further, in [216, Th. 1], the best upper bound on the capacity $\mathbb{C}(\lambda, P_{\text{ave}}, \infty)$ for a positive dark current term, i.e., $\lambda \geq 0$ is given by

$$\mathbb{C}(\lambda, P_{\text{ave}}, \infty) \leq \ln \left(\delta_\lambda + \frac{1}{\sqrt{2e}} \left(\frac{1}{\sqrt{1 - q_{\lambda, P_{\text{ave}}}}} - 1 \right) \right) - (P_{\text{ave}} + \lambda) \ln q_{\lambda, P_{\text{ave}}} \tag{D.10}$$

where $\delta_\lambda = e^{-\lambda e^\lambda E_1(\lambda)}$, with $E_1(z)$ being the exponential integral function and $q_{\lambda, P_{\text{ave}}}$ given by

$$q_{\lambda, P_{\text{ave}}} = 1 - \frac{1}{1 + e^{1+\gamma} (P_{\text{ave}} + \lambda) + \frac{2 - e^{1+\gamma}}{1 + P_{\text{ave}} + \lambda} (P_{\text{ave}} + \lambda)^2} \tag{D.11}$$

where γ is the Euler-Mascheroni constant.

Poisson channel under an average power constraint alone is considered in [208, 225] where it was assumed that both P_{ave} and λ tend to infinity while their ratio, $\frac{P_{\text{ave}}}{\lambda}$, defined as SNR, is kept constant. Namely, for every $\epsilon > 0$ capacity is lower bounded by

$$\mathbb{C}(\lambda, P_{\text{ave}}, \infty) \geq \frac{1}{2} \log \frac{P_{\text{ave}}}{2\pi} - \frac{1}{2} \log \left(1 + \frac{1}{\text{SNR}} \right) - \epsilon \,, \tag{D.12}$$

and upper bounded by

$$\mathbb{C}(\lambda, P_{\text{ave}}, \infty) \leq \frac{1}{2} \log \frac{P_{\text{ave}}}{2\pi} + \log \left(\sqrt{\text{SNR}} \left(1 + \frac{1}{P_\epsilon} \right) + \frac{1}{\sqrt{\text{SNR}}} \right) + 1 + \log \frac{3}{2} + \epsilon \,, \tag{D.13}$$

where the $P_\epsilon < P_{\text{ave}}$ is a large constant depending on ϵ but strictly less that the average power constraint P_{ave}.

An upper bound on the capacity of a point-to-point DTPC $W(y|x)$ under an average power constraint P_{ave} and dark current λ was proposed in [33, Example 2] where it was shown that the Shannon capacity satisfies

$$\begin{aligned}\mathbb{C}\left(\lambda, P_{\text{ave}}, \infty\right) &= \max_{X\,:\,\mathbb{E}[X]\leq P_{\text{ave}}} I(X,Y) \\ &\leq \max_{X\,:\,\mathbb{E}[X]\leq P_{\text{ave}}} \operatorname{Cov}\left(X+\lambda\,,\,\log\left(X+\lambda\right)\right)\,, \end{aligned} \tag{D.14}$$

with $Y\sim\text{Pois}(\lambda+X)$ and $\operatorname{Cov}(X,Y)=\mathbb{E}\left[XY\right]-\mathbb{E}\left[X\right]\mathbb{E}\left[Y\right]$. Further, for a DTPC with an average input constraint P_{ave} and peak power constraint P_{max}, the following upper bound was reported

$$\begin{aligned}\mathbb{C}\left(\lambda, P_{\text{ave}}, P_{\text{max}}\right) &= \max_{\substack{X\,:\,\mathbb{E}[X]\leq P_{\text{ave}}\,,\\ 0\leq X\leq P_{\text{max}}}} I(X,Y) \\ &\leq \begin{cases} \frac{P_{\text{ave}}}{P_{\text{max}}}\left(P_{\text{max}}-P_{\text{ave}}\right)\log\left(\frac{P_{\text{max}}}{\lambda}+1\right) & P_{\text{ave}}\leq\frac{P_{\text{max}}}{2}\,, \\ \frac{P_{\text{max}}}{4}\log\left(\frac{P_{\text{max}}}{\lambda}+1\right) & P_{\text{ave}}\geq\frac{P_{\text{max}}}{2}\,. \end{cases}\end{aligned} \tag{D.15}$$

The capacity of the direct-detection Poisson photon-counting channel with fading (Poisson fading channel) is established in [226] where a single-letter characterization of the capacity, assuming a perfect CSI at the receiver is provided with perfect and no CSI at the transmitter. Also, the limiting behavior of the capacity in the high and low peak-signal-to-dark-noise ratio (SNR) regimes, namely in the limits as $\lambda\to 0$ and $\lambda\to\infty$ is addressed. The capacity for perfect CSI at the transmitter is reported to be

$$\mathbb{C}\left(\lambda, P_{\text{ave}}, P_{\text{max}}\right) = \max_{\substack{\mu\,:\,\mathbb{R}_0^*\to[0,1]\,,\\ \mathbb{E}\left[\mu(S)\right]\leq\sigma}} \mathbb{E}\left[\mu(S)\zeta\left(S\alpha,\lambda\right)-\zeta\left(\mu(S)S\alpha,\lambda\right)\right]\,, \tag{D.16}$$

where $\zeta(x,y)\coloneqq(x+y)\ln(x+y)-y\ln y$ for $x,y>0$ with $0\ln 0\coloneqq 0$. The $0\leq\sigma\leq 1$ is the ratio of average to peak power constraint, and S is a distribution satisfying following conditions

$$\Pr\left[S>0\right]=1\,, \tag{D.17}$$

$$\mathbb{E}\left[S\right]<\infty\,, \tag{D.18}$$

$$\mathbb{E}\left[\left|\zeta(S\alpha,\lambda)\right|\right]<\infty\,. \tag{D.19}$$

Finally, the capacity for no CSI at the transmitter is given by

$$\mathbb{C}\left(\lambda, P_{\text{ave}}, P_{\text{max}}\right) = \max_{0\leq\mu\leq\sigma} \mathbb{E}\left[\mu\zeta\left(S\alpha,\lambda\right)-\zeta\left(\mu S\alpha,\lambda\right)\right]\,. \tag{D.20}$$

Lower and upper bounds on the constrained capacity in diffusion-based MC is provided in [227]. Capacities and optimal input distributions for particle-intensity channels (PIC) as a more general formulation of the DTPC is discussed in [31]. In [36, Lem. 4], it has been shown that the capacity $\mathbb{C}$ of a diffusion channel under the PAM modulation with symbol duration Δ and peak and average power constraints A and μA ($\mu \in [0,1]$) is lower bounded as

$$\mathbb{C} \geq \frac{1}{\Delta} \max_{P_\lambda \in \mathcal{P}(A,\mu)} I(\gamma; y)\,, \tag{D.21}$$

where $\gamma\Delta/h_0$ is distributed as $P_\lambda \in \mathcal{P}(A,\mu)$ and y is the output of W_Δ with input $\gamma + \tilde{L}/\Delta$, where $\tilde{L}$ is the maximum inter-symbol interference (ISI) term given by

$$\tilde{L} \stackrel{\text{def}}{=} gA\Delta \int_\Delta^\infty \Gamma(\mathbf{r}, s)\, \mathrm{d}s\,, \tag{D.22}$$

with g being coefficient in function $f(x) = gx$ describing how the rate of Poisson reception process at the receiver is related to the particle concentration $\rho(\mathbf{r}, t)$, that is,

$$\gamma(t) = f(\rho(\mathbf{r}, t)) = g\rho(\mathbf{r}, t)\,. \tag{D.23}$$

Observe that $\mathbf{r}$ here is the receiver coordination and $\Gamma(\mathbf{x}, t)$ is the fundamental solution (free space and impulse release) to the Fick's second law of diffusion and is given by

$$\Gamma(\mathbf{x}, t) = \frac{1}{\sqrt{(4\pi D t)^3}} \cdot \exp\left\{-\frac{|\mathbf{x}|^2}{4Dt}\right\}\,, \tag{D.24}$$

where $\mathbf{r}$ is the coordination. In [36, Th. 5] further it was shown that for any $\epsilon, \Delta > 0$, there exists $A_0 \in \mathbb{R}^+$ such that the capacity $\mathbb{C}$ of the diffusion channel with peak and average power constraints of A and μA, respectively, is lower bounded

$$\mathbb{C} \geq \left(\frac{1}{2} - \epsilon\right) \frac{\log A}{\Delta}\,, \tag{D.25}$$

For $A \geq A_0$. Furthermore, the lower bound can be achieved by deploying a PAM modulation scheme with symbol duration Δ at the input. A closed form solution for capacity of a diffusion-based molecular communication system with channel memory and molecular noise is addressed and obtained in [30] as follows

$$\mathbb{C} = 2W\left(1 + \log\left(\frac{\overline{P}_{\mathcal{H}}}{3WK_bT}\right) - 2\log\left(\pi D d\right)\right) - \frac{4d}{3\ln 2}\sqrt{\frac{\pi W}{D}}$$

$$- 2W\frac{2\overline{P}_{\mathcal{H}}R_{V_R}}{9W^2dK_bT} - 2W\ln\left(W\tau_p\right) - 2W\ln\left(\Gamma\left(\frac{2\overline{P}_{\mathcal{H}}R_{V_R}}{9W^2dK_bT}\right)\right)$$
$$- 2W\left(1-\frac{2\overline{P}_{\mathcal{H}}R_{V_R}}{9W^2dK_bT}\right)\psi\left(\frac{2\overline{P}_{\mathcal{H}}R_{V_R}}{9W^2dK_bT}\right), \tag{D.26}$$

where $\overline{P}_{\mathcal{H}}$ is the average thermodynamic power spent by the transmitter, K_b is the Boltzmann constant, T is the absolute temperature of the system, W is the bandwidth of the transmitted signal X, τ_p is the time interval in which we consider a quasi-constant particle distribution, $\psi(.)$ is the digamma function, D is the diffusion coefficient, d is the distance between the transmitter and the receiver, R_{V_R} and is the radius of the spherical receiver volume V_R. In this model, the environment is assumed to be three dimensional space with an infinite extension to all directions. Transmitter is point-wise and transmitted signal is the number of particles that are emitted by transmitter into the space. Receiver is an ideal type where the received signal is defined as time varying number of particles that are present inside its volume.

Further, lower and upper bounds for the constrained capacity of a diffusive MC system in the case where the information is associated with the concentration of molecules released by the transmitter is investigated in [227].

Appendix E

PROOF OF REDUCTION LEMMA – DMC

Let $\mathcal{W}$ be a given DMC, with a stochastic matrix $W : \mathcal{X} \to \mathcal{Y}$ and its reduced version $W_r : \mathcal{X}_r \to \mathcal{Y}$ as defined in Definition 2.3.1. Observe that the capacity of the original channel is lower bounded by that of the reduced channel, i.e.,

$$\mathbb{C}_{DI}(\mathcal{W}, L) \geq \mathbb{C}_{DI}(\mathcal{W}_r, L) \,, \tag{E.1}$$

since every code for W_r can also be used for W. Hence, it remains to be shown that

$$\mathbb{C}_{DI}(\mathcal{W}_r, L) \geq \mathbb{C}_{DI}(\mathcal{W}, L) \,. \tag{E.2}$$

Assume without loss of generality that the input alphabet of the original channel $\mathcal{W}$ is given by $\mathcal{X} = \{1, 2, \cdots, |\mathcal{X}|\}$. Let $\sigma : \mathcal{X} \to \mathcal{X}_r$ denote the projection of the input alphabet onto the equivalent classes,

$$\sigma[x] = z(\ell) \quad \text{iff} \quad x \in \mathcal{X}(\ell) \,. \tag{E.3}$$

Now let $(\mathcal{U}, \mathcal{D})$ be an $(L(n,R), n, \lambda_1, \lambda_2)$ code for $\mathcal{W}$. Then the type I probability of error can be expressed as

$$P_{e,1}^{W}(i) = \sum_{y^n \notin \mathcal{D}_i} W^n(y^n|u_i) = \sum_{y^n \notin \mathcal{D}_i} \prod_{t=1}^{n} W(y(t)|u_i(t)) \,, \tag{E.4}$$

where we use the notation $y^n = \Big(y(t)\Big)_{t=1}^{n}$ and $u_i = \Big(u_i(t)\Big)_{t=1}^{n}$. Next we define a code $(\widetilde{\mathcal{U}}, \mathcal{D})$ for the channel $\mathcal{W}_r$, where the codebook consists of the following codewords,

$$\tilde{u}_i = \Big(\sigma[u_i(t)]\Big)_{t=1}^{n} \,. \tag{E.5}$$

Now recall that we have defined the equivalence classes such that input letters in the same equivalence class correspond to identical rows in the channel matrix W (see Definition 2.3.1). Thus, by definition,

$$W_r(y|\sigma[x]) = W(y|x) \,, \tag{E.6}$$

for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Hence, the error probability of type I for the reduced channel $\mathcal{W}_r$ satisfies

$$\begin{aligned} P_{e,1}^{W_r}(i) &\overset{(a)}{=} \sum_{y^n \notin \mathcal{D}_i} W_r(y^n|\tilde{u}_i) \\ &\overset{(b)}{=} \sum_{y^n \notin \mathcal{D}_i} \prod_{t=1}^{n} W_r\Big(y(t)|\sigma[u_i(t)]\Big) \\ &\overset{(c)}{=} \sum_{y^n \notin \mathcal{D}_i} \prod_{t=1}^{n} W(y(t)|u_i(t)) \\ &\overset{(d)}{=} \sum_{y^n \notin \mathcal{D}_i} W^n(y^n|u_i) \\ &\overset{(e)}{=} P_{e,1}^{W}(i)\,, \end{aligned} \tag{E.7}$$

for all i, where (a) and (e) are due to (2.7); (b) and (d) hold since the channel is memoryless, and (c) follows from (E.6). By the same considerations, we also have $P_{e,2}^{W_r}(i,j) = P_{e,2}^{W}(i,j)$ for all $j \neq i$. That is, the error probabilities of the code $(\widetilde{\mathcal{U}}, \mathcal{D})$ are the same as those of the original code for $\mathcal{W}$. Therefore, the code constructed above for $\mathcal{W}_r$ is also an $(L(n,R), n, \lambda_1, \lambda_2)$ code, and the proof of Lemma 2.3.1 follows. □

Appendix F

PROOF OF MISCELLANEOUS CODEBOOK SIZE LEMMA – FAST FADING CHANNEL

In the following, a rigorous proof for comparing the capacity value in different scale is given. In particular, we assume that given a known scale for the codebook size used in the course of the capacity theorem, we try to set a scale for the codebook which has a higher scale and then, we try to calculate the capacity value for the higher scale. More specifically, we would show that if the capacity is finite and positive in the scale of L_1, then the capacity is zero in the scale L_2, where L_2 dominate the L_1 for asymptotic n, i.e., when $n \to \infty$.

Proof. The proof is straightforward. Let $\mathbb{C}_{DI}(\mathscr{G}_{\text{fast}}, L_0) = c_0$, where $c_0 > 0$ is a finite number. Then, for every λ_1, λ_2, and sufficiently large n, there exists an $(M_n, n, \lambda_1, \lambda_2)$ code where the number of messages is

$$M_n = L_0(n, c_0 - \epsilon) \tag{F.1}$$

where $\epsilon > 0$ is arbitrarily small.

Assume to the contrary that in the L^--scale, the DI capacity $\mathbb{C}_{DI}(\mathscr{G}_{\text{fast}}, L^-) = c^-$ is also finite. Then, the converse for this claim implies that the number of messages is bounded by

$$M_n \leq L^-(n, c^- + \epsilon) \tag{F.2}$$

Hence, by (F.1) and (F.2),

$$L_0(n, c_0 - \epsilon) \leq L^-(n, c^- + \epsilon) \tag{F.3}$$

which contradicts the assumption that $L^- \prec L_0$ (see Definition 2.2.1). Hence, $\mathbb{C}_{DI}(\mathscr{G}_{\text{fast}}, L^-)$ is infinite.

Assume to the contrary that in the L^+-scale, the DI capacity is positive, i.e.,

$$\mathbb{C}_{DI}(\mathscr{G}_{\text{fast}}, L^+) = c^+ > 0\,,$$

Then, for every λ_1, λ_2, and sufficiently large n, there exists an $(M_n^+, n, \lambda_1, \lambda_2)$ code where the number of messages is

$$M_n^+ = L^+(n, c^+ - \epsilon) \tag{F.4}$$

Now, by the converse part for the C_0-scale,

$$M_n^+ \leq L_0(n, c_0) \tag{F.5}$$

Hence, by (F.4) and (F.5),

$$L^+(n, c^+ - \epsilon) \leq L_0(n, c_0) \tag{F.6}$$

which contradicts the assumption that $L_0 \prec L^+$ (see Definition 2.2.1). Hence, $\mathbb{C}_{DI}(\mathscr{G}_{\text{fast}}, L^+)$ is zero. □

UPPER BOUND ON THE VARIANCE OF POISSON–SQUARED RANDOM VARIABLE

Let $Y_t(i) \sim \text{Pois}\left(\rho_0 c_{i,t} + I_t^{\mathbf{c}_i}\right)$ denote the channel output at time t given that $\mathbf{x} = \mathbf{c}_i$. Recall that $\bar{y}_t(i) \stackrel{\text{def}}{=} y_t(i) - (\rho_0 c_{i,t} + \lambda)$, then we have

$$
\begin{aligned}
\text{Var}\left[(\bar{Y}_t(i))^2\right] &= \text{Var}\left[\left(Y_t - (\rho_0 c_{i,t} + \lambda)\right)^2\right] \\
&\stackrel{(a)}{\leq} \mathbb{E}\left[\left(Y_t(i) - (\rho_0 c_{i,t} + \lambda)\right)^4\right] \\
&\stackrel{(b)}{=} \mathbb{E}\left[Y_t^4(i) - 4Y_t^3(i)\left(\rho_0 c_{i,t} + \lambda\right) + 6Y_t^2(i)\left(\rho_0 c_{i,t} + \lambda\right)^2 - 4Y_t(i)\left(\rho_0 c_{i,t} + \lambda\right)^3 + \left(\rho_0 c_{i,t} + \lambda\right)^4\right] \\
&\stackrel{(c)}{\leq} \mathbb{E}\left[Y_t^4(i)\right] - 4\lambda\mathbb{E}\left[Y_t^3(i)\right] + 6(\rho_0 c_{i,t} + \lambda)^2\mathbb{E}\left[Y_t^2(i)\right] - 4(\rho_0 c_{i,t} + \lambda)^3\mathbb{E}\left[Y_t(i)\right] + (\rho_0 c_{i,t} + \lambda)^4 \\
&\stackrel{(d)}{\leq} \mathbb{E}\left[Y_t^4(i)\right] - 4\lambda\mathbb{E}\left[Y_t^3(i)\right] + 6(A + \lambda)^2\mathbb{E}\left[Y_t^4(i)\right] - 4(\rho_0 c_{i,t} + \lambda)^3\mathbb{E}\left[Y_t(i)\right] + (A + \lambda)^4 \\
&\stackrel{(e)}{\leq} \mathbb{E}\left[Y_t^4(i)\right] + 4\lambda\mathbb{E}\left[Y_t^4(i)\right] + 6(A + \lambda)^2\mathbb{E}\left[Y_t^4(i)\right] + 4(A + \lambda)^3\mathbb{E}\left[Y_t^4(i)\right] + (A + \lambda)^4 \\
&\leq \mathbb{E}\left[Y_t^4(i)\right]\left(1 + 4\lambda + 6(A + \lambda)^2 + 4(A + \lambda)^3\right) + (A + \lambda)^4 \\
&\stackrel{(f)}{\leq} (A + \lambda)^4 e^{\frac{8}{\lambda}}\left(1 + 4\lambda + 6(A + \lambda)^2 + 4(A + \lambda)^3\right) + (A + \lambda)^4
\end{aligned}
$$

$$= (A+\lambda)^4 \left(1 + e^{\frac{8}{\lambda}} \left(1 + 4\lambda + 6(A+\lambda)^2 + 4(A+\lambda)^3\right)\right)$$
$$= 6(A+\lambda)^4 \left(1 + e^{\frac{8}{\lambda}} \left(1 + (A+\lambda) + (A+\lambda)^2 + (A+\lambda)^3\right)\right), \tag{G.1}$$

where (a) follows from $\mathrm{Var}\{Z\} \leq \mathbb{E}[Z_t^2]$ with $Z_t = \left(Y_t - \left(\rho_0 c_{i,t} + \lambda\right)\right)^2$, (b) holds by the 8-th order binomial expansion, (c) follows by the linearity of the expectation operator, (d) and (e) follows from $c_{i,t} \leq A\,, \forall t \in [\![n]\!]$, and since expectation is an increasing function, that is, for integers p, q we have

$$Y_t^p(i) < Y_t^q(i) \Rightarrow \mathbb{E}\left[Y_t^p(i)\right] < \mathbb{E}\left[Y_t^p(i)\right], \tag{G.2}$$

(f) holds by employing an upper bound on the non-central moment of a Poisson random variable with mean λ_Z as follows (see [228, Coroll. 1])

$$\mathbb{E}\left[Z^k\right] \leq \lambda_Z^k \exp\left\{\frac{k^2}{2\lambda_Z}\right\}. \tag{G.3}$$

Appendix H

VOLUME OF A HYPER SPHERE WITH GROWING RADIUS – POISSON CHANNEL WITH MEMORY

To solidify the idea of packing spheres within a hyper cube, we explain about the packing of hyper spheres with growing radius in the codeword length n. Despite the fact that radius of the hyper sphere's diverges to infinity as $n \to \infty$ as $\sim n^{\frac{1+\kappa}{4}}$, still the associated volume converges to zero super-exponentially inverse as of order $\sim n^{-\frac{(1+\kappa)n}{4}}$. This makes an accommodation of super-exponential number of such hyper spheres inside the hyper cube possible. The ratio of the spheres in our construction grows with n, as $\sim n^{\frac{1+\kappa}{4}}$. Volume of an n-dimensional *unit*-hyper sphere, i.e., with a radius of $r_0 = 1$, tends to zero, as $n \to \infty$ [138, Ch. 1, Eq. (18)]. Nonetheless, we observe that the volume still tends to zero for a radius of $r_0 = n^c$, where $0 < c < \frac{1}{2}$. More precisely,

$$\begin{aligned}\lim_{n\to\infty} \mathrm{Vol}\left(\mathcal{S}_{\mathbf{c}_1}(n, r_0)\right) &= \lim_{n\to\infty} \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} \cdot r_0^n \\ &= \lim_{n\to\infty} \frac{\pi^{\frac{n}{2}}}{\frac{n}{2}!} \cdot r_0^n \\ &= \lim_{n\to\infty} \left(\sqrt{\frac{2\pi}{n}} r_0\right)^n , \end{aligned} \tag{H.1}$$

where the last equality follows by Stirling's approximation [173, P. 52], that is, $\log n! = n \log n - n \log e + o(n)$. The last expression in (H.1) tends to zero for all $r_0 = n^c$ with $c \in (0, \frac{1}{2})$. Observe that when $n \to \infty$, the volume of a hyper cube $\mathcal{Q}_{\mathbf{0}}(n, A)$ with edge length A when $A < 1$ tends to zeros, that is,

$$\lim_{n\to\infty} \mathrm{Vol}\left(\mathcal{Q}_{\mathbf{0}}(n, A)\right) = \lim_{n\to\infty} A^n = 0 .$$

Now, to count the number of spheres that can be packed inside the hyper cube

$\mathcal{Q}_{\mathbf{0}}(n, A)$, we derive the log-ratio of the volumes as follows

$$\begin{aligned}\log\left(\frac{\mathrm{Vol}\left(\mathcal{Q}_{\mathbf{0}}(n,A)\right)}{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{c}_1}(n,r_0)\right)}\right) &= \log\left(\frac{A^n}{\pi^{\frac{n}{2}} r_0^n}\cdot\frac{n}{2}!\right)\\ &= n\log\left(\frac{A}{\sqrt{\pi}r_0}\right)+\log\left(\frac{n}{2}!\right)\\ &= n\log A - n\log r_0 - n\log\sqrt{\pi}+\frac{1}{2}n\log\frac{n}{2}-\frac{n}{2}\log e+o(n)\\ &= \left(\frac{1}{2}-c\right)n\log n + n\left(\log\left(\frac{A}{\sqrt{\pi e}}\right)-\frac{3}{2}\right)+o(n)\,, \end{aligned} \tag{H.2}$$

where the last equality follows from $r_0 = n^c$. Now, since the dominant term in (H.2) involves $n\log n$, we deduce that codebook size should be $M(n,R) = 2^{(n\log n)R}$, thereby by (9.19) we obtain

$$\begin{aligned} R &\geq \frac{1}{n\log n}\left[\log\left(\frac{\mathrm{Vol}\left(\mathcal{Q}_{\mathbf{0}}(n,A)\right)}{\mathrm{Vol}\left(\mathcal{S}_{\mathbf{c}_1}(n,r_0)\right)}\right)-n\right]\\ &= \frac{1}{n\log n}\left[\left(\frac{1}{2}-c\right)n\log n + n\left(\log\left(\frac{A}{\sqrt{\pi e}}\right)-\frac{3}{2}\right)+o(n)\right]\,, \end{aligned} \tag{H.3}$$

which tends to $\frac{1}{2}-c$ when $n\to\infty$. As a result, (H.3) induces that condition $c<\frac{1}{2}$ with c not being arbitrary approaching $\frac{1}{2}$ to derive a meaningful (non-zero) lower bound. Since $c = \frac{1+\kappa}{4}$ we obtain

$$\frac{1+\kappa}{4} < \frac{1}{2} \Rightarrow \kappa < 1\,. \tag{H.4}$$

Appendix I

LOWER BOUND ON THE VOLUME OF THE HAMMING BALL

Lemma I.0.1. *Let $n, q \geq 2$ be positive integers and assume a real p where $0 \leq \lfloor n\varepsilon \rfloor / n \leq 1 - 1/q$. Then, volume of the Hamming ball in the q-ary alphabet is lower bounded as follows*

$$Vol\left(\mathcal{B}_{\mathbf{x}_0}(n,r)\right) \triangleq \sum_{j=0}^{\lfloor n\varepsilon \rfloor} \binom{n}{j} (q-1)^j \geq q^{H_q\left(\frac{\lfloor n\varepsilon \rfloor}{n}\right) - o\left(\log_q n\right)}. \tag{I.1}$$

Proof. Observe that the **Stirling's approximation** [229] gives the following double bound on $n!$

$$\sqrt{2n\pi}\left(\frac{n}{e}\right)^n e^{\lambda_1(n)} \leq n! \leq \sqrt{2n\pi}\left(\frac{n}{e}\right)^n e^{\lambda_2(n)}. \tag{I.2}$$

Now, we have

$$\begin{aligned}
&\binom{n}{\lfloor n\varepsilon \rfloor} \\
&= \frac{n!}{\lfloor n\varepsilon \rfloor! \left(n - \lfloor n\varepsilon \rfloor\right)!} \\
&> \frac{\sqrt{2n\pi} \cdot \left(\frac{n}{e}\right)^n \cdot e^{\lambda_1(n)}}{\left[\sqrt{2\lfloor n\varepsilon \rfloor \pi} \cdot \left(\frac{\lfloor n\varepsilon \rfloor}{e}\right)^{\lfloor n\varepsilon \rfloor} \cdot e^{\lambda_1(n)}\right] \left[\sqrt{2\left(n\left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)\right)\pi} \cdot \left(\frac{n\left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)}{e}\right)^{n\left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)}\right]} \\
&= \left[\frac{\left(\frac{n}{e}\right)^n}{\left(\frac{\lfloor n\varepsilon \rfloor}{e}\right)^{\lfloor n\varepsilon \rfloor} \cdot \left(\frac{n\left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)}{e}\right)^{n\left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)}}\right] \cdot \left[\frac{e^{\lambda_1(n) - \lambda_2(\lfloor n\varepsilon \rfloor) - \lambda_2\left(n\left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)\right)}}{\sqrt{2\pi \lfloor n\varepsilon \rfloor \left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)}}\right] \\
&\overset{(a)}{=} \left[\frac{1}{\left(\frac{\lfloor n\varepsilon \rfloor}{n}\right)^{\lfloor n\varepsilon \rfloor} \cdot \left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)^{n\left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)}}\right] \cdot \left[\frac{e^{\lfloor n\varepsilon \rfloor} \cdot e^{n\left(1 - \frac{\lfloor n\varepsilon \rfloor}{n}\right)}}{e^n}\right] \cdot \mathrm{Res}(n)
\end{aligned}$$

$$\overset{(b)}{=} \frac{\mathrm{Res}(n)}{\left(\frac{\lfloor n\varepsilon\rfloor}{n}\right)^{\lfloor n\varepsilon\rfloor}\cdot\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)^{n\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)}} \tag{I.3}$$

where (a) holds since we let

$$\mathrm{Res}(n) \triangleq \frac{e^{\lambda_1(n)-\lambda_2(\lfloor n\varepsilon\rfloor)-\lambda_2\left(n\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)\right)}}{\sqrt{2\pi\,\lfloor n\varepsilon\rfloor\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)}}, \tag{I.4}$$

and (b) holds since

$$\frac{e^{\lfloor n\varepsilon\rfloor}\cdot e^{n\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)}}{e^{n}} = 1\,. \tag{I.5}$$

Next, we proceed to bound the Hamming ball as follows: Observe that the volume of Hamming ball as provided in (I.1) is lower bounded by the Binomial coefficient for the largest index, i.e., $j = \lfloor n\varepsilon\rfloor$. Therefore,

$$\begin{aligned}
\mathrm{Vol}\left(\mathcal{B}_{\mathbf{x}_0}(n,r)\right) &\triangleq \sum_{j=0}^{\lfloor n\varepsilon\rfloor}\binom{n}{j}(q-1)^j \\
&\geq \binom{n}{\lfloor n\varepsilon\rfloor}(q-1)^{\lfloor n\varepsilon\rfloor} \\
&> \frac{(q-1)^{\lfloor n\varepsilon\rfloor}}{\left(\frac{\lfloor n\varepsilon\rfloor}{n}\right)^{\lfloor n\varepsilon\rfloor}\cdot\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)^{n\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)}}\cdot\mathrm{Res}(n) \\
&= q^{\log_q\left(\frac{(q-1)^{\lfloor n\varepsilon\rfloor}}{\left(\frac{\lfloor n\varepsilon\rfloor}{n}\right)^{\lfloor n\varepsilon\rfloor}\cdot\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)^{n\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)}}\right)+\log_q\mathrm{Res}(n)} \\
&= q^{\lfloor n\varepsilon\rfloor\log_q(q-1)-\lfloor n\varepsilon\rfloor\log_q\frac{\lfloor n\varepsilon\rfloor}{n}-n\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)\log_q\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)+\log_q\mathrm{Res}(n)} \\
&= q^{n\left(\frac{\lfloor n\varepsilon\rfloor}{n}\log_q(q-1)-\frac{\lfloor n\varepsilon\rfloor}{n}\log_q\frac{\lfloor n\varepsilon\rfloor}{n}-\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)\log_q\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)+\log_q\mathrm{Res}(n)\right)} \\
&= q^{nH_q\left(\frac{\lfloor n\varepsilon\rfloor}{n}\right)+\log_q\mathrm{Res}(n)}\,.
\end{aligned} \tag{I.6}$$

Now by letting $\lambda_1(n) = 0$ and $\lambda_2(n) = 1/(12n)$, we obtain

$$\mathrm{Res}(n) = \frac{e^{-\frac{1}{12\lfloor n\varepsilon\rfloor}-\frac{1}{n-\lfloor n\varepsilon\rfloor}}}{\sqrt{2\pi\,\lfloor n\varepsilon\rfloor\left(1-\frac{\lfloor n\varepsilon\rfloor}{n}\right)}}$$

$$\begin{aligned}
&\overset{(a)}{\leq} \frac{e^{-\frac{1}{12\lfloor n\varepsilon\rfloor}-\frac{1}{n-\lfloor n\varepsilon\rfloor}}}{\sqrt{2\pi\,\lfloor n\varepsilon\rfloor\,(1-\varepsilon)}} \\
&\overset{(b)}{=} K(\varepsilon)\,\lfloor n\varepsilon\rfloor^{-\frac{1}{2}}\,e^{-\frac{1}{12\lfloor n\varepsilon\rfloor}-\frac{1}{n-\lfloor n\varepsilon\rfloor}}\,, \qquad \text{(I.7)}
\end{aligned}$$

where (b) follows for sufficiently large n, since $\lfloor n\varepsilon\rfloor$ (b) holds for $K(\varepsilon) = \frac{1}{\sqrt{2\pi(1-\varepsilon)}}$. Therefore,

$$\begin{aligned}
\log_q \mathrm{Res}(n) &= \log_q K(\varepsilon) - \frac{1}{2}\log_q \lfloor n\varepsilon\rfloor - \frac{1}{12\lfloor n\varepsilon\rfloor} - \frac{1}{n-\lfloor n\varepsilon\rfloor} \\
&= o\left(\log_q n\right)\,, \qquad \text{(I.8)}
\end{aligned}$$

which implies that

$$\lim_{n\to\infty} \frac{\log_q \mathrm{Res}(n)}{\log_q n} = 0\,. \qquad \text{(I.9)}$$

Thereby,

$$\begin{aligned}
\mathrm{Vol}\left(\mathcal{B}_{\mathbf{x}_0}(n,r)\right) &\triangleq \sum_{j=0}^{\lfloor n\varepsilon\rfloor}\binom{n}{j}(q-1)^j \\
&\geq q^{nH_q\left(\frac{\lfloor n\varepsilon\rfloor}{n}\right)+o\left(\log_q n\right)}\,. \qquad \text{(I.10)}
\end{aligned}$$

□

Appendix J

UPPER BOUND ON THE VOLUME OF THE HAMMING BALL

Lemma J.0.1 (see [230, Lem. 16.19]). *Let integer $n \geq 1$ and $0 < \varepsilon \leq \frac{1}{2}$ with $n > \lfloor n\varepsilon \rfloor \geq 1$. Then, volume of the Hamming ball in the binary alphabet is upper bounded as follows*

$$\mathrm{Vol}\left(\mathcal{B}_{\mathbf{x}_0}(n,r)\right) \triangleq \sum_{j=0}^{\lfloor n\varepsilon \rfloor} \binom{n}{j} \leq 2^{nH(\varepsilon)}\,, \tag{J.1}$$

Proof. Observe that for $0 < \varepsilon \leq \frac{1}{2}$, the logit function, i.e., $H(\varepsilon) \triangleq \log\left(\frac{\varepsilon}{1-\varepsilon}\right)$ is non-positive. That is,

$$H(\varepsilon) = \log\left(\frac{\varepsilon}{1-\varepsilon}\right) = \log\varepsilon - \log(1-\varepsilon) \leq 0\,. \tag{J.2}$$

Next, notice that for $i \in [0, \lfloor n\varepsilon \rfloor]$ we obtain the following

$$i\log\varepsilon + (n-i)\log(1-\varepsilon) \geq -nH(\varepsilon)\,, \tag{J.3}$$

where $H(\varepsilon)$ is the binary entropy function. Hence, $\varepsilon^i(1-\varepsilon)^{n-i} \geq 2^{-nH(\varepsilon)}$. Now,

$$\begin{aligned} 1 &= (\varepsilon + (1-\varepsilon))^n \\ &= \sum_{i=0}^{n} \binom{n}{i} \varepsilon^i (1-\varepsilon)^{n-i} \\ &\geq \sum_{i=0}^{\lfloor n\varepsilon \rfloor} \varepsilon^i (1-\varepsilon)^{n-i} \\ &\geq 2^{-nH(\varepsilon)} \sum_{i=0}^{\lfloor n\varepsilon \rfloor} \binom{n}{i}\,. \end{aligned} \tag{J.4}$$

□

Therefore, we obtain

$$\mathrm{Vol}\left(\mathcal{B}_{\mathbf{x}_0}(n,r)\right) \triangleq \sum_{j=0}^{\lfloor n\varepsilon \rfloor} \binom{n}{j} \leq 2^{nH(\varepsilon)}\,, \tag{J.5}$$

Appendix K

BOUND ON THE UPPER TAIL OF THE BINOMIAL CUMULATIVE DISTRIBUTION FUNCTION – PART 1

Lemma K.0.1 (see [200, Probl. 5.8 – (c)]). *Let* $0 < \varepsilon < 1$ *and* $\varepsilon < \frac{k}{n} < 1$. *Then,*

$$\binom{n}{k}\varepsilon^j(1-\varepsilon)^{n-k} \leq \sum_{j=k}^{n}\binom{n}{j}\varepsilon^j(1-\varepsilon)^{n-j} \leq \binom{n}{k}\varepsilon^k(1-\varepsilon)^{n-k}\left[\frac{k(1-\varepsilon)}{k(1-\varepsilon)-(n-k)\varepsilon}\right]. \tag{K.1}$$

Proof. The proof for the lower bound is trivial and obvious. For proving the upper bound, we employ the provided hints given in [200, P. 531] as follows: Observe that

$$\binom{n}{j+1} = \binom{n}{j}\binom{n-k}{k+1} < \binom{n}{j}\binom{n-j}{j}, \tag{K.2}$$

and

$$\binom{n}{k+m} = \binom{n}{k+m-1}\binom{n-(k+m-1)}{k+m-1} < \binom{n}{k+m-1}\binom{n-k}{k}, \tag{K.3}$$

Using the induction, we obtain

$$\binom{n}{k+m} < \binom{n}{k}\binom{n-k}{k}^m, \tag{K.4}$$

Now let us sum over j by using a geometric series. Next, we combine this results with the result of part (a) in the Problem 5.8 of [200, Probl. 5.8], and we obtain the desired upper bound. That is,

$$\begin{aligned}
&\sqrt{\frac{n}{8k(n-k)}}e^{nH(k/n)+k\log\varepsilon+(n-k)\log(1-\varepsilon)} \\
&\leq \sum_{j=k}^{n}\binom{n}{j}\varepsilon^j(1-\varepsilon)^{n-j} \\
&< \sqrt{\frac{n}{2\pi k(n-k)}}\cdot\frac{k(1-\varepsilon)}{k(1-\varepsilon)-(n-k)\varepsilon}\cdot e^{nH(k/n)+k\log\varepsilon+(n-k)\log(1-\varepsilon)},
\end{aligned} \tag{K.5}$$

□

Appendix L

BOUND ON THE UPPER TAIL OF THE BINOMIAL CUMULATIVE DISTRIBUTION FUNCTION – PART 2

Lemma L.0.1. *Let* $0 < \varepsilon < 1$ *and* $\varepsilon < \frac{k}{n} < 1$. *Then,*

$$\sum_{j=k}^{n} \binom{n}{j} \varepsilon^j (1-\varepsilon)^{n-j} \leq 2^{n\left[H\left(\frac{k}{n}\right) - T_\varepsilon\left(\frac{k}{n}\right)\right]} \left[\frac{k(1-\varepsilon)}{k(1-\varepsilon) - (n-k)\varepsilon}\right] . \tag{L.1}$$

Proof. Recall that the equation of the tangent line to the binary entropy function $H(\delta)$ at the specific point $\delta = \varepsilon$ is given by

$$\begin{aligned}
&T_\varepsilon(\delta) \\
&\overset{(a)}{=} H(\varepsilon) + (\delta - \varepsilon) \frac{\mathrm{d}H(\delta)}{\mathrm{d}\delta}\bigg|_{\delta=\varepsilon} \\
&\overset{(b)}{=} H(\varepsilon) + (\delta - \varepsilon) \log\left(\frac{1-\varepsilon}{\varepsilon}\right) \\
&= H(\varepsilon) + (\delta - \varepsilon)\left[\log(1-\varepsilon) - \log\varepsilon\right] \\
&\overset{(c)}{=} -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon) + \delta\log(1-\varepsilon) - \delta\log\varepsilon - \varepsilon\log(1-\varepsilon) + \varepsilon\log\varepsilon \\
&= -\varepsilon\log\varepsilon - \log(1-\varepsilon) + \varepsilon\log(1-\varepsilon) + \delta\log(1-\varepsilon) - \delta\log\varepsilon - \varepsilon\log(1-\varepsilon) + \varepsilon\log\varepsilon \\
&= -\log(1-\varepsilon) + \delta\log(1-\varepsilon) - \delta\log\varepsilon \\
&= -\log(1-\varepsilon) + \delta\log(1-\varepsilon) - \delta\log\varepsilon \\
&= -\delta\log(\varepsilon) - (1-\delta)\log(1-\varepsilon) ,
\end{aligned} \tag{L.2}$$

where (a) holds by definition of a tangent line to a function at specific point, (b) follows since derivative of the entropy function reads the negative of the logit function, i.e.,

$$\frac{\mathrm{d}H(\delta)}{\mathrm{d}\delta} = -\mathrm{logit}\,(\delta) \triangleq -\log\left(\frac{\delta}{1-\delta}\right) , \tag{L.3}$$

for $0 < \delta < 1$, and (c) holds by definition of the entropy function, i.e.,

$$H(\varepsilon) \triangleq -\varepsilon\log\varepsilon - (1-\varepsilon)\log(1-\varepsilon) . \tag{L.4}$$

Therefore exploiting (L.2) we obtain,

$$T_\varepsilon\left(\frac{k}{n}\right) = -\frac{k}{n}\log(\varepsilon) - (1-\frac{k}{n})\log(1-\varepsilon)\,, \tag{L.5}$$

which implies $-nT_\varepsilon(\frac{k}{n}) = k\log(\varepsilon) + (n-k)\log(1-\varepsilon)$. Thereby,

$$2^{-nT_\varepsilon(\frac{k}{n})} = \varepsilon^k(1-\varepsilon)^{n-k}\,. \tag{L.6}$$

Now, observe that the Binomial coefficient $\binom{n}{k}$ where $k \geq 1$ and $n-k \geq 1$, can be upper bounded as follows [120, see P. 353]

$$\binom{n}{k} \leq 2^{nH(\frac{k}{n})}\,. \tag{L.7}$$

Therefore,

$$\begin{aligned}\frac{k(1-\varepsilon)}{k(1-\varepsilon)-(n-k)\varepsilon}\cdot\binom{n}{k}\varepsilon^k(1-\varepsilon)^{n-k} &\overset{(a)}{\leq} \frac{k(1-\varepsilon)}{k(1-\varepsilon)-(n-k)\varepsilon}\cdot 2^{nH(\frac{k}{n})}\cdot\varepsilon^k(1-\varepsilon)^{n-k}\\ &\overset{(b)}{\leq} \frac{k(1-\varepsilon)}{k(1-\varepsilon)-(n-k)\varepsilon}\cdot 2^{nH(\frac{k}{n})}\cdot 2^{-nT_\varepsilon(\frac{k}{n})}\\ &= \left[\frac{k(1-\varepsilon)}{k(1-\varepsilon)-(n-k)\varepsilon}\right]\cdot 2^{n\left[H(\frac{k}{n})-T_\varepsilon(\frac{k}{n})\right]}\,,\end{aligned} \tag{L.8}$$

where (a) holds by (L.6) and (b) follows by exploiting (L.6). Now, recalling (L.1), we obtain

$$\sum_{j=k}^{n}\binom{n}{j}\varepsilon^j(1-\varepsilon)^{n-j} \leq \frac{k(1-\varepsilon)}{k(1-\varepsilon)-(n-k)\varepsilon}2^{n\left[H(\frac{k}{n})-T_\varepsilon(\frac{k}{n})\right]} \tag{L.9}$$

This completes the proof of Lemma L.0.1 □

Appendix M

BOUND ON THE BINOMIAL CUMULATIVE DISTRIBUTION FUNCTION

Lemma M.0.1. *Let $0 < \varepsilon < 1$ and $k < n$ with $\frac{k}{n} < \varepsilon$. Then,*

$$\sum_{j=0}^{k} \binom{n}{j} \varepsilon^j (1-\varepsilon)^{n-j} \leq \frac{\varepsilon(n-k)}{\varepsilon n - k} \cdot 2^{n\left[H(\frac{k}{n}) - T_\varepsilon(\frac{k}{n})\right]}. \tag{M.1}$$

Proof. Let define

$$\begin{aligned} k' &\triangleq n - k, \\ \varepsilon' &\triangleq 1 - \varepsilon. \end{aligned} \tag{M.2}$$

i.e., $k \leftrightarrow k'$ and $\varepsilon \leftrightarrow \varepsilon'$ or equivalently

$$\begin{aligned} k &\leftrightarrow n - k, \\ \varepsilon &\leftrightarrow 1 - \varepsilon. \end{aligned} \tag{M.3}$$

Now, observe that

$$\frac{k}{n} > \varepsilon \Rightarrow \frac{k'}{n} < \varepsilon'. \tag{M.4}$$

Furthermore, by definition of the binary entropy function and its tangent line, we have

$$H\left(\frac{k}{n}\right) = H\left(\frac{n-k}{n}\right), \tag{M.5}$$

$$T_\varepsilon\left(\frac{k}{n}\right) = T_{1-\varepsilon}\left(\frac{n-k}{n}\right). \tag{M.6}$$

where (M.5) follows by (L.4) and (M.6) holds by (L.5).

Now applying the variable exchange of $j \leftrightarrow n - j$ unto (L.1), we obtain

$$\sum_{n-j=k}^{n-j=n} \binom{n}{n-j} \varepsilon^{n-j} (1-\varepsilon)^{n-(n-j)} \leq 2^{n\left[H(\frac{k}{n}) - T_\varepsilon(\frac{k}{n})\right]} \left[\frac{k(1-\varepsilon)}{k(1-\varepsilon) - (n-k)\varepsilon}\right]. \tag{M.7}$$

Observe that since the index of sum in (L.1) runs form k to n, i.e., $k \leq j \leq n$, in the new system we have $k \leq n-j \leq n$ which is equivalent to $0 \leq j \leq n-k$. Further, the Binomial coefficient satisfy

$$\binom{n}{n-j} = \binom{n}{j}, \tag{M.8}$$

for $0 \leq j \leq n$. Thereby,

$$\sum_{j=0}^{n-k} \binom{n}{j} \varepsilon^{n-j}(1-\varepsilon)^j \leq 2^{n\left[H(\frac{k}{n})-T_\varepsilon(\frac{k}{n})\right]} \left[\frac{k(1-\varepsilon)}{k(1-\varepsilon)-(n-k)\varepsilon}\right]. \tag{M.9}$$

Now, applying the exchange of variables given in (M.3) unto (M.9), we obtain

$$\begin{aligned}\sum_{j=0}^{k} \binom{n}{j} (1-\varepsilon)^{n-j}\varepsilon^j &\leq 2^{n[H(\frac{n-k}{n})-T_{1-\varepsilon}(\frac{n-k}{n})]} \left[\frac{(n-k)\varepsilon}{(n-k)\varepsilon-k(1-\varepsilon)}\right] \\ &= 2^{n\left[H(\frac{k}{n})-T_\varepsilon(\frac{k}{n})\right]} \left[\frac{(n-k)\varepsilon}{(n-k)\varepsilon-k(1-\varepsilon)}\right],\end{aligned} \tag{M.10}$$

where the equality holds by (M.5) and (M.6). Therefore,

$$\sum_{j=0}^{k} \binom{n}{j} (1-\varepsilon)^{n-j}\varepsilon^j \leq 2^{n\left[H(\frac{k}{n})-T_\varepsilon(\frac{k}{n})\right]} \left[\frac{(n-k)\varepsilon}{(n-k)\varepsilon-k(1-\varepsilon)}\right]. \tag{M.11}$$

Now we focus on the bracket in (M.10) which can be simplified as follows

$$\begin{aligned}\frac{(n-k)\varepsilon}{(n-k)\varepsilon-k(1-\varepsilon)} &= \frac{\left(\frac{n-k}{n}\right)\varepsilon}{\left(\frac{n-k}{n}\right)\varepsilon-\frac{k}{n}(1-\varepsilon)} \\ &= \frac{\varepsilon-\frac{k}{n}\varepsilon}{\varepsilon-\frac{k}{n}} \\ &= \frac{\varepsilon(n-k)}{\varepsilon n-k},\end{aligned} \tag{M.12}$$

where the first equality follows by dividing both sides in the left side by factor n. Thereby,

$$\sum_{j=0}^{k} \binom{n}{j} (1-\varepsilon)^{n-j}\varepsilon^j \leq \frac{\varepsilon(n-k)}{\varepsilon n-k} \cdot 2^{n\left[H(\frac{k}{n})-T_\varepsilon(\frac{k}{n})\right]}. \tag{M.13}$$

This completes the proof of Lemma M.0.1. □

REFERENCES

[1] N. Farsad, H. B. Yilmaz, A. Eckford, C.-B. Chae, and W. Guo, "A comprehensive survey of recent advancements in molecular communication," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1887–1919, 2016.

[2] S. Bush, J. Paluh, G. Piro, V. S. Rao, R. Prasad, and A. Eckford, "Defining communication at the bottom," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, pp. 90–96, 2015.

[3] T. Nakano, M. J. Moore, F. Wei, A. V. Vasilakos, and J. Shuai, "Molecular communication and networking: Opportunities and challenges," *IEEE Trans. Nanobiosci.*, vol. 11, no. 2, pp. 135–148, 2012.

[4] T. Nakano, A. W. Eckford, and T. Haraguchi, *Molecular Communication*. Cambridge University Press, 2013.

[5] V. Jamali, A. Ahmadzadeh, W. Wicke, A. Noel, and R. Schober, "Channel Modeling For Diffusive Molecular Communication - A Tutorial Review," *Proc. IEEE*, vol. 107, no. 7, pp. 1256–1301, 2019.

[6] V. Jamali, "Design and analysis of molecular communication systems," Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), 2019.

[7] M. Kuscu, E. Dinc, B. A. Bilgin, H. Ramezani, and O. B. Akan, "Transmitter and receiver architectures for molecular communications: A survey on physical design with modulation, coding, and detection techniques," *Proc. IEEE*, vol. 107, no. 7, pp. 1302–1341, 2019.

[8] C. A. Söldner, E. Socher, V. Jamali, W. Wicke, A. Ahmadzadeh, H.-G. Breitinger, A. Burkovski, K. Castiglione, R. Schober, and H. Sticht, "A survey of biological building blocks for synthetic molecular communication systems," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2765–2800, 2020.

[9] A. Gohari, M. Mirmohseni, and M. Nasiri-Kenari, "Information theory of molecular communication: Directions and challenges," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 2, no. 2, pp. 120–142, 2016.

[10] C. Rose, I. S. Mian, and M. Ozmen, "Capacity bounds on point-to-point communication using molecules," *Proc. IEEE*, vol. 107, no. 7, pp. 1342–1355, 2019.

[11] Y.-P. Hsieh and P.-C. Yeh, "Mathematical foundations for information theory in diffusion-based molecular communications," *arXiv:1311.4431*, 2013.

[12] N. Farsad, D. Pan, and A. Goldsmith, "A novel experimental platform for in-vessel multi-chemical molecular communications," in *Proc. IEEE Global Commun. Conf.*, 2017, pp. 1–6.

[13] S. Giannoukos, A. Marshall, S. Taylor, and J. Smith, "Molecular communication over gas stream channels using portable mass spectrometry," *J. Amer. Soc. Mass Spectrom.*, vol. 28, no. 11, pp. 2371–2383, 2017.

[14] H. Unterweger, J. Kirchner, W. Wicke, A. Ahmadzadeh, D. Ahmed, V. Jamali, C. Alexiou, G. Fischer, and R. Schober, "Experimental molecular communication testbed based on magnetic nanoparticles in duct flow," in *Proc. IEEE Int. Works. Sig. Process. Advances Wireless Commun.*, 2018, pp. 1–5.

[15] L. Grozinger, M. Amos, T. E. Gorochowski, P. Carbonell, D. A. Oyarzún, R. Stoof, H. Fellermann, P. Zuliani, H. Tas, and A. Goñi-Moreno, "Pathways to cellular supremacy in biocomputing," *Nat. Commun.*, vol. 10, no. 1, pp. 1–11, 2019.

[16] I. F. Akyildiz, M. Pierobon, S. Balasubramaniam, and Y. Koucheryavy, "The internet of Bio-Nano things," *IEEE Commun. Mag.*, vol. 53, pp. 32–40, 2015.

[17] S. Senturk, I. Kok, and F. Senturk, "Internet of nano, bio-nano, biodegradable and ingestible things: A survey," *arXiv:2202.12409*, 2022.

[18] C. McBride, R. Shah, and D. Del Vecchio, "The effect of loads in molecular communications," *Proc. IEEE*, vol. 107, no. 7, pp. 1369–1386, 2019.

[19] Y. Liu, J. Li, T. Tschirhart, J. L. Terrell, E. Kim, C.-Y. Tsao, D. L. Kelly, W. E. Bentley, and G. F. Payne, "Connecting biology to electronics: Molecular communication via redox modality," *Adv. Healthc. Mater.*, vol. 6, no. 24, p. 1700789, 2017.

[20] G. Simmons, "A survey of information authentication," *Proc. of the IEEE*, vol. 76, no. 5, pp. 603–620, 1988.

[21] S. Zafar, W. Aman, M. M. U. Rahman, A. Alomainy, and Q. H. Abbasi, "Channel impulse response-based physical layer authentication in a diffusion-based molecular communication system," in *UK/China Emerg. Technol.* IEEE, 2019, pp. 1–2.

[22] F. Dressler and S. Fischer, "Connecting in-body nano communication with body area networks: Challenges and opportunities of the internet of nano things," *Nano Commun. Networks*, vol. 6, pp. 29–38, 2015.

[23] R. Ahlswede and G. Dueck, "Identification Via Channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 15–29, 1989.

[24] G. Calhoun and G. M. Calhoun, *3rd Gen. Wireless Sys.: Post-Shannon Signal Architectures*. Artech House, 2003, vol. 1.

[25] P. Schwenteck, G. T. Nguyen, H. Boche, W. Kellerer, and F. H. P. Fitzek, "6G Perspective of Mobile Network Operators, Manufacturers, and Verticals," *IEEE Networking Letters*, pp. 1–1, 2023.

[26] W. Haselmayr, A. Springer, G. Fischer, C. Alexiou, H. Boche, P. A. Hoeher, F. Dressler, and R. Schober, "Integration of Molecular Communications Into Future Generation Wireless Networks," in *Proc. 1st 6G Wireless Summit., Levi, Finland*, 2019.

[27] J. A. Cabrera, H. Boche, C. Deppe, R. F. Schaefer, C. Scheunert, and F. H. Fitzek, "6G and The Post-Shannon Theory," in *Shaping Future 6G Networks: Needs, Impacts and Technologies*, N. O. Frederiksen and H. Gulliksen, Eds. Hoboken, NJ, United States: Wiley-Blackwell, 2021.

[28] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Sys. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[29] N. Farsad, Y. Murin, A. W. Eckford, and A. Goldsmith, "Capacity limits of diffusion-based molecular timing channels with finite particle lifetime," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 4, no. 2, pp. 88–106, 2018.

[30] M. Pierobon and I. F. Akyildiz, "Capacity of a diffusion-based molecular communication system with channel memory and molecular noise," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 942–954, 2012.

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

[32] G. Aminian, H. Arjmandi, A. Gohari, M. N. Kenari, and U. Mitra, "Capacity of LTI-Poisson channel for diffusion based molecular communication," in *Proc. IEEE Intl. Conf. Commun.*, 2015, pp. 1060–1065.

[33] G. Aminian, H. Arjmandi, A. Gohari, M. Nasiri-Kenari, and U. Mitra, "Capacity of diffusion-based molecular communication networks over LTI-Poisson channels," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 1, no. 2, pp. 188–201, 2015.

[34] A. Etemadi, H. Arjmandi, P. Azmi, and N. Mokari, "Capacity bounds for diffusive molecular communication over discrete-time compound Poisson channels," *IEEE Commun. Lett.*, vol. 23, no. 5, pp. 793–796, 2019.

[35] A. Etemadi, P. Azmi, H. Arjmandi, and N. Mokari, "Compound Poisson noise sources in diffusion-based molecular communication," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4104–4116, 2019.

[36] H. Mahdavifar and A. Beirami, "Diffusion channel with Poisson reception process: capacity results and applications," in *Proc. IEEE Int. Symp. Inf. Theory.*

IEEE, 2015, pp. 1956–1960.

[37] F. Ratti, F. Vakilipoor, H. Awan, and M. Magarini, "Bounds on the constrained capacity for the diffusive Poisson molecular channel with memory," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 7, no. 2, pp. 100–105, 2021.

[38] T. S. Han, *Information-Spectrum Methods in Information Theory*, ser. Stochastic Modelling and Applied Probability. Springer-Verlag Berlin Heidelberg, 2014.

[39] A. C. Yao, "Some Complexity Questions Related to Distributive Computing," in *Proc. Ann. ACM Symp. Theory Comp.*, 1979, pp. 209–213.

[40] L. Lovász, "Communication complexity: A survey," in *Paths, Flows and VLSI-Layout*, 1989, pp. 235–265.

[41] A. Y. Anup Rao, *Communication Complexity: and Applications*. Cambridge University Press, 2020.

[42] Y. S. Abu-Mostafa, *Complexity in Information Theory*. Springer, 1988.

[43] R. Ahlswede, "Elimination of correlation in random codes for arbitrarily varying channels," *Zs. Wahrscheinlichkeitstheorie Verw. Geb.*, vol. 44, no. 2, pp. 159–175, 1978.

[44] R. Ahlswede and G. Dueck, "Identification in the presence of feedback-a discovery of new capacity formulas," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 30–36, Jan 1989.

[45] H. Boche, R. F. Schaefer, and H. Vincent Poor, "On the computability of the secret key capacity under rate constraints," in *IEEE Int. Conf. Acoust. Speech Sig. Proc. (ICASSP)*, May 2019, pp. 2427–2431.

[46] M. V. Burnashev, "On identification capacity of infinite alphabets or continuous-time channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2407–2414, 2000.

[47] H. Boche and C. Deppe, "Secure identification for wiretap channels; robustness, super-additivity and continuity," *IEEE Trans. Inf. Foren. Secur.*, vol. 13, no. 7, pp. 1641–1655, 2018.

[48] A. Winter, "Quantum and classical message identification via quantum channels," *Quantum Inf. Comput.*, *arxiv:quant-ph/0401060*, vol. 5, no. 7, pp. 605–606, 2005.

[49] A. Bracher and A. Lapidoth, "Identification via the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3480–3501, 2017.

[50] C. Shannon, "The zero error capacity of a noisy channel," *IRE Trans. Inf. Theory*, vol. 2, no. 3, pp. 8–19, 1956.

[51] M. J. Salariseddigh, U. Pereg, H. Boche, C. Deppe, V. Jamali, and R. Schober, "Deterministic Identification For Molecular Communications Over The Poisson Channel," *arXiv:2203.02784*, 2022. [Online]. Available: https://arxiv.org/pdf/2203.02784.pdf

[52] J. JáJá, "Identification is Easier Than Decoding," in *Proc. Ann. Symp. Found. Comp. Scien.*, 1985, pp. 43–50.

[53] R. Ahlswede and N. Cai, "Identification Without Randomization," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2636–2642, 1999.

[54] K. Mehlhorn and E. M. Schmidt, "Las Vegas is Better Than Determinism in VLSI and Distributed Computing," in *Proc. of The 14th Ann. ACM symp. on Theory of Comp.*, 1982, pp. 330–337.

[55] S. Derebeyoğlu *et al.*, "Performance analysis of identification codes," *Entropy*, vol. 22, no. 10, p. 1067, 2020.

[56] S. Verdú and V. K. Wei, "Explicit construction of optimal constant-weight codes for identification via channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 30–36, 1993.

[57] K. Kurosawa and T. Yoshida, "Strongly universal hashing and identification codes via channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 6, pp. 2091–2095, 1999.

[58] J. Bringer *et al.*, "Private interrogation of devices via identification codes," in *Proc. Int. Conf. Cryptol. in India*. Springer, 2009, pp. 272–289.

[59] ——, "Identification codes in cryptographic protocols," in *Proc. IEEE Inf. Theory Workshop*, 2010, pp. 1–5.

[60] O. Günlü, J. Kliewer, R. F. Schaefer, and V. Sidorenko, "Code constructions and bounds for identification via channels," *IEEE Trans. Commun.*, vol. 70, no. 3, pp. 1486–1496, 2021.

[61] S. Verdu and V. Wei, "Explicit construction of optimal constant-weight codes for identification via channels," *IEEE Trans. Inf. Theory*, vol. 39, no. 1, pp. 30–36, 1993.

[62] W. Labidi, "Secure Identification for Gaussian Channels," Master's thesis, LNT, Technical University of Munich (TUM), June 2019.

[63] W. Labidi, C. Deppe, and H. Boche, "Secure identification for Gaussian channels," in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Process.*, 2020, pp. 2872–2876.

[64] W. Labidi, H. Boche, C. Deppe, and M. Wiese, "Identification over the Gaussian channel in the presence of feedback," in *Proc. IEEE Int. Symp. Inf. Theory*, 2021, pp. 278–283.

[65] R. Ezzine, W. Labidi, H. Boche, and C. Deppe, "Common randomness generation and identification over Gaussian channels," in *Proc. IEEE Global Commun. Conf.* IEEE, 2020, pp. 1–6.

[66] Z. Brakerski, Y. T. Kalai, and R. R. Saxena, "Deterministic and efficient interactive coding from hard-to-decode tree codes," in *Proc. IEEE Ann. Symp. Found. Comp. Scien.*, 2020, pp. 446–457.

[67] R. L. Bocchino, V. Adve, S. Adve, and M. Snir, "Parallel programming must be deterministic by default," *Usenix HotPar*, vol. 6, no. 10.5555, pp. 1 855 591–1 855 595, 2009.

[68] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[69] M. Wiese, W. Labidi, C. Deppe, and H. Boche, "Identification Over Additive Noise Channels in The Presence of Feedback," *IEEE Trans. Inf. Theory*, pp. 1–1, 2022.

[70] T. Nakano, T. Suda, Y. Okaie, M. J. Moore, and A. V. Vasilakos, "Molecular communication among biological nanomachines: A layered architecture and research issues," *IEEE Trans. Nanobiosci.*, vol. 13, no. 3, pp. 169–197, 2014.

[71] S. Ghavami, "Anomaly detection in molecular communications with applications to health monitoring networks," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 6, no. 1, pp. 50–59, 2020.

[72] R. H. Muller and C. M. Keck, "Challenges and solutions for the delivery of biotech drugs–a review of drug nanocrystal technology and lipid nanoparticles," *J. Biotech.*, vol. 113, no. 1-3, pp. 151–170, 2004.

[73] S. K. Hobbs, W. L. Monsky, F. Yuan, W. G. Roberts, L. Griffith, V. P. Torchilin, and R. K. Jain, "Regulation of transport pathways in tumor vessels: role of tumor type and microenvironment," *Proc. Natl. Acad. Sci.*, vol. 95, no. 8, pp. 4607–4612, 1998.

[74] R. K. Jain, "Transport of molecules, particles, and cells in solid tumors," *Annu. Biomed. Eng. Rev.*, vol. 1, no. 1, pp. 241–263, 1999.

[75] S. Wilhelm, A. J. Tavares, Q. Dai, S. Ohta, J. Audet, H. F. Dvorak, and W. C. Chan, "Analysis of nanoparticle delivery to tumours," *Nat. Rev. Mater.*, vol. 1, no. 5, pp. 1–12, 2016.

[76] M. Liu, Y. Zhang, J. Wang, N. Qin, H. Yang, K. Sun, J. Hao, L. Shu, J. Liu, Q. Chen, P. Zhang, and T. H. Tao, "A star-nose-like tactile-olfactory bionic sensing array for robust object recognition in non-visual environments," *Nat. Commun.*, vol. 13, no. 1, pp. 1–10, 2022.

[77] T. D. Wyatt, *Pheromones and Animal Behaviour*. Cambridge University Press, Cambridge, 2003.

[78] U. B. Kaupp, "Olfactory signalling in vertebrates and insects: Differences and commonalities," *Nature Rev. Neuroscience*, vol. 11, no. 3, pp. 188–200, 2010.

[79] L. B. Buck, "Unraveling the sense of smell (Nobel lecture)," *Angew. Chem. Int. Ed.*, vol. 44, no. 38, pp. 6128–6140, 2005.

[80] A. Buettner, *Springer Handbook of Odor*. Springer, 2017.

[81] V. Jamali, H. M. Loos, A. Buettner, R. Schober, and H. V. Poor, "Olfaction-inspired MCs: Molecule mixture shift keying and cross-reactive receptor arrays," *arXiv:2203.04225, in revision for IEEE Trans. Commun.*, 2022.

[82] H. Boche and C. Deppe, "Robust and secure identification," in *Proc. IEEE Int. Symp. Inf. Theory.*, 2017, pp. 1539–1543.

[83] T. S. Han and S. Verdú, "New results in the theory of identification via channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 14–25, Jan 1992.

[84] U. Tamm, "Communication complexity and orthogonal polynomials," in *Codes and Association Schemes*, 1999.

[85] J. Bringer, H. Chabanne, G. Cohen, and B. Kindarji, "Private interrogation of devices via identification codes," in *Progress in Cryptology - INDOCRYPT 2009*, B. Roy and N. Sendrier, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 272–289.

[86] G. Fettweis, H. Boche, T. Wiegand, E. Zielinski, H. Schotten, P. Merz, S. Hirche, A. Festag, W. Häffner, M. Meyer, and E. Steinbach, "The tactile internet-itu-t technology watch report," *Int. Telecom. Union (ITU), Geneva*, 2014.

[87] K. Guan, B. Ai, M. Liso Nicolás, R. Geise, A. Möller, Z. Zhong, and T. Körner, "On the influence of scattering from traffic signs in vehicle-to-x communications," *IEEE Trans. Vehicl. Tech.*, vol. 65, no. 8, pp. 5835–5849, 2016.

[88] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 160–167, 2016.

[89] P. Moulin, "The role of information theory in watermarking and its application to image watermarking," *Signal Process.*, vol. 81, no. 6, pp. 1121–1139, 2001.

[90] Y. Steinberg and N. Merhav, "Identification in the presence of side information with application to watermarking," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1410–1422, 2001.

[91] R. Ahlswede and N. Cai, *Watermarking Identification Codes with Related Topics on Common Randomness*. Berlin, Heidelberg: Springer-Verlag, 2006, p. 107–153.

[92] R. Gay, A. Charlesworth, and R. Esen, *Online marketing: A customer-led approach*. Oxford University Press, 2007.

[93] C. Gurău, "Integrated online marketing communication: implementation and management," *J. Commun. Manag.*, 2008.

[94] H. Lasi, P. Fettke, H.-G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & information systems engineering*, vol. 6, no. 4, pp. 239–242, 2014.

[95] L. D. Xu, E. L. Xu, and L. Li, "Industry 4.0: state of the art and future trends," *International Journal of Production Research*, vol. 56, no. 8, pp. 2941–2962, 2018.

[96] Y. Lu, "Industry 4.0: A survey on technologies, applications and open research issues," *J. Indust. Inf. Integ.*, vol. 6, pp. 1–10, 2017.

[97] R. Kaylor, D. Everhart, and J. Lindsay, "Healthcare networks with biosensors," Apr. 22 2004, US Patent App. 10/277,170.

[98] M. J. Salariseddigh, U. Pereg, H. Boche, and C. Deppe, "Deterministic Identification Over Channels With Power Constraints," in *Proc. IEEE Int. Conf. Commun., arXiv:2010.04239*, 2021, pp. 1–6. [Online]. Available: http://arxiv.org/abs/2010.04239.pdf

[99] ——, "Deterministic Identification Over Channels With Power Constraints," *IEEE Trans. Inf. Theory*, vol. 68, no. 1, pp. 1–24, 2022.

[100] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. USA: Academic Press, Inc., 1982.

[101] I. Csiszár, "The Method of Types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2505–2523, 1998.

[102] G. Kramer, "Topics in Multi-user Information Theory," *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 4–5, pp. 265–444, 2008.

[103] I. Csiszár and P. C. Shields, "Information Theory and Statistics: A Tutorial," *Foundations and Trends® in Communications and Information Theory*, vol. 1, no. 4, pp. 417–528, 2004.

[104] P. Brémaud, *Discrete Probability Models and Methods: Probability on Graphs and Trees, Markov Chains and Random Fields, Entropy and Coding*. Springer, 2017, vol. 78.

[105] M. J. Salariseddigh, U. Pereg, H. Boche, and C. Deppe, "Deterministic Identification Over Fading Channels," in *2020 IEEE Inf. Theory Workshop (ITW)*, 2021, pp. 1–5.

[106] ——, "Deterministic Identification Over Fading Channels," *arXiv:2010.10010*, 2020. [Online]. Available: https://arxiv.org/pdf/2010.10010.pdf

[107] M. J. Salariseddigh, M. Spahovic, and C. Deppe, "Deterministic K-Identification For Slow Fading Channel," *arXiv preprint arXiv:2212.02732, accepted for publication in IEEE Inf. Theory Workshop 2023*, 2022. [Online]. Available: https://arxiv.org/abs/2212.02732

[108] M. J. Salariseddigh, U. Pereg, H. Boche, C. Deppe, and R. Schober, "Deterministic Identification Over Poisson Channels," in *Proc. IEEE Global Comm. Conf., arXiv:2107.06061*, 2021, pp. 1–6.

[109] ——, "Deterministic Identification Over Poisson Channels," *arXiv:2107.06061*, 2021. [Online]. Available: http://arxiv.org/abs/2107.06061.pdf

[110] M. J. Salariseddigh, V. Jamali, U. Pereg, H. Boche, C. Deppe, and R. Schober, "Deterministic Identification For MC ISI-Poisson Channel," *arXiv:2010.04239*, 2022. [Online]. Available: http://arxiv.org/abs/2211.11024.pdf

[111] M. J. Salariseddigh, V. Jamali, H. Boche, C. Deppe, and R. Schober, "Deterministic Identification For MC Binomial Channel," *arXiv preprint arXiv:2304.12493, accepted for publication in Int. Symp. Inf. Theory 2023*, 2023. [Online]. Available: http://arxiv.org/abs/2304.12493.pdf

[112] O. Dabbabi, M. J. Salariseddigh, C. Deppe, and H. Boche, "Deterministic K-Identification For Binary Symmetric Channel," *arXiv preprint arXiv:2305.04260*, 2023. [Online]. Available: http://arxiv.org/abs/2305.04260.pdf

[113] M. V. Burnashev, "On the method of types and approximation of output measures for channels with finite alphabets," *Prob. Inf. Trans.*, vol. 36, no. 3, pp. 195–212, 2000.

[114] R. Ahlswede, "A Method of Coding and Its Application to Arbitrarily Varying Channels," *J. Comb. Inf. Sys. Scien*, vol. 5, no. 1, 1980.

[115] N. Cai, Pers. Commun.: Prof. Cai noted that a straightforward application of the proof techniques from [114] does not yield the DI capacity result., July. 08 2020.

[116] K. Eswaran, "Identification via channels and constant-weight codes," https://people.eecs.berkeley.edu/~ananth/229BSpr05/Reports/KrishEswaran.pdf, 2005.

[117] A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm (Eds.), *Identification and Other Probabilistic Models, Rudolf Ahlswede's Lectures on Information Theory 6*, 1st ed., ser. Found. Signal Process., Commun. Netw. Springer Verlag, 2020, vol. 15, to appear.

[118] H. Cohn, "Order and Disorder in Energy Minimization," in *Proc. Int. Congr. Mathn.* World Scientific, 2010, pp. 2416–2443.

[119] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2009.

[120] T. Cover and J. Thomas, *Elements of Information Theory*, ser. Wiley Series Telecomm. New York: John Wiley & Sons, 1991.

[121] G. Kramer, *Topics in multi-user information theory*. Now Found. Tren. Commun. Inf. Theory, 2008.

[122] T. S. Han and S. Verdú, "New results in the theory of identification via channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 14–25, 1992.

[123] M. V. Burnashev and S. Verdú, "Measures separated in l_1 metrics and id-codes," *Prob. Inf. Trans.*, vol. 30, no. 3, pp. 3–14, 1994.

[124] J. Bringer *et al.*, "Private interrogation of devices via identification codes," in *Int. Conf. Crypto. India*. Springer, 2009, pp. 272–289.

[125] J. A. Stankovic, "Wireless sensor networks," *Computer*, vol. 41, no. 10, pp. 92–95, 2008.

[126] X. Tian, P. M. Lee, Y. J. Tan, T. L. Y. Wu, H. Yao, M. Zhang, Z. Li, K. A. Ng, B. C. K. Tee, and J. S. Ho, "Wireless body sensor networks based on metamaterial textiles," *Nature Electronics*, vol. 2, no. 6, pp. 243–251, 2019.

[127] H. S. Dhillon, H. Huang, and H. Viswanathan, "Wide-area wireless communication challenges for the internet of things," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 168–174, 2017.

[128] M. Li, H. Yin, Y. Huang, and Y. Wang, "Impact of correlated fading channels on cognitive relay networks with generalized relay selection," *IEEE Access*, vol. 6, pp. 6040–6047, 2018.

[129] A. Goldsmith, *Wireless Communications*. Cambridge university press, 2005.

[130] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Vehicl. Tech.*, vol. 43, no. 2, pp. 359–378, 1994.

[131] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[132] R. Zhang, S. Cui, and Y. Liang, "On ergodic sum capacity of fading cognitive multiple-access and broadcast channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5161–5178, 2009.

[133] R. Masmoudi, "Ergodic capacity for fading channels in cognitive radio networks," in *4th Int. Conf. Adv. Tech. Sig. Image Proc. (ATSIP 2018)*, 2018, pp. 1–6.

[134] X. Yang, "Capacity of Fading Channels Without Channel Side Information," *arXiv:1903.12360*, 2019.

[135] A. Nafkha and N. Demni, "Closed-form expressions of ergodic capacity and mmse achievable sum rate for mimo jacobi and rayleigh fading channels," *IEEE Access*, vol. 8, pp. 149 476–149 486, 2020.

[136] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546–3557, 2020.

[137] M. J. Salariseddigh, U. Pereg, H. Boche, and C. Deppe, "Deterministic Identification Over Channels With Power Constraints," *arXiv:2010.04239*, 2020. [Online]. Available: http://arxiv.org/abs/2010.04239.pdf

[138] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices and Groups*. Springer Science & Business Media, 2013.

[139] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. USA: Cambridge University Press, 2012.

[140] R. L. Goodstein, "Transfinite ordinals in recursive number theory," *J. Symb. Log.*, vol. 12, no. 4, pp. 123–129, 1947.

[141] S. Arora and B. Barak, *Computational complexity: a modern approach*. Cambridge University Press, 2009.

[142] Q. Zhang and V. Tan, "Covert identification over binary-input discrete memoryless channels," *arXiv:2007.13333*, 2021.

[143] B. A. Bash *et al.*, "Limits of reliable communication with low probability of detection on AWGN channels," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1921–1930, 2013.

[144] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge university press, 2005.

[145] Z. Li, R. Yates, and W. Trappe, "Achieving secret communication for fast rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 9, pp. 2792–2799, 2010.

[146] F. Hosseinigoki and O. Kosut, "Capacity of Gaussian arbitrarily-varying fading channels," in *53rd Ann. Conf. Inf. Scien. Syst. (CISS 2019)*, 2019, pp. 1–6.

[147] K. Besser and E. A. Jorswieck, "Bounds on the ergodic secret-key capacity for dependent fading channels," in *WSA 2020; 24th Int. ITG Workshop Smart Anten.*, 2020, pp. 1–5.

[148] K. Niu and Y. Li, "Polar codes for fast fading channel: Design based on polar spectrum," *IEEE Trans. Vehicl. Tech.*, pp. 1–1, 2020.

[149] A. J. Goldsmith and P. P. Varaiya, "Capacity of Fading Channels With Channel Side Information," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.

[150] V. K. N. Lau and T. Wu, "Optimal transmission and limited feedback design for ofdm/mimo systems in frequency selective block fading channels," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1569–1573, 2007.

[151] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, "Capacity results for binary fading interference channels with delayed csit," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6093–6130, 2014.

[152] S. Shamai, "A broadcast strategy for the Gaussian slowly fading channel," in *Proc. IEEE Int. Symp. Inf. Theory.*, 1997, pp. 150–.

[153] S. Shamai and A. Steiner, "A broadcast approach for a single-user slowly fading MIMO channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2617–2635, 2003.

[154] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *Proc. IEEE Int. Symp. Inf. Theory.*, 2006, pp. 356–360.

[155] H. Khoshnevis, I. Marsland, H. Jafarkhani, and H. Yanikomeroglu, "Space–time signal design for multilevel polar coding in slow fading broadcast channels," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 5940–5952, 2019.

[156] H. Arjmandi, A. Gohari, M. N. Kenari, and F. Bateni, "Diffusion-based nanonetworking: A new modulation technique and performance analysis," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 645–648, 2013.

[157] V. Jamali, N. Farsad, R. Schober, and A. Goldsmith, "Diffusive molecular communications with reactive molecules: Channel modeling and signal design," *IEEE Trans. Mol. Biol. Multi-Scale Commun.*, vol. 4, no. 3, pp. 171–188, 2018.

[158] C. T. Chou, "Chemical reaction networks for computing logarithm," *Synth. Biol.*, vol. 2, no. 1, 2017.

[159] R. M. Gagliardi and S. Karp, "Optical communications," *New York*, 1976.

[160] S. S. Shamai, "Capacity of a pulse amplitude modulated direct detection photon channel," *Proc. I-Commun. Speech and Vision*, vol. 137, no. 6, pp. 424–430, 1990.

[161] S. Verdú, "Poisson communication theory," presented at the International Technion Communication Day in Honor of Israel Bar-David, Haifa, Israel, Mar. 25, 1999, Invited Talk, 1999.

[162] H. B. Yilmaz and C.-B. Chae, "Arrival modelling for molecular communication via diffusion," *Electronics Lett.*, vol. 50, no. 23, pp. 1667–1669, 2014.

[163] T. Thomas-Danguin, C. Sinding, S. Romagny, F. El Mountassir, B. Atanasova, E. Le Berre, A.-M. Le Bon, and G. Coureaud, "The perception of odor objects in everyday life: a review on the processing of odor mixtures," *Front. Psychol.*, vol. 5, p. 504, 2014.

[164] H. B. Yilmaz, A. C. Heren, T. Tugcu, and C.-B. Chae, "Three-dimensional channel characteristics for molecular communications with an absorbing receiver," *IEEE Commun. Lett.*, vol. 18, no. 6, pp. 929–932, 2014.

[165] A. Ahmadzadeh, H. Arjmandi, A. Burkovski, and R. Schober, "Comprehensive reactive receiver modeling for diffusive molecular communication systems: Reversible binding, molecule degradation, and finite number of receptors," *IEEE Trans. Nanobiosci.*, vol. 15, no. 7, pp. 713–727, 2016.

[166] A. Noel, K. C. Cheung, and R. Schober, "Improving receiver performance of diffusive molecular communication with enzymes," *IEEE Trans. Nanobiosci.*, vol. 13, no. 1, pp. 31–43, 2014.

[167] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 5, pp. 2334–2354, 2016.

[168] M. Egan, V. Loscrì, T. Q. Duong, and M. Di Renzo, "Strategies for coexistence in molecular communication," *IEEE Trans. Nanobiosci.*, vol. 18, no. 1, pp. 51–60, 2018.

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

[170] R. Urbanke and B. Rimoldi, "Lattice codes can achieve capacity on the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 44, no. 1, pp. 273–278, 1998.

[171] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[172] R. Beals and R. Wong, *Special functions: a graduate text*. Cambridge University Press, 2010, vol. 126.

[173] W. Feller, *An Introduction to Probability Theory and its Applications*. John Wiley & Sons, 1966.

[174] D. S. Mitrinovic, J. Pecaric, and A. M. Fink, *Classical and New Inequalities in Analysis*. Springer Sci. & Bus. Media, 2013, vol. 61.

[175] J. Cao, S. Hranilovic, and J. Chen, "Capacity-achieving distributions for the discrete-time Poisson channel—part i: General properties and numerical techniques," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 194–202, 2013.

[176] ——, "Capacity-achieving distributions for the discrete-time Poisson channel—part ii: Binary inputs," *IEEE Trans. Commun.*, vol. 62, no. 1, pp. 203–213, 2013.

[177] N. Ahmadypour and A. Gohari, "Transmission of a bit over a discrete Poisson channel with memory," *IEEE Trans. Inf. Theory*, vol. 67, no. 7, pp. 4710–4727, 2021.

[178] R. Mosayebi, H. Arjmandi, A. Gohari, M. Nasiri-Kenari, and U. Mitra, "Receivers for diffusion-based molecular communication: Exploiting memory and sampling rate," *IEEE J. Selec. Areas Commun.*, vol. 32, no. 12, pp. 2368–2380, 2014.

[179] S. Galmés and B. Atakan, "Performance analysis of diffusion-based molecular communications with memory," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3786–3793, 2016.

[180] F. Vakilipoor, F. Ratti, M. Magarini, and H. Awan, "Linear receiver design for time-varying Poisson molecular communication channels with memory," in *Intl. Wksp. Signal Process. Adv. Wireless Commun.* IEEE, 2020, pp. 1–5.

[181] A. O. Kislal, H. B. Yilmaz, A. E. Pusane, and T. Tugcu, "Isi-aware channel code design for molecular communication via diffusion," *IEEE Trans. Nanobiosci.*, vol. 18, no. 2, pp. 205–213, 2019.

[182] N. Ahmadypour and A. Gohari, "Transmission of a bit over a discrete Poisson channel with memory," *arXiv:2011.05931*, 2020. [Online]. Available: https://arxiv.org/pdf/2011.05931.pdf

[183] T. D. Ahle, "Sharp and Simple Bounds For The Raw Moments of The Binomial and Poisson Distributions," *Stat. Probab. Lett.*, vol. 182, p. 109306, 2022.

[184] C. Komninakis, L. Vandenberghe, and R. D. Wesel, "Capacity of The Binomial Channel, or Minimax Redundancy For Memoryless Sources," in *Intl. Symp. Inf. Theory*, 2001, pp. 127–127.

[185] R. D. Wesel, E. E. Wesel, L. Vandenberghe, C. Komninakis, and M. Medard, "Efficient Binomial Channel Capacity Computation With an Application to Molecular Communication," in *2018 Inf. Theory Appl. Workshop (ITA)*, 2018, pp. 1–5.

[186] N. Farsad, C. Rose, M. Médard, and A. Goldsmith, "Capacity of Molecular Channels With Imperfect Particle-Intensity Modulation and Detection," in *2017 IEEE Intl. Symp. Inf. Theory (ISIT)*, 2017, pp. 2468–2472.

[187] M. Damrath, "Channel Coding in Molecular Communication," Ph.D. dissertation, Technischen Fakultät der Christian-Albrechts-Universität zu Kiel, 2020. [Online]. Available: https://macau.uni-kiel.de/servlets/MCRFileNodeServlet/macau_derivate_00001814/Martin_Damrath.pdf

[188] A. Noel, K. C. Cheung, and R. Schober, "Improving Receiver Performance of Diffusive Molecular Communication With Enzymes," *IEEE Trans. Nanobiosci.*, vol. 13, no. 1, pp. 31–43, 2014.

[189] N. Farsad and A. Goldsmith, "A Molecular Communication System Using Acids, Bases and Hydrogen Ions," in *2016 IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, 2016, pp. 1–6.

[190] F. Qi, "Bounds For The Ratio of Two Gamma Functions–From Wendel's and Related Inequalities to Logarithmically Completely Monotonic Functions," *arXiv preprint arXiv:0904.1048*, 2009. [Online]. Available: https://arxiv.org/abs/0904.1048

[191] E. Biglieri, J. Proakis, and S. Shamai, "Fading Channels: Information-Theoretic and Communications Aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, 1998.

[192] A. T. Asyhari and A. G. Fàbregas, "Nearest neighbor decoding in MIMO block-fading channels with imperfect csir," *IEEE Trans. Info. Theory.*, vol. 58, no. 3, pp. 1483–1517, 2012.

[193] R. Ahlswede, "General Theory of Information Transfer: Updated," *Discrete Appl. Math.*, vol. 156, no. 9, pp. 1348–1388, 2008.

[194] Y. Li, X. Wang, H. Zhang, J. Wang, W. Tong, G. Yan, and Z. Ma, "Deterministic Identification Over Channels Without CSI," in *IEEE Inf. Theory Workshop*, 2022, pp. 332–337.

[195] F. Topsøe, "Some bounds for the logarithmic function," *RGMIA Res. Rep. Collection*, vol. 7, no. 2, pp. 1–20, 2004.

[196] P. Elias, "Coding For Two Noisy Channels," in *Proc. 3rd London Symp. Inf. Theory*, 1955.

[197] M. J. Salariseddigh, M. Spahovic, and C. Deppe, "Deterministic K-Identification For Slow Fading Channel," *arXiv:2212.02732, accepted for publication in Inf. Theory Workshop 2023*, 2022. [Online]. Available: https://arxiv.org/pdf/2212.02732.pdf

[198] R. Ahlswede, L. Bäumer, N. Cai, H. Aydinian, V. Blinovsky, C. Deppe, and H. Mashurian, *General Theory of Information Transfer and Combinatorics*. Springer, 2006, vol. 68.

[199] H. Yamamoto and M. Ueda, "Multiple Object Identification Coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4269–4276, 2015.

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

[201] E. N. Gilbert, "A Comparison of Signalling Alphabets," *Bell Sys. Tech. J.*, vol. 31, no. 3, pp. 504–522, 1952.

[202] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[203] J. H. Van Lint, *Introduction to Coding Theory*. Springer Science & Business Media, 1998, vol. 86.

[204] J. Rosenberger, A. Ibrahim, C. Deppe, and R. Ferrara, "Deterministic Identification Over Multiple-Access Channels," in *IEEE Int. Symp. Inf. Theory 2023*. IEEE, Jun 2023.

[205] H. Boche, Y. Böck, and C. Deppe, "On Effective Convergence in Fekete's Lemma and Related Combinatorial Problems in Information Theory," *arXiv:2010.09896*, 2020.

[206] R. Ahlswede, "On Concepts of Performance Parameters For Channels," in *General Theory of Information Transfer and Combinatorics*. Springer, 2006, pp. 639–663.

[207] H. Boche and C. Deppe, "Computability of The Channel Reliability Function and Related Bounds," *arXiv:2101.09754*, 2021.

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

[209] A. Lapidoth and S. M. Moser, "Bounds on the capacity of the discrete-time Poisson channel," in *Proc. Allerton Conf. Comm., Control, Computing*, vol. 41, no. 1. The University; 1998, 2003, pp. 201–210.

[210] A. Lapidoth, J. H. Shapiro, V. Venkatesan, and L. Wang, "The discrete-time Poisson channel at low input powers," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3260–3272, 2011.

[211] L. Wang and G. W. Wornell, "A refined analysis of the Poisson channel in the high-photon-efficiency regime," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4299–4311, 2014.

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

[213] J. Cao, S. Hranilovic, and J. Chen, "Lower bounds on the capacity of discrete-time Poisson channels with dark current," in *25th Biennial Symp. Commun.* IEEE, 2010, pp. 357–360.

[214] Y. Yu, Z. Zhang, L. Wu, and J. Dang, "Lower bounds on the capacity for Poisson optical channel," in *Proc. Int. Conf. Wireless Commun. and Sig.* IEEE, 2014, pp. 1–5.

[215] M. Cheraghchi and J. Ribeiro, "Improved upper bounds and structural results on the capacity of the discrete-time Poisson channel," *IEEE Trans. Inf. Theory*,

vol. 65, no. 7, pp. 4052–4068, 2019.

[216] ——, "Non-asymptotic capacity upper bounds for the discrete-time Poisson channel with positive dark current," *arXiv:2010.14858*, 2020.

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

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

[219] A. ElMoslimany and T. M. Duman, "On the discreteness of capacity-achieving distributions for fading and signal-dependent noise channels with amplitude-limited inputs," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1163–1177, 2017.

[220] J. Fahs and I. Abou-Faycal, "On properties of the support of capacity-achieving distributions for additive noise channel models with input cost constraints," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1178–1198, 2017.

[221] A. Dytso, M. Goldenbaum, H. V. Poor, and S. S. Shitz, "When are discrete channel inputs optimal?—optimization techniques and some new results," in *52nd Annu. Conf. Inf. Scien. Sys. (CISS)*. IEEE, 2018, pp. 1–6.

[222] A. Dytso, L. Barletta, and S. Shamai, "Bounds on the number of mass points of the capacity achieving distribution of the amplitude constraint Poisson noise channel," *arXiv:2104.14431*, 2021.

[223] J. Cao, "Discrete-time Poisson channel: capacity and signalling design," Ph.D. dissertation, McMaster University, 2013.

[224] A. Lapidoth and S. M. Moser, "The asymptotic capacity of the discrete-time Poisson channel," *Proc. Win. Sch. Coding and Inf. Theory, Monte Verita, Ascona, Switzerland*, 2003.

[225] D. Brady, "The analysis of optical, direct detection communication systems with point process observations," Ph.D. dissertation, Princeton University, 1990. [Online]. Available: https://search.proquest.com/docview/303849140

[226] K. Chakraborty and P. Narayan, "The Poisson fading channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2349–2364, 2007.

[227] F. Ratti, F. Vakilipoor, M. Magarini, and H. Awan, "Upper and lower bounds of constrained capacity in diffusion-based molecular communication," in *2020 IEEE 21st Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*. IEEE, 2020, pp. 1–5.

[228] T. D. Ahle, "Sharp and simple bounds for the raw moments of the binomial and Poisson distributions," *Stat. Probab. Lett.*, vol. 182, p. 109306, 2022.

[229] H. Robbins, "A Remark On Stirling's Formula," *The American mathematical monthly*, vol. 62, no. 1, pp. 26–29, 1955.

[230] J. Flum and M. Grohe, *Parameterized Complexity Theory*, ser. Texts in Theoretical Comput. Sci. An EATCS Series. Springer, 2006.