ABSTRACT

Title of dissertation: INFORMATION THEORETIC SECRET KEY GENERATION: STRUCTURED CODES AND TREE PACKING

Sirin Nitinawarat, Doctor of Philosophy, 2010

Dissertation directed by: Professor Prakash Narayan
Department of Electrical and Computer Engineering
Institute for Systems Research

This dissertation deals with a multiterminal source model for secret key generation by multiple network terminals with prior and privileged access to a set of correlated signals complemented by public discussion among themselves. Emphasis is placed on a characterization of secret key capacity, i.e., the largest rate of an achievable secret key, and on algorithms for key construction. Various information theoretic security requirements of increasing stringency: weak, strong and perfect secrecy, as well as different types of sources: finite-valued and continuous, are studied. Specifically, three different models are investigated.

First, we consider strong secrecy generation for a discrete multiterminal source

model. We discover a connection between secret key capacity and a new source coding concept of "minimum information rate for signal dissemination," that is of independent interest in multiterminal data compression. Our main contribution is to show for this discrete model that structured linear codes suffice to generate a strong secret key of the best rate.

Second, strong secrecy generation is considered for models with continuous observations, in particular jointly Gaussian signals. In the absence of suitable analogs of source coding notions for the previous discrete model, new techniques are required for a characterization of secret key capacity as well as for the design of algorithms for secret key generation. Our proof of the secret key capacity result, in particular the converse proof, as well as our capacity-achieving algorithms for secret key construction based on structured codes and quantization for a model with two terminals, constitute the two main contributions for this second model.

Last, we turn our attention to perfect secrecy generation for fixed signal observation lengths as well as for their asymptotic limits. In contrast with the analysis of the previous two models that relies on probabilistic techniques, perfect secret key generation bears the essence of "zero-error information theory," and accordingly, we rely on mathematical techniques of a combinatorial nature. The model under consideration is the "Pairwise Independent Network" (PIN) model in which every pair of terminals share a random binary string, with the strings shared by distinct pairs of terminals being mutually independent. This model, which is motivated by practical aspects of a wireless communication network in which terminals communicate on the same frequency, results in three main contributions. First, the concept

of perfect omniscience in data compression leads to a single-letter formula for the perfect secret key capacity of the PIN model; moreover, this capacity is shown to be achieved by linear noninteractive public communication, and coincides with strong secret key capacity. Second, taking advantage of a multigraph representation of the PIN model, we put forth an efficient algorithm for perfect secret key generation based on a combinatorial concept of maximal packing of Steiner trees of the multigraph. When all the terminals seek to share perfect secrecy, the algorithm is shown to achieve capacity. When only a subset of terminals wish to share perfect secrecy, the algorithm is shown to achieve at least half of it. Additionally, we obtain nonasymptotic and asymptotic bounds on the size and rate of the best perfect secret key generated by the algorithm. These bounds are of independent interest from a purely graph theoretic viewpoint as they constitute new estimates for the maximum size and rate of Steiner tree packing of a given multigraph. Third, a particular configuration of the PIN model arises when a lone "helper" terminal aids all the other "user" terminals generate perfect secrecy. This model has special features that enable us to obtain necessary and sufficient conditions for Steiner tree packing to achieve perfect secret key capacity.

INFORMATION THEORETIC SECRET KEY GENERATION: STRUCTURED CODES AND TREE PACKING

by

Sirin Nitinawarat

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Professor Prakash Narayan, Chair/Advisor
Professor Alexander Barg
Professor John Baras
Professor Sennur Ulukus
Professor Jonathan Katz (Dean's Representative)

Acknowledgements

I express my heartfelt gratitude to my advisor, Professor Prakash Narayan, for teaching me information theory. I thank him not only for his direction of and collaboration on this research, and his encouragement and patience, but especially also for being a role model of an academic researcher. It is a true privilege to have had the opportunity to work with him.

I owe special thanks also to Professor Alexander Barg for teaching me coding theory and for many instructive discussions on the subject of this dissertation as well as on other subjects in coding theory and information theory.

I am grateful to Dr. Chuxuan Ye and Dr. Alex Reznik of InterDigital for introducing me to the "Pairwise Independent Network" model and the role of tree packing in secret key generation. Our collaborative investigation of this model was a rewarding experience.

I thank my committee members especially Professors John Baras and Jonathan Katz for their stimulating comments and questions, and Professor Katz for pointers to relevant references on "group secret key distribution" based on the approach of computational theoretic security.

Special thanks are due again to Professors Narayan and Barg for the many benefits of their "Information and Coding Theory Research Seminar." I appreciate all their very helpful questions and comments during each seminar. The experience in this seminar constitutes a significant part of my development as a researcher.

Because of all my friends at Maryland: Ravi Tandon, Arya Mazumdar, Prasanth Anthapadmanabhan, Punarbasu Purkayastha, Himanshu Tyagi and Vishal

Patel, there has never been a short supply of stimulating discussions on various subjects including but not limited to my research. I thank them again for enriching my tenure as a graduate student.

With my deepest gratitude, I dedicate this dissertation to my family. I am especially indebted to my parents for their continuing support, encouragement and love. Special thanks are due to my grandmother and sister for their love and patience.

I gratefully acknowledge that this research has been supported by the National Science Foundation under grants ITR/SI(SPIII) Grant 0112560, CCF0515124, CCF0635271, CCF0830697 and by InterDigital.

Table of Contents

Chapter 1

Introduction: Information Theoretic Secret Key Generation

1.1 Overview of Prior Work

Information security is a crucial requirement in current and emerging communication networks, and issues of secure communication have thrust themselves to the forefront of network operation, and of research in information theory. These developments emphasize the need for the study of network or multiterminal models of information security, which are of significantly greater scope and complexity than their point-to-point predecessors.

The security of all currently used cryptosystems is reliant on the difficulty currently faced in solving an underlying computational problem, e.g., factoring large numbers into prime numbers or computing discrete logarithms. Such a notion of computational security can offer guarantees only under assumptions of restricted computational power available to an adversary. Thus, it is desirable – from both a theoretical as well as a practical standpoint – to design cryptosystems that are based on a rigorously provable notion of security which does not assume any restrictions on an adversary's computational power. The notion of information theoretic secrecy or unconditional security meets this requirement. It affords the guarantee that a secret key and, thereby, a legitimate plaintext message are, in effect, statistically independent of the observations of an adversary with eavesdropping and wiretap-

ping capabilities, and not limited in terms of computational resources. The first information theoretic study of a secret key cryptosystem was Shannon's classical work [47].

An important and popular class of models in the study of information theoretic security, that fall in the category of "keyless" cryptosystems, involve reliable and secure message transmission over insecure channels. Wyner's pioneering wiretap channel [51] has one input and two outputs; a legitimate transmitter controls the input, while a legitimate receiver and a wiretapper have access to each output. The wiretapper's channel output is a degraded version of that of the legitimate terminal. Subsequently, Csiszár and Körner [12] generalized Wyner's result for a model in which the eavesdropper's channel output need only be more noisy than that of the legitimate channel. These initial works have inspired numerous important extensions in several new directions in recent years; a survey can be found in [33].

This dissertation deals with secret key generation for a multiple source model by multiple network terminals based on prior and privileged access to a set of correlated signals followed by public discussion among themselves. Separate terminals that observe the outputs of distinct albeit correlated sources can generate a secret key (SK) by means of public communication. Specifically, these terminals are able to generate "common randomness" (CR) regarding which an eavesdropper, with access to the public communication, can glean only a negligibly small amount of mutual information. Typical applications arise in sensor networks and satellite-terrestrial networks. This phenomenon, first observed by Bennett et al [4] and Maurer [39], followed by Ahlswede and Csiszár [1], for a model with two terminals, has been

investigated later by many researchers. A model which includes a helper terminal, observing the output of another source and assisting in generating SK, was investigated by Ahlswede and Csiszár [1], and by Csiszár and Narayan [14]. These works were followed by that of Csiszár and Narayan [15, 16] in which SK generation was studied for models with arbitrary numbers of terminals and arbitrary number of helpers. These models of SK generation are broadly referred to as "multiterminal source models."

Another class of SK generation models namely "multiterminal channel models" are considered in [39, 1] for models with two terminals, and in [16, 17] for models with arbitrary numbers of terminals and arbitrary numbers of helpers. In a channel model, a set of terminals can transmit information over a secure multi-input multi-output channel to another set of terminals. Additionally, all the terminals are allowed to communicate over a public noiseless channel of unlimited capacity, which is observed by an eavesdropper.

1.2 The Multiterminal Source Model

Our focus is on the multiterminal source model introduced in [15]. In this model, each of the terminals observes a distinct component of a memoryless multiple source; we consider sources with discrete as well as continuous alphabets. A set of the terminals then wish to generate a SK with the cooperation of the remaining terminals. To this end, the terminals are allowed to communicate publicly with each other, possibly interactively in many rounds. No rate constraint is imposed on

the public communication. Randomization may be permitted at each terminal.

We assume that an eavesdropper has full access to the public interterminal communication, but that it is passive, i.e., it cannot tamper with the public communication. No restrictions are assumed on the eavesdropper's computational power.

The SK capacity—the largest rate at which a SK can be generated—for a model with a discrete memoryless multiple source (DMMS) is determined in [15]. This capacity result holds in a strong sense: the mutual information of the SK and the public communication vanishes exponentially in the observation length. A concept of strong SK capacity was introduced in [40] in which the mentioned mutual information is only required to go to 0, and the stronger version we use here was first considered in [11, 14, 15].

The SK capacity for a model with a DMMS derived in [15] reveals an innate connection between SK generation and lossless distributed data compression without any secrecy constraints. In particular, consider m terminals each observing independent and identically distributed (i.i.d.) repetitions of discrete random variables (rvs) $X_1, \ldots, X_m$, respectively. A set of terminals $A \subseteq \{1, \ldots, m\}$ seek a SK with the help of the remaining terminals. The SK capacity for this model can be computed by subtracting from the total joint entropy $H(X_1, \ldots, X_m)$ the smallest rate of communication which enables each terminal in A to reconstruct near-losslessly all the m components of the DMMS, i.e., for the terminals in A to become omniscient. The problem of determining the latter smallest rate is one of multiterminal data compression and does not involve any secrecy constraints.

The mentioned connection also suggests a means of generating a SK of op-

timum rate for such a model with a DMMS by decomposing the problem of SK generation into two parts. First, the terminals publicly communicate at the most parsimonious rate to enable all the terminals in A to become omniscient. Second, each terminal in A generates a SK by extracting from this omniscience part that is nearly independent of the public communication. It is also shown in [15, 16] that the SK capacity can be achieved, based on the decomposition, by noninteractive communication and without randomization.

1.3 Motivation

Algorithms for SK generation for multiterminal source models constitute a largely unexplored domain; preliminary results are available for models consisting of two terminals and for special cases with multiple terminals [53, 56]. For example, for a source model, can a SK of optimum rate be generated using structured codes, e.g., linear codes?

Next, turning to a multiterminal source model in which the multiple source observed by the terminals is continuous-valued, e.g., a memoryless jointly Gaussian multiple source, the SK capacity is not immediate, in general, since there is no meaningful analog of the concept of minimum communication for omniscience. However, from the discussion above, we can expect an inherent connection between the problems of SK generation and lossy data compression. What are the characterizations of SK capacities for source models with memoryless jointly Gaussian multiple sources? Can a SK of optimum rate be generated using structured codes,

e.g., lattice and linear codes?

SK generation for a special incarnation of the source model has an intriguing connection with combinatorial problems, in theoretical computer science, of tree packing in multigraphs. This connection is seen in SK generation for a "pairwise independent network" (PIN) source model. The PIN model is motivated by practical aspects of a wireless communication channel in which the transmitters and receivers operate on the same frequency in a multipath environment. The SK capacity for the PIN model will be seen to depend on the joint distribution of the underlying rvs only through the best rates of pairwise SKs. First, this fact suggests that a (globally) optimum secret key for the terminals in A can be built from locally generated keys for pairs of terminals. Second, it hints at the possibility of a connection to algorithms for tree packing in order to propagate optimum pairwise keys to form a globally optimum key. Can tree packing algorithms be used, in general, to generate global keys for a set of terminals from locally generated keys? Under what conditions are such global keys of optimum rate? Furthermore, can information theoretic SK generation provide tools for investigating problems of tree packing in multigraphs?

This dissertation strives to answer the questions raised above.

1.4 Outline and Summary of Contributions

We begin in Section 2.1 with a description of a general multiterminal source model for information theoretic secrecy generation by multiple terminals based on privileged and correlated observations of signals at the terminals, followed by public

interterminal communication. Various information theoretic security requirements of increasing stringency: weak, strong and perfect secrecy, are discussed followed by the operational definitions of the corresponding notions of SK capacity.

Section 2.2 deals with secrecy generation for the discrete multiterminal source model. Specifically, in Section 2.2.1, we present an existing and motivating result on a connection between SK capacity and a source coding concept of "minimum rate of communication for omniscience." Section 2.2.2 describes our contribution on a new connection between SK capacity and a new source coding concept of "minimum information rate for signal dissemination." Section 2.2.3 contains our results on the structural properties of optimal codes for secrecy generation; specifically, we show that linear codes suffice to generate a strong SK of the best rate.

Chapter 3 investigates secrecy generation for models with continuous observations focusing on jointly Gaussian signals. In the absence of analogs to source coding notions for the discrete model of Chapter 2, new techniques are required for the characterization of SK capacity as well as for the design of algorithms for SK generation. The proof of the SK capacity result in Section 3.2, in particular the converse proof, as well as the design of capacity-achieving algorithms for SK generation based on structured codes and quantization, for a model with two terminals in Section 3.3, constitute the two main contributions of this chapter.

In Chapter 4, we turn our attention to perfect SK generation for fixed signal observation lengths as well as for their asymptotic limits. In contrast with the substance of Chapters 2 and 3 that relies on probabilistic techniques, Chapter 4 bears the essence of "zero-error information theory," and accordingly, we rely on

mathematical techniques of a combinatorial nature. The model under consideration in this chapter is the "Pairwise Independent Network (PIN)" model in which every pair of terminals share a random binary string, with the strings shared by distinct pairs of terminals being mutually independent. Chapter 4 offers three main contributions. First, the concept of perfect omniscience leads to a single-letter formula for the perfect SK capacity of the PIN model; moreover, this capacity is shown to be achieved by linear noninteractive communication, and coincides with the strong SK capacity. Second, taking advantage of a multigraph representation of the PIN model, we put forth an efficient algorithm for perfect SK generation based on a combinatorial concept of a maximal packing of Steiner trees of the multigraph. When all the terminals seek to share perfect secrecy, the algorithm is shown to achieve perfect SK capacity. However, when only a subset of terminals wish to share perfect secrecy, the algorithm can fall short of achieving capacity; nonetheless, it is shown to achieve at least half of it. Additionally, we obtain nonasymptotic and asymptotic bounds on the size and rate of the best perfect SK generated by the algorithm. These bounds are of independent interest from a purely graph theoretic viewpoint as they constitute new estimates for the maximum size and rate of Steiner tree packing of a given multigraph. Third, a particular configuration of the PIN model arises when a lone "helper" terminal aids all the other "user" terminals generate perfect secrecy. This model has special features that enable us to obtain necessary and sufficient conditions for Steiner tree packing to achieve perfect SK capacity, as also a further sufficient condition that posits a "weak" role for the helper terminal.

In the concluding Chapter 5, we first compile in Section 5.1 specific open

problems emerging from our work in Chapters 3 and 4 that are yet to be resolved. Finally, in Section 5.2 we point out broader research directions that are motivated by this dissertation.

1.5 Apropos Group Secret Key Cryptosystems

There exists a rich body of work on group secret key distribution based on the approach of computational theoretic security, on which existing cryptosystems are based (cf. e.g., [6, 29, 36, 35, 37]). It is our hope that our work on information theoretic group secret key generation will serve as a useful complementary step.

Chapter 2

The General Multiterminal Source Model

This dissertation deals with models for secrecy generation by multiple terminals based on privileged and correlated observations of signals at the terminals, followed by public interterminal communication. Typical applications arise in sensor networks and satellite-terrestrial networks. For example, in a sensor network, multiple sensor nodes are deployed to measure a parameter of the environment; the distinct measurements at the various sensor nodes usually exhibit certain correlation structures depending on their locations. In a satellite-terrestrial network, various ground stations can observe different noisy versions of a common broadcast signal. In this setting, if the terminals are afforded a means to publicly exchange messages, then it transpires that the terminals can devise a secret key (SK) [39, 1]. In other words, these terminals are able to generate common randomness regarding which an eavesdropper with access to the public communication can glean only a negligibly small amount of mutual information.

2.1 Description and Secret Key (SK) Capacity

We begin with a description of the "multiterminal source model" for SK generation. Our model builds on the discrete model introduced in [15, 16]. Terminals $1, \ldots, m$ represent legitimate parties that cooperate in SK generation. Let

$X_1, \ldots, X_m$ be discrete or $\mathbb{R}$-valued rvs with alphabets denoted by $\mathcal{X}_1, \ldots, \mathcal{X}_m$, and with (known) joint probability distribution. Let $\mathcal{M} = \{1, \ldots, m\}$. Terminal $i \in \mathcal{M}$ observes n independent identically distributed (i.i.d.) repetitions of the rv X_i, namely $\mathbf{X}_i = \mathbf{X}_i^{(n)} = X_i^n = (X_{i1}, \ldots, X_{in})$. We use the notation $X_{\mathcal{M}} \triangleq (X_1, \ldots, X_m)$ and $\mathbf{X}_{\mathcal{M}} \triangleq (\mathbf{X}_1, \ldots, \mathbf{X}_m)$. Following these observations, the terminals are allowed to communicate over a public noiseless channel, possibly interactively in multiple rounds. We assume without loss of generality that the public communication, which may be interactive, takes place in consecutive time slots in r rounds. Specifically, following the formulation in [15], it is depicted by the mappings $f_1, \ldots, f_{mr}$ with f_ν corresponding to the transmission in slot ν by terminal $i \equiv \nu \bmod m$; we allow f_ν to yield any function of the source sequence ($\mathbf{X}_i = \mathbf{x}_i$) observed at terminal i and of all previous communication $f_{[1,\nu-1]} = (f_1, \ldots, f_{\nu-1})$. The corresponding rvs representing the communication are denoted by $F_1, \ldots, F_{mr}$, with $F_\nu = f_\nu(\mathbf{X}_i, F_{[1,\nu-1]})$. We denote the communication collectively by $\mathbf{F} = F_{[1,mr]}$. The goal is for a set of terminals $A \subseteq \mathcal{M}$ to generate *secret* common randomness with the cooperation of the remaining terminals in $\mathcal{M} \backslash A$, which is concealed from an eavesdropper with access to the public communication $\mathbf{F}$. This is formalized next.

Following [15], given $\epsilon > 0$, a rv L will be said to be ϵ-*recoverable* from a rv Z if there exists a function $f(Z)$ so that $Pr\{L \neq f(Z)\} \leq \epsilon$. If such a function exists so that $Pr\{L \neq f(Z)\} = 0$, then L is said to be *perfectly recoverable* from Z.

A function L of $\mathbf{X}_{\mathcal{M}}$ is ϵ-*common randomness* (ϵ-CR) for a set of terminals $A \subseteq \mathcal{M}$, achievable with communication $\mathbf{F}$, if L is ϵ-recoverable from $(\mathbf{X}_i, \mathbf{F})$, for each $i \in A$.

A function K of $\mathbf{X}_{\mathcal{M}}$ with values in a *finite* set $\mathcal{K}$ constitutes an *ϵ-secret key* (ϵ-SK) for a set of terminals $A \subseteq \mathcal{M}$, achievable with communication $\mathbf{F}$, if K is ϵ-CR for A and, in addition, K has a *security index*[1]

$$s(K;\mathbf{F}) \triangleq \log|\mathcal{K}| - H(K|\mathbf{F}) \leq \epsilon, \tag{2.1}$$

where $|\mathcal{K}|$ denotes the cardinality of $\mathcal{K}$. Observe that if K is an ϵ-SK, then both

$$\log|\mathcal{K}| - H(K) \leq \epsilon \tag{2.2}$$

and

$$I(K \wedge \mathbf{F}) \leq \epsilon \tag{2.3}$$

hold so that K is nearly uniformly distributed and is nearly independent of $\mathbf{F}$, since ϵ is typically small.

Definition 2.1: A nonnegative number R is an *achievable SK rate* for a set of terminals $A \subseteq \mathcal{M}$ if there exist[2] ϵ_n-SKs $K^{(n)}$ with values in finite sets $\mathcal{K}^{(n)}$ that are achievable with suitable public communication (with the number of rounds possibly depending on n), such that $\lim_{n\to\infty}\epsilon_n = 0$ and $\lim_{n\to\infty}\frac{1}{n}\log|\mathcal{K}^{(n)}| = R$. The supremum of achievable SK rates for A is called the *SK capacity* $C(A)$. An ϵ_n-SK is termed a *strong* SK if ϵ_n vanishes exponentially in n; the corresponding SK capacity is called the *strong SK capacity*.

Remark: In the earlier works on the source models for SK generation, a weaker notion of SK was adopted [39, 1]. In particular, the corresponding notion of SK

[1] All logarithms are natural unless stated otherwise.

[2] The requirement is only for an infinite sequence of $K^{(n)}$ (infinitely many n), and not necessarily for all n sufficiently large.

capacity therein, which we shall call the *weak* SK capacity, is defined as the largest rate of a sequence of ϵ_n-SKs with ϵ_n being required to satisfy only the condition $\lim_{n\to\infty} n\epsilon_n = 0$. In other words, the secrecy requirement on the SK is relaxed from that one in Definition 2.1 to $\lim_{n\to\infty} \frac{1}{n} s(K;\mathbf{F}) = 0$. An obvious drawback of this weak notion of SK is that $s(K;\mathbf{F})$ can grow with n (as long as it grows slower than n) and, hence, the *weak* SK K may satisfy $\lim_{n\to\infty} I(K \wedge \mathbf{F}) = \infty$. In effect, the eavesdropper can gather an unbounded amount of information regarding the SK with increasing signal observation length. However, it was shown in [41, 15] that in almost all source models for SK generation studied previously, weak SK capacities coincide with (strong) SK capacities.

Next, for certain multiterminal source models for SK generation, e.g., the Pairwise Independent Network model of Chapter 4, it is interesting to investigate an even stronger notion of SK, namely a perfect SK. It will also be clear that this notion affords the strongest possible form of information theoretic secrecy. The notion of perfect secrecy dates back to the work of Shannon [47] in the context of a secrecy model involving a noiseless channel. Here, we investigate perfect secrecy in the context of multiterminal source models for SK generation.

Definition 2.2: A nonnegative number R is an *achievable perfect SK rate* for a set of terminals $A \subseteq \mathcal{M}$ if there exist 0-SKs $K^{(n)}$ with values in $\mathcal{K}^{(n)}$ that are achievable with suitable public communication (with the number of rounds possibly depending on n), such that $\lim_{n\to\infty} \frac{1}{n} \log |\mathcal{K}^{(n)}| = R$. The supremum of achievable perfect SK rates for A is called the *perfect SK capacity* $C_p(A)$.

Observe that if K is a 0-SK, i.e., a perfect SK, then K is independent of $\mathbf{F}$

and is uniformly distributed.

Remarks: (i) The stringent requirement of a perfect SK K, namely $s(K;\mathbf{F}) = 0$ makes the study of perfect SK generation a subject in zero-error information theory. Zero-error information theoretic problems are typically of a combinatorial nature [30], and this is true of perfect SK generation as well. Proof techniques involved in results concerning perfect SK generation are of a combinatorial kind, and not probabilistic such as those involved in the study of (strong) SK generation.

(ii) In the formulation of models for SK generation or perfect SK generation, it is possible to allow the terminals, in addition to engaging public communication, to also randomize. The randomization at each terminal $i \in \mathcal{M}$ is represented by a rv M_i such that the rvs $M_1, M_2, \ldots, M_m, (\mathbf{X}_1, \ldots, \mathbf{X}_m)$ are mutually independent. By virtue of this mutual independence, it transpires that randomization does not enhance SK capacities [15]; this is why we chose to exclude randomization in our general model for SK generation. On the other hand, randomization can facilitate SK construction and in some of our specific models below, we explicitly use randomization in code constructions for SK generation (without any enhancement in SK capacity).

2.2 SK Generation for the Discrete Model

In this section, we focus on SK generation for a model in which each X_i takes values in a finite set $\mathcal{X}_i,\ i \in \mathcal{M}$. We discuss certain connections between the SK generation and related problems of multiterminal data compression. We begin with

a summary of an existing and motivating result on a connection between the SK capacity and a source coding concept of "minimum rate of communication for omniscience." Then we proceed to describe our contribution on a new connection between the SK capacity and a new source coding concept of "minimum information rate for signal dissemination." Lastly, we discuss our results on the structural properties of optimal codes for SK generation.

2.2.1 SK Generation from Omniscience

The SK capacity for the model in this Section 2.2 was characterized in [15], where a connection was established between SK capacity and a new concept in data compression of *"minimum rate of communication for omniscience"* (that did not involve any secrecy constraint).

Definition 2.3 [15]: A number R is called an *achievable rate of communication for omniscience* (CO rate) for a set of terminals A if there exists communication $\mathbf{F}^{(n)}$ as described in the second pararaph of Section 2.1 such that $\mathbf{X}_{\mathcal{M}}$ is ϵ_n-CR achievable with $\mathbf{F}^{(n)}$, with $\lim_{n\to\infty} \epsilon_n = 0$ and $\lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{F}^{(n)}\| = R$, where $\|\mathbf{F}^{(n)}\|$ denotes the cardinality of the range of $\mathbf{F}^{(n)}$. The infimum of achievable CO rates for A is denoted by $OMN(A)$ and is termed the *minimum rate of communication for omniscience.*

Proposition 2.1 [15]: *The minimum rate of communication for omniscience for a set of terminals $A \subseteq \mathcal{M}$ equals*

$$OMN(A) = \min_{R_{\mathcal{M}} \in \mathcal{R}(A)} \sum_{i=1}^{m} R_i, \tag{2.4}$$

where

$$\mathcal{R}(A) = \{R_{\mathcal{M}} = (R_1, \ldots, R_m) : \sum_{i \in B} R_i \geq H(X_B|X_{B^c}), B \subset \mathcal{M}, B \not\supseteq A\}. \quad (2.5)$$

Furthermore, $OMN(A)$ is achievable by noninteractive communication and with an exponentially vanishing error in recovery at each terminal $i \in A$.

Theorem 2.2 [15]: *The (strong) SK capacity for the terminals $A \subseteq \mathcal{M}$ equals*

$$C(A) = H(X_1, \ldots, X_m) - OMN(A). \quad (2.6)$$

Furthermore, $C(A)$ is achievable by noninteractive communication.

Remark: The formula (2.6) not only serves as a single-letter characterization of SK capacity but also carries an interpretation that suggests a specific recipe for generating a key of maximum rate. Specifically, it follows from (2.6) that there will be no loss of optimality (in terms of the maximum achievable SK rate) by a restriction to the class of schemes for SK generation that proceed in the following two stages. In the first stage, the terminals engage in the most parsimonious interterminal communication for the sole purpose of allowing each terminal in A to become omniscient, i.e., to reconstruct losslessly the signals observed by all the other terminals. This step involves no secrecy requirement. Then, the second stage entails SK extraction from omniscience, i.e., the maximal CR $\mathbf{X}_{\mathcal{M}}$ with the SK being independent of the communication in the first stage. The maximum rate of the SK that can be so generated by this two-stage approach is the entropy rate of $\mathbf{X}_{\mathcal{M}}$, i.e., $H(X_{\mathcal{M}})$, less the minimum rate of communication for omniscience.

2.2.2 SK Generation from Less-Than-Omniscience

In order to attain SK capacity, it suffices, in effect, for the terminals in A to obtain omniscience through the most parsimonious communication. It is also interesting to note that omniscience is not necessary to achieve SK capacity, as is known already for the case of $m = 2$ terminals (cf. [39, 1]). In fact, the following result illustrates this observation.

Theorem 2.3 [16]: *The (strong) SK capacity for $A \subseteq \mathcal{M}$ can be achieved with noninteractive communication and with each terminal $i \in \mathcal{M}$ publicly communicating a single message $f_i(\mathbf{X}_i)$. Further, SK capacity can be achieved with the key generated at any particular terminal $k \in A$ obliviously of the public communication.*

A useful way of interpreting the result of Theorem 2.3 is by establishing a new connection, reminiscent of Theorem 2.2, between the SK capacity and a new concept in source coding of the *"minimum information rate of communication for disseminating among A the signal of a member terminal* (in A). This new connection, which is our contribution, is discussed next.

Definition 2.4: For each $k \in A$, a number $I \geq 0$ is called an *achievable information rate of communication for disseminating among A the signal of terminal k* (ICD-k rate), if there exists communication $\mathbf{F}^{(n)}$ as described in the second paragraph of Section 2.1, such that $\mathbf{X}_k$ is ϵ_n-CR achievable with communication $\mathbf{F}^{(n)}$, and with $\lim_{n\to\infty} \epsilon_n = 0$ and $\lim_{n\to\infty} \frac{1}{n} I(\mathbf{X}_k \wedge \mathbf{F}) \leq I$. The infimum of achievable ICD-k rates for A is denoted by $CD_k(A)$ and is termed the *minimum information rate of communication for disseminating among A the signal of terminal k.*

Theorem 2.4: *For each $k \in A$, the (strong) SK capacity equals*

$$C(A) = H(X_k) - CD_k(A). \tag{2.7}$$

Corollary 2.5: *For each $k \in A$, the minimum information rate of communication for disseminating among A the signal of terminal k equals*

$$CD_k(A) = OMN(A) - H(X_{\mathcal{M}\setminus\{k\}} \,|\, X_k). \tag{2.8}$$

Remarks: (i) In disseminating among A the signal of a particular terminal (in A), the characterization of the minimum rate of the needed communication (rather than the information rate as in Definition 2.4) is an open problem in multiterminal data compression that belongs to the longstanding class of open problems collectively known as the "helper problems" [13, Chapter 3]. However, a characterization of the minimum "information" rate of such communication in Theorem 2.4 and Corollary 2.5 follows almost immediately from Theorems 2.2 and 2.3, and is also of independent interest in the context of multiterminal data compression.

(ii) It is seen readily from Theorem 2.4 that for every $k \in A$,

$$C(A) = \lim_{n\to\infty} \frac{1}{n} \max_{\mathbf{F}_k} H\left(\mathbf{X}_k \middle| \mathbf{F}_k\right) \tag{2.9}$$

where the maximum is over all communication $\mathbf{F}_k = \mathbf{F}_k^{(n)}$ for disseminating among A the signal of terminal k. The invariance with $k \in A$ of the operational term in the right side of (2.9) which does not involve any secrecy requirement, could be of independent interest in multiterminal data compression.

2.2.3 Linear Codes for SK Generation

Theorems 2.2 and 2.4 suggest two different schemes that suffice to achieve the SK capacity. The first scheme involves attaining omniscience with the most parsimonious communication, followed by the extraction of an optimum-rate SK from omniscience. The second scheme entails a dissemination among A of the signal of a member terminal of A in the most "informationally" parsimonious manner, followed by SK extraction of optimum rate from the mentioned signal. Our next results establish that, in fact, these two schemes can be implemented with *linear* communication and subsequent *linear* secrecy extraction without sacrificing the rate-optimality of the resulting SK.

In this Section 2.2.3, we shall assume that $\mathcal{X}_1 = \mathcal{X}_2 = \ldots = \mathcal{X}_m = \mathbb{F}_q$, where $\mathbb{F}_q$ is a Galois field; each $\mathcal{X}_i^n = \mathbb{F}_q^n$ is regarded as a vector space over this field and each realization $x_i^n \in \mathbb{F}_q^n$ of an $\mathbb{F}_q^n$-valued rv is thought of as a column vector in $\mathbb{F}_q^n,\ i = 1, \ldots, m$.

Definition 2.5: The communication $\mathbf{F} = \mathbf{F}^{(n)}$ is termed *linear noninteractive communication* (LC) if $\mathbf{F} = (F_1, \ldots, F_m)$ with[3] $F_i = L_i \mathbf{X}_i$, where L_i is a $b_i \times n$ matrix with $\mathbb{F}_q$-valued entries for some positive integer $b_i,\ i = 1, \ldots, m$.

Theorem 2.6: *For any $R < C(A)$, there exists linear noninteractive communication $\mathbf{F} = \mathbf{F}^{(n)}$ and a $\lfloor \frac{nR}{m \log q} \rfloor \times n$ matrix K with $\mathbb{F}_{q^m}$-valued entries such that $\mathbf{X}_{\mathcal{M}}$ is ϵ_n-recoverable from $(\mathbf{X}_j, \mathbf{F})$ at each terminal $j \in A$, with ϵ_n vanishing exponentially in n; further, $K\mathbf{X}_{\mathcal{M}}$ constitutes a strong SK of rate R, where*

[3] All additions and multiplications are in $\mathbb{F}_q$.

$\mathbf{X}_{\mathcal{M}} = (X_{\mathcal{M},1}, \ldots, X_{\mathcal{M},n})$ *is regarded as a column vector*[4] in $\mathbb{F}_{q^m}^n$.

Theorem 2.7: *For any $R < C(A)$ and for any fixed $k \in A$, there exists linear noninteractive communication $\mathbf{F} = \mathbf{F}^{(n)}$ and a $\lfloor \frac{nR}{\log q} \rfloor \times n$ matrix K_k with $\mathbb{F}_q$-valued entries such that $\mathbf{X}_k$ is ϵ_n-recoverable from $(\mathbf{X}_j, \mathbf{F})$ at each terminal $j \in A\backslash\{k\}$ with ϵ_n vanishing exponentially in n; further, $K_k\mathbf{X}_k$ constitutes a strong SK of rate R.*

2.2.4 Proofs

Proof of Theorem 2.4: First, we show that

$$C(A) \leq H(X_k) - CD_k(A). \tag{2.10}$$

For an arbitrary $\delta > 0$, let $\mathbf{F} = \mathbf{F}^{(n)}$ be communication for omniscience of minimum rate for A, i.e., $\mathbf{X}_{\mathcal{M}}$ is ϵ_n-CR for A for some ϵ_n decaying to 0, and

$$\lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{F}\| \leq OMN(A) + \delta; \tag{2.11}$$

the existence of such $\mathbf{F}$ is asserted by Proposition 2.1. Noting that $\mathbf{F}$ is a function of $\mathbf{X}_{\mathcal{M}}$, it follows from Theorem 2.2 that

$$\lim_{n\to\infty} \frac{1}{n} H(\mathbf{X}_{\mathcal{M}}|\mathbf{F}) \geq H(X_{\mathcal{M}}) - \lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{F}\| \geq C(A) - \delta. \tag{2.12}$$

Since $k \in A$, $\mathbf{X}_{\mathcal{M}\backslash\{k\}}$ is ϵ_n-recoverable from $(\mathbf{X}_k, \mathbf{F})$. Then

$$\begin{aligned} \frac{1}{n} I(\mathbf{X}_k \wedge \mathbf{F}) &= H(X_k) - \frac{1}{n} H(\mathbf{X}_k|\mathbf{F}) \\ &= H(X_k) - \frac{1}{n} H(\mathbf{X}_{\mathcal{M}}|\mathbf{F}) + \frac{1}{n} H(\mathbf{X}_{\mathcal{M}\backslash\{k\}}|\mathbf{X}_k, \mathbf{F}) \\ &\leq H(X_k) - C(A) + \delta + \alpha\epsilon_n, \end{aligned} \tag{2.13}$$

[4]All additions and multiplications in the computation of $K\mathbf{X}_{\mathcal{M}}$ are in $\mathbb{F}_{q^m}$

for some $\alpha > 0$, by (2.12) and Fano's inequality. Since $\mathbf{X}_k$ is ϵ_n-recoverable from $(\mathbf{F}, \mathbf{X}_i)$ for every $i \in A\backslash\{k\}$, $\mathbf{F}$ is also communication for disseminating $\mathbf{X}_k$ among A with the ICD-k rate at most $H(X_k) - C(A) + \delta$ by (2.13). Since δ is arbitrarily small, it follows from Definition 2.4 that $CD_k(A) \leq H(X_k) - C(A)$ which is (2.10).

In order to show that

$$C(A) \geq H(X_k) - CD_k(A), \tag{2.14}$$

for an arbitrary $\delta > 0$, consider a communication $\mathbf{F}$ for disseminating $\mathbf{X}_k$ among A with ICD-k rate $\frac{1}{n} I(\mathbf{X}_k \wedge \mathbf{F}) \leq CD_k(A) + \delta$ for all n sufficiently large. Then, $\mathbf{X}_k$ is ϵ_n-CR for A for some ϵ_n decaying to zero in n and

$$\frac{1}{n} H(\mathbf{X}_k|\mathbf{F}) = H(X_k) - \frac{1}{n} I(\mathbf{X}_k \wedge \mathbf{F}) \geq H(X_k) - CD_k(A) - \delta. \tag{2.15}$$

It suffices to prove assertion (2.14) for blocklengths that are integer multiples of n. To this end, consider N i.i.d. repetitions of $(\mathbf{X}_{\mathcal{M}}, \mathbf{F})$. For each $j \in A\backslash\{k\}$, since $\mathbf{X}_k$ is ϵ_n-recoverable from $(\mathbf{X}_j, \mathbf{F})$, it holds that

$$H(\mathbf{X}_k|\mathbf{X}_j, \mathbf{F}) \leq n\epsilon_n \log |\mathcal{X}_k| + h(\epsilon_n). \tag{2.16}$$

A consequence of the Slepian-Wolf theorem [16, Lemma 3.1], we get that $\mathbf{X}_k^N$ is η_N-recoverable from $\left(\mathbf{X}_j^N, \mathbf{F}^N, g_j\left(\mathbf{X}_k^N\right)\right)$ for a suitable
$g_j : \mathcal{X}_k^{nN} \to \{1, \ldots, \lfloor e^{N(n\epsilon_n \log |\mathcal{X}_k| + h(\epsilon_n))} \rfloor\}$ where η_N decays to 0 exponentially rapidly in N.

It remains to show that there exist SKs $K^{(nN)}$ with rate $\lim_{N\to\infty} \frac{1}{nN} \log \|K^{(nN)}\|$ arbitrarily close to the right side of (2.14). To this end, we apply [15, Lemma B3] with $U = \mathbf{X}_k$, $V = \mathbf{F}$, and $R = (|A| - 1)\left(n\epsilon_n \log |\mathcal{X}_k| + h(\epsilon_n)\right)$. With this choice,

and by (2.16),

$$\begin{aligned} H(U|V) - R &= H(\mathbf{X}_k|\mathbf{F}) - (|A|-1)\left(n\epsilon_n \log|\mathcal{X}_k| + h(\epsilon_n)\right) \\ &\geq n\left[H(X_k) - CD_k(A) - \delta - (|A|-1)\left(\epsilon_n \log|\mathcal{X}_k| + \frac{1}{n}h(\epsilon_n)\right)\right]. \end{aligned}$$

Since ϵ_n decays to 0 in n, [15, Lemma B3] gives that there exists $K^{(nN)}$ with $\log\|K\| \geq nN\left(H(X_k) - CD_k(A) - \delta\right)$, achievable with appropriate communiction $\tilde{\mathbf{F}}^{(nN)}$, such that $s(K;\tilde{\mathbf{F}})$ decays to 0 exponentially fast in N. Since δ was arbitrary, this establishes (2.14), thereby completing the proof of the theorem. The corollary is immediate. ∎

The proofs of Theorems 2.6 and Theorem 2.7 rely on the following technical lemma which can be regarded as an extension of [15, Lemma B3]. Its proof is provided in Appendix A.1.

Lemma 2.1: *Let A be a rv with values in a Galois field $\mathbb{F}_q$ and let U be an $\mathbb{R}^n$-valued rv (with or without a density with respect to the Lebesque measure). Consider N i.i.d. repetitions of (A,U), namely $\left(A^N,U^N\right) = \left((A_1,U_1),\ldots,(A_N,U_N)\right)$ and let $B = B^{(N)} \in \mathcal{B} = \mathcal{B}^{(N)}$ be a finite-valued rv with a given joint distribution with (A^N,U^N). Then, for every $\delta > 0$ and every $R < H(A|U) - \frac{1}{N}\log|\mathcal{B}| - 2\delta$, there exists a $\lfloor \frac{NR}{\log q} \rfloor \times N$ matrix L with $\mathbb{F}_q$-valued entries such that $s(LA^N;U^N,B)$ vanishes exponentially in N.*

Remark: Lemma 2.1 extends [15, Lemma B3] in the following specific ways. First, [15, Lemma B3] deals with a finite-valued U while the U in Lemma 2.1 can have an arbitrary alphabet. Second, [15, Lemma B3] asserts the existence of strong secrecy from $\left(U^N,B\right)$ obtained as a function of A^N but, unlike in Lemma 2.1,

without the guarantee that the function is linear.

Proof of Theorem 2.6: We begin by invoking a known result in source coding from [10] that asserts the existence of linear noninteractive communication for omniscience (cf. Definition 2.3), yielding that $\mathbf{X}_{\mathcal{M}}$ is ϵ_n-CR for A achievable with $\mathbf{F}$ for an exponentially vanishing ϵ_n, with the optimum rate $\lim_{n\to\infty}\frac{1}{n}\log\|\mathbf{F}\| = OMN(A)$. Theorem 2.6 now follows from the fact that (cf. (2.6))

$$C(A) = H\left(X_{\mathcal{M}}\right) - OMN(A) = \frac{1}{n}H\left(\mathbf{X}_{\mathcal{M}}\right) - \lim_{n\to\infty}\frac{1}{n}\log\|\mathbf{F}\| = \lim_{n\to\infty}\frac{1}{n}H\left(\mathbf{X}_{\mathcal{M}}|\mathbf{F}\right),$$

and by the use of Lemma 2.1 with $N = n, A = X_{\mathcal{M}}, B = B^{(n)} = \mathbf{F}$ and with U being constant, respectively. ∎

Proof of Theorem 2.7: We start in a similar manner as in the proof of Theorem 2.6. Specifically, by invoking the result of [10], we get that for an arbitrary but fixed $\delta > 0$ and for all t sufficiently large, there exists linear noninteractive communication $\mathbf{F}^{(t)} = \left\{F_i\mathbf{X}_i^{(t)},\ i = 1,\ldots,m\right\}$ with $\log\|\mathbf{F}^{(t)}\| = \left(\sum_{i=1}^m r_i\right)\log q \leq t(OMN(A)+\delta)$ where F_i is an $r_i \times t$ matrix with $\mathbb{F}_q$-valued entries, $i = 1,\ldots,m$, such that $\mathbf{X}_{\mathcal{M}}^{(t)}$ is ϵ-recoverable from $\left(\mathbf{X}_i^{(t)}, \mathbf{F}^{(t)}\right)$ for each $i \in A$. Note that for $k \in A$,

$$\begin{aligned}
H\left(\mathbf{X}_k^{(t)}|\mathbf{F}^{(t)}\right) &= H\left(\mathbf{X}_k^{(t)}|\mathbf{F}^{(t)}\right) - H\left(\mathbf{X}_{\mathcal{M}}^{(t)}|\mathbf{X}_k, \mathbf{F}^{(t)}\right) \\
&= tH\left(X_{\mathcal{M}}\right) - H\left(\mathbf{F}^{(t)}\right) - H\left(\mathbf{X}_{\mathcal{M}}|\mathbf{X}_k, \mathbf{F}^{(t)}\right) \\
&\geq tH\left(X_{\mathcal{M}}\right) - H\left(\mathbf{F}^{(t)}\right) - \epsilon t\sum_{i=1}^m \log|\mathcal{X}_i| - h(\epsilon) \\
&\geq t\left(C(A) - \delta - \epsilon\sum_{i=1}^m \log|\mathcal{X}_i|\right) - h(\epsilon). \qquad (2.17)
\end{aligned}$$

Here the first inequality is a consequence of the ϵ-recoverability of $\mathbf{X}_{\mathcal{M}}^{(t)}$ from $\left(\mathbf{X}_k^{(t)}, \mathbf{F}^{(t)}\right)$

and Fano's inequality, and the second inequality follows from

$$H\left(\mathbf{F}^{(t)}\right) \leq \log \|\mathbf{F}^{(t)}\| \leq t\left(OMN(A) + \delta\right)$$

since $H\left(X_{\mathcal{M}}\right) - OMN(A) = C(A)$ by Theorem 2.2.

It suffices to prove the assertion for blocklengths equal to integer multiples of t. To this end, consider N i.i.d. repetitions of $\left(\mathbf{X}_{\mathcal{M}}^{(t)}, \mathbf{F}^{(t)}\right)$, namely $\left(\mathbf{X}_{\mathcal{M}}^{(t)N}, \mathbf{F}^{(t)N}\right)$. Then, for each $j \in A\backslash\{k\}$, since $H\left(\mathbf{X}_k^{(t)}|\mathbf{X}_j^{(t)}, \mathbf{F}^{(t)}\right) \leq t\epsilon \log|\mathcal{X}_k| + h(\epsilon)$, the result of [10] gives that $\mathbf{X}_k^{(t)N}$ is η_N-recoverable from $\left(\mathbf{X}_j^{(t)N}, \mathbf{F}^{(t)N}, H_j\mathbf{X}_k^{(t)N}\right)$ for a $\lfloor \frac{N(t\epsilon \log|\mathcal{X}_k| + h(\epsilon) + \delta)}{t \log q} \rfloor \times N$ matrix H_j with $\mathbb{F}_{q^t}$-valued entries, and η_N decays to 0 exponentially rapidly in N. It follows upon setting $\tilde{\mathbf{F}}^{(tN)} = \left(\mathbf{F}^{(t)N}, \left\{H_j\mathbf{X}_k^{(t)N}\right\}_{j\in A\backslash\{k\}}\right)$ that $\tilde{\mathbf{F}}^{(tN)}$ satisfies the recoverability assertion of the theorem with $n = tN$, $N = 1, 2, \ldots$.

It remains to show that there exists a linear function $K^{(tN)}$ of $\mathbf{X}_k^{(t)N}$ that satisfies the assertion on $K^{(n)}$ with $n = tN$. To this end, we apply Lemma 2.1 with $A = \mathbf{X}_k^{(t)}, B = \mathbf{F}^{(t)}$ and $B = \left\{H_j\mathbf{X}_k^{(t)N}\right\}_{j\in A\backslash\{k\}}$. With this choice, by (2.17)

$$\begin{aligned} H(A|B) - \frac{1}{N}\log|\mathcal{B}| &= H\left(\mathbf{X}_k^{(t)}|\mathbf{F}^{(t)}\right) - (|A|-1)(t\epsilon\log|\mathcal{X}_k| + h(\epsilon) + \delta) \\ &\geq t\Big[C(A) - \delta - \epsilon\left(|A|\log|\mathcal{X}_k| + \sum_{i=1}^{m}\log|\mathcal{X}_i|\right)\Big] \\ &\quad -|A|(h(\epsilon)+\delta) \\ &> t(C(A) - 2\delta) \end{aligned}$$

upon choosing $\epsilon > 0$ sufficiently small and t sufficiently large. Hence, Lemma 2.1 gives that there exists a $\lfloor \frac{Nt(C(A)-2\delta)}{t\log q} \rfloor \times N$ matrix K_k such that for $K^{(tN)} = K_k\mathbf{X}_k^{(t)N}$, $s\left(K^{(tN)}; \tilde{\mathbf{F}}^{(tN)}\right)$ vanishing to 0 exponentially rapidly in N. The proof of

Theorem 2.7 is completed by the observation that a linear mapping from $\mathbb{F}_{q^t}^N$ to $\mathbb{F}_{q^t}^M$, for some M, that corresponds to left-multiplication of a vector in $\mathbb{F}_{q^t}^N$ by each of the matrices $K_k, \{H_i\}_{i\in A\setminus\{k\}}$ as above, has an alternative representation as another linear mapping from $\mathbb{F}_q^{tN}$ to $\mathbb{F}_q^{tN}$ which can be represented then by a $tM \times tN$ matrix with $\mathbb{F}_q$-valued entries (instead of with $\mathbb{F}_{q^t}$-valued entries in its current form). ■

Chapter 3

SK Generation for the Gaussian Model

In the general model described in Section 2.1, let $X_1, \ldots, X_m$ be $\mathbb{R}$-valued jointly Gaussian rvs with

$$\mathbb{E}\begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix} = \mathbf{0}, \quad Q = \mathbb{E}\left[\begin{bmatrix} X_1 \\ \vdots \\ X_m \end{bmatrix} \begin{bmatrix} X_1, \ldots, X_m \end{bmatrix}\right] > 0, \tag{3.1}$$

where $Q_{ij} = \sigma_i \sigma_j \rho_{ij}$ with $\sigma_i^2 = \mathbb{E}\left[X_i^2\right],\ 1 \leq i, j \leq m$. It follows as a consequence of (3.1) that

$$-\infty < h\left(X_B\right) \leq h\left(X_{\mathcal{M}}\right) < \infty \text{ for every } B \subseteq \mathcal{M}. \tag{3.2}$$

This model describes a situation in which the terminals have prior and privileged access to jointly Gaussian signals. The SK capacity of the Gaussian model cannot be inferred from the counterpart result for the discrete model in Chapter 2. Specifically, the central role of omniscience in the attainment of SK capacity Theorem 2.2, cannot be replayed directly now as the minimum rate of public communication for omniscience is unbounded. We characterize the SK capacity for the Gaussian model. Our achievability proof is based on a suitably refined quantization of the signals at the terminals combined with our results in Chapter 2. The converse proof, which constitutes the first main contribution of this chapter, provides a technique that is applicable to the discrete model of [15] as well as to models

with $\mathbb{R}$-valued rvs under suitable technical conditions. Our SK capacity formula acquires a simple form for the special case of "symmetrically correlated" Gaussian signals. Of special interest is the model with two terminals with signals that are *a fortiori* symmetrically correlated. Considering schemes that involve quantization at one terminal, we characterize the best rate of an achievable SK as a function of quantization rate; SK capacity is attained as the quantization rate tends to infinity. Structured codes are shown to attain the optimum tradeoff between SK rate and quantization rate, constituting the second main contribution of this chapter. This result shows how SK rate increases optimally with processing complexity (as measured by quantization rate).

3.1 SK Capacity

We begin with the observation that the SK capacity for the model above will depend on the joint distribution of $X_1, \ldots, X_m$ only through the correlation coefficients $\{\rho_{ij},\ 1 \leq i \neq j \leq n\}$. This is obvious since replacing X_i by $\frac{X_i}{\sigma_i}$ where $\sigma_i > 0$ (by (3.1)), $i = 1, \ldots, m$, does not alter SK capacity.

As in [16], for $A \subseteq \mathcal{M}$, let

$$\mathcal{B}(A) = \{B \subset \mathcal{M} :\ B \neq \emptyset,\ B \not\supseteq A\} \tag{3.3}$$

and $\mathcal{B}_i(A)$ be its subset consisting of those $B \in \mathcal{B}(A)$ that contain i, $i \in \mathcal{M}$. Let $\Lambda(A)$ be the set of all collections $\lambda = \{\lambda_B :\ B \in \mathcal{B}(A)\}$ of weights $0 \leq \lambda_B \leq 1$, satisfying

$$\sum_{B \in \mathcal{B}_i(A)} \lambda_B = 1,\ \text{for all } i = 1, \ldots, m. \tag{3.4}$$

Theorem 3.1: *The (strong) SK capacity equals*

$$C(A) = h(X_1, \ldots, X_m) - \max_{\lambda \in \Lambda(A)} \sum_{B \in \mathcal{B}(A)} \lambda_B h(X_B | X_{B^c}). \tag{3.5}$$

Corollary 3.2: *The (strong) SK capacity for a "symmetric" Gaussian model with*

$$Q = \begin{cases} Q(i,i) = \sigma_i^2, & 1 \leq i \leq m \\ Q(i,j) = \rho \sigma_i \sigma_j, & 1 \leq i \neq j \leq m, \end{cases} \tag{3.6}$$

with $-\frac{1}{m-1} < \rho < 1$, *and with* $A = \mathcal{M}$, *equals*

$$C(\mathcal{M}) = \frac{1}{2} \log \left[\frac{1}{(1-\rho)\,(1+(m-1)\rho)^{\frac{1}{m-1}}} \right]. \tag{3.7}$$

In particular, when $m = 2$,

$$C(\mathcal{M} = \{1,2\}) = \frac{1}{2} \log \frac{1}{1-\rho^2}. \tag{3.8}$$

3.2 Proof of the SK Capacity Theorem

Proof of Theorem 3.1:

Achievability: The idea is to use scalar quantization of X_i at terminal i, $i = 1, \ldots, m$, followed by SK generation for the resulting finite-alphabet source model along the lines of [15]. By appropriately choosing the scalar quantizer, the claimed rate of (3.5) will be shown to be achievable in the limit of infinite quantization rates.

In particular, for each positive integer q, consider a quantizer $f_q : \mathbb{R} \to \{0, 1, \ldots, 2q^2\}$, where

$$f_q(x) = \begin{cases} 0, & \text{if } x > q \text{ or } x \leq -q \\ \lceil q(x+q) \rceil, & \text{if } -q < x \leq q. \end{cases} \tag{3.9}$$

At each terminal i, consider the $\{0, 1, \ldots, 2q^2\}$-valued rv $Y_i^{(q)} = f_q(X_i),\ i = 1, \ldots, m$. Define the $\{0,1\}^m$-valued rv $Y_{m+1}^{(q)} = \left(1_{(f_q(X_i) \neq 0)}\right)_{i=1}^m$.

Next, consider a fictitious (finite-alphabet) source model for "private key" generation iwth $m+1$ terminals consisting of legitimate terminals $1, \ldots, m$ that observe respectively $\mathbf{Y}_1^{(q)}, \ldots, \mathbf{Y}_m^{(q)}$, and a (compromised helper) terminal $m+1$ that observes $\mathbf{Y}_{m+1}^{(q)}$. Now the terminals in the set $A \subseteq \mathcal{M} = \{1, \ldots, m\}$ seek to generate a *private key* (PK), say K, with the help of all the remaining terminals including terminal $m+1$, using public communication, say $\mathbf{F}$, so that the security condition (2.1) is satisfied with $\left(\mathbf{Y}_{m+1}^{(q)}, \mathbf{F}\right)$ in the role of $\mathbf{F}$, i.e.,

$$s\left(K; \mathbf{Y}_{m+1}^{(q)}, \mathbf{F}\right) \leq \epsilon. \tag{3.10}$$

Such a PK K is concealed from terminal $m+1$ as well as from an eavesdropper that observes $\mathbf{F}$. The corresponding largest rate of such a PK, namely PK capacity was characterized in [15]; it was shown therein that PK capacity is achievable by allowing the compromised terminal $m+1$ to fully reveal its observations $\mathbf{Y}_{m+1}^{(q)}$ prior to public communication by the various terminals. From [15], the (strong) PK capacity for this finite-alphabet source model equals

$$\min_{\lambda \in \Lambda(A)} \left[H\left(Y_{\mathcal{M}}^{(q)} | Y_{m+1}^{(q)}\right) - \sum_{B \in \mathcal{B}(A)} \lambda_B H\left(Y_B^{(q)} | Y_{\mathcal{M}\setminus B}^{(q)}, Y_{m+1}^{(q)}\right) \right]. \tag{3.11}$$

Returning to the Gaussian model at hand, terminals $1, \ldots, m$ can simulate the mentioned model for PK generation by using the scalar quantizer f_q at each terminal and letting each terminal i reveal publicly the i.i.d. repetitions of the rv $1_{(f_q(X_i) \neq 0)},\ i = 1, \ldots, m$. Consequently, in the limit of infinite quantization,

$$\lim_{q \to \infty} \min_{\lambda \in \Lambda(A)} \left[H\left(Y_{\mathcal{M}}^{(q)} | Y_{m+1}^{(q)}\right) - \sum_{B \in \mathcal{B}(A)} \lambda_B H\left(Y_B^{(q)} | Y_{\mathcal{M}\backslash B}^{(q)}, Y_{m+1}^{(q)}\right) \right], \tag{3.12}$$

is an achievable (strong) SK rate for the Gaussian model, by (3.11).

Next, for a fixed $\lambda \in \Lambda(A)$, using (3.4), we get that

$$\begin{aligned} \sum_{B \in \mathcal{B}(A)} \lambda_B |\mathcal{B}\backslash B| &= \sum_{B \in \mathcal{B}(A)} \lambda_B \left(m - |B|\right) \\ &= m \sum_{B \in \mathcal{B}(A)} \lambda_B - \sum_{B \in \mathcal{B}(A)} \lambda_B \sum_{i=1,\ldots,m} 1_{(i \in B)} \\ &= m \left(\sum_{B \in \mathcal{B}(A)} \lambda_B - 1 \right). \end{aligned} \tag{3.13}$$

Consequently, we have that

$$H\left(Y_{\mathcal{M}}^{(q)} | Y_{m+1}^{(q)} = \mathbf{1}\right) - \sum_{B \in \mathcal{B}(A)} \lambda_B H\left(Y_B^{(q)} | Y_{\mathcal{M}\backslash B}^{(q)}, Y_{m+1}^{(q)} = \mathbf{1}\right)$$

$$\begin{aligned} &= \sum_{B \in \mathcal{B}(A)} \lambda_B H\left(Y_{\mathcal{M}\backslash B}^{(q)} | Y_{m+1}^{(q)} = \mathbf{1}\right) - \left(\sum_{B \in \mathcal{B}(A)} \lambda_B - 1 \right) H\left(Y_{\mathcal{M}}^{(q)} | Y_{m+1}^{(q)} = \mathbf{1}\right) \\ &= \sum_{B \in \mathcal{B}(A)} \lambda_B \left[H\left(Y_{\mathcal{M}\backslash B}^{(q)} | Y_{m+1}^{(q)} = \mathbf{0}\right) - |\mathcal{M}\backslash B| \log q \right] \\ &\quad - \left(\sum_{B \in \mathcal{B}(A)} \lambda_B - 1 \right) \left[H\left(Y_{\mathcal{M}}^{(q)} | Y_{m+1}^{(q)} = \mathbf{1}\right) - m \log q \right], \end{aligned} \tag{3.14}$$

by (3.13).

We proceed by using the following technical lemma whose proof is relegated to the end of this proof of achievability.

Lemma 3.1: *For the Gaussian rvs $X_1 \ldots, X_m$ in the statement of Theorem 3.1, a quantizer f_q as described in (3.9), and every $B \subseteq \mathcal{M} = \{1, \ldots, m\}$, we get that*

$$\lim_{q\to\infty} \left[H\left(Y_B^{(q)} | Y_{m+1}^{(q)} = \mathbf{1}\right) - |B| \log q \right] = h(X_B). \tag{3.15}$$

Furthermore,

$$\lim_{q\to\infty} Pr\left\{ Y_{m+1}^{(q)} = \mathbf{1} \right\} = 1. \tag{3.16}$$

Continuing with (3.14) upon using (3.15) of Lemma 3.1, we get that for every $\lambda \in \Lambda(A)$,

$$\begin{aligned}
&\lim_{q\to\infty} \left[H\left(Y_{\mathcal{M}}^{(q)} | Y_{m+1}^{(q)} = \mathbf{1}\right) - \sum_{B\in\mathcal{B}(A)} \lambda_B H\left(Y_B^{(q)} | Y_{\mathcal{M}\setminus B}^{(q)}, Y_{m+1}^{(q)} = \mathbf{1}\right) \right] \\
&\qquad = \sum_{B\in\mathcal{B}(A)} \lambda_B h\left(X_{\mathcal{M}\setminus B}\right) - \left(\sum_{B\in\mathcal{B}(A)} \lambda_B - 1 \right) h\left(X_{\mathcal{M}}\right) \\
&\qquad = h\left(X_{\mathcal{M}}\right) - \sum_{B\in\mathcal{B}(A)} \lambda_B h\left(X_B | X_{B^c}\right).
\end{aligned} \tag{3.17}$$

In [15], it was shown using the duality of linear programming that the minimization in the right side of the expression for the PK capacity (3.11) can be taken over a finite subset $\Lambda'(A)$ (of $\Lambda(A)$) that depends only on $\mathcal{M}$ and A. Consequently, the following achievable (strong) SK rate in (3.12) can be bounded below further as follows

$$\lim_{q\to\infty} \min_{\lambda\in\Lambda(A)} \left[H\left(Y_{\mathcal{M}}^{(q)} | Y_{m+1}^{(q)}\right) - \sum_{B:\ B\in\mathcal{B}(A)} \lambda_B H\left(Y_B^{(q)} | Y_{\mathcal{M}\setminus B}^{(q)}, Y_{m+1}^{(q)}\right) \right]$$

$$
\begin{aligned}
&= \lim_{q\to\infty} \min_{\lambda\in\Lambda'(A)\subset\Lambda(A)} \left[H\left(Y_{\mathcal{M}}^{(q)}|Y_{m+1}^{(q)}\right) - \sum_{B\in\mathcal{B}(A)} \lambda_B H\left(Y_B^{(q)}|Y_{\mathcal{M}\setminus B}^{(q)}, Y_{m+1}^{(q)}\right)\right] \\
&\geq \lim_{q\to\infty} \min_{\lambda\in\Lambda'(A)\subset\Lambda(A)} Pr\left\{Y_{m+1}^{(q)} = \mathbf{1}\right\} \times \\
&\qquad\qquad \left[H\left(Y_{\mathcal{M}}^{(q)}|Y_{m+1}^{(q)} = \mathbf{1}\right) - \sum_{B\in\mathcal{B}(A)} \lambda_B H\left(Y_B^{(q)}|Y_{\mathcal{M}\setminus B}^{(q)}, Y_{m+1}^{(q)} = \mathbf{1}\right)\right] \\
&= \min_{\lambda\in\Lambda'(A)\subset\Lambda(A)} \left[h\left(X_{\mathcal{M}}\right) - \sum_{B\in\mathcal{B}(A)} \lambda_B h\left(X_B|X_{B^c}\right)\right], \\
&\quad \text{by (3.16), (3.17) and by the fact that } \Lambda'(A), \text{ which does not depend on } q, \text{ is finite} \\
&= \min_{\lambda\in\Lambda(A)} \left[h(X_{\mathcal{M}}) \; - \sum_{B\in\mathcal{B}(A)} \lambda_B h(X_B|X_{B^c})\right],
\end{aligned}
$$

which is (3.5).

It remains to prove Lemma 3.1.

$$
\begin{aligned}
Pr\left\{Y_{m+1}^{(q)} \neq \mathbf{1}\right\} &= Pr\left\{(X_1,\ldots,X_m) \in ([-q,q]^m)^c\right\} \\
&= \int_{([-q,q]^m)^c} \frac{\exp\left(-\frac{1}{2}\mathbf{x}^T Q^{-1}\mathbf{x}\right)}{(2\pi)^{n/2}\,|Q|^{1/2}} d\mathbf{x} \\
&\leq \int_{\left\{\mathbf{x}:\|\mathbf{x}\|\geq q\right\}} \frac{\exp\left(-\frac{1}{2}\mathbf{x}^T Q^{-1}\mathbf{x}\right)}{(2\pi)^{n/2}\,|Q|^{1/2}} d\mathbf{x} \\
&\leq \int_{\left\{\mathbf{x}:\|\mathbf{x}\|\geq q\right\}} \frac{\exp\left(-\frac{1}{2}\frac{\|\mathbf{x}\|^2}{\lambda_{max}}\right)}{(2\pi)^{n/2}\,|Q|^{1/2}} d\mathbf{x} \qquad (3.18) \\
&= \frac{mC_m}{(2\pi)^{n/2}\,|Q|^{1/2}} \int_q^\infty r^{m-1} \exp\left(-\frac{r^2}{2}\right) dr \qquad (3.19) \\
&= O\Big(q^{(m-2)} \exp\left(-\frac{q^2}{2\lambda_{max}}\right)\Big) \to 0, \text{as } q\to\infty, \qquad (3.20)
\end{aligned}
$$

where $\lambda_{max} > 0$ in (3.18) is the largest eigenvalue of Q and C_m in (3.19) is a constant that depends only on m. This establishes (3.16).

Next, for each $B \subseteq \mathcal{M}$, let

$$
\epsilon_q \triangleq Pr\left\{\left(1_{(Y_i\neq 0)}\right)_{i\in B} \neq \mathbf{1}\right\} \leq Pr\left\{Y_{m+1}^{(q)} \neq \mathbf{1}\right\} = o_q(1). \qquad (3.21)
$$

Further, for any collection of $|B|$ integers $k_B,\ k_i \in \{1, 2, \ldots, 2q^2\},\ i \in B$, let $P_{k_B} = Pr\{Y_B = k_B\}$. It now follows by the uniform continuity of the density function f_{X_B} of X_B and the mean value theorem that for any k_B, there exists an $x_B(k_B)$ satisfying $Y_i = f_q(x_i(k_B)) = k_i,\ i \in B$ so that $P_{k_B} = f_{X_B}(x_B(k_B))q^{-|B|}$. Consequently,

$$
\begin{aligned}
& H\left(Y_B | \left(1_{(Y_i \neq 0)}\right)_{i \in B} = \mathbf{1}\right) \\
&= -\sum_{k_B \in \{1\ldots,2q^2\}^{|B|}} \left(\frac{P_{k_B}}{1-\epsilon_q}\right) \log\left(\frac{P_{k_B}}{1-\epsilon_q}\right) \\
&= -\sum_{k_B \in \{1\ldots,2q^2\}^{|B|}} \left(\frac{P_{k_B}}{1-\epsilon_q}\right) \log\left(\frac{P_{k_B}}{1-\epsilon_q}\right) \\
&= -\sum_{k_B \in \{1\ldots,2q^2\}^{|B|}} \left(\frac{f_{X_B}(x_B(k_B))q^{-|B|}}{1-\epsilon_q}\right) \log\left(f_{X_B}(x_B(k_B))q^{-|B|}\right) \\
&\quad + \log(1-\epsilon_q) \\
&= \frac{1}{1-\epsilon_q}\left[-\sum_{k_B \in \{1\ldots,2q^2\}^{|B|}} \left(f_{X_B}(x_B(k_B))q^{-|B|}\right) \log\left(f_{X_B}(x_B(k_B))\right)\right] \\
&\quad + \log(1-\epsilon_q) + |B|\log q.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& H\left(Y_B | \left(1_{(Y_i \neq 0)}\right)_{i \in B} = \mathbf{1}\right) + |B|\log q \\
&= \frac{1}{1-\epsilon_q}\left[-\sum_{k_B \in \{1\ldots,2q^2\}^{|B|}} \left(f_{X_B}(x_B(k_B))q^{-|B|}\right) \log\left(f_{X_B}(x_B(k_B))\right)\right].
\end{aligned}
$$

Taking limit as $q \to \infty$, we obtain (3.15); this concludes the proof of Lemma 3.1. ■

Converse: Our converse constitutes our first main contribution of this chapter. The main technical tools are supplied by Lemmas 3.2 and 3.3 that follow.

Lemma 3.2: *Consider the i.i.d. repetitions of the jointly Gaussian rvs* $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ *in the statement of Theorem 3.1, namely,* $\mathbf{X}_{\mathcal{M}} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$, *and*

let Z be a rv with a joint distribution with $\mathbf{X}_{\mathcal{M}}$. For any $\lambda \in \Lambda(A)$, $i = 1, \ldots, m$, and any U_i that is a function of $(\mathbf{X}_i, Z)$, i.e., $U_i = u_i(\mathbf{X}_i, Z)$, it holds that

$$\begin{aligned} h(\mathbf{X}_{\mathcal{M}}|Z) \;-\; & \sum_{B \in \mathcal{B}(A)} \lambda_B h(\mathbf{X}_B|\mathbf{X}_{B^c}, Z) \\ = \; & \sum_{B \in \mathcal{B}(A): B \ni i} \lambda_B I(U_i \wedge \mathbf{X}_{B^c}|Z) \\ & + \Big[h(\mathbf{X}_{\mathcal{M}}|Z, U_i) - \sum_{B \in \mathcal{B}(A)} \lambda_B h(\mathbf{X}_B|\mathbf{X}_{B^c}, Z, U) \Big]. \end{aligned} \tag{3.22}$$

Proof:

$$\begin{aligned} h(\mathbf{X}_{\mathcal{M}}|Z) \;-\; & \sum_{B \in \mathcal{B}(A)} \lambda_B h(\mathbf{X}_B|\mathbf{X}_{B^c}, Z) \\ = \; & h(\mathbf{X}_{\mathcal{M}}|Z, U_i) + I(U_i \wedge \mathbf{X}_{\mathcal{M}}|Z) \\ & - \sum_{B \in \mathcal{B}(A)} \lambda_B \left[h(\mathbf{X}_B|\mathbf{X}_{B^c}, Z, U_i) + I(U_i \wedge \mathbf{X}_B|\mathbf{X}_{B^c}, Z) \right] \\ = \; & \left[h(\mathbf{X}_{\mathcal{M}}|Z, U_i) - \sum_{B \in \mathcal{B}(A)} \lambda_B h(\mathbf{X}_B|\mathbf{X}_{B^c}, Z, U_i) \right] \\ & + \left[I(U_i \wedge \mathbf{X}_{\mathcal{M}}|Z) - \sum_{B \in \mathcal{B}(A)} \lambda_B I(U_i \wedge \mathbf{X}_B|\mathbf{X}_{B^c}, Z) \right] \\ = \; & \left[h(\mathbf{X}_{\mathcal{M}}|Z, U_i) - \sum_{B \in \mathcal{B}(A)} \lambda_B h(\mathbf{X}_B|\mathbf{X}_{B^c}, Z, U_i) \right] \\ & + \left[\left(\sum_{B \in \mathcal{B}(A): B \ni i} \lambda_B \right) I(U_i \wedge \mathbf{X}_{\mathcal{M}}|Z) - \sum_{B \in \mathcal{B}(A): B \ni i} \lambda_B I(U_i \wedge \mathbf{X}_B|\mathbf{X}_{B^c}, Z) \right] \\ = \; & \left[h(\mathbf{X}_{\mathcal{M}}|Z, U_i) - \sum_{B \in \mathcal{B}(A)} \lambda_B h(\mathbf{X}_B|\mathbf{X}_{B^c}, Z, U_i) \right] \\ & + \sum_{B \in \mathcal{B}(A): B \ni i} \lambda_B I(U_i \wedge \mathbf{X}_{B^c}|Z). \end{aligned}$$

■

Lemma 3.3: *Consider the i.i.d. repetitions of the jointly Gaussian rvs* $X_{\mathcal{M}} = (X_1, \ldots, X_m)$ *in the statement of Theorem 3.1, namely,* $\mathbf{X}_{\mathcal{M}} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$*, and let* Z *be a rv with a joint distribution with* $\mathbf{X}_{\mathcal{M}}$*. For any* $\lambda \in \Lambda(A)$*, it holds that*

$$h\left(\mathbf{X}_{\mathcal{M}}|Z\right) - \sum_{B \in \mathcal{B}(A)} \lambda_B h\left(\mathbf{X}_B|\mathbf{X}_{B^c}, Z\right) \geq 0. \tag{3.23}$$

Proof:

$$\begin{aligned} h\left(\mathbf{X}_{\mathcal{M}}|Z\right) &= \sum_{i \in \mathcal{M}} \left(\sum_{B \in \mathcal{B}(A): B \ni i} \lambda_B \right) h\left(\mathbf{X}_i|\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, Z\right) \\ &= \sum_{B \in \mathcal{B}(A)} \lambda_B \sum_{i \in B} h\left(\mathbf{X}_i|\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, Z\right) \\ &\geq \sum_{B \in \mathcal{B}(A)} \lambda_B \sum_{i \in B} h\left(\mathbf{X}_i|\mathbf{X}_{\{1,\ldots,i-1\} \cap B}, \mathbf{X}_{B^c}, Z\right) \\ &= \sum_{B \in \mathcal{B}(A)} \lambda_B h\left(\mathbf{X}_B|\mathbf{X}_{B^c}, Z\right). \end{aligned}$$

■

Suppose that $K^{(n)}$ represents an ϵ_n-SK for A achievable with (possibly interactive) communication $\mathbf{F}^{(n)}$ with, say, r rounds (as described in the second paragraph of Section 2.1), where $\lim_{n \to 0} \epsilon_n = 0$ (see Definition 2.1).

For $j = 1, \ldots, mr$, by repeated application of Lemma 3.2 with $F_{[1,j]}, F_j$ and $j \bmod m$ in the roles of Z, U_i and i, respectively, and the fact that $\sum_{B \in \mathcal{B}(A): B \ni i} \lambda_B I\left(U \wedge \mathbf{X}_{B^c}|Z\right) \geq 0$, we obtain

$$h\left(\mathbf{X}_{\mathcal{M}}\right) - \sum_{B \in \mathcal{B}(A)} \lambda_B h\left(\mathbf{X}_B|\mathbf{X}_{B^c}\right) \;\geq\; h\left(\mathbf{X}_{\mathcal{M}}|\mathbf{F}\right) - \sum_{B \in \mathcal{B}(A)} \lambda_B h\left(\mathbf{X}_B|\mathbf{X}_{B^c}, \mathbf{F}\right). \tag{3.24}$$

Next, for some $i \in A$, let $K_i = k_i^{(n)}\left(\mathbf{X}_i, \mathbf{F}\right)$ be such that $Pr\left\{K_i^{(n)} = K^{(n)}\right\} \leq \epsilon_n$. Continuing from (3.24) by using Lemma 3.2 again but now with $\mathbf{F}$ and K_i in the

roles of Z and U_i, respectively, we obtain that

$$
\begin{aligned}
& h\left(\mathbf{X}_{\mathcal{M}}|\mathbf{F}\right) - \sum_{B\in\mathcal{B}(A)} \lambda_B h\left(\mathbf{X}_B|\mathbf{X}_{B^c}|\mathbf{F}\right) \\
&\quad = \sum_{B\in\mathcal{B}(A):B\ni i} \lambda_B I\left(K_i \wedge \mathbf{X}_{B^c}|\mathbf{F}\right) \\
&\qquad + \Big[h\left(\mathbf{X}_{\mathcal{M}}|\mathbf{F},K_i\right) - \sum_{B\in\mathcal{B}(A)} \lambda_B h\left(\mathbf{X}_B|\mathbf{X}_{B^c},\mathbf{F},K_i\right)\Big]. \\
&\quad \geq \sum_{B\in\mathcal{B}(A):B\ni i} \lambda_B I\left(K_i \wedge \mathbf{X}_{B^c}|\mathbf{F}\right), \text{by Lemma 3.3} \\
&\quad \geq \sum_{B\in\mathcal{B}(A):B\ni i} \lambda_B \big[H\left(K_i|\mathbf{F}\right) - H(K_i|\mathbf{X}_{B^c},\mathbf{F})\big] \\
&\quad \geq \sum_{B\in\mathcal{B}(A):B\ni i} \lambda_B \big[H\left(K|\mathbf{F}\right) - H(K|K_i,\mathbf{F}) - H\left(K_i|\mathbf{X}_{B^c},\mathbf{F}\right)\big] \\
&\quad \geq H\left(K|\mathbf{F}\right) - 2\big[\log|\mathcal{K}|\epsilon_n + 1\big], \text{ by (3.4) and Fano's inequality} \\
&\quad \geq (\log|\mathcal{K}| - \epsilon_n) - 2\big[\log|\mathcal{K}|\epsilon_n + 1\big], \text{ by (2.2)} \\
&\quad \geq (1-2\epsilon_n)\log|\mathcal{K}| - \epsilon_n - 2. \qquad (3.25)
\end{aligned}
$$

Consequently, by (3.25) and (3.24), we have that for every $\lambda \in \Lambda(A)$ that

$$\lim_{n\to\infty} \frac{1}{n} \log|\mathcal{K}| \leq h(X_{\mathcal{M}}) - \sum_{B\in\mathcal{B}(A)} \lambda_B h(X_B|X_{B^c}).$$

The converse proof now follows by minimization over the set of $\lambda \in \Lambda(A)$. ∎

Proof of Corollary 3.2:

For a set $B = \{i_1,\ldots,i_b\} \subset \mathcal{M},\ i_1 < i_2 < \ldots < i_b$, and a permutation π on $\{1,\ldots,m\}$, let $\pi(B)$ denote $\{\pi(i_1),\ldots,\pi(i_b)\}$. For $\lambda^* \in \Lambda(\mathcal{M})$ attaining the maximization in the right side of (3.5) for the symmetric Gaussian model, consider $\lambda^{**} = \left\{\frac{1}{m!}\sum_\pi \lambda^*_{\pi(B)},\ B \in \mathcal{B}(\mathcal{M})\right\}$ where the summation is over all permutations on $\{1,\ldots,m\}$. It is readily seen that λ^{**} is in $\Lambda(\mathcal{M})$. By virtue of the fact that $h\left(X_B|X_{B^c}\right)$ depends on B only through $|B|$, it is clear that λ^{**} also attains the

maximization in (3.7). Note that λ^{**} has the property that $\lambda^{**}_B = \lambda^{**}_{B'}$, for any B, B' such that $|B| = |B'|$. Consequently, the optimization in right side of (3.5) can be reduced to the following easier one:

$$\max_{(\gamma_i)_{i=1}^{m-1}:\ \sum_{i=1}^{m-1}\gamma_i \binom{m-1}{i-1} = 1} \sum_{i=1}^{m-1} \gamma_i \binom{m}{i} H_i \tag{3.26}$$

where $H_i = h\left(X_{\{1,\ldots,i\}} | X_{\{i+1,\ldots,m\}}\right) = h\left(X_B | X_{B^c}\right)$ for any B with $|B| = i$. Let K_i denote the $i \times i$ matrix with diagonal entries being 1 and with all off-diagonal entries being ρ. It now follows from (3.26) that

$$\begin{aligned}
C(\mathcal{M}) &= h\left(X_{\mathcal{M}}\right) - \max_{i=1,\ldots,m-1} \frac{\binom{m}{i} H_i}{\binom{m-1}{i-1}} = h\left(X_{\mathcal{M}}\right) - \max_{i=1,\ldots,m-1} \frac{m}{i} H_i. \\
&= \min_{i=1,\ldots,m-1} h\left(X_1,\ldots,X_m\right) - \frac{m}{i}\left[h\left(X_1,\ldots,X_m\right) - h\left(X_{i+1},\ldots,X_m\right)\right] \\
&= \min_{i=1,\ldots,m-1} -\left(\frac{m-i}{i}\right) h\left(X_1,\ldots,X_m\right) + \frac{m}{i} h\left(X_{i+1},\ldots,X_m\right) \\
&= \min_{i=1,\ldots,m-1} -\left(\frac{m-i}{i}\right) \frac{1}{2} \log\left((2\pi e)^m \det\left(K_m\right)\right) \\
&\quad + \frac{m}{i}\frac{1}{2} \log\left((2\pi e)^{m-i} \det\left(K_{m-i}\right)\right) \\
&= \min_{i=1,\ldots,m-1} -\frac{1}{i}\left[\frac{1}{2}\log\left(\frac{(2\pi e)^{m(m-i)} \det\left(K_m\right)^{m-i}}{(2\pi e)^{(m-i)m} \det\left(K_{m-i}\right)^m}\right)\right] \\
&= \min_{i=1,\ldots,m-1} -\frac{1}{i}\left[\frac{1}{2}\log\left(\frac{\det\left(K_m\right)^{m-i}}{\det\left(K_{m-i}\right)^m}\right)\right] \\
&= \min_{i=1,\ldots,m-1} -\frac{1}{i}\left[\frac{1}{2}\log\left(\frac{\left(\frac{1}{(1-\rho)^{m-1}(1+(m-1)\rho)}\right)^{m-i}}{\left(\frac{1}{(1-\rho)^{m-i-1}(1+(m-i-1)\rho)}\right)^m}\right)\right] \\
&= \min_{i=1,\ldots,m-1} -\frac{1}{i}\left[\frac{1}{2}\log\left(\frac{(1+(m-i-1)\rho)^m}{(1-\rho)^i(1+(m-1)\rho)^{m-i}}\right)\right].
\end{aligned} \tag{3.27}$$

By a simple calculation, the minimum in (3.26) is always attained by $i^* = m-1$, from which (3.7) follows. ∎

3.3 Trading SK Rate off Quantization Rate by Structured Codes

The achievability proof of Theorem 3.1 involves scalar quantization of X_i at terminal i, $i = 1, \ldots, m$, followed by SK generation for the resulting finite-alphabet source model along the lines of [15, 16]; SK capacity is attained in the limit of infinite quantization rates. The SK, extracted from omniscience at all the terminals in $\mathcal{M}$, involves public communication by said terminals. As underlying the proof of achievability of SK capacity for the finite-alphabet source model with two terminals [39, 1], communication from a single terminal, say terminal 1, suffices to generate an optimum-rate SK from less-than-omniscience.

In the context of a Gaussian source model with two terminals, this motivates the following questions.

- Suppose that quantization at a rate R is permitted at terminal 1, what is the largest rate of SK that can be generated from the quantized source at terminal 1 and the original Gaussian source at terminal 2 using public communication?
- Does the rate of SK thereby generated tend to the SK capacity $C(\mathcal{M} = \{1, 2\}) = \frac{1}{2} \log \frac{1}{1-\rho^2}$ (by (3.8) of Corollary 3.2), as $R \to \infty$?
- Can an explicit code structure be identified for quantization, communication as well as SK extraction?

In order to address these questions, we pose the following formulation, and provide answers that involve structured codes, namely, nested lattice codes and linear codes, combined with randomization at the terminals.

Let M_1, M_2 be independent $\mathcal{M}_1$- and $\mathcal{M}_2$-valued rvs with (M_1, M_2) being independent of $(\mathbf{X}_1, \mathbf{X}_2)$. For each $R > 0$, let $q_R : \mathcal{M}_1 \times \mathbb{R}^n \to \mathcal{Q}_R$ be a (vector) random quantizer of rate R, where $\mathcal{Q}_R \subset \mathbb{R}^n$ with $\frac{1}{n} \log |\mathcal{Q}_R| \leq R$. Let $C(R)$ be the largest rate of a SK that can be generated from $q_R(M_1, \mathbf{X}_1)$ at terminal 1 and $(M_2, \mathbf{X}_2)$ at terminal 2 by public communication (cf. the second paragraph of Section 2.1, with $(M_1, \mathbf{X}_1)$ and $(M_2, \mathbf{X}_2)$ in the roles of $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively) among all choice of q_R, M_1, M_2 as above.

Theorem 3.3: *For every $R > 0$, we have*

$$C(R) = \frac{1}{2} \log \frac{1}{e^{-2I(X_1 \wedge X_2)} + \left(1 - e^{-2I(X_1 \wedge X_2)}\right) e^{-2R}}. \tag{3.28}$$

In particular,

$$\lim_{R \to \infty} C(R) = I(X_1 \wedge X_2) = \frac{1}{2} \log \frac{1}{1 - \rho^2} = C\left(\mathcal{M} = \{1, 2\}\right). \tag{3.29}$$

We present first the converse proof. The proof of achievability using structured lattice codes and linear codes constitutes a second main contribution of this chapter and is presented in Section 3.3.2 below.

3.3.1 The Converse Proof

The proof uses the following technical lemma, the first part of which provides an alternative expression for $C(R)$ in Theorem 3.3.

Lemma 3.4: *(i) For every $R \geq 0$, it holds that*

$$C(R) = \max_{\substack{U \multimap X_1 \multimap X_2 \\ I(U \wedge X_1) \leq R}} I(U \wedge X_2). \tag{3.30}$$

Further, $C(R)$ is nondecreasing, concave and continuous for $R \geq 0$.

(ii) For each $n \geq 1$ and for every quantizer $q : \mathbb{R}^n \to \mathcal{Q}$ where $\mathcal{Q}$ is a finite set, and $R' = \frac{1}{n} \log |\mathcal{Q}|$, it holds that

$$\frac{1}{n} I(q(\mathbf{X}_1) \wedge \mathbf{X}_2) \ \leq \ C(R'). \tag{3.31}$$

Proof: Without loss of generality, we assume that $X_2 = X_1 + N$, where N is independent of $X_1 \sim \mathcal{N}(0,1)$. Consequently,

$$\rho = \frac{\mathbb{E}[X_1 X_2]}{\sqrt{\mathbb{E}[X_1^2]}\sqrt{\mathbb{E}[{X_2}^2]}} = \frac{\mathbb{E}[X_1(X_1+N)]}{\sqrt{1+\sigma_N^2}} = \frac{1}{\sqrt{1+\sigma_N^2}}$$

and, hence,

$$\sigma_N^2 = \frac{1}{\rho^2} - 1. \tag{3.32}$$

Let (U, X_1, X_2) be such that $U \multimap X_1 \multimap X_2$ and $I(U \wedge X_1) \leq R$. We get that

$$\begin{aligned} I(U \wedge X_2) &= h(X_2) - h(X_2|U) = h(X_2) - h(X_1 + N|U) \\ &\leq h(X_2) - \frac{1}{2} \log \left[e^{2h(X_1|U)} + e^{2h(N|U)} \right] \qquad (3.33) \\ &= h(X_2) - \frac{1}{2} \log \left[e^{2h(X_1|U)} + e^{2h(N)} \right], \qquad (3.34) \end{aligned}$$

where (3.33) follows from the conditional entropy power inequality, and (3.34) follows from

$$\begin{aligned} 0 &= I(U \wedge X_2|X_1) = I(U \wedge X_1 + N|X_1) \\ &= I(U \wedge N|X_1) = I(U, X_1 \wedge N) \\ &\geq I(U \wedge N), \end{aligned}$$

where the fourth equality is by the fact that X_1 and N are independent.

Let I denote $I(X_1 \wedge X_2) = \frac{1}{2}\log\frac{1}{1-\rho^2}$. Following from (3.34), we have that

$$\begin{aligned} I(U \wedge X_2) &\leq h(X_2) - \frac{1}{2}\log\left[e^{2h(X_1|U)} + e^{2h(N)}\right] \qquad (3.35) \\ &= \frac{1}{2}\log\left((2\pi e)\frac{1}{\rho^2}\right) - \frac{1}{2}\log\left[e^{2h(X_1|U)} + (2\pi e)\frac{1-\rho^2}{\rho^2}\right] \\ &= \frac{1}{2}\log\left((2\pi e)\frac{1}{\rho^2}\right) - \frac{1}{2}\log\left[e^{-2I(U\wedge X_1)}e^{2h(X_1)} + (2\pi e)\frac{1-\rho^2}{\rho^2}\right] \\ &\leq \frac{1}{2}\log\left((2\pi e)\frac{1}{\rho^2}\right) - \frac{1}{2}\log\left[e^{-2R}(2\pi e) + (2\pi e)\frac{1-\rho^2}{\rho^2}\right], \qquad (3.36) \\ &= \frac{1}{2}\log\frac{1}{e^{-2R}\rho^2 + (1-\rho^2)} \\ &= \frac{1}{2}\log\frac{1}{e^{-2R}(1-e^{-2I}) + e^{-2I}} = C(R). \end{aligned}$$

The first equality follows from (3.32); the second inequality follows from $I(U \wedge X_1) \leq R$. Consequently,

$$\max_{U \multimap X_1 \multimap X_2,\ I(U\wedge X_1)\leq R} I(U \wedge X_2) \leq C(R).$$

To show equality, i.e., to establish (3.30), we shall select a rv U that satisfies $U \multimap X_1 \multimap X_2$ and achieves equalities in both (3.33) (and, hence, (3.35)) and (3.36).

To this end, we shall find a zero-mean Gaussian rv U satisfying $I(U \wedge X_2|X_1) = 0$, $I(X_1 \wedge N|U) = 0$ and $I(U \wedge X_1) = R$. By the conditional entropy power inequality, the condition $I(X_1 \wedge N|U) = 0$ will give equality in (3.33) and the condition $I(U \wedge X_1) = R$ will give equality in (3.36). Specifically, let $U = X_1 + \tilde{N}$, where $\tilde{N}$ is independent of (X_1, X_2) and $\tilde{N} \sim \mathcal{N}(0, \left(\frac{e^{-2R}}{1-e^{-2R}}\right))$. Clearly, $I(U \wedge X_2|X_1) = 0$. Also,

$$\begin{aligned} I(X_1 \wedge U) &= h(U) - h(U|X_1) = \frac{1}{2}\log\frac{\sigma_U^2}{\sigma_{\tilde{N}}^2} \\ &= \frac{1}{2}\log\left(\frac{1}{\sigma_{\tilde{N}}^2} + 1\right) = R. \end{aligned}$$

Next,

$$\begin{aligned} E[NU] &= E[(X_2 - X_1)(X_1 + \tilde{N})] = E[(X_2 - X_1)X_1] \\ &= E[NX_1] = 0. \end{aligned}$$

The second and last equalities are by the facts that $\tilde{N}$ is independent of (X_1, X_2) and that N is independent of X_1, respectively. Since (U, N) are jointly Gaussian with both means being zero, U is independent of N. Consequently,

$$\begin{aligned} I(X_1 \wedge N|U) &= I(X_1, U \wedge N) = I(X_1, \tilde{N} \wedge N) \\ &= I(X_1 \wedge N) + I(\tilde{N} \wedge N|X_1) \\ &\leq I(X_1 \wedge N) + I(\tilde{N} \wedge N, X_1) \\ &= I(X_1 \wedge N) + I(\tilde{N} \wedge X_1, X_2) = 0, \end{aligned}$$

also by the facts that X_1 is independent of N and that $\tilde{N}$ is independent of (X_1, X_2). With this choice of U, (3.30) is established.

It is clear from the definition of $C(R)$ in Theorem 3.3 that it is increasing and continuous for $R \geq 0$. Concavity of $C(R)$ follows from the fact that

$$\frac{dC(R)}{dR} = \frac{(1-e^{-2I})e^{-2R}}{e^{-2R}(1-e^{-2I})+e^{-2I}} = \frac{1}{1+e^{2R}\frac{e^{-2I}}{1-e^{-2I}}},$$

which is positive and decreasing for $R \geq 0$.

The proof of part (ii) is similar to the converse proof in [52] and it is given in Appendix B.1. ∎

Let K be an ϵ_n-SK generated by a scheme in Theorem 3.3 using a randomized quantizer q of rate at most R together with public communication $\mathbf{F}$ and randomization M_1, M_2. Then,

$$\begin{aligned}
H(K) &= I(K \wedge M_2, \mathbf{F}, \mathbf{X}_2) \ + \ H(K|M_2, \mathbf{F}, \mathbf{X}_2) \\
&= I(K \wedge \mathbf{F}) + I(K \wedge M_2, \mathbf{X}_2|\mathbf{F}) + \ H(K|M_2, \mathbf{F}, \mathbf{X}_2) \\
&\leq \epsilon_n + I(K, M_1, q(M_1, \mathbf{X}_1) \wedge M_2, \mathbf{X}_2|\mathbf{F}) + \ H(K|M_2, \mathbf{F}, \mathbf{X}_2) \\
&= \epsilon_n \ + \ I(M_1, q(M_1, \mathbf{X}_1) \wedge M_2, \mathbf{X}_2|\mathbf{F}) + \ I(K \wedge M_2, \mathbf{X}_2|M_1, q(M_1, \mathbf{X}_1), \mathbf{F}) \\
&\quad + H(K|M_2, \mathbf{F}, \mathbf{X}_2) \\
&\leq \epsilon_n \ + \ I(M_1, q(M_1, \mathbf{X}_1) \wedge M_2, \mathbf{X}_2|\mathbf{F}) + \ H(K|M_1, q(M_1, \mathbf{X}_1), \mathbf{F}) \\
&\quad + H(K|M_2, \mathbf{F}, \mathbf{X}_2), \\
&\leq \epsilon_n \ + \ I(M_1, q(M_1, \mathbf{X}_1) \wedge M_2, \mathbf{X}_2|\mathbf{F}) + \ 2(\epsilon_n \log|\mathcal{K}| + 1), \\
&\leq \epsilon_n \ + \ I(M_1, q(M_1, \mathbf{X}_1) \wedge M_2, \mathbf{X}_2) + \ 2(\epsilon_n \log|\mathcal{K}| + 1), \qquad (3.37) \\
&\leq \epsilon_n \ + \ I(q(M_1, \mathbf{X}_1) \wedge \mathbf{X}_2|M_1) + \ 2(\epsilon_n \log|\mathcal{K}| + 1), \qquad (3.38)
\end{aligned}$$

where the third line follows from (2.3); the sixth line follows from the fact that K is recoverable from $(M_1, q(M_1, \mathbf{X}_1), \mathbf{F})$ as also from $(M_2, \mathbf{F}, \mathbf{X}_2)$; (3.37) follows from

[1, Lemma 2.2] (equivalently, this is tantamount to the repeated use of Lemma 3.2 in the manner similar to the attainment of (3.24)); and the last line follows from the mutual independence of $M_1, M_2, (\mathbf{X}_1, \mathbf{X}_2)$.

Using Lemma 3.4 and the fact that M_1, and $(\mathbf{X}_1, \mathbf{X}_2)$ are mutually independent, it follows that

$$\frac{1}{n} I(q(M_1, \mathbf{X}_1) \wedge \mathbf{X}_2 | M_1) \leq C(\frac{1}{n} \log |\mathcal{Q}|).$$

Continuing from (3.38) using (2.2), we have that

$$\frac{1}{n} \log |\mathcal{K}| \leq C(\frac{1}{n} \log |\mathcal{Q}|) + o_n(1).$$

The converse proof is completed by the fact that $C(R)$ is continuous for $R \geq 0$. ■

3.3.2 SK Generation Scheme Using Nested Lattice and Linear Codes

In order to describe our scheme for achieving $C(R)$ and, hence, the SK capacity in Theorem 3.3 using nested lattice codes and linear codes, we first compile, in Section 3.3.1.1, pertinent definitions and facts from [60]. Our scheme and results are presented in Section 3.3.1.2.

3.3.2.1 Nested Lattice Codes: Definitions and Facts

Definition 3.1: Consider n basis (column) vectors $\mathbf{g}_1, \ldots, \mathbf{g}_n$ in $\mathbb{R}^n$. An *n-dimensional lattice code* Λ is the set of all integral combinations of these basis vectors, i.e.,

$$\Lambda \triangleq \{\lambda : \lambda = G\,\mathbf{i} \ \text{ for some } \ \mathbf{i} \in \mathbb{Z}^n\},$$

with $n \times n$ generating matrix $G = [\mathbf{g}_1, \ldots, \mathbf{g}_n]$. Clearly, Λ contains the zero vector $\mathbf{0}$ in $\mathbb{R}^n$.

• The *Voronoi region* of a lattice code Λ, denoted by $\nu(\Lambda)$, is the nearest neighbor set of $\mathbf{0}$ in $\mathbb{R}^n$, i.e.,

$$\nu(\Lambda) \triangleq \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\| < \min_{\lambda \in \Lambda,\, \lambda \neq \mathbf{0}} \|\mathbf{u} - \lambda\|\},$$

where $\| \, . \, \|$ denotes Euclidean norm. Let $|\nu(\Lambda)|$ denote the volume in $\mathbb{R}^n$ of $\nu(\Lambda)$.

• The *second moment per dimension* of a lattice code Λ, denoted by $\sigma^2(\Lambda)$, is

$$\sigma^2(\Lambda) \triangleq \frac{1}{n}\text{Var}[\text{ rv distributed uniformly in } \nu(\Lambda)].$$

• The *covering radius of a lattice code* Λ, denoted by $r_\Lambda^{cov}(n)$, is the infimum of all positive numbers r such that $\mathbb{R}^n \subseteq \Lambda + r\mathcal{B}$, where $\mathcal{B}$ be the n-dimensional sphere with unit radius.

• The operation of *quantization* by a lattice code Λ, denoted by Q_Λ, is

$$Q_\Lambda(\mathbf{x}) \triangleq \arg\min_{\lambda \in \Lambda} \|\mathbf{x} - \lambda\|, \ \mathbf{x} \in \mathbb{R}^n,$$

where ties are broken arbitrarily.

• The *mod* operation of a lattice code Λ is

$$\mathbf{x} \bmod \Lambda \triangleq \mathbf{x} - Q_\Lambda(\mathbf{x}), \ \mathbf{x} \in \mathbb{R}^n,$$

and corresponds to the quantization error.

The following property of the mod operation (cf. [60]) will be useful:

$$((\mathbf{x} \bmod \Lambda) + \mathbf{y}) \bmod \Lambda = (\mathbf{x} + \mathbf{y}) \bmod \Lambda, \ \ \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{3.39}$$

A pair of lattice codes Λ_1, Λ_2 are *nested*, i.e., $\Lambda_1 \supset \Lambda_2$, if there exists an $n \times n$ matrix H with $\mathbb{Z}$-valued entries and with $det(H) > 1$, such that $G_2 = G_1 H$, where G_1 and G_2 are the generating matrices of Λ_1 and Λ_2, respectively. It follows that $|\nu(\Lambda_2)|/|\nu(\Lambda_1)| = det(H)$.

For $\lambda \in \Lambda_1$, the set $\lambda + \Lambda_2 \subset \Lambda_1$ is called a coset of Λ_2 relative to Λ_1; it turns out that there are exactly $|\nu(\Lambda_2)|/|\nu(\Lambda_1)|$ distinct such cosets. For $\lambda' \neq \lambda''$ belonging to both Λ_1 and $\nu(\Lambda_2)$, the cosets $\lambda' + \Lambda_2$ and $\lambda'' + \Lambda_2$ are disjoint. It transpires that we can always find a set $\mathcal{S}$ of $|\nu(\Lambda_2)|/|\nu(\Lambda_1)|$ lattice points of Λ_1, comprising all the lattice points of Λ_1 in $\nu(\Lambda_2)$ and some of the lattice points of Λ_1 on the boundary of $\nu(\Lambda_2)$ such that for distinct $\mathbf{v} \in \mathcal{S}$, the sets $\mathbf{v} + \Lambda_2$ are disjoint and furthermore

$$\Lambda_1 = \bigsqcup_{\mathbf{v} \in \mathcal{S}} \{\mathbf{v} + \Lambda_2\}.$$

The set $\mathcal{S} \subset \Lambda_1$ is called a set of coset leaders of Λ_2 relative to Λ_1; note that there can be several such sets $\mathcal{S}$ since the lattice points of Λ_1 on the boundary of $\nu(\Lambda_2)$ can be selected in many ways. Upon fixing one such set $\mathcal{S}$, the ties of the quantization operation $Q_{\Lambda_2}(\mathbf{x})$, $\mathbf{x} \in \Lambda_1$, can be broken systematically in a unique manner by requiring that $\mathbf{x} \bmod \Lambda_2 = \mathbf{x} - Q_{\Lambda_2}(\mathbf{x})$ coincides with the unique coset leader in $\mathcal{S}$ of the coset containing $\mathbf{x}$.

In the *dithered* quantization of a source using a lattice code (cf. [58, 59]), a rv distributed uniformly in its Voronoi region and independent of the source, is added to the source sequence prior to quantization. This procedure, in effect, decorrelates the "quantization error" from the source, as formalized in the following result of [58, 59].

Lemma 3.5 *[58, 59]: For a $\mathbb{R}^n$-valued rv $\mathbf{X}$ and any given lattice code Λ, let $\mathbf{U}$ be the "dither" rv distributed uniformly in $\nu(\Lambda)$ and independent of $\mathbf{X}$. Then, the quantization error $(\mathbf{X}+\mathbf{U})-Q_\Lambda(\mathbf{X}+\mathbf{U})$ is independent of $\mathbf{X}$ and is distributed as $\mathbf{U}$.*

3.3.2.2 The Scheme

Our scheme for SK generation consists of two steps—analog, followed by digital. It is motivated by, and partly follows, the work of Zamir et al. [60].

1) Analog Part: In the first (analog) step, terminals 1 and 2 agree upon three n-dimensional nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ to be specified below. The following operations are performed on N i.i.d. repetitions $(\mathbf{X}_{1,i}, \mathbf{X}_{2,i})$, $i = 1, \ldots, N$, of $(\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^n$. We remark that the reason for the use of N i.i.d. repetitions is to obtain strong secrecy in step (2.2) below.

• (1.1) *Dithered quantization at terminal 1:* Terminal 1 generates i.i.d. rvs $\mathbf{U}_i$, $i = 1, \ldots, N$, where $\mathbf{U}_i$ is uniformly distributed in $\nu(\Lambda_1)$, and $\{\mathbf{U}_i, \mathbf{X}_{1,i}, \mathbf{X}_{2,i}\}_{i=1}^N$ are mutually independent. This is followed by dithered quantization of $\alpha\mathbf{X}_{1,i}$, $i = 1, \ldots, N$, and a mod operation of the lattice code Λ_3 to yield

$$\mathbf{L}_i = Q_{\Lambda_1}(\alpha\mathbf{X}_{1,i} + \mathbf{U}_i) \bmod \Lambda_3, \tag{3.40}$$

for $\alpha > 0$ to be specified below. Each $\mathbf{L}_i$ takes values in the set of coset leaders of Λ_3 relative to Λ_1, denoted by $\mathcal{L}$, where $|\mathcal{L}| = \frac{|\nu(\Lambda_3)|}{|\nu(\Lambda_1)|}$. The associated quantization rate is $\cong \frac{1}{n}\log\frac{|\nu(\Lambda_3)|}{|\nu(\Lambda_1)|}$.

• (1.2) *Public communication from terminal 1 to terminal 2:* Terminal 1 com-

putes

$$\mathbf{P}_i = \mathbf{L}_i \bmod \Lambda_2 = Q_{\Lambda_1}(\alpha \mathbf{X}_{1,i} + \mathbf{U}_i) \bmod \Lambda_2,\ i = 1, \ldots, N, \tag{3.41}$$

since $\Lambda_2 \supset \Lambda_3$, and publicly communicates $(\mathbf{P}^N, \mathbf{U}^N) = (\mathbf{P}_1, \ldots, \mathbf{P}_N, \mathbf{U}_1, \ldots, \mathbf{U}_N)$ to terminal 2. Observe that each $\mathbf{P}_i$ takes values in the set of coset leaders of Λ_2 relative to Λ_1, denoted by $\mathcal{P}$, with $|\mathcal{P}| = \frac{|\nu(\Lambda_2)|}{|\nu(\Lambda_1)|}$.

• (1.3) *Reconstruction of quantized rvs at terminal 2:* Terminal 2 reconstructs $\mathbf{L}_i$ as $\hat{\mathbf{L}}_i$, $i = 1, \ldots, N$, where

$$\hat{\mathbf{L}}_i = [(\mathbf{P}_i - \alpha \mathbf{X}_{2,i} - \mathbf{U}_i) \bmod \Lambda_2 + \alpha \mathbf{X}_{2,i} + \mathbf{U}_i] \bmod \Lambda_3. \tag{3.42}$$

For $R > 0$ and an arbitrary but fixed $D > 0$, we select α as

$$\alpha(R, D) = \frac{\sqrt{D}\sqrt{e^{2R} - 1}}{\sigma_{X_1}}, \tag{3.43}$$

whereby

$$R = \frac{1}{2} \log \frac{\alpha^2 \sigma_{X_1}^2 + D}{D}. \tag{3.44}$$

Let $I \triangleq \frac{1}{2} \log \frac{1}{1-\rho^2}$ (cf. Theorem 3.3). Our following main technical lemma summarizes the outcome of the first step of the algorithm.

Lemma 3.6: *For $R > 0$, let*

$$R_p = R_p(R) \triangleq \frac{1}{2} \log \left(\left(e^{2R} - 1 \right) e^{-2I} + 1 \right). \tag{3.45}$$

For every $\epsilon > 0$ and all n sufficiently large, there exist n-dimensional nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ such that, for $i = 1, \ldots, N$,

$$\frac{1}{n} \log |\mathcal{L}| \leq R + \epsilon,\ \frac{1}{n} \log |\mathcal{P}| \leq R_p + \epsilon, \tag{3.46}$$

$$Pr\{\hat{\mathbf{L}}_i \neq \mathbf{L}_i\} = o_n(1), \tag{3.47}$$

$$\textit{and} \quad R - \frac{1}{n} H(\mathbf{L}_i|\mathbf{U}_i) = o_n(1). \tag{3.48}$$

2) Digital Part:

Before describing the (digital) part 2, we note that the (finite) set $\mathcal{L}$ in step 1.1 above can be shown to be in 1-1 correspondence with a (finite) field $\mathbb{F}_{|\mathcal{L}|}$ through a mapping f (see Lemma 3.7 and Appendix B.2 below); the rvs $f(\mathbf{L}_i)$ will then take values in $\mathbb{F}_{|\mathcal{L}|}$, $i = 1, \ldots, N$. Part 2 of the scheme entails the following.

• (2.1) *CR generation at terminals 1 and 2 by Slepian-Wolf data compression:* The i.i.d. sequence $f(\mathbf{L}_i)$, $i = 1, \ldots, N$, at terminal 1 is reconstructed near-losslessly at terminal 2 with $f(\hat{\mathbf{L}}_i)$, $i = 1, \ldots, N$, as side information using Slepian-Wolf data compression. This reconstruction is performed with error probability vanishing exponentially in N. Specifically, a linearly encoded Slepian-Wolf codeword $A_1 f(\mathbf{L})^N$, where $f(\mathbf{L})^N = (f(\mathbf{L}_1), \ldots, f(\mathbf{L}_N))$ and A_1 is a matrix with entries taking values in $\mathbb{F}_{|\mathcal{L}|}$, is transmitted publicly. Terminal 2 produces an estimate

$$\widehat{f(\mathbf{L})^N} = (f(\mathbf{L}_1), \widehat{\ldots, f(\mathbf{L}_N)})$$

based on the codeword $A_1 f(\mathbf{L})^N$ and the side information

$$f(\hat{\mathbf{L}})^N = (f(\hat{\mathbf{L}}_1), \ldots, f(\hat{\mathbf{L}}_N)).$$

The rate of the codeword $A_1 f(\mathbf{L})^N$ is

$$\frac{1}{n} H(f(\mathbf{L}_1)|f(\hat{\mathbf{L}}_1)) = o_n(1)$$

by Fano's inequality since, from (3.47), $Pr\{\mathbf{L}_1 \neq \hat{\mathbf{L}}_1\} = o_n(1)$.

• (2.2) *SK generation by linear operation on CR:* Lastly, terminals 1 and 2 generate a SK K, by means of a linear operation on the CR $f(\mathbf{L})^N$, viz. $K = A_2 f(\mathbf{L})^N$, with A_2 having entries in $\mathbb{F}_{|\mathcal{L}|}$, of rate arbitrarily close to

$$C(R) = R - R_p = \frac{1}{2} \log \frac{1}{e^{-2I} + (1 - e^{-2I})e^{-2R}},$$

the optimum tradeoff between SK rate and the quantization rate in Theorem 3.3.

3.3.2.3 Achievability Proof

Using the two-step scheme described in the previous section, our second main contribution in this chapter establishes the existence of a strong SK of rate arbitrarily close to $C(R)$. In particular, we show now that for any $R > 0$ and any $R_s < C(R)$ (cf. Theorem 3.3), there exist n-dimensional nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$, mapping f and matrices A_1, A_2, such that the scheme above produces a rv $f(\mathbf{L})^N$ of rate arbitrarily close to R from which a strong SK $K^{(n)} = A_2 f(\mathbf{L})^N$ can be extracted of rate arbitrarily close to R_s.

Without loss of generality, we can write

$$\mathbf{X}_1 = \mathbf{X}_2 + \mathbf{Z}, \tag{3.49}$$

where $\mathbf{Z}$ is independent of $\mathbf{X}_2$ and consists of n i.i.d. repetitions of the rv $Z \sim \mathcal{N}(0, \sigma_Z^2)$ with $\sigma_Z^2 = \sigma_{X_1}^2 - \sigma_{X_2}^2$ so that, from (3.44), $R_p = R_p(R)$ can be written also

as

$$\begin{aligned} R_p &= \frac{1}{2}\log\left((e^{2R}-1)e^{-2I}+1\right) \\ &= \frac{1}{2}\log\left(\frac{(e^{2R}-1)\sigma_Z^2}{\sigma_{X_1}^2}+1\right) \\ &= \frac{1}{2}\log\frac{\alpha^2\sigma_Z^2+D}{D}. \end{aligned} \tag{3.50}$$

Further, $\mathbf{U}$ is taken to be independent of $(\mathbf{X}_2, \mathbf{Z})$ and, hence, $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Z})$.

Proof of Lemma 3.6: We suppress the symbol i. The proof relies on the existence of three "good" n-dimensional nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ with the properties stated below in Lemma 3.7; the proof of existence is obtained by suitably generalizing ideas from [21] (see the proof in Appendix B.2).

Lemma 3.7: *("Good" lattice codes): For each $R > 0$ and $D > 0$, let $R_p = R_p(R)$ and $\alpha = \alpha(R, D)$ as in (3.43), (3.45). For every $\epsilon > 0$ and all n sufficiently large, there exist n-dimensional nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ with*

$$\frac{1}{n}\log\frac{|\nu(\Lambda_3)|}{|\nu(\Lambda_1)|} \leq R+\epsilon, \quad \frac{1}{n}\log\frac{|\nu(\Lambda_2)|}{|\nu(\Lambda_1)|} \leq R_p+\epsilon, \tag{3.51}$$

$$Pr\{\alpha\mathbf{Z}-\mathbf{U} \notin \nu(\Lambda_2)\} = o_n(1) \tag{3.52}$$

$$Pr\{\alpha\mathbf{X}_1-\mathbf{U} \notin \nu(\Lambda_3)\} = o_n(1), \tag{3.53}$$

$$\sigma^2(\Lambda_1) = D, \tag{3.54}$$

$$\text{and} \quad r_{\Lambda_1}^{cov}(n) = O(\sqrt{n}). \tag{3.55}$$

Upon using such "good" lattice codes, the claimed rates in (3.46) follow from

(3.51). We next consider (3.47). By (3.41), upon using (3.39), we get

$$\begin{aligned}(\mathbf{P}-\alpha\mathbf{X}_2-\mathbf{U})\bmod\Lambda_2 &= (Q_{\Lambda_1}(\alpha\mathbf{X}_1+\mathbf{U})\bmod\Lambda_2-\alpha\mathbf{X}_2-\mathbf{U})\bmod\Lambda_2\\ &= (Q_{\Lambda_1}(\alpha\mathbf{X}_1+\mathbf{U})-\alpha\mathbf{X}_2-\mathbf{U})\bmod\Lambda_2\\ &= (\alpha\mathbf{Z}-\mathbf{E})\bmod\Lambda_2,\end{aligned}$$

where $\mathbf{E}=(\alpha\mathbf{X}_1+\mathbf{U})\bmod\Lambda_1$ is a quantization error with respect to Λ_1. Therefore, from (3.42),

$$\begin{aligned}\hat{\mathbf{L}} &= [\,[(\mathbf{P}-\alpha\mathbf{X}_2-\mathbf{U})\bmod\Lambda_2]+\alpha\mathbf{X}_2+\mathbf{U}]\bmod\Lambda_3\\ &= [(\alpha\mathbf{Z}-\mathbf{E})\bmod\Lambda_2+\alpha\mathbf{X}_2+\mathbf{U}]\bmod\Lambda_3.\end{aligned}$$

Defining the event $\mathcal{E}=\{\alpha\mathbf{Z}-\mathbf{E}\notin\nu(\Lambda_2)\}$, we see that in $\mathcal{E}^c$,

$$\begin{aligned}\hat{\mathbf{L}} &= (\alpha\mathbf{Z}-\mathbf{E}+\alpha\mathbf{X}_2+\mathbf{U})\bmod\Lambda_3\\ &= (\alpha\mathbf{X}_1+\mathbf{U}-(\alpha\mathbf{X}_1+\mathbf{U})\bmod\Lambda_1)\bmod\Lambda_3\\ &= Q_{\Lambda_1}(\alpha\mathbf{X}_1+\mathbf{U})\bmod\Lambda_3\\ &= \mathbf{L},\end{aligned}$$

so that $\{\mathbf{L}\neq\hat{\mathbf{L}}\}\subseteq\mathcal{E}$. Now, observe that $\mathbf{E}$ is conditionally independent of $\mathbf{Z}$ conditioned on $\mathbf{X}_2=\mathbf{x}_2,\ \mathbf{x}_2\in\mathbb{R}^n$, which, combined with the independence of $\mathbf{X}_2$ and $\mathbf{Z}$, gives that $\mathbf{E}$ is independent of $\mathbf{Z}$. Further, $\mathbf{E}$ is distributed as $\mathbf{U}$ by Lemma 3.5 and $\mathbf{U}$ is independent of $\mathbf{Z}$, so that

$$Pr\{\mathbf{L}\neq\hat{\mathbf{L}}\}\leq Pr\{\mathcal{E}\}=Pr\{\alpha\mathbf{Z}-\mathbf{E}\notin\nu(\Lambda_2)\}$$

$$Pr\{\alpha\mathbf{Z}-\mathbf{U}\notin\nu(\Lambda_2)\}=o_n(1),$$

by Lemma 3.7, thereby establishing (3.47).

Lastly, in order to establish (3.48), the idea is to show that $\mathbf{L}$ serves as a codeword of an optimum Gaussian rate distortion code for the source $\mathbf{X}_1$, with CR $\mathbf{U}$ at the encoder and decoder. Since $\mathbf{L}$ can be selected to have rate arbitrarily close to R, it will possess the mentioned attribute if there exists a decoder for reconstructing $\mathbf{X}_1$ from $(\mathbf{L}, \mathbf{U})$ with mean-squared error distortion $\cong e^{-2R}\sigma^2_{X_1}$. Then, with $\mathbf{U}$ (independent of $\mathbf{X}_1$) being known to the encoder and decoder, the codeword $\mathbf{L}$—at optimality—must be nearly independent of $\mathbf{U}$ and nearly uniformly distributed, thereby establishing (3.48). Firstly, we show that, upon using the nested lattice codes above with a suitable decoder, we can reconstruct $\mathbf{X}_1$ from $(\mathbf{L}, \mathbf{U})$ with the distortion above. To this end, consider the decoder that reconstruct $\mathbf{X}_1$ as

$$\hat{\mathbf{X}}_1 = c((\mathbf{L} - \mathbf{U}) \bmod \Lambda_3),$$

where $c > 0$ is to be chosen later so as to minimize the mean-squared error distortion.

Using (3.39), we have that

$$\begin{aligned}
(\mathbf{L} - \mathbf{U}) \bmod \Lambda_3 &= (\, Q_{\Lambda_1}(\alpha\mathbf{X}_1 + \mathbf{U}) \bmod \Lambda_3 - \mathbf{U}) \bmod \Lambda_3 \\
&= (\, Q_{\Lambda_1}(\alpha\mathbf{X}_1 + \mathbf{U}) - \mathbf{U}) \bmod \Lambda_3 \\
&= (\alpha\mathbf{X}_1 - \mathbf{E}) \bmod \Lambda_3,
\end{aligned}$$

so that $\hat{\mathbf{X}}_1 = c((\alpha\mathbf{X}_1 - \mathbf{E}) \bmod \Lambda_3)$. Observe next that by Lemma 3.5, $\mathbf{E}$ is independent of $\mathbf{X}_1$ and is distributed as $\mathbf{U}$ and hence by (3.53),

$$Pr\{(\alpha\mathbf{X}_1 - \mathbf{E}) \bmod \Lambda_3 \neq \alpha\mathbf{X}_1 - \mathbf{E}\} = o_n(1). \tag{3.56}$$

It readily follows, as shown in Appendix B.3, that

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\|\mathbf{X}_1-\hat{\mathbf{X}}_1\|^2]\le(1-c\alpha)^2\sigma_{X_1}^2+c^2D+o_n(1). \tag{3.57}$$

The significant sum on the right-side above is minimized by the choice $c=\frac{\alpha\sigma_{X_1}^2}{\alpha^2\sigma_{X_1}^2+D}=\frac{\alpha\sigma_{X_1}^2}{De^{2R}}$ by (3.44), and so

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\|\mathbf{X}_1-\hat{\mathbf{X}}_1\|^2]=\sigma_{X_1}^2e^{-2R}+o_n(1). \tag{3.58}$$

Now we are ready to prove (3.48). With

$$R_{X_1}(D)\ \triangleq\ \min_{\mathbb{E}[(\hat{X}_1-X_1)^2]\le D} I(\hat{X}_1\wedge X_1)\ =\ \frac{1}{2}\log\frac{\sigma_{X_1}^2}{D},\ D\ge 0,$$

$$\begin{aligned}
\frac{1}{n}H(\mathbf{L}|\mathbf{U}) &= \frac{1}{n}I(\mathbf{L}\wedge\mathbf{X}_1|\mathbf{U})\\
&= \frac{1}{n}I(\mathbf{L},\mathbf{U}\wedge\ \mathbf{X}_1)\\
&\ge \frac{1}{n}I(\hat{\mathbf{X}}_1\wedge\mathbf{X}_1)\\
&\ge \frac{1}{n}\sum_{t=1}^{n}I(\hat{X}_{1,t}\wedge X_{1,t})\\
&\ge \frac{1}{n}\sum_{t=1}^{n}R_{X_1}(\mathbb{E}[(\hat{X}_{1,t}-X_{1,t})^2])\\
&\ge R_{X_1}(\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}[(\hat{X}_{1,t}-X_{1,t})^2]).
\end{aligned} \tag{3.59}$$

where the second equality above is by the independence of $\mathbf{U}$ and $\mathbf{X}_1$; the first inequality follows from $\hat{\mathbf{X}}_1$ being a function of $\mathbf{L}$ and $\mathbf{U}$; the second inequality is from $\mathbf{X}_1$ having independent components; and the last inequality is by the convexity of $R_{X_1}(\centerdot)$. Finally, combining (3.58) and (3.59), and noting that $R_{X_1}(\centerdot)$ is nonincreasing and uniformly continuous, we get that

$$\frac{1}{n}H(\mathbf{L}|\mathbf{U})\ge R-o_n(1),$$

which is (3.48). ∎

Fix $R > 0$ and $\epsilon > 0$, the latter to be specified later. For every $\eta > 0$ and all n sufficiently large, Lemma 3.6 provides for the existence of n-dimensional lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ such that

$$\log|\mathcal{L}| < n(R+\eta/2), \quad \log|\mathcal{P}| < n(R_p+\eta/2), \tag{3.60}$$

$$Pr\{\hat{\mathbf{L}}_1 \neq \mathbf{L}_1\} < \epsilon \tag{3.61}$$

and

$$H(\mathbf{L}_1|\mathbf{U}_1) > n(R-\epsilon). \tag{3.62}$$

By Fano's inequality and (3.60), (3.61),

$$H(\mathbf{L}_1|\hat{\mathbf{L}}_1) \leq \epsilon n(R+\eta/2) + h(\epsilon) \tag{3.63}$$

where $h(\centerdot)$ is the binary entropy function. With f being a $\mathbb{F}_{|\mathcal{L}|}$-valued mapping as above, the existence of the matrix A_1, follows from [10, Theorem 1] on the adequacy of linear encoding for the Slepian-Wolf data compression of the i.i.d. rvs $f(\mathbf{L}_i)$, $i = 1,\ldots,N$, with decoder side-information $f(\hat{\mathbf{L}}_i)$, $i = 1,\ldots,N$, and decoding error probability vanishing to zero exponentially in N. Specifically, from [10, Theorem 1] and using (3.61), there exists a $\lceil \frac{N[\epsilon n(R+\eta/2)+h(\epsilon)]}{\log|\mathcal{L}|} \rceil \times N$-matrix A_1 with $\mathbb{F}_{|\mathcal{L}|}$-valued entries such that $f(\mathbf{L})^N$ can be reconstructed from $A_1 f(\mathbf{L})^N$ and $f(\hat{\mathbf{L}})^N$ with the probability of error vanishing exponentially in N.

It remains to show the existence of a matrix A_2 with the asserted property. To this end, we shall use Lemma 2.1, as for the Section 2.2.4, but now with the U in the lemma being continuous valued rv.

Apply Lemma 2.1 with $A = f(\mathbf{L}_1),\ U = \mathbf{U}_1$ and $B = (\mathbf{P}^N, A_1 f(\mathbf{L})^N)$ with

$$\frac{1}{n}\log|\mathcal{B}| \le n(R_p + \eta/2) + \epsilon n(R + \eta/2) + h(\epsilon),$$

and consequentially, using (3.62),

$$\begin{aligned} H(A|U) - \frac{1}{n}\log|\mathcal{B}| &= H(\mathbf{L}_1|\mathbf{U}_1) - \\ &\quad [n(R_p + \eta/2) + \epsilon n(R + \eta/2) + h(\epsilon)] \\ &\ge n(R - R_p - \eta) \\ &= n(C(R) - \eta) \end{aligned} \tag{3.64}$$

if (the yet unspecified) $\epsilon > 0$ is chosen to be sufficiently small. Hence, Lemma 2.1 gives that there exists a matrix A_2 such that for $K^{(nN)} = A_2 f(\mathbf{L})^N$ of range $\mathcal{K}^{(nN)}$,

$$\begin{aligned} \log|\mathcal{K}^{(nN)}| &= \log|\mathcal{L}|^{\frac{\lceil NR_s\rceil}{\lceil \log|\mathcal{L}|\rceil}} \\ &\ge NR_s(1 - o_n(1)) \\ &\ge nN(C(R) - \eta)(1 - o_n(1)) \end{aligned} \tag{3.65}$$

and $s(K^{(nN)}; \mathbf{U}^N, \mathbf{P}^N, A_1 f(\mathbf{L})^N)$ vanishes exponentially in N. Since η is arbitrary, the rate of $f(\mathbf{L})$ and hence $f(\mathbf{L})^N$ can be chosen arbitrary close to R, and the rate of $K^{(nN)}$ can be chosen arbitrarily close to $C(R)$ for all n sufficiently large. This completes the proof. ■

Chapter 4

Perfect SK Generation for the Pairwise Independent Network Model

4.1 Motivation and the Model

In this chapter, we turn our attention to the problem of perfect SK generation. In contrast with materials in the previous two Chapters 2 and 3, this chapter bears the essence of "zero-error information theory," and accordingly, we rely on mathematical techniques of a combinatorial nature [30]. Our emphasis here is on perfect SK generation for fixed signal observation lengths as well as for their asymptotic limits [44]. It is of interest also to consider strong (rather than perfect) SK generation for a natural variant of the model studied in this chapter. This variant model [45] which entails a probabilistic analog of the blueprint below, is deferred to Applendix C.3 so as to retain the combinatorial flavor of the chapter.

We consider a "Pairwise Independent Network (PIN)" model in which every pair i, j of terminals, $1 \leq i < j \leq m$, share a random binary string of length e_{ij} (bits), with the strings shared by distinct pairs of terminals being mutually independent.

The PIN model is motivated by practical aspects of a wireless communication network in which terminals communicate on the same frequency. In a typical multipath environment, the wireless channel between each pair of terminals produces a random mapping between the transmitted and received signals which is time-varying

and location-specific. For a fixed time and location, this mapping is reciprocal, i.e., effectively the same in both directions. Also, the mapping decorrelates over different time-coherence intervals as well as over distances of the order of a few wavelengths.

Our three main contributions in this chapter described below are motivated by a known general connection between (not necessarily perfect) SK generation at the maximum rate and the minimum communication for (not necessarily perfect) omniscience [15, 16].

First, the concept of perfect omniscience enables us to obtain a single-letter formula for the perfect SK capacity of the PIN model; moreover, this capacity is shown to be achieved by linear noninteractive communication, and coincides with the strong SK capacity. This result establishes a connection between perfect SK capacity and the minimum rate of communication for perfect omniscience, thereby particularizing to the PIN model a known general link between these notions *sans* the requirement of the omniscience or secrecy being perfect [15].

Second, the PIN model can be represented by a multigraph. Taking advantage of this representation, we put forth an efficient algorithm for perfect SK generation using a maximal packing of Steiner trees of the multigraph. This algorithm involves public communication that is linear as well as noninteractive, and produces a perfect SK of length equal to the maximum size of such Steiner tree packing. When all the terminals in $\mathcal{M}$ seek to share a perfect SK, the algorithm is shown to achieve perfect SK capacity. However, when only a subset of terminals in $A \subset \mathcal{M}$ wish to share a perfect SK, the algorithm can fall short of achieving capacity; nonetheless, it is shown to achieve at least half of it. Additionally, we obtain nonasymptotic and

asymptotic bounds on the size and rate of the best perfect SKs generated by the algorithm. *These bounds are of independent interest from a purely graph theoretic viewpoint as they constitute new estimates for the maximum size and rate of Steiner tree packing of a given multigraph.*

Third, a special configuration of the PIN model arises when a lone "helper" terminal m aids the "user" terminals in $A = \mathcal{M}\backslash\{m\}$ generate a perfect SK. This model has two special features: firstly, (a single) terminal m possesses all the bit strings that are not in A; secondly, a Steiner tree for A is a spanning tree for either A or $\mathcal{M}$. These features enable us to obtain necessary and sufficient conditions for Steiner tree packing to achieve perfect SK capacity, as also a further sufficient condition that posits a "weak" role for the helper terminal m.

The PIN model is a special case of the multiterminal source model for SK generation in Section 2.1 in which each rv $X_i,\ i = 1, \ldots m$, is finite-valued and is of the form $X_i = (Y_{ij},\ j \in \mathcal{M}\backslash\{i\})$ with $m-1$ components, with the "reciprocal pairs" of rvs $\{(Y_{ij}, Y_{ji}),\ 1 \leq i < j \leq m\}$ being mutually independent. We assume further that $Y_{ij} = Y_{ji},\ 1 \leq i \neq j \leq m$, where Y_{ij} is uniformly distributed over the set of all binary strings of length e_{ij} (bits). Thus, every pair of terminals is associated with a random binary string that is independent of all other random binary strings associated with all other pairs of terminals. The assumption is tantamount to every pair of terminals i, j sharing at the outset privileged and pairwise "perfect secrecy" of e_{ij} bits.

Now, the alphabet[1] $\mathcal{X}_i = \{0,1\}^{\sum_{j\neq i} e_{ij}},\ i = 1, \ldots, m$ and, hence, the alpha-

[1]It is assumed that $\sum_{j\neq i} e_{ij} \geq 1$.

bets at different terminals can be different, Definition 2.5 of linear noninteractive communication is suitably modified here for the PIN model as follows.

Definition 4.1: The communication $\mathbf{F} = \mathbf{F}^{(n)}$ is termed *linear noninteractive communication* (LC) if $\mathbf{F} = (F_1, \ldots, F_m)$ with[2] $F_i = L_i \tilde{X}_i^n$, where L_i is a $b_i \times \left(\sum_{j \neq i} n\, e_{ij}\right)$ matrix[3] with $\{0,1\}$-valued entries, $i = 1, \ldots, m$. The integer $b_i \geq 0,\ i = 1, \ldots, m$, represents the length (in bits) of the communication F_i from terminal i; the overall communication $\mathbf{F}$ has length $\sum_{i=1}^m b_i$ (bits).

A central role is played by the notion of *perfect omniscience* which is a strict version of the concept of *omniscience* introduced in [15]. *This notion does not involve any secrecy requirements.*

Definition 4.2: The communication $\mathbf{F}$ is *communication for perfect omniscience* for A if $(\tilde{X}_1^n, \ldots, \tilde{X}_m^n)$ is perfectly recoverable from $(\tilde{X}_i^n, \mathbf{F})$ for every $i \in A$. Further, $\mathbf{F}$ is *linear noninteractive communication for perfect omniscience* ($\mathrm{LCO}^{(n)}(A)$) if $\mathbf{F}$ is an LC and satisfies the previous perfect recoverability condition. The minimum length (in bits) of an $\mathrm{LCO}^{(n)}(A)$, i.e., $\min_{\mathrm{LCO}^{(n)}(A)} \sum_{i=1}^m b_i$, will be denoted by $\mathrm{LCO}_m^{(n)}(A)$. The *minimum rate* of $\mathrm{LCO}^{(n)}(A)$ is $OMN_p(A) \triangleq \limsup_n \frac{1}{n}\mathrm{LCO}_m^{(n)}(A)$.

4.2 Perfect SK Capacity

Our first main contribution in this chapter is a (single-letter) characterization of the perfect SK capacity for the PIN model, which brings forth a connection with

[2]All additions and multiplications are modulo 2.

[3]It is assumed that $\sum_{j \neq i} e_{ij} \geq 1,\ i = 1, \ldots, m$.

the minimum rate of communication for perfect omniscience.

Theorem 4.1: *The perfect SK capacity for a set of terminals $A \subseteq \mathcal{M}$ is*

$$C_p(A) \;=\; \sum_{i,j} e_{ij} \;-\; OMN_p(A) \tag{4.1}$$

where

$$OMN_p(A) \;=\; \min_{(R_1,\ldots,R_m)\;\in\;\mathcal{R}(A)} \sum_{i=1}^{m} R_i, \tag{4.2}$$

with

$$\mathcal{R}(A) = \left\{ \begin{array}{c} (R_1,\ldots,R_m) \in \mathbb{R}^m :\; R_i \geq 0,\; i=1,\ldots,m, \\ \sum_{i\in B} R_i \;\geq\; \sum_{1\leq i<j\leq m,\; i\in B,\; j\in B} e_{ij}, \\ \forall B \not\supseteq A,\; \emptyset \neq B \subset \mathcal{M} \end{array} \right\}. \tag{4.3}$$

Furthermore, this perfect SK capacity can be achieved with linear noninteractive communication.

Remarks: (i) Clearly, the perfect SK capacity, by definition, cannot exceed the SK capacity Indeed, Theorem 1 implies that the latter is attained by a perfect SK.

(ii) In the same vein, the minimum rate of communication for (asymptotic) omniscience [15] can be attained for the PIN model with perfect recoverability at A of $(X_1^n,\ldots,X_m^n)$ for all n sufficiently large, and with linear noninteractive communication. We mention that noninteractive communication, without a claim of linearity, was shown to suffice for (asymptotic) omniscience in [15].

4.3 Perfect SK Generation by Steiner Tree Packing

Theorem 4.1 serves to establish the sufficiency of an LC in achieving perfect SK capacity through the intermediate attainment of perfect omniscience for A, as seen in its proof below. However, as also evident from the proof, decoding is by exhaustive search of prohibitive complexity.

The PIN model can be represented by a multigraph. This representation leads us to an efficient algorithm for perfect SK generation, not necessarily through perfect omniscience, by a maximal packing of Steiner trees of the multigraph. In particular, this algorithm will be seen to entail public communication in the form of an LC. On the other hand, such an algorithm based on maximal Steiner tree packing need not attain perfect SK capacity. The size of the largest perfect SK that is thus generated can be estimated in terms of the minimum length of an $\mathrm{LCO}^{(n)}(A)$.

Definition 4.3: A *multigraph* $G = (V, E)$ with vertex set V and edge set E is a connected undirected graph with no self loops and with multiple edges possible between any pair of vertices. Given $G = (V, E)$ and a positive integer n, let $G^{(n)} = \left(V, E^{(n)}\right)$ denote the multigraph with vertex set V and edge set $E^{(n)}$ wherein every vertex pair is connected by n times as many edges as in E; in particular, $G^{(1)} = G$. Furthermore, $|E^{(n)}|$ will denote the total number of edges in $E^{(n)}$.

To the PIN model $X_1, \ldots, X_m$, we can associate a multigraph $G = (\mathcal{M}, E)$ with $\mathcal{M} = \{1, \ldots, m\}$ and the number of edges connecting a vertex pair (i, j) in E equal to e_{ij}; in particular, the edge connecting (i, j) will be associated with the random binary string Y_{ij}.

By this association, it will be convenient to represent (4.2) and (4.3) as

$$OMN_G(A) \;=\; \min_{(R_1,\ldots,R_m)\,\in\,\mathcal{R}_G(A)} \sum_{i=1}^{m} R_i, \tag{4.4}$$

with

$$\mathcal{R}_G(A) \;=\; \left\{ \begin{array}{c} (R_1,\ldots,R_m) \in \mathbb{R}^m :\; R_i \geq 0,\; i=1,\ldots,m, \\ \\ \sum_{i\,\in\,B} R_i \geq \sum_{1\leq i<j\leq m,\; i\in B,\; j\in B} e_{ij}, \\ \\ \forall B \not\supseteq A,\; \emptyset \neq B \subset \mathcal{M} \end{array} \right\}, \tag{4.5}$$

whereupon (4.1) can be restated as

$$C_p(A) \;=\; |E| - OMN_G(A). \tag{4.6}$$

Furthermore, it is easy and useful to note that for every $n \geq 1$,

$$OMN_{G^{(n)}}(A) \;=\; nOMN_G(A). \tag{4.7}$$

Definition 4.4: For $A \subseteq V$, a *Steiner tree* (for A) of $G = (V, E)$ is a subgraph of G that is a tree, i.e., containing no cycle, and whose vertex set contains A; such a Steiner tree is said to *cover* A. A *Steiner tree packing* of G is any collection of edge-disjoint Steiner trees of G. Let $\mu(A, G)$ denote the *maximum* size of such a packing (cf. [26]), i.e., the maximum number of trees in the packing. The *maximum rate*[4] of Steiner tree packing of G is $\limsup_{n\to\infty} \frac{1}{n}\mu(A, G^{(n)})$. When $A = V$, a Steiner tree becomes a *spanning tree*, with corresponding notions of *spanning tree packing*, maximum size and rate.

Given a PIN model, the notion of Steiner tree packing of the associated multigraph leads to an efficient algorithm for constructing an $\mathrm{LCO}^{(n)}(A)$ and thereby

[4]In fact, $\lim_{n\to\infty} \frac{1}{n}\mu(A, G^{(n)})$ exists, as shown later in Proposition 4.4.

generating a perfect SK. The next Theorem 4.2 indicates that the largest size of a perfect SK that the algorithm generates is the maximum size of the Steiner tree packing. Furthermore, Theorem 4.2 and its corollary, and Theorem 4.5 provide nonasymptotic and asymptotic bounds on the size and rate, respectively, of the best perfect SKs generated by the algorithm. *Of independent interest from a purely graph theoretic viewpoint, these results also constitute new bounds for the maximum size and rate of Steiner tree packing of a given multigraph.*

Theorem 4.2: *For the multigraph $G = (\mathcal{M}, E)$ associated with a PIN model and for $A \subseteq \mathcal{M}$, it holds for every $n \geq 1$ that*

(i) *the terminals in $\mathcal{M}$ can devise an $LCO^{(n)}(A)$ of total length $n|E^{(1)}| - \mu(A, G^{(n)})$ and subsequently generate a perfect SK $K^{(n)}$ with $\log|\mathcal{K}^{(n)}| = \mu(A, G^{(n)})$;*

(ii) $$\mu(A, G^{(n)}) \;\leq\; n|E^{(1)}| - LCO_m^{(n)}(A); \tag{4.8}$$

(iii) *furthermore, $LCO_m^{(n)}(A)$ is bounded below by the value of an integer linear program according to*

$$LCO_m^{(n)}(A) \;\geq\; INT_{G^{(n)}}(A)$$

where

$$INT_{G^{(n)}}(A) \;=\; \min_{(I_1,\ldots,I_m)\;\in\;\mathcal{I}_{G^{(n)}}(A)} \sum_{i=1}^{m} I_i, \tag{4.9}$$

with

$$\mathcal{I}_{G^{(n)}}(A) \;=\; \left\{ \begin{array}{c} (I_1,\ldots,I_m) \in \mathbb{Z}^m :\; I_i \geq 0,\; i = 1,\ldots,m, \\ \sum_{i\in B}\, I_i \;\geq\; n \sum_{1\leq i<j\leq m,\; i\in B,\; j\in B}\; e_{ij}, \\ \forall B \not\supseteq A,\; \emptyset \neq B \subset \mathcal{M} \end{array} \right\}. \tag{4.10}$$

Corollary 4.3: *For every $n \geq 1$, the maximum size of Steiner tree packing of a multigraph $G^{(n)}$ satisfies*

$$\mu(A, G^{(n)}) \ \leq \ n \ |E^{(1)}| - INT_{G^{(n)}}(A), \tag{4.11}$$

with equality when $A = \mathcal{M}$.

Remarks: (i) Note that the bounds in Theorem 4.2 are nonasymptotic, i.e., valid for every n. Also, note in the bound in Theorem 4.2 (ii) for $\mu(A, G^{(n)})$ that $\mathrm{LCO}_m^{(n)}(A)$ is defined in terms of its *operational significance.*

(ii) Further, Theorem 4.2 provides a nonasymptotic *computable* lower bound for $\mathrm{LCO}_m^{(n)}(A)$ in terms of an integer linear program. The optimum value of its linear programming relaxation constitutes a further lower bound which equals $OMN_{G^{(n)}}(A) = nOMN_G(A)$, by (4.7).

Next, we turn to connections between perfect SK capacity $C_p(A)$ and the maximum rate of Steiner tree packing of $G = (\mathcal{M}, E)$. The following concept of "fractional" Steiner tree packing will be relevant.

For $A \subseteq \mathcal{M} = \{1, \ldots, m\}$, consider the collection $\{S_1, \ldots, S_k\}$ of *all distinct* Steiner trees (for A) of G, where $k = k(G)$. Consider the region

$$\mathcal{T}_G(A) = \left\{ \begin{array}{c} (T_1, \ldots, T_k) \in \mathbb{R}^k : \ T_l \geq 0, \ l = 1, \ldots, k, \\ \sum_{l:(i,j) \in S_l} T_l \leq \ e_{ij} \\ \forall (i,j), \ 1 \leq i < j \leq m \end{array} \right\}. \tag{4.12}$$

Definition 4.5: For a multigraph $G = (\mathcal{M}, E)$ and $A \subseteq \mathcal{M}$, the *maximal*

"fractional" Steiner tree packing of G, denoted $\mu_f(A,G)$, is $\mu_f(A,G) \triangleq \max_{\mathcal{T}_G(A)} \sum_{l=1}^{k} T_l$.

Remarks: (i) Clearly, $\mu_f(A,G)$ corresponds to a linear program with finite optimum value, and the maximum is attained. Furthermore, it is readily verified that for every $n \geq 1$,

$$\mu_f(A, G^{(n)}) \;=\; n\,\mu_f(A,G). \tag{4.13}$$

(ii) We observe that in Definition 4.4, $\mu(A,G) \triangleq \max_{\mathcal{T}_G(A)\cap\mathbb{Z}^k} \sum_{l=1}^{k} T_l$.

Proposition 4.4: *For a multigraph $G=(\mathcal{M},E)$ and $A \subseteq \mathcal{M}$, it holds that the maximum rate of Steiner tree packing (for A) of G satisfies*

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n}\mu(A,G^{(n)}) &= \liminf_{n\to\infty} \frac{1}{n}\mu(A,G^{(n)}) \\ &= \lim_{n\to\infty} \frac{1}{n}\mu(A,G^{(n)}) \\ &= \mu_f(A,G). \end{aligned} \tag{4.14}$$

Theorem 4.5: *For the multigraph $G=(\mathcal{M},E)$ associated with the PIN model and for $A \subseteq \mathcal{M}$, it holds that*

$$\frac{1}{2}C_p(A) \;\leq\; \lim_{n\to\infty} \frac{1}{n}\mu(A,G^{(n)}) \;\leq\; C_p(A). \tag{4.15}$$

Furthermore, when $A=\mathcal{M}$,

$$\lim_{n\to\infty} \frac{1}{n}\mu(\mathcal{M},G^{(n)}) \;=\; C_p(\mathcal{M}). \tag{4.16}$$

Remark: For the PIN model with m terminals, every Steiner tree has at most $m-1$ edges. Also, from (4.15), $\mu(A,G^{(n)}) \lesssim nC(A)$ for all large n. Hence, the overall complexity of the perfect SK generation algorithm based on Steiner tree packing

is linear (in n).

The upper bound on $\lim_{n\to\infty} \frac{1}{n}\mu(A, G^{(n)})$ in Theorem 4.5 is not tight, in general, as seen by the following example.

Example: Consider the multigraph [32] in Figure 1 with $|\mathcal{M}| = 7$ and $|A| = 4$; the terminals in A are represented by the solid circles and every shown edge is single. Computations give that $C_p(A) = 2.0$ by (4.6), (4.4), while $\lim_{n\to\infty} \frac{1}{n}\mu(G^{(n)}, A) = 1.8$ by Proposition 4.4 and the scheme in Lemma 4.1 below.

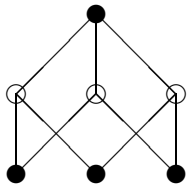

Figure 1: Example

4.4 The Single Helper Case

As observed after Theorem 4.5, the maximum rate of Steiner tree packing can fail to achieve perfect SK capacity. A natural question that remains open is whether the maximum rate of Steiner tree packing equals perfect SK capacity for the special case of the PIN model in which a lone "helper" terminal m assists the "user" terminals in $A = \{1, \ldots, m-1\}$ generate a perfect SK. In this section, we provide partial answers.

First, we derive necessary and sufficient conditions for the maximum rate of Steiner tree packing to equal perfect SK capacity in (4.15) and, analogously, the

(nonasymptotic) maximum size of Steiner tree packing to meet its upper bound in (4.11). These conditions entail the notion of a *fractional* multigraph. Throughout this section, we shall assume that $A = \{1, \ldots, m-1\} \subset \mathcal{M} = \{1, \ldots, m\}$.

Definition 4.6: Given a multigraph $G = (\mathcal{M}, E)$ as in Definition 4.3, a *fractional multigraph* $\tilde{G} = (A, \tilde{E})$ in A (with vertex set A) has edge set $\tilde{E} = \{\tilde{e}_{ij} \in \mathbb{R},\ 0 \leq \tilde{e}_{ij} \leq e_{ij},\ 1 \leq i < j \leq m-1\}$. For any such $\tilde{G}$, the *complementary fractional multigraph* $G\backslash\tilde{G} = (\mathcal{M}, E\backslash\tilde{E})$ has vertex set $\mathcal{M}$ and edge set $E\backslash\tilde{E} \triangleq \{e_{ij} - \tilde{e}_{ij},\ 1 \leq i < j \leq m-1;\ e_{im},\ 1 \leq i \leq m-1\}$. The definitions of $\mathcal{R}_G(A)$ in (4.5), $OMN_G(A)$ in (4.4), $\mathcal{T}_G(A)$ in (4.12) and $\mu_f(A, G)$ in Definition 4.5 all have obvious extensions to $\tilde{G}$ and $G\backslash\tilde{G}$ as well. Further, (4.7) and (4.13) also hold for $\tilde{G}$ and $G\backslash\tilde{G}$.

Proposition 4.6: *For the multigraph $G = (\mathcal{M}, E)$ associated with the PIN model, the following hold:*

(i)

$$\mu_f(A, G) \geq \max_{\tilde{G}} \mu_f(A, \tilde{G}) + \mu_f(\mathcal{M}, G\backslash\tilde{G});$$

(ii)

$$OMN_G(A) \leq \min_{\tilde{G}} OMN_{\tilde{G}}(A) + OMN_{G\backslash\tilde{G}}(\mathcal{M});$$

(iii)

$$\mu(A, G) \geq \max_{\tilde{G}_I} \mu(A, \tilde{G}_I) + \mu(\mathcal{M}, G\backslash\tilde{G}_I);$$

(iv)

$$INT_G(A) \leq \min_{\tilde{G}_I} INT_{\tilde{G}_I}(A) + INT_{G\backslash\tilde{G}_I}(\mathcal{M}),$$

where the optima in (i) and (ii) are over all fractional multigraphs $\tilde{G} = (A, \tilde{E})$ in A, and the optima in (iii) and (iv) are over all multigraphs $\tilde{G}_I = (A, \tilde{E})$ in A for

which $\tilde{E}$ consists of only integer-valued $\tilde{e}_{ij}$s.

Theorem 4.7: *For the multigraph $G = (\mathcal{M}, E)$ associated with the PIN model,*

(i)

$$\lim_{n\to\infty} \frac{1}{n}\mu(A, G^{(n)}) \;=\; C_p(A) \tag{4.17}$$

iff

$$OMN_G(A) = \min_{\tilde{G}} \; OMN_{\tilde{G}}(A) + OMN_{G\backslash\tilde{G}}(\mathcal{M}), \tag{4.18}$$

where the minimum is over all fractional multigraphs $\tilde{G} = (A, \tilde{E})$ in A;

(ii)

$$\mu(A, G^{(n)}) = |E| - INT_G(A)$$

iff

$$INT_G(A) = \min_{\tilde{G}_I} \; INT_{\tilde{G}_I}(A) + INT_{G\backslash\tilde{G}_I}(\mathcal{M}), \tag{4.19}$$

where the minimum is over all multigraphs $\tilde{G}_I = (A, \tilde{E})$ for which $\tilde{E}$ consists of only integer-valued $\tilde{e}_{ij}$s.

Our final result provides another sufficient condition for the maximum rate of Steiner tree packing to equal perfect SK capacity. Recall from Theorem 4.1 that, in general, perfect SK capacity for A can be attained with public communication that corresponds to the minimum communication for perfect omniscience. If the latter can be accomplished with the sole helper terminal m communicating "sparingly," then it transpires that maximal Steiner tree packing attains the best perfect SK rate. An analogous nonasymptotic version of this claim also holds. Heuristically, a

sufficient "weak" role of the helper terminal m turns the Steiner tree packing of A, in effect, into a spanning tree packing of A.

Let $d_i \triangleq \sum_{j \neq i} e_{ij}$ denote the degree of vertex i, $i = 1, \ldots, m$. Clearly, any $(R_1^*, \ldots, R_m^*)$ (resp. $(I_1^*, \ldots, I_m^*)$) that attains the minimum corresponding to $OMN_G(A)$ (cf. (4.4)) (resp. $INT_G(A)$ (cf. (4.9))) must satisfy $R_i^* \leq d_i$ (resp. $I_i^* \leq d_i$), $i = 1, \ldots, m$.

Theorem 4.8: *For the multigraph $G = (\mathcal{M}, E)$ associated with the PIN model,*

(i) *if there exists $(R_1^*, \ldots, R_m^*)$ that attains $OMN_G(A)$ (cf. (4.4)) with $R_m^* \leq d_m/2$, then*

$$\lim_{n \to \infty} \frac{1}{n} \mu(A, G^{(n)}) \;=\; C_p(A) \;=\; |E| - OMN_G(A).$$

(ii) *if there exists $(I_1^*, \ldots, I_m^*)$ that attains $INT_G(A)$ (cf. (4.9)) with $I_m^* \leq \lfloor d_m/2 \rfloor$, then*

$$\mu(A, G) \;=\; |E| - INT_G(A).$$

4.5 Proofs

Proof of Theorem 4.1: From Remark (i) following Theorem 1, we need prove only the achievability part. The main step is to show, using a random coding argument, the existence with large probability of an $\mathrm{LCO}^{(n)}(A)$ of small length under appropriate conditions; the terminals in A then extract from the corresponding perfect omniscience a perfect SK of optimum rate.

Let $X^n_{\mathcal{M}} = (X^n_1, \ldots, X^n_m)$ take values in $\mathcal{X}^n_{\mathcal{M}} = \mathcal{X}^n_1 \times \ldots \times \mathcal{X}^n_m$, where $\mathcal{X}^n_i = \{0,1\}^{\sum_{j \neq i} n e_{ij}}$. We denote a realization of $X^n_{\mathcal{M}}$ by $x^n_{\mathcal{M}} = (x^n_1, \ldots, x^n_m)$. Fix $b_1, \ldots, b_m$. Let $\mathbf{L} = (\mathbf{L}_1, \ldots, \mathbf{L}_m)$ consist of mutually independent random matrices of appropriate dimensions as in Definition 4.1. Furthermore, the rv $\mathbf{L}_i$ consists of i.i.d. equiprobable components, $i = 1, \ldots, m$. Clearly, $\mathbf{L}_1, \ldots, \mathbf{L}_m$ makes for a random LC.

Since for $\mathbf{L}_1, \ldots, \mathbf{L}_m$ to constitute an $\mathrm{LCO}^{(n)}(A)$, it suffices that the mapping

$$x^n_{\mathcal{M}} \quad \rightarrow (x^n_i, \mathbf{L}_1 x^n_1, \ldots, \mathbf{L}_m x^n_m)$$

be one-to-one for every $i \in A$, we have

$$Pr\{\ \mathbf{L} \text{ does not constitute an } \mathrm{LCO}^{(n)}(A)\ \}$$

$$\begin{aligned}
&= Pr\left\{ \begin{array}{c} \exists\ x^n_{\mathcal{M}} \neq x'^n_{\mathcal{M}} \in \mathcal{X}^n_{\mathcal{M}} \text{ satisfying} \\ x^n_j = x'^n_j \text{ for some } j \in A \text{ such that} \\ \mathbf{L}_i x^n_i = \mathbf{L}_i x'^n_i \text{ for each } i = 1, \ldots, m \end{array} \right\} \\
&= Pr\left\{ \begin{array}{c} \exists\ x^n_{\mathcal{M}} \neq \mathbf{0} \in \mathcal{X}^n_{\mathcal{M}} \text{ satisfying} \\ x^n_j = \mathbf{0} \text{ for some } j \in A \text{ such that} \\ \mathbf{L}_i x^n_i = \mathbf{0} \text{ for each } i = 1, \ldots, m \end{array} \right\} \qquad (4.20) \\
&\leq \sum_{\substack{B \neq \emptyset, \\ B \not\supseteq A}} Pr\left\{ \begin{array}{c} \exists\ x^n_{\mathcal{M}} \in \mathcal{X}^n_{\mathcal{M}} \text{ satisfying} \\ x^n_j \neq \mathbf{0}\ \forall j \in B, \text{ and } x^n_j = \mathbf{0}\ \forall j \in B^c \\ \text{such that } \mathbf{L}_i x^n_i = \mathbf{0} \\ \text{for each } i = 1, \ldots, m \end{array} \right\}, \qquad (4.21)
\end{aligned}$$

where (4.20) is by the linearity of the communication and (4.21) is obtained by applying the union bound to the event in (4.20).

Now, we note by the assumed independence of $\mathbf{L}_1, \ldots \mathbf{L}_m$ and the fact that the components of $\mathbf{L}_i$ are i.i.d. and equiprobable, $i = 1, \ldots, m$, that for each nonempty $B \not\supseteq A$, and any $x^n_{\mathcal{M}}$ satisfying $x^n_j \neq \mathbf{0}\ \forall j \in B,$ and $x^n_j = \mathbf{0}\ \forall j \in B^c$, we have

$$\begin{aligned} Pr\{\mathbf{L}_i x^n_i = \mathbf{0} \text{ for every } i = 1, \ldots, m\} &= Pr\{\mathbf{L}_i x^n_i = \mathbf{0} \text{ for every } i \in B\} \\ &= \prod_{i \in B} 2^{-b_i} = 2^{-\sum_{i \in B} b_i}. \end{aligned} \tag{4.22}$$

Continuing with (4.21) upon using (4.22), we obtain

$$\begin{aligned} & Pr\{\ \mathbf{L} \text{ does not constitute an } \mathrm{LCO}^{(n)}(A)\ \} \\ &\leq \sum_{\substack{B \neq \emptyset, \\ B \not\supseteq A}} \left| \left\{ \begin{array}{l} x^n_{\mathcal{M}} \in \mathcal{X}^n_{\mathcal{M}} : x^n_j \neq \mathbf{0} \\ \\ \forall j \in B,\ x^n_j = \mathbf{0}\ \forall j \in B^c \end{array} \right\} \right| 2^{-\sum_{i \in B} b_i} \\ &\leq \sum_{\substack{B \neq \emptyset, \\ B \not\supseteq A}} 2^{n\left(\sum_{l,k \in B} e_{lk}\right)} 2^{-\sum_{i \in B} b_i} \\ &= \sum_{\substack{B \neq \emptyset, \\ B \not\supseteq A}} 2^{-n\left(\frac{1}{n}\sum_{i \in B} b_i - \sum_{l,k \in B} e_{lk}\right)}. \end{aligned} \tag{4.23}$$

We note that in this proof, the special structure of the PIN model is used for the first time in the second inequality above.

Now, let $(R^*_1, \ldots, R^*_m)$ achieve the minimum in the right side of (4.2). Pick an arbitrary $\epsilon > 0$ and choose b_i in (4.23) as $b_i = \lceil n(R^*_i + \epsilon) \rceil,\ i = 1, \ldots, m.$ Then, by the definition of $\mathcal{R}(A)$, the right side of (4.23) decays to zero exponentially rapidly in n; in particular, we get that for all n sufficiently large, $\mathbf{L}$ constitutes an

$\text{LCO}^{(n)}(A)$ with large probability. This implies the existence of a (deterministic) $L = (L_1, \ldots, L_m)$ that constitutes an $\text{LCO}^{(n)}(A)$ for all n sufficiently large.

It remains to extract a perfect SK from the perfect omniscience obtained above. By the definition of the PIN model, observe that

$$Pr\{X_{\mathcal{M}}^n = x_{\mathcal{M}}^n\} = 2^{-\sum_{l,k} n e_{lk}} \text{ for all } x_{\mathcal{M}}^n \in \mathcal{X}_{\mathcal{M}}^n.$$

By the linearity of the $\text{LCO}^{(n)}(A)$ above, it is readily seen that the cardinality $|\{x_{\mathcal{M}}^n \in \mathcal{X}_{\mathcal{M}}^n : L_i x_i^n = a_i,\, i = 1, \ldots, m\}|$ is the same for all feasible $(a_1, \ldots, a_m)$ where $a_i \in \{0,1\}^{b_i},\ i = 1, \ldots, m$, and that this common number is at least

$$N = 2^{\left(\sum_{l,k} n e_{lk}\right) - \left(\sum_{i=1}^m b_i\right)}.$$

For each communication message $(a_1, \ldots, a_m)$, we index the elements of the *coset* $\{x_{\mathcal{M}}^n : L_i x_i^n = a_i,\ i = 1, \ldots, m\}$ in a fixed manner. Then, for a realization $x_{\mathcal{M}}^n \in \mathcal{X}_{\mathcal{M}}^n$, every terminal in A (which knows $x_{\mathcal{M}}^n$ by omniscience) picks as the perfect SK the index of $x_{\mathcal{M}}^n$ in its coset, as in [55]. Since $X_{\mathcal{M}}^n$ takes values in $\mathcal{X}_{\mathcal{M}}^n$ and since each coset has the same size, it follows that this random index is uniformly distributed and independent of the coset (the communication message), thereby constituting a perfect SK. Lastly, the rate of this perfect SK is at least

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \log N &= \sum_{l,k} e_{lk} - \sum_{i=1}^m R_i^* - m\epsilon \\ &= \sum_{l,k} e_{lk} - OMN_p(A) - m\epsilon, \end{aligned}$$

where $\epsilon > 0$ is arbitrary. ∎

Proof of Theorem 4.2: The proof will rely on the technical Lemma 4.1 which is stated next and established in Appendix C.1.

Lemma 4.1: *Let $G = (V, T)$ be a tree, and associate with each edge a bit. Then the terminals in V can devise a (noninteractive) LC of length $|T| - 1$ bits enabling every terminal in V to recover all the edges of T, i.e., all the bits associated with the edges of T.*

(i,ii) If $\mu(A, G^{(n)}) = k$, say, then $E^{(n)}$ is the disjoint union of k Steiner trees $T_1, \ldots, T_k$ (each of which covers A) and the remaining edge set R, so that

$$|E^{(n)}| \ = \ n|E^{(1)}| \ = \ \sum_{i=1}^{k} |T_i| \ + \ |R|, \tag{4.24}$$

where $|T_i|$ denote the number of edges in T_i.

Apply Lemma 4.1 to every Steiner tree $T_i,\ i = 1, \ldots, k$, in (4.24) to get k LCs that enable every terminal in A to recover the edges of all the $T_i,\ i = 1, \ldots, k$. An additional communication of $|R|$ bits will lead to the recovery of the leftover edges in R. Thus, there exists an $\mathrm{LCO}^{(n)}(A)$ of length

$$\sum_{i=1}^{k} |T_i| \ - \ k \ + \ |R| = n|E^{(1)}| \ - \ k \text{ (bits)},$$

which establishes the first assertion of (i); also, clearly, $\mathrm{LCO}_m^{(n)}(A) \leq n|E^{(1)}| \ - \ k$, thereby proving (ii). To establish the second assertion of (i), it remains to extract a perfect SK from the perfect omniscience obtained using the $\mathrm{LCO}^{(n)}(A)$ above of total length $n|E^{(1)}| \ - \ \mu(A, G^{(n)})$ (bits). This is accomplished exactly as in the proof of Theorem 4.1, whereby the terminals in A extract a perfect SK $K^{(n)}$ with $\log |\mathcal{K}^{(n)}| = \mu(A, G^{(n)})$.

(iii) Consider an $\mathrm{LCO}^{(n)}(A) = (L_1, \ldots, L_m)$ achieving $\mathrm{LCO}_m^{(n)}(A)$ with $(b_1, \ldots, b_m)$ (bits), respectively. Fix $B \subset \mathcal{M},\ B \not\supseteq A$, and consider $\mathcal{S} = \{x_{\mathcal{M}}^n : \ x_j^n =$

$\mathbf{0}$ for every $j \in B^c\}$ with cardinality $2^{n\sum_{1\leq i<j\leq m,\ i\in B,\ j\in B}\ e_{ij}}$. For every $k \in B^c \cap A$ and every $x_{\mathcal{M}}^n \in \mathcal{S}$, it holds that $x_k^n = \mathbf{0}$. Consequently, by the perfect recoverability property of an $\mathrm{LCO}^{(n)}(A)$, such a terminal k must be able to discern all the sequences in $\mathcal{S}$ using *only* $(L_1, \ldots, L_m)$. Note also that for every $x_{\mathcal{M}}^n \in \mathcal{S}$ and every $i \in B^c$, it follows that $L_i(x_i^n) = \mathbf{0}$; therefore, the set of all communication messages corresponding to $\mathcal{S}$ has cardinality at most $2^{\sum_{i\in B} b_i}$. From the mentioned condition on perfect recoverability at terminal $k \in B^c \cap A$ of all sequences in $\mathcal{S}$, it must hold that $2^{\sum_{i\in B} b_i} \geq 2^{n\sum_{1\leq i<j\leq m,\ (i,j)\in B}\ e_{ij}}$. Since this argument is valid for every $B \subset \mathcal{M},\ B \not\supseteq A$, we have that $(b_1, \ldots, b_m) \in \mathcal{I}_{G^{(n)}}(A)$ and, hence, $\mathrm{LCO}_m^{(n)}(A)$ is at least $\min_{(I_1,\ldots,I_m)\in\mathcal{I}_{G^{(n)}}(A)}\ \sum_{i=1}^m I_i$. ∎

Proof of Corollary 4.3: The inequality in the Corollary 4.3 is immediate from (4.8) and (4.10). Equality when $A = \mathcal{M}$ relies on Lemma 4.2 and 4.3 below; Lemma 4.2 is a classic result of Nash-Williams [43] and Tutte [48] on the maximal size of spanning tree packing of a multigraph, and Lemma 4.3 [15] provides an upper bound for strong SK capacity

Lemma 4.2 *[43], [48]: For a multigraph $G = (\mathcal{M}, E)$,*

$$\mu(\mathcal{M}, G) \ = \ \Big\lfloor \min_{\mathcal{P}} \ \frac{1}{|\mathcal{P}| - 1} \ \Big|\{e \in E : e \text{ crosses } \mathcal{P}\}\Big| \Big\rfloor,$$

where the minimum is over all partitions $\mathcal{P}$ of $\mathcal{M}$.

Lemma 4.3 *[15]: For the multigraph $G = (\mathcal{M}, E)$ associated with the PIN model and for $A \subseteq \mathcal{M}$,*

$$C_p(A) \ = \ |E| - \min_{(R_1,\ldots,R_m)\ \in\ \mathcal{R}_G(A)} \ \sum_{i=1}^m R_i \ \leq \ \min_{\mathcal{P}} \ \frac{1}{|\mathcal{P}| - 1} \ \Big|\{e \in E : e \text{ crosses } \mathcal{P}\}\Big|,$$

where the minimum is over all partitions $\mathcal{P}$ of $\mathcal{M}$ such that each atom of $\mathcal{P}$ intersects A .

By (4.5) and (4.10), $\mathcal{R}_{G^{(n)}}(\mathcal{M}) \supset \mathcal{I}_{G^{(n)}}(\mathcal{M})$ with $G^{(n)}$ and $\mathcal{M}$ in the roles of G and A in (4.5), it is clear that

$$\left\lceil \min_{(R_1,\ldots,R_m)\ \in\ \mathcal{R}_{G^{(n)}}(\mathcal{M})} \sum_{i=1}^{m} R_i \right\rceil \quad \leq \min_{(I_1,\ldots,I_m)\ \in\ \mathcal{I}_{G^{(n)}}(\mathcal{M})} \sum_{i=1}^{m} I_i, \tag{4.25}$$

noting that the value on the right side above is an integer.

Then the claimed equality follows since

$$\begin{aligned}
\mu(\mathcal{M}, G^{(n)}) &\leq n\,|E^{(1)}| - \min_{(I_1,\ldots,I_m)\ \in\ \mathcal{I}_{G^{(n)}}(\mathcal{M})} \sum_{i=1}^{m} I_i \\
&\leq \left\lfloor n\,|E^{(1)}| - \min_{(R_1,\ldots,R_m)\ \in\ \mathcal{R}_{G^{(n)}}(\mathcal{M})} \sum_{i=1}^{m} R_i \right\rfloor, \quad \text{by (4.25)} \\
&\leq \left\lfloor \min_{\mathcal{P}} \frac{1}{|\mathcal{P}|-1} \left|\{e \in E^{(n)} : e \text{ crosses } \mathcal{P}\}\right| \right\rfloor \qquad (4.26) \\
&= \mu(\mathcal{M}, G^{(n)}), \quad \text{by Lemma 4.2,}
\end{aligned}$$

where (4.26) is by Lemma 4.3. ∎

Proof of Proposition 4.4: By Remark (ii) after Definition 4.5 in Section 4.3, we have that

$$\frac{1}{n}\mu(A, G^{(n)}) \;=\; \frac{1}{n} \max_{\mathcal{T}_G^{(n)}(A)\cap\mathbb{Z}^k} \sum_{l=1}^{k} T_l \;=\; \max_{\mathcal{T}_G(A)\cap\frac{1}{n}\mathbb{Z}^k} \sum_{l=1}^{k} T_l.$$

Since

$$\lim_{n\to\infty} \max_{\mathcal{T}_G(A)\cap\frac{1}{n}\mathbb{Z}^k} \sum_{l=1}^{k} T_l \;=\; \max_{\mathcal{T}_G(A)} \sum_{l=1}^{k} T_l \;=\; \mu_f(A, G),$$

the assertion follows. ∎

Proof of Theorem 4.5: The second inequality of the theorem is immediate by Theorem 4.2 (i) and the definition of $C_p(A)$.

The proof of the first inequality takes recourse to the following result.

Lemma 4.4 *[34, 31]: For a multigraph $G = (\mathcal{M}, E)$ that is Eulerian*[5] *and $A \subseteq \mathcal{M}$,*

$$\mu(A, G) \geq \left\lfloor \frac{1}{2} \min_{C \subset \mathcal{M}: C \cap A \neq \emptyset} \left| \{e \in E : e \text{ crosses } C, C^c\} \right| \right\rfloor.$$

Now, for every n, $\mathcal{R}_{G^{(n)}}(A) \supset \mathcal{I}_{G^{(n)}}(A)$, and so

$$\min_{\mathcal{I}_{G^{(n)}}(A)} \sum_{i=1}^{m} I_i \geq \min_{\mathcal{R}_{G^{(n)}}(A)} \sum_{i=1}^{m} R_i.$$

By Lemma 4.3,

$$\begin{aligned} n|E^{(1)}| - \min_{\mathcal{R}_{G^{(n)}}(A)} \sum_{i=1}^{m} R_i &\leq \min_{\mathcal{P}} \frac{1}{|\mathcal{P}| - 1} \left| \{e \in E^{(n)} : e \text{ crosses } \mathcal{P}\} \right| \\ &\leq \min_{C \subset \mathcal{M}: C \cap A \neq \emptyset} \left| \{e \in E^{(n)} : e \text{ crosses } C, C^c\} \right|. \quad (4.27) \end{aligned}$$

Restricting ourselves to n even, note that $G^{(n)}$ is Eulerian, i.e., each vertex has even degree. Then since the term within $\lfloor\ \rfloor$ in the right side in Lemma 4.4 is clearly an integer, we have that

$$\begin{aligned} \mu(A, G^{(n)}) &\geq \frac{1}{2} \min_{\emptyset \neq C \subset \mathcal{M}: C \cap A \neq \emptyset} \left| \{e \in E^{(n)} : e \text{ crosses } C, C^c\} \right| \\ &\geq \frac{1}{2} \left[n|E^{(1)}| - \min_{\mathcal{R}_{G^{(n)}}(A)} \sum_{i=1}^{m} R_i \right], \quad \text{by (4.27)} \\ &= \frac{1}{2} \left[n|E^{(1)}| - OMN_{G^{(n)}}(A) \right] \\ &= \frac{1}{2} n \left[|E^{(1)}| - OMN_G(A) \right], \quad \text{by (4.7)} \\ &= \frac{1}{2} n C_p(A), \end{aligned}$$

thereby establishing the left inequality of the theorem. ∎

[5]The number of edges incident on each vertex is even.

Proof of Proposition 4.6: We prove (i) and (ii). The proofs of (iii) and (iv) are similar but simpler, and are omitted.

(i) Similarly as in Remark (i) following Definition 4.5, we note that the right side of (i) corresponds to a linear program with finite optimum value, and the maximum is attained. Let $\tilde{G}^*$, $(T_1^*, \ldots, T_{k_1}^*)$, $(T_1^{**}, \ldots, T_{k_2}^{**})$ attain the maximum in the right side of (i), where $(T_1^*, \ldots, T_{k_1}^*)$ and $(T_1^{**}, \ldots, T_{k_2}^{**})$ attain the respective maxima in $\mu_f(A, \tilde{G}^*)$ and $\mu_f(\mathcal{M}, G \backslash \tilde{G}^*)$, with k_1 (resp. k_2) being the number of all distinct spanning trees in A (resp. $\mathcal{M}$) of G. Clearly, $(T_1^*, \ldots, T_{k_1}^*, T_1^{**}, \ldots, T_{k_2}^{**})$ is feasible for $\mu_f(A, G)$, noting that a Steiner tree for A of G is either a spanning tree in A or a spanning tree in $\mathcal{M}$.

(ii) Similarly as in the proof of (i), we let $\tilde{G}^*$ $(R_1^*, \ldots, R_{m-1}^*)$, $(R_1^{**}, \ldots, R_m^{**})$ attain the minimum in the right side of (ii), where $(R_1^*, \ldots, R_{m-1}^*)$ and $(R_1^{**}, \ldots, R_m^{**})$ attain the respective minima in $OMN_{\tilde{G}^*}(A)$ and $OMN_{G \backslash \tilde{G}^*}(\mathcal{M})$. Clearly, $(R_1^* + R_1^{**}, \ldots, R_{m-1}^* + R_{m-1}^{**}, R_m^{**})$ is feasible for $OMN_G(A)$, thereby proving (ii).

Similar arguments considering the corresponding integer linear programs lead to (iii) and (iv). ∎

Proof of Theorem 4.7: We shall prove only (i); the proof of (ii) is similar and is omitted.

First, we show that (4.18) implies (4.17), i.e.,

$$\lim_{n \to \infty} \frac{1}{n} \mu(A, G^{(n)}) \geq C(A) = |E| - OMN_G(A), \tag{4.28}$$

(since the reverse inequality always hold by Theorem 4.5). Let a fractional multi-

graph $\tilde{G}^* = (A, \tilde{E}^*)$ achieve the minimum in the right side of (4.18). Then,

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\mu(A, G^{(n)}) &= \mu_f(A, G), \quad \text{by (4.14)} \\
&\geq \max_{\tilde{G}} \ \mu_f(A, \tilde{G}) + \mu_f(\mathcal{M}, G\backslash\tilde{G}), \\
&\qquad\qquad \text{by Proposition 4.6 (i)} \\
&\geq \mu_f(A, \tilde{G}^*) + \mu_f(\mathcal{M}, G\backslash\tilde{G}^*). \qquad (4.29)
\end{aligned}$$

Next, because the linear program in the right side of (4.18) involves a cost and linear constraints with only integer-valued coefficients, $\tilde{G}^* = (A, \tilde{E}^*)$ can always be taken to be *rational*, i.e., all $\tilde{e}^*_{ij}$s in $\tilde{E}^*$ are rational. Next, let l be the least common multiple of all $\tilde{e}^*_{ij}$s so that $\tilde{G}^{*(l)} = (A, \tilde{E}^{*(l)})$ is a multigraph with edge set $\tilde{E}^{*(l)} = \{l\ \tilde{e}^*_{ij},\ 1 \leq i < j \leq m-1\}$. Then,

$$\begin{aligned}
\mu_f(A, \tilde{G}^*) &= \frac{1}{l}\mu_f(A, \tilde{G}^{*(l)}), \quad \text{by (4.13)} \\
&= \frac{1}{l}(|\tilde{E}^{*(l)}| - OMN_{\tilde{G}^{*(l)}}(A)) \\
&= |\tilde{E}^*| - OMN_{\tilde{G}^*}(A), \quad \text{by (4.7)}; \qquad (4.30)
\end{aligned}$$

the second equality is by Proposition 4.4 and the second assertion of Theorem 4.5 noting that the vertex set of $\tilde{G}^{*(l)}$ is A. By a similar argument, we have that

$$\mu_f(\mathcal{M}, G\backslash\tilde{G}^*) = |E\backslash\tilde{E}^*| - OMN_{G\backslash\tilde{G}^*}(\mathcal{M}). \qquad (4.31)$$

Substituting (4.30) and (4.31) in (4.29),

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n}\mu(A, G^{(n)}) &\geq |\tilde{E}^*| + |E\backslash\tilde{E}^*| \\
&\qquad -(OMN_{\tilde{G}^*}(A) + OMN_{G\backslash\tilde{G}^*}(\mathcal{M})) \\
&= |E| - OMN_G(A), \quad \text{by (4.18)}
\end{aligned}$$

thereby giving (4.28).

Conversely, to prove that (4.17) implies (4.18), i.e.,

$$OMN_G(A) \geq \min_{\tilde{G}} OMN_{\tilde{G}}(A) + OMN_{G\backslash\tilde{G}}(\mathcal{M})$$

(since the reverse inequality always holds by Proposition 4.6 (ii)), we can assume similarly as above that $\mu_f(A, G)$ is attained by $(T_1^*, \ldots, T_k^*)$ with rational components, where $k = k(G)$ is the number of distinct Steiner trees (for A) of G (see passage preceding (4.12)). Next, since $A = \{1, \ldots, m-1\} \subset \mathcal{M}$, the collection of all distinct Steiner trees of (for A) of G, namely $\{S_1, \ldots, S_k\}$ can be decomposed as $\mathcal{S}_1 \sqcup \mathcal{S}_2$, where $\mathcal{S}_1$ (resp. $\mathcal{S}_2$) comprises all spanning trees in A (resp. $\mathcal{M}$). Consider the fractional multigraph in A defined by

$$\tilde{\tilde{G}}^* = (A, \tilde{\tilde{E}}^*),\ \tilde{\tilde{E}}^* = \{\tilde{\tilde{e}}_{ij}^* = \sum_{\substack{l:(i,j)\in S_l,\\ S_l\in\mathcal{S}_1}} T_l^*, 1 \leq i < j \leq m-1\})$$

Then, it follows that

$$\mu_f(A, G) = \mu_f(A, \tilde{\tilde{G}}^*) + \mu_f(\mathcal{M}, G\backslash\tilde{\tilde{G}}^*) \tag{4.32}$$

since

$$\begin{aligned} \mu_f(A, G) &= \sum_{l=1}^{k} T_l^* \\ &= \sum_{l:\ S_l\in\mathcal{S}_1} T_l^* + \sum_{l:\ S_l\in\mathcal{S}_2} T_l^* \\ &\leq \mu_f(A, \tilde{\tilde{G}}^*) + \mu_f(A, G\backslash\tilde{\tilde{G}}^*), \end{aligned}$$

by the definition of μ_f; the reverse inequality is always true. Finally, the right side of (4.17) satisfies

$$
\begin{aligned}
& OMN_{\tilde{G}^*}(A) + OMN_{G\backslash\tilde{G}^*}(\mathcal{M}) \\
& \qquad \leq \; OMN_{\tilde{\tilde{G}}^*}(A) + OMN_{G\backslash\tilde{\tilde{G}}^*}(\mathcal{M}) \\
& \qquad = \; (|\tilde{\tilde{E}}^*| - \mu_f(A, \tilde{\tilde{G}}^*)) + \\
& \qquad \quad (|E\backslash\tilde{\tilde{E}}^*| - \mu_f(\mathcal{M}, G\backslash\tilde{\tilde{G}}^*)), \\
& \qquad \quad \text{as in (4.30), (4.31)} \\
& \qquad = \; |E| - \mu_f(A, G), \quad \text{by (4.32)} \\
& \qquad = \; OMN_G(A),
\end{aligned}
$$

by (4.17), (4.14) and (4.6). ■

Proof of Theorem 4.8: First, we prove (ii), and then (i) by applying (ii) to $G^{(n)} = (\mathcal{M}, E^{(n)})$ and taking appropriate limits.

The proof of (ii) entails considering a modification of $G = (\mathcal{M}, E)$ obtained by "edge-splitting" at the helper vertex m. Specifically, if G has more than one vertex in A connecting to m, then for any two such vertices $u, v \in A$, let $G^{uv} = (\mathcal{M}, E^{uv})$ denote the multigraph obtained from G by splitting off the edges (u, m) and (v, m), i.e., by reducing e_{um} and e_{vm} each by unity and increasing e_{uv} by unity; note that $|E^{uv}| = |E| - 1$.

The following claim, whose proof is relegated to Appendix C.2, will be used to establish the theorem.

Claim: For a multigraph $G = (\mathcal{M}, E)$,

(a) if m is connected to at most one vertex in A or if there exists $(I_1^*, \ldots, I_m^*)$

attaining $INT_G(A)$ *with* $I_m^* = 0$*, then*

$$\mu(A, G) = |E| - INT_G(A); \tag{4.33}$$

(b) if m *is connected to more than one vertex in* A *and if there exists* $(I_1^*, \ldots, I_m^*)$ *attaining* $INT_G(A)$ *with* $0 < I_m^* \leq \lfloor d_m/2 \rfloor$*, then for* $u \in A$ *connecting to* m *there exists* $v = v(u) \in A$*,* $v \neq u$*, also connecting to* m*, such that* $(I_1^*, \ldots, , I_{m-1}^*, I_m^* - 1)$ *attains* $INT_{G^{uv}}(A)$*, and so*

$$|E| - INT_G(A) = |E^{uv}| - INT_{G^{uv}}(A); \tag{4.34}$$

(c) if m *is connected to more than one vertex in* A*, then for* $u, v \in A$ *both connecting to* m*,*

$$\mu(A, G) \geq \mu(A, G^{uv}).$$

In order to prove (ii), we observe first that it holds if the hypothesis of Claim (a) is met. It remains to consider the realm of Claim (b). Let $(I_1^*, \ldots, I_m^*)$ be as in Claim (b). Then we obtain $G_2 = (\mathcal{M}, E_2) = G^{uv}$ for some $u, v \in A$ connecting to m, and with $(I_1^*, \ldots, I_m^* - 1)$ attaining $INT_{G_2}(A)$. If $I_m^* - 1 = 0$ or m connects to at most one vertex in A in G_2, then by (4.33) (4.34),

$$\mu(A, G_2) \;=\; |E_2| - INT_{G_2}(A) \;=\; |E| - INT_G(A).$$

Else, $G_2 = (\mathcal{M}, E_2)$ is back in the realm of Claim (b), noting that the degree of m in G_2 is $d_m - 2$ and $I_m^* - 1 \leq \lfloor (d_m - 2)/2 \rfloor$ as $2 \leq I_m^* \leq \lfloor d_m/2 \rfloor$. Thus, we obtain a finite number of multigraphs $G_1 = G, G_2, \ldots, G_q$, such that $G_i = (\mathcal{M}, E_i) = G_{i-1}^{uv}$

for some $(u, v) = (u, v)(i)$ in A, and satisfying

$$|E_{i-1}| - INT_{G_{i-1}}(A) = |E_i| - INT_{G_i}(A),\ i = 2, \ldots, q \tag{4.35}$$

and

$$\mu(A, G_q) = |E_q| - INT_{G_q}(A). \tag{4.36}$$

Using Claim (c) repeatedly,

$$\begin{aligned} \mu(A, G) &= \mu(A, G_1) \geq \mu(A, G_q) \\ &= |E_q| - INT_{G_q}(A). \text{ by (4.36)} \\ &= |E| - INT_G(A) \end{aligned} \tag{4.37}$$

by the repeated use of (4.35). Then, (ii) is immediate from (4.37) and Corollary 4.3.

To establish (i), the hypothesis implies (with a slight abuse of notation) that

$$\min_{\mathcal{R}_G(A) \bigcap \{R_m \leq d_m/2\}} \sum_{i=1}^{m} R_i = OMN_G(A). \tag{4.38}$$

Pick $(R_1^*, \ldots, R_m^*)$ that attains the left side with all rational components, and let l be the least common multiple of their denominators. Thus, for every integer $n \geq 1$, $(nlR_1^*, \ldots, nlR_m^*)$ attains $INT_{G^{(nl)}}(A)$. As $nlR_m^* \leq nl\frac{d_m}{2}$, it follows from (ii) that

$$\begin{aligned} \mu(A, G^{(nl)}) &= nl|E| - INT_{G^{(nl)}}(A) \\ &= nl|E| - nlOMN_G(A), \text{ by (4.38).} \end{aligned}$$

Upon dividing both sides by nl and taking limits as $n \to \infty$ (with l fixed), we obtain (i). ∎

Chapter 5

Conclusion

This dissertation, in a nutshell, investigates secret key (SK) generation for specific multiterminal source models with emphasis on a single-letter characterization of SK capacities and algorithms for SK construction. Various security requirements: weak, strong and perfect secrecy, as well as different types of sources: finite-valued and continuous, are studied. In this concluding chapter, we first compile *specific* open problems emerging from our work that are yet to be resolved. Finally, we point out broader research directions that are motivated by this dissertation

5.1 Specific Open Problems

5.1.1 SK for Gaussian Sources

In the Gaussian multiterminal model of Chapter 3, it remains to extend the formulation of Theorem 3.3 to models with arbitrary number $m > 2$ of terminals; the resulting model will involve quantization at one terminal in the secrecy-seeking set A, say terminal k, with randomization allowed at all the terminals. In particular, following the paragraph preceding the statement of Theorem 3.3, the model under consideration can be described as follows. Let M_i be a $\mathcal{M}_i$-valued rv, $i = 1, \ldots, m$, with $M_1, M_2, \ldots, M_m, \mathbf{X}_{\mathcal{M}}$ being mutually independent. For each $R > 0$, let $q_R : \mathcal{M}_k \times \mathbb{R}^n \to \mathcal{Q}_R$ be a (vector) random quantizer at terminal k of rate R, where

$\mathcal{Q}_R \subset \mathbb{R}^n$ with $\frac{1}{n}\log|\mathcal{Q}_R| \leq R$. Let $C(R)$ be the largest rate of a SK that can be generated from $q_R(M_k, \mathbf{X}_k)$ at terminal k and $(M_i, \mathbf{X}_i)$ at each of the other terminals $i \in \mathcal{M}\backslash\{k\}$. A characterization of $C(R)$, and devising algorithms for SK generation that attains the optimum tradeoff $C(R)$, will constitute the main objectives of this effort. Similarly as in Theorem 3.3, these questions are connected to problems in multiterminal Gaussian lossy data compression (cf. e.g., [49]).

5.1.2 Perfect SK for the PIN Model

Consider the PIN model of Chapter 4. When all the terminals in $\mathcal{M}$ seek to share a perfect SK, i.e., $A = \mathcal{M}$, we see from Theorem 4.5 that maximal spanning tree packing attains perfect SK capacity; this is no longer true, in general, when $A \subset \mathcal{M}$ (cf. the example in Section 4.3). However, the single helper model in Section 4.4 possesses the special feature that a Steiner tree for A is a spanning tree for either A or $\mathcal{M}$. In spite of this, it is unresolved whether a maximal Steiner tree packing of A attains perfect SK capacity (i.e., if the second inequality in (4.15) is tight) or if (4.11) holds with equality (whereupon the sufficient conditions of Theorem 4.8 become superfluous). We note that the optimality of maximal spanning tree packing in (4.11) and (4.16), constitutes, in effect, a reformulation of the classic graph-theoretic results of Nash-Williams [43] and Tutte [48]. A better *information theoretic* understanding of (4.11) and (4.16) is desirable, and might suggest alternative interpretations of related results in combinatorial tree packing.

Perfect SK capacity in Theorem 4.1 was shown to be achievable by way of the

attainment of perfect omniscience at a minimum communication rate $OMN_p(A)$. However, when $A = \mathcal{M}$, Theorem 4.5 asserts that maximal spanning tree packing attains capacity; an examination of its proof (cf. Lemma 4.1) shows the corresponding rate of communication to be $(m-1)C_p(\mathcal{M})$ which can be less than $OMN_p(\mathcal{M})$. It remains open to characterize the minimum rate of public communication needed to attain perfect SK capacity.

Maximal Steiner tree packing is guaranteed by Theorem 4.5 to attain a fraction of at least half of the capacity $C_p(A)$. What is the best feasible value of this fraction?

A natural generalization of the PIN model involves signals observed at various terminals that go beyond being correlated in pairs. For example, in a sensor network a group of proximate sensors can be assumed to observe (nearly) identical signals. An apt generalization of the PIN model is the Groupwise Independent Network (GIN) model: Consider a collection of mutually independent binary strings $E_1, E_2, \ldots, E_e$, where each E_i is shared by a group of terminals $G_i \subseteq \mathcal{M},\ i = 1, \ldots, e$, with overlaps allowed among the groups. Consequently, each terminal i observes i.i.d. repetitions of the rv $X_i = \{E_j,\ j = 1, \ldots, e :\ i \in G_j\}$. Perfect SK capacity for the GIN model can be characterized using methods similar to those in the proof of Theorem 4.1. An appropriate combinatorial representation of the GIN model is now a hypermultigraph with hyperedges replacing edges in the multigraph representation of the PIN model. The definition of Steiner tree can be suitably modified for hypermultigraphs, and can be used to generate a perfect SK in an efficient manner. However, similarly as in the case of the PIN model, these efficient algorithms are not expected to achieve perfect SK capacity.

This raises the following question: Can ideas from network coding be exploited to construct efficient algorithms for generating rate-optimal perfect SKs? Prompting this question are two findings in Chapter 4: A connection between perfect SK generation and perfect omniscience, and the adequacy of linear coding for attaining perfect SK capacity.

5.2 Future Research Directions

The results in this dissertation represent only a small step towards realizing network cryptosystems that guarantee information theoretic security. Several issues are left unexplored and remain rich research areas for future exploration.

One avenue of future investigations deals with certain fundamental unanswered questions concerning SK capacity. First, the results in Sections 2.2.1 and 2.2.2 regarding connections between SK generation and related multiterminal data compression problems raise the following interesting question: What are the smallest (entropy) rates of common randomness (rather than full omniscience) and the associated communication that are required to attain SK capacity? Answers can be useful in studying practical cryptographic systems with limited communication or storage resources.

Second, the feasibility of perfect SK in a general discrete multiterminal source model merits further investigation. The Pairwise Independent Network (PIN) model in Chapter 4 possesses a special structure that allows the terminals to devise a perfect SK instead of a strong SK without sacrificing rate optimality. This will not

be the case for models with general discrete sources. Thus, what are the correlation structures of sources that enable perfect SK generation?

Third, can there be an intermediate notion of SK capacity between strong SK capacity and perfect SK capacity? In particular, for a SK K with key-space $\mathcal{K}$ and the eavesdropper's observation $\mathbf{F}$, the security index in (2.1) can be interpreted as a measure of closeness, *on the average*, of the conditional entropy of K, conditioned on $\mathbf{F} = \mathbf{f}$, to $\log|\mathcal{K}|$, the entropy of a rv uniformly distributed on $\mathcal{K}$. A more stringent criterion can require that the SK is close to being uniformly distributed for *every* realization $\mathbf{f}$ of $\mathbf{F}$. In this regard, it makes sense to define a new security index as $\bar{s}(K;\mathbf{F}) = \max_{\mathbf{f}} [\log|\mathcal{K}| - H(K|\mathbf{F}=\mathbf{f})]$; a new notion of SK capacity can be defined suitably with respect to this new security index. Clearly, this new SK capacity is bounded below by perfect SK capacity and is bounded above by the strong SK capacity. What are the ramifications of this new security index $\bar{s}$?

A second avenue of related future research involves code constructions for SK generation for various multiterminal source models. Such constructions are connected closely to code constructions for multiterminal data compression (lossless or lossy). These, in turn, are closely related to multiterminal channel code constructions which also constitute a largely open research field. Further, the nature and structure of optimal codes may vary significantly for different types of source models as well as with different levels of security requirements. The primary goal in this direction is to obtain explicit constructions of optimal codes for SK generation that admit efficient implementations. Preliminary works can be found in [56] (see also [54]). In general, constructing optimal codes for SK generation involves more

complex models than those for constructing optimal codes for other information theoretic purposes. One reason is that the problem of SK generation involves two simultaneous constraints: common randomness (recoverability) and secrecy, while other information theoretic problems usually deal with a single constraint (probability of error or distortion).

Appendix A

Appendix for Chapter 2

A.1 Proof of Lemma 2.1

Consider a random $\lfloor \frac{NR}{\log q} \rfloor \times N$ matrix $\mathbf{L}$ with entries taking values mutually independently and uniformly in $\mathbb{F}_q$, i.e., $\mathbf{L}$ is uniformly distributed on the set of all $M \times N = \lfloor \frac{NR}{\log q} \rfloor \times N$ matrices with $\mathbb{F}_q$-valued entries. Further, assume that $\mathbf{L}$ is independent of (A^N, U^N, B) and, hence, the average of $H(LA^N|U^N, B)$ over the set of all $M \times N$ matrices L can be written as $H(\mathbf{L}A^N|\mathbf{L}, U^N, B)$.

The proof of Lemma 2.1 involves a series of steps that result in successive lower bounds for $H(\mathbf{L}A^N \,|\, \mathbf{L}, U^N, B)$, yielding eventually that under the assumptions of the lemma,

$$H(\mathbf{L}A^N \,|\, \mathbf{L}, U^N, B) \geq M \log q - \epsilon_N, \tag{A.1}$$

for an exponentially vanishing ϵ_N, whereupon the assertion of the lemma follows. These steps in the proof will follow, in a similar manner, the recipe in the proof of Lemma 7 of [41] which established an analogous version of Lemma of 2.1 but with an extra assumption that U is also a finite-valued rv.

Note that the set of all $M \times N$ matrices with $\mathbb{F}_q$-valued entries corresponds, in a one-to-one manner, to the set of all linear functions $\mathcal{G} = \{g : \mathbb{F}_q^N \to \mathbb{F}_q^M\}$. Let G denote a rv distributed uniformly on $\mathcal{G}$. Then, it holds that, for any $a_1^N \neq$

$a_2^N,\ a_1^N, a_2^N \in \mathbb{F}_q^N$,

$$Pr\{G(a_1^N) = G(a_2^N)\} = Pr\{\mathbf{L}a_1^N = \mathbf{L}a_2^N\} = q^{-M} = \frac{1}{|\mathbb{F}_q^M|}. \tag{A.2}$$

For a set of functions, not necessarily linear, with a common domain and a common range, this very property (A.2) of the set that for a random function distributed uniformly on the set, the reciprocal of the cardinality of the size of the common range equals the probability that the two values of the random function applied to *any* two distinct inputs coincide, is referred to as the "universal" property in [3], and is used therein to prove the following result.

Lemma A.1 [3, Theorem 3]: *For a finite-valued rv $X \in \mathcal{X}$, let G be the rv uniformly distributed on a universal set of functions $\mathcal{G}$ from $\mathcal{X}$ to a finite set $\mathcal{B}$. Then it holds that*

$$H(G(X)|G) \geq \log|\mathcal{B}| - e^{\log|\mathcal{B}| - H_c(X)},$$

where

$$H_c(X) \triangleq -\log \sum_{x \in \mathcal{X}} P_X(x)^2. \tag{A.3}$$

The proof of Lemma 2.1 relies, additionally, on the following three lemmas; the first two of these lemmas are new with their proofs being relegated to the end of this appendix, while the third lemma was shown in [7].

Lemma A.2: *For $\delta > 0$, let*

$$\mathcal{G}_N(\delta) = \left\{ \begin{array}{l} (a^N, u^N) \in \mathbb{F}_q^N \times \mathbb{R}^{nN} : \\ P_{A^N|U^N}(a^N|u^N) \leq e^{-N(H(A|U)-\delta)} \end{array} \right\},$$

where $H(A|U) \triangleq \mathbb{E}[-\log P_{A|U}(A|U)]$ *and* $P_{A|U}(.|u),\ u \in \mathbb{R}^n$, *is the conditional pmf of* A *conditioned on* $U = u$. *Then,* $Pr\Big((A^N, U^N) \notin \mathcal{G}_N(\delta)\Big) = o_N(e^{-\alpha N})$, *for some* $\alpha > 0$.

Lemma A.3: *For* $\delta > 0$, *let*

$$\mathcal{F}_N(\delta) = \mathcal{G}_N(\delta) \cap \left[\begin{array}{l} \mathbb{F}_q^N \times \\ \\ \{u^N : \sum_{a^N:(a^N,u^N)\in\mathcal{G}_N(\delta)} P_{A^N|U^N}(a^N|u^N) > e^{-\delta N}\} \end{array} \right].$$

Then $Pr\Big((A^N, U^N) \notin \mathcal{F}_N(\delta)\Big) = o_N(e^{-\beta N})$ *for some* $\beta > 0$. *Furthermore, for every* u^N *in a subset of* $\mathbb{R}^{nN}$ *of* $P_{U^N|\mathcal{F}_N(\delta)}$*-measure 1, it holds that*[1]

$$H_c\left(A^N|U^N = u^N, (A^N, U^N) \in \mathcal{F}_N(\delta)\right) \geq N(H(A|U) - 2\delta)$$

for all N *sufficiently large.*

Lemma A.4 [7]: *Let* A *and* B *be finite-valued rvs, and let* $\epsilon > 0$ *be given. Then*

$$P_B\Big(\ \{b \in \mathcal{B} :\ H_c(A) - H_c(A|B = b) \leq \log|\mathcal{B}| + \epsilon\}\ \Big) \geq 1 - 2e^{-\epsilon/2}.$$

Fix $\epsilon > 0$. By Lemma A.4, for every $u^N \in \mathbb{R}^{nN}$,

$$P_{B|U^N,\mathcal{F}_N(\delta)}\left(\left\{ b \in \mathcal{B} : \begin{array}{l} H_c\Big(A^N|U^n = u^N, B = b, \mathcal{F}_N(\delta)\Big) \geq \\ H_c\Big(A^N|U^n = u^N, \mathcal{F}_N(\delta)\Big) - \log|\mathcal{B}| - \epsilon \end{array} \right\} \Bigg| u^N, \mathcal{F}_N(\delta) \right)$$
$$\geq 1 - 2e^{-\epsilon/2}. \qquad \text{(A.4)}$$

[1]Here, $H_c\Big(A^N|U^N = u^N, (A^N, U^N) \in \mathcal{F}_N(\delta)\Big)$ is computed according to (A.3) but with the conditional pmf of A^N conditioned on $\big\{U^n = u^N, (A^N, U^N) \in \mathcal{F}_N(\delta)\big\}$ instead of according to its marginal probability.

Then with $\mathcal{S} = \mathcal{S}^{(N)} \subset \mathbb{R}^{nN}$ denoting the special subset from Lemma A.3 of $P_{U^N|\mathcal{F}_N(\delta)}$-measure 1, we have

$$P_{U^N,B|\mathcal{F}_N(\delta)}\left(\left\{\big(u^N,b\big) \in \mathcal{S}\times\mathcal{B} : H_c\Big(A^N|U^n=u^N, B=b, \mathcal{F}_N(\delta)\Big) \geq Q\right\}\middle|\mathcal{F}_N(\delta)\right) \geq 1-2e^{-\epsilon/2} \tag{A.5}$$

for all $N = N(\delta)$ sufficiently large, where

$$Q \triangleq N\left(H(A|U)-2\delta\right) - \log|\mathcal{B}| - \epsilon.$$

By Lemma A.1, for (u^N, b) satisfying the condition of (A.5), we have

$$H\Big(\mathbf{L}A^N|\mathbf{L}, U^N=u^N, B=b, 1\Big(\big(A^N,U^N\big)\in\mathcal{F}_N(\delta)\Big)=1\Big) \geq M\log q - e^{M\log q - Q}. \tag{A.6}$$

Since

$$H(\mathbf{L}A^N|\mathbf{L},U^N,B) \;\;\geq\;\; P_{A^N,U^N}(\mathcal{F}(\delta))H\left(\mathbf{L}A^N|\mathbf{L},U^N,B,1\Big(\big(A^N,U^N\big)\in\mathcal{F}_N(\delta)\Big)=1\right),$$

and furthermore by (A.5), (A.6),

$$H\left(\mathbf{L}A^N|\mathbf{L},U^N,B,1\Big(\big(A^N,U^N\big)\in\mathcal{F}_N(\delta)\Big)=1\right) \geq (1-2e^{-\epsilon/2})(M\log q - e^{M\log q-Q}), \tag{A.7}$$

for all $N = N(\delta)$ sufficiently large, we obtain that

$$H(\mathbf{L}A^N|\mathbf{L},U^N,B) \geq Pr(\mathcal{F}(\delta))(1-2e^{-\epsilon/2})(M\log q - e^{M\log q-Q}) \tag{A.8}$$

for all $N = N(\delta)$ sufficiently large.

Upon selecting δ sufficiently small and $\gamma = \epsilon/N$ such that

$$R < [H(A|U) - 2\delta - \frac{1}{N}\log|\mathcal{B}| - \gamma,$$

we obtain from (A.8) that

$$H(\mathbf{A}U^N|\mathbf{L}, U^N, B) \geq (1 - o_N(e^{-\beta N}))(1 - 2e^{-\gamma N/2})(M \log q - e^{-\eta N}),$$

for some $\eta > 0$. This proves (A.1) and, hence, the assertion of the lemma. ∎

Proof of Lemma A.2:

The proof uses results from large deviations theory (see, e.g., [18]) to prove the lemma. Let $X_i = -\log P_{A|U}(A_i|U_i),\ i = 1, \ldots, N$, where $\left(A^N, U^N\right)$ are N i.i.d. repetitions of (A, U). Observe that X_i is nonnegative and $\mathbb{E}[X_i] = H(A|U),\ i = 1, \ldots, N$, where $0 < H(A|U) \leq \log |\mathbb{F}_q| < \infty$.

Then, from Theorem 2.2.3 of [18], we have that for $0 < \delta < H(A|U)$,

$$\limsup \frac{1}{N} \log P(\frac{1}{N} \sum_{i=1}^{N} X_i \in [0, H(A|U) - \delta]) \leq - \inf_{x \in [0, H(A|U) - \delta]} \Lambda^*(x),$$

where

$$\begin{aligned} \Lambda^*(x) &\triangleq \sup_{\lambda \in \mathbb{R}} \lambda x - \Lambda(\lambda) \text{ and} \\ \Lambda(\lambda) &\triangleq \log \mathbb{E}[e^{\lambda(-\log P_{A|U}(A|U))}]. \end{aligned}$$

As in [18], we define $\mathcal{D}_\Lambda \triangleq \{\lambda : \Lambda(\lambda) < \infty\}$. Next, we have that

$$\begin{aligned} \Lambda(1) &= \log \mathbb{E}[\frac{1}{P_{A|U}(A|U)}] \\ &= \log \Big(\int \sum_{a: P_{A|U}(a|u) > 0} P_{A|U}(a|u) \frac{1}{P_{A|U}(a|u)} dF_U(u) \Big) \\ &\leq \log |\mathbb{F}_q| < \infty. \end{aligned}$$

In addition, because $-\log P_{A|U}(A|U) \geq 0$, $\Lambda(\centerdot)$ is nondecreasing. We then have that $(-\infty, 0]$ is in the interior of $\mathcal{D}_\Lambda$. It follows from Lemma 2.2.5(b) in [18] that

$\Lambda^*(H(A|U)) = 0$ and that $\Lambda^*(x)$ is nonincreasing for $x < H(A|U)$. Consequently,

$$\begin{aligned} \limsup \frac{1}{N} \log P(\frac{1}{N}\sum_{i=1}^{N} X_i \in [0, H(A|U) - \delta]) &\leq -\inf_{x\in[0,H(A|U)-\delta]} \Lambda^*(x) \\ &= \Lambda^*(H(A|U) - \delta). \end{aligned}$$

Also, from Lemma 2.2.5(c) in [18] and the fact that 0 is in the the interior of $\mathcal{D}_\Lambda$, $\Lambda(\lambda)$ is differentiable at $\lambda = 0$ and $\Lambda'(0) = H(A|U)$. It suffices to prove that for $0 < \delta < H(A|U)$, $\Lambda^*(H(A|U)-\delta) > 0$. Suppose this is not the case, i.e., there exists $0 < \delta < H(A|U)$ such that $\Lambda^*(H(A|U)-\delta) = 0$. Then, from the definition of $\Lambda^*(x)$, it is necessarily true that for every $\lambda < 0$, $\lambda(H(A|U) - \delta) \leq \Lambda(\lambda)$. Consequently,

$$\lim_{\lambda\to 0^-} \frac{\Lambda(0) - \Lambda(\lambda)}{0 - \lambda} = \lim_{\lambda\to 0^-} \frac{\Lambda(\lambda)}{\lambda} \leq H(A|U) - \delta,$$

thereby contradicting the fact that $\Lambda'(0) = H(A|U)$. This completes the proof of Lemma A.3. ■

Proof of Lemma A.3:

We have that

$$\begin{aligned} Pr\Big((A^N, U^N) \notin \mathcal{F}_N(\delta)\Big) &\leq Pr\Big((A^N, U^N) \notin \mathcal{G}_N(\delta)\Big) \\ &\quad + P_{U^N}\left(\left\{ \begin{array}{c} u^N : Pr\Big((A^N, U^N) \notin \mathcal{G}_N(\delta)|U^N = u^N\Big) \\ > 1 - e^{-\delta N} \end{array} \right\}\right). \end{aligned}$$

In the right side above, the first term $= o_N(e^{-\alpha N})$ by Lemma A.2, while the second term $= \frac{o_N(e^{-\alpha N})}{1-e^{-\delta N}} = o_N(e^{-\beta N})$ for some $\beta < \alpha$. Thus, $Pr\Big((A^N, U^N) \notin \mathcal{F}_N(\delta)\Big) = o_N(e^{-\beta N})$, which is the first assertion of the lemma.

Next, for every $\big(a^N, u^N\big) \in \mathcal{F}_N(\delta)$,

$$Pr\Big(A^N = a^N | U^N = u^N, \big(A^N, U^N\big) \in \mathcal{F}_N(\delta)\Big)$$

$$
\begin{aligned}
&= \frac{P_{A^N|U^N}(a^N|u^N)}{Pr\Big(\,(A^N,U^N)\in\mathcal{F}_N(\delta)|U^N=u^N\Big)}\\
&= \frac{P_{A^N|U^N}(a^N|u^N)}{\sum_{a^N:(a^N,u^N)\in\mathcal{G}_N(\delta)}P_{A^N|U^N}(a^N|u^N)}\\
&< \frac{e^{-N(H(A|U)-\delta)}}{e^{-N\delta}},\ \text{since } \mathcal{F}_N(\delta)\subseteq\mathcal{G}_N(\delta)\\
&= e^{-N(H(A|U)-2\delta)}.
\end{aligned}
$$

Hence, for every $u^N\in\mathcal{S}$, it holds that

$$
\begin{aligned}
&H_c\Big(A^N|U^N=u^N,\big(A^N,U^N\big)\in\mathcal{F}_N(\delta)\Big)\\
&= -\log\sum_{a^N}\Big[Pr\Big(A^N=a^N|U^N=u^N,\big(A^N,U^N\big)\in\mathcal{F}_N(\delta)\Big)\Big]^2\\
&\geq -\log\Big[\sum_{a^N}Pr\Big(A^N=a^N|U^N=u^N,\big(A^N,U^N\big)\in\mathcal{F}_N(\delta)\Big)\centerdot e^{-N(H(A|U)-2\delta)}\Big]\\
&= N(H(A|U)-2\delta),
\end{aligned}
$$

thereby establishing the second assertion of the lemma. ∎

Appendix B

Appendix for Chapter 3

B.1 Proof of Part (ii) of Lemma 3.4

Note that $\mathbf{X}_1 = (X_{1,1}, \ldots, X_{1,n}) = X_1^n$ and $\mathbf{X}_2 = (X_{2,1}, \ldots, X_{2,n}) = X_2^n$, where $(X_{1,i}, X_{2,i})$, $i = 1, \ldots, n$, are i.i.d. repetitions of the jointly Gaussian random variables (X_1, X_2) with both means being zero and with correlation coefficient ρ.

$$\begin{aligned}
\frac{1}{n} I(q(\mathbf{X}_1) \wedge \mathbf{X}_2) &= \frac{1}{n} I(q(X_1^n) \wedge X_2^n) \\
&= \frac{1}{n} h(X_2^n) - \frac{1}{n} h(X_2^n | q(X_1^n)) \\
&= \frac{1}{n} \sum_{i=1}^{n} h(X_{2,i}) - \frac{1}{n} \sum_{i=1}^{n} h(X_{2,i} | q(X_1^n), X_{2,1}^{i-1}) \\
&\leq \frac{1}{n} \sum_{i=1}^{n} h(X_{2,i}) - \frac{1}{n} \sum_{i=1}^{n} h(X_{2,i} | q(X_1^n), X_{1,1}^{i-1}, X_{2,1}^{i-1}) \\
&= \frac{1}{n} \sum_{i=1}^{n} I(q(X_1^n), X_{1,1}^{i-1}, X_{2,1}^{i-1} \wedge X_{2,i}). \qquad \text{(B-1)}
\end{aligned}$$

Next,

$$\begin{aligned}
\frac{1}{n} \log |\mathcal{Q}| &\geq \frac{1}{n} I(q(X_1^n) \wedge X_1^n) \\
&= \frac{1}{n} h(X_1^n) - \frac{1}{n} h(X_1^n | q(X_1^n)) \\
&= \frac{1}{n} \sum_{i=1}^{n} h(X_{1,i}) - \frac{1}{n} \sum_{i=1}^{n} h(X_{1,i} | q(X_1^n), X_{1,1}^{i-1}) \\
&= \frac{1}{n} \sum_{i=1}^{n} h(X_{1,i}) - \frac{1}{n} \sum_{i=1}^{n} h(X_{1,i} | q(X_1^n), X_{1,1}^{i-1}, X_{2,1}^{i-1}), \\
&= \frac{1}{n} \sum_{i=1}^{n} I(q(X_1^n), X_{1,1}^{i-1}, X_{2,1}^{i-1} \wedge X_{1,i}), \qquad \text{(B-2)}
\end{aligned}$$

where the last equality is from

$$\begin{aligned} & I(X_{1,i} \wedge X_{2,1}^{i-1} | q(X_1^n), X_{1,1}^{i-1}) \\ & \qquad \leq \; I(X_{1,i}, q(X_1^n) \wedge X_{2,1}^{i-1} | X_{1,1}^{i-1}) \\ & \qquad \leq \; I(X_{1,i}, X_{1,i+1}^n \wedge X_{2,1}^{i-1} | X_{1,1}^{i-1}) \\ & \qquad \leq \; I(X_{1,i}, X_{1,i+1}^n \wedge X_{1,1}^{i-1}, X_{2,1}^{i-1}) \; = \; 0, \end{aligned}$$

by the assumption that $(X_{1,i}, X_{2,i}),\ i = 1, \ldots, n$, are i.i.d.

Let $U_i = (q(X_1^n), X_{1,1}^{i-1}, X_{2,1}^{i-1})$. Then,

$$\begin{aligned} I(U_i \wedge X_{2,i} | X_{1,i}) \; &= \; I(q(X_1^n), X_{1,1}^{i-1}, X_{2,1}^{i-1} \wedge X_{2,i} | X_{1,i}) \\ &\leq \; I(X_{1,1}^{i-1}, X_{1,i}^n, X_{2,1}^{i-1} \wedge X_{2,i} | X_{1,i}) \\ &\leq \; I(X_{1,1}^{i-1}, X_{1,i}^n, X_{2,1}^{i-1} \wedge X_{1,i}, X_{2,1}^{i-1}) \; = \; 0, \end{aligned} \tag{B-3}$$

by the assumption that $(X_{1,i}, X_{2,i}),\ i = 1, \ldots, n$, are i.i.d., so that $U_i \multimap X_{1,i} \multimap X_{2,i}$.

Consequently, from (B-1) and (3.30), we get

$$\frac{1}{n} I(q(\mathbf{X}_1) \wedge \mathbf{X}_2) \leq \frac{1}{n} \sum_{i=1}^{n} C(I(U_i \wedge X_{1,i})) \tag{B-4}$$

and, from (B-2), we get

$$\frac{1}{n} \sum_{i=1}^{n} I(U_i \wedge X_{1,i}) \leq \frac{1}{n} \log |\mathcal{Q}|. \tag{B-5}$$

Continuing from (B-4), we have that

$$\begin{aligned} \frac{1}{n} I(q(\mathbf{X}_1) \wedge \mathbf{X}_2) \; &\leq \; \frac{1}{n} \sum_{i=1}^{n} C(I(U_i \wedge X_{1,i})) \\ &\leq \; C\left(\frac{1}{n} \sum_{i=1}^{n} I(U_i \wedge X_{1,i}) \right) \\ &\leq \; C\left(\frac{1}{n} \log |\mathcal{Q}| \right), \end{aligned}$$

where the second inequality is by the concavity of $C(\bullet)$ and the last inequality is from (B-5) and the fact that $C(\bullet)$ is increasing.

B.2 Proof of Lemma 3.7

For the proof of Lemma 3.7, we shall need the following definitions.

Definition B.1: The *effective radius* of an n-dimensional lattice code Λ, denoted by $r_{\Lambda}^{eff}(n)$, is the radius of an n-dimensional sphere with the same volume as the Voronoi region of the lattice code.

Definition B.2: A sequence of lattice codes is *good for covering* if the ratio of its covering radius to effective radius approaches 1 as the dimension of the lattice codes tends to ∞.

Definition B.3: Let Z be a Gaussian rv with mean 0 and variance σ_Z^2 and let $\mathbf{Z}$ be n i.i.d. repetitions of Z. For each $\delta > h(Z) = \frac{1}{2}\log(2\pi e\sigma_Z^2)$, a sequence of lattice codes Λ is said to be *exponentially good for AWGN channel coding for noise Z without power constraint and with parameter δ* if there exists a mapping $E(\bullet)$ such that $E(u) > 0$ for every $u > 0$,

$$\lim_{n\to\infty} \frac{1}{n} \log |\nu(\Lambda)| = \delta,$$

and

$$Pr\{\mathbf{Z} \notin \nu(\Lambda)\} < e^{-n(E(\delta - h(Z)) - o_n(1))}.$$

The exponent of the error probability $Pr\{\mathbf{Z} \notin \nu(\Lambda)\}$ can be expressed in terms of the ratio, $\rho > 1$, of the effective radius of the lattice code to the (approximated)

radius of the Gaussian noise vector $\sqrt{n}\sigma_Z$. In [46], Poltyrev characterizes an achievable error exponent $E_P(\rho)$. In particular, he shows the existence of a sequence of lattice codes $\Lambda = \Lambda^{(n)}$ with the property that

$$\lim_{n\to\infty} \frac{r_\Lambda^{eff}(n)}{\sqrt{n}\sigma_Z^2} \to \rho$$

and

$$Pr\{\mathbf{Z} \notin \nu(\Lambda)\} < e^{-n(E_P(\rho)-o_n(1))},$$

where $E_P(\rho)$ is the *Poltyrev* exponent given by

$$E_P(\rho) \;=\; \begin{cases} \frac{1}{2}[(\rho^2-1)-\ln\rho], & 1 \le \rho^2 \le 2 \\ \frac{1}{2}\ln\frac{e\rho^2}{4}, & 2 \le \rho^2 \le 4 \\ \frac{\rho^2}{8}, & \rho^2 \ge 4. \end{cases}$$

Observe that the properties of being good for covering and being exponentially good for AWGN channel coding and achieving the Poltyrev exponent are invariant under scaling. To prove the existence of nested lattice codes with the required properties, we shall use the results of [21]. In [21], the authors considered the random lattice ensemble constructed from a random linear code $\mathcal{C}$ by the following procedure described in x below (The construction is known as Construction A in the theory of lattices, see, e.g., [9]). The random lattice ensemble is denoted by $\mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}$, where $p(n)$ is a sequence of primes and $\mathcal{C}$ is the ensemble of uniform random linear code over $\mathbb{Z}_{p(n)}$.

Definition B.4:

• Let $\mathcal{C}$ denote the uniform random $(n, k(n))$ linear code ensemble over $\mathbb{Z}_{p(n)}$. Specifically, the random generating matrix G of the code in the ensemble is obtained

by drawing each element of G independently and uniformly from $\mathbb{Z}_{p(n)}$ and letting $\mathcal{C} \triangleq \{\mathbf{x} = \mathbf{i}G, \mathbf{i} \in \mathbb{Z}_{p(n)}^{k(n)}\}$, where all the operations are over $\mathbb{Z}_{p(n)}$ (i.e., modulo-p).

• Transform each codeword of $\mathcal{C}$ into a point in $[0,1]^n \subset \mathbb{R}^n$ by dividing all the coordinates by $p(n)$. Denote such a random constellation by $\frac{1}{p(n)}\mathcal{C} \subset [0,1]^n$.

• Replicate $\frac{1}{p(n)}\mathcal{C}$ over all of $\mathbb{R}^n$ by performing $\mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}$. It is easy to check that $\mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}$ is indeed a lattice.

Note that if G is nonsingular then the volume of the Voronoi region is $p(n)^{-k(n)}$. The probability of G being singular can be easily shown to be at most $p(n)^{-n}(p^k(n)-1)$ (see (24) of [21]). Following [21], we shall consider only lattice ensembles such that $k(n) \leq \beta n$ for some $0 < \beta < 1$, so that this mentioned probability goes to zero in n at least exponentially (p may also grow with n). Consequently, there is a relationship among the parameters $p(n)$, $k(n)$ and $r_\Lambda^{eff}(n)$ for typical lattice codes in the ensemble which can be stated as:

$$\begin{aligned} p(n)^{k(n)} &= \frac{1}{V_{\mathcal{B}}(r_\Lambda^{eff}(n))} = \frac{\Gamma(\frac{n}{2}+1)}{\pi^{n/2}(r_\Lambda^{eff}(n))^n} \\ &\approx \sqrt{n\pi}(\frac{n}{2\pi(r_\Lambda^{eff}(n))^2})^{\frac{n}{2}}, \end{aligned} \tag{B-6}$$

where $V_{\mathcal{B}}(r_\Lambda^{eff}(n))$ denotes the volume of the ball of radius $r_\Lambda^{eff}(n)$ in $\mathbb{R}^n$.

As in [21], we shall hold $r_\Lambda^{eff}(n)$ approximately constant as $n \to \infty$. Specifically, since $p(n)$ is prime and $k(n)$ is an integer, $r_\Lambda^{eff}(n)$ cannot be a constant. For a suitably chosen $k(n)$, it suffices to pick $p(n)$ such that $r_\Lambda^{eff}(n)$ as defined in (B-6) satisfies, for all sufficiently large n,

$$r_{min} < r_\Lambda^{eff}(n) < 2r_{min}, \tag{B-7}$$

for a constant r_{min}. By the fact that $k(n) \leq \beta n$ for some $\beta < 1$, it transpires that

for a fixed r_{min} and for all n sufficiently large there exist a prime $p(n)$ and $r_{\Lambda}^{eff}(n)$ satisfying both (B-6) and (B-7). The results of [21], restated here, give constraints on the ranges of r_{min} and $k(n)$ of the random lattice ensemble $\mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}$, with $p(n)$ appropriately picked as above, such that with probability approaching 1, a lattice code in the ensemble has full dimension and is good for covering or is exponentially good for AWGN channel coding, respectively. These constraints are summarized below.

Lemma B.1 *(Goodness for covering):* [21, Theorem 2] *For any fixed r_{min} such that $0 < r_{min} < \frac{1}{4}$ and any $k(n)$ such that $\log^2 n < k(n) \leq \beta n$, for some $\beta < 1$, let $p(n)$ and $r_{\Lambda}^{eff}(n)$ be such that both (B-6) and (B-7) are satisfied (for all n sufficiently large). Then, for such parameters $(n, p(n), k(n))$, with probability approaching 1, the random lattice code in the lattice ensemble $\mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}$ is good for covering and $\mathcal{C}$ has dimension exactly $k(n)$.*

Lemma B.2 *(Exponentially goodness for AWGN channel coding achieving the Poltyrev exponent evaluated at ρ):* [21, Theorem 4] *For any fixed r_{min} such that $0 < r_{min} < \min(\frac{\rho^2}{32E_p(\rho)}, \frac{1}{4})$ and any $k(n)$ such that $k(n) \leq \beta n$ for some $\beta < \frac{1}{2}$, let $p(n)$ and $r_{\Lambda}^{eff}(n)$ be such that both (B-6) and (B-7) are satisfied (for all n sufficiently large). Then, for such parameters $(n, p(n), k(n))$, with probability approaching 1, the random lattice code in the lattice ensemble $\mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}$ is exponentially good for AWGN channel coding and achieving the Poltyrev exponent evaluated at ρ, and $\mathcal{C}$ has dimension exactly $k(n)$.*

Lemma B.3: *For any D, R_2, R_3, $\rho_2 > 1$ and $\rho_3 > 1$, there exists a sequence*

of three level nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ *such that* $\frac{|\nu(\Lambda_3)|}{|\nu(\Lambda_2)|}^{\frac{1}{n}} = e^{R_3+o_n(1)}$ *and* $\frac{|\nu(\Lambda_2)|}{|\nu(\Lambda_1)|}^{\frac{1}{n}} = e^{R_2+o_n(1)}$. *Furthermore,* Λ_1 *is good for covering with second moment per dimension* D *and* Λ_2 *(*Λ_3*) is exponentially good for AWGN channel coding and achieving the Poltyrev exponents evaluated at* ρ_2 *(*ρ_3*), respectively.*

Proof of Lemma B.3: We shall consider the nested ensembles of lattice codes $\Lambda_1 = \mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}_1 \supset \Lambda_2 = \mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}_2 \supset \Lambda_3 = \mathbb{Z}^n + \frac{1}{p(n)}\mathcal{C}_3$, where $\mathcal{C}_1 \supset \mathcal{C}_2 \supset \mathcal{C}_3$ denote the nested uniform random $(n, k_1(n))$, $(n, k_2(n))$ and $(n, k_3(n))$ linear codes over $\mathbb{Z}_{p(n)}$, respectively. In particular, we first draw a uniform random $k_1(n) \times n$ matrix (with entries taking values uniformly and independently in $\mathbb{Z}_{p(n)}$) to be the random generating matrix of $\mathcal{C}_1$. Then the first $k_2(n)$, $k_3(n)$ rows of the random matrix constitute the random generating matrices of $\mathcal{C}_2$, $\mathcal{C}_3$, respectively. It is then left to pick $(p(n), k_1(n), k_2(n), k_3(n))$ appropriately to obtain the required nested lattice codes.

To this end, we first select $r_{min} = \min(\frac{{\rho_2}^2}{32E_P(\rho_2)}, \frac{{\rho_3}^2}{32E_P(\rho_3)}, \frac{1}{4})$. Next we select r_{min3} small enough and $r_{min1} < r_{min2} < r_{min3} < r_{min}$ so that $\frac{r_{min2}}{r_{min1}} = e^{R_2}$ and $\frac{r_{min3}}{r_{min2}} = e^{R_3}$. We then select $k_1(n)$ as growing linearly in n, say $\frac{1}{4}n$. Then, $k_2(n)$ and $k_3(n)$ are constrained by the ratios of the volume of the Voronoi regions of the three lattice codes, namely $k_2(n) = k_1(n) - \lfloor \frac{nR_2}{\log p(n)} \rfloor$ and $k_3(n) = k_2(n) - \lfloor \frac{nR_3}{\log p(n)} \rfloor$. Lastly, we shall select a prime number $p(n)$ so that $r_{min1} < r_{\Lambda_1}^{eff}(n) < 2r_{min1}$. Observe that for every large enough n, we can find a prime $p(n)$ so that $r_{min1} < r_{\Lambda_1}^{eff}(n) < 2r_{min1}$. To see this, let $p^* \in \mathbb{R}$ satisfy (B-6) for a radius $2r_{min1}$, i.e., $p^{*k_1(n)} = \frac{1}{V_{\mathcal{B}}(2r_{min1})}$. From (B-6) and by $r_{min1} < r_{\Lambda_1}^{eff}(n) < 2r_{min1}$, our claim is true if we can find a

prime $p(n) \in [p^*, 2^{\frac{n}{k_1(n)}}p^*]$. Since $k_1(n) = \frac{1}{4}n$, it follows that there exists such a prime for every large enough n, because there is a prime number between i and $2i$ for every integer i (Bertrand's postulate, see, e.g. [27]). By (B-6), the choice of $k_1(n)$ above and the fact that $r_{min1} < r_{\Lambda_1}^{eff}(n) < 2r_{min1}$, it is clear that $p(n)$ grows subexponentially in n. It then follows, for all n sufficiently large, by the manner of selection of $k_2(n)$, $k_3(n)$, r_{min2} and r_{min3} that $r_{min2} < r_{\Lambda_2}^{eff}(n) < 2r_{min2}$ and $r_{min3} < r_{\Lambda_3}^{eff}(n) < 2r_{min3}$. It is clear that $(r_{min1}, k_1(n))$ satisfy the constraints in Lemma B.1 above for Λ_1 to be good for covering and $(r_{min2}, k_2(n))$ $((r_{min3}, k_3(n)))$ satisfy the constraints in Lemma B.2 above for Λ_2, Λ_3 to be exponentially good for AWGN channel coding and achieving the Poltyrev exponent evaluated at ρ_2, ρ_3, respectively. Specifically, by Lemma B.1 and Lemma B.2 above, we have that the events

$$A = \{\Lambda_1 \text{ is good for covering and } dim(\mathcal{C}_1) = k_1(n)\},$$

$$B = \left\{ \begin{array}{l} \Lambda_2 \text{ is good for AWGN channel coding} \\ \\ \text{achieving } E_P(\rho_2) \text{ and } dim(\mathcal{C}_2) = k_2(n) \end{array} \right\}$$

and

$$C = \left\{ \begin{array}{l} \Lambda_3 \text{ is good for AWGN channel coding} \\ \\ \text{achieving } E_P(\rho_3) \text{ and } dim(\mathcal{C}_3) = k_3(n) \end{array} \right\}$$

satisfy $Pr\{A\} = 1 - o_n(1)$, $Pr\{B\} = 1 - o_n(1)$ and $Pr\{C\} = 1 - o_n(1)$, respectively. Consequently,

$$Pr\{A \cap B \cap C\} = 1 - Pr\{A^c \cup B^c \cup C^c\} > 1 - o_n(1).$$

Therefore, there exists a sequence of three level nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ such that Λ_1 is good for covering, and Λ_2, Λ_3 are exponentially good for AWGN

channel coding and achieving the Poltyrev exponents evaluated at ρ_2, ρ_3, respectively. The claims regarding the ratio of volume of the Voronoi region of Λ_3 to that of Λ_2 and the ratio of volume of the Voronoi region of Λ_2 to that of Λ_1 follow from $k_2(n) - k_3(n) = \lfloor \frac{nR_3}{\log p(n)} \rfloor$, $k_1(n) - k_2(n) = \lfloor \frac{nR_2}{\log p(n)} \rfloor$ and the fact that $p(n)$ grows subexponentially, respectively. Lastly, we shall scale all lattice codes so that the second moment per dimension of Λ_1 is D. ■

Now, returning to Lemma 3.7, we have the following Lemma.

Lemma B.4: *For $R > 0$ and an arbitrary but fixed $D > 0$, let α be selected as in (3.43). Then, for any $P > \alpha^2\sigma_Z^2 + D$ and any $Q > \alpha^2\sigma_{X_1}^2 + D$, there exists a sequence of nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ such that $\sigma^2(\Lambda_1) = D$, Λ_1 is good for covering,*

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{|\nu(\Lambda_3)|}{|\nu(\Lambda_1)|} = \frac{1}{2} \log Q/D, \tag{B-8}$$

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{|\nu(\Lambda_2)|}{|\nu(\Lambda_1)|} = \frac{1}{2} \log P/D, \tag{B-9}$$

$$Pr\{\alpha\mathbf{Z} - \mathbf{U} \notin \nu(\Lambda_2)\} \ \to 0 \text{ exponentially in } n \tag{B-10}$$

and

$$Pr\{\alpha\mathbf{X}_1 - \mathbf{U} \notin \nu(\Lambda_3)\} \ \to 0 \text{ exponentially in } n. \tag{B-11}$$

Proof of Lemma B.4: Let $r > 1$ be sufficiently close to 1 such that $\frac{P}{r^2} > \alpha^2\sigma_Z^2 + D$ and $\frac{Q}{r^2} > \alpha^2\sigma_{X_1}^2 + D$. By Lemma B.3, there exists a sequence of nested

lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$ such that

$$\frac{|\nu(\Lambda_2)|}{|\nu(\Lambda_1)|} = e^{\frac{n}{2}(\log \frac{P}{D} + o_n(1))};\ \frac{|\nu(\Lambda_3)|}{|\nu(\Lambda_2)|} = e^{\frac{n}{2}(\log \frac{Q}{P} + o_n(1))}, \tag{B-12}$$

Λ_1 is good for covering with second moment per dimension D, and Λ_2, Λ_3 are exponentially good for AWGN channel coding and achieving the Poltyrev exponent evaluated at ρ_2, ρ_3, respectively, where

$$\rho_2 = \sqrt{\frac{P}{r^2(\alpha^2\sigma_Z^2 + D)}}, \qquad \rho_3 = \sqrt{\frac{Q}{r^2(\alpha^2\sigma_{X_1}^2 + D)}}. \tag{B-13}$$

It is then left to prove (B-10) and (B-11) for the sequence of nested lattice codes $\Lambda_1 \supset \Lambda_2 \supset \Lambda_3$.

First, we claim that

$$\frac{2}{n}\log|\nu(\Lambda_1)| = \log(2\pi e)D + o_n(1). \tag{B-14}$$

The normalized second moment of $\nu(\Lambda_1)$, denoted by $G(\nu(\Lambda_1))$, is defined as (see, e.g., [9])

$$G(\nu(\Lambda_1)) \triangleq \frac{\sigma^2(\Lambda_1)}{|\nu(\Lambda_1)|^{2/n}}. \tag{B-15}$$

It is known that the normalized second moment is invariant under scaling and that the normalized second moment of a sphere, denoted by G_n^*, converges to $\frac{1}{2\pi e}$ as $n \to \infty$ (see, e.g. [9]). By the fact that Λ_1 is good for covering, we have that (see [21, Proposition 1])

$$G(\nu(\Lambda_1)) \to \frac{1}{2\pi e}. \tag{B-16}$$

Then (B-14) follows from $\sigma^2(\Lambda_1) = D$, (B-15) and (B-16).

The following lemma, from Lemma 6 and 11 in [22], gives upper bounds for the two probabilities (B-10) and (B-11) in terms of those for i.i.d. Gaussian rvs with asymptotically equal variances per source symbol.

Lemma B.5: *If Λ_1 is good for covering and $\sigma^2(\Lambda_1) = D$, then there exists ϵ_1^n and ϵ_2^n depending only on Λ_1 and going to 0 and 1 in n, respectively such that*

$$Pr\{\alpha\mathbf{Z} - \mathbf{U} \notin \nu(\Lambda_2)\} \leq Pr\{\hat{\mathbf{Z}} \notin \nu(\Lambda_2)\}e^{n\epsilon_1^n}$$

and

$$Pr\{\alpha\mathbf{X}_1 - \mathbf{U} \notin \nu(\Lambda_3)\} \leq Pr\{\hat{\mathbf{X}}_1 \notin \nu(\Lambda_3)\}e^{n\epsilon_1^n},$$

where $\hat{\mathbf{Z}}$ and $\hat{\mathbf{X}}_1$ are n i.i.d. repetitions of Gaussian rvs with mean 0 and variances $\epsilon_2^n(\alpha^2\sigma_Z^2 + D)$ and $\epsilon_2^n(\alpha^2\sigma_{X_1}^2 + D)$, respectively. Specifically,

$$\frac{n}{n+2} \leq \epsilon_1^n < \left(\frac{r_{\Lambda_1}^{cov}(n)}{r_{\Lambda_1}^{eff}(n)}\right)^2$$

and

$$\epsilon_2^n = \log\left(\frac{r_{\Lambda_1}^{cov}(n)}{r_{\Lambda_1}^{eff}(n)}\right) + \frac{1}{2}\log 2\pi e G_n^* + \frac{1}{n}.$$

Let $\tilde{\rho}_2$ denote the the ratio of the effective radius of Λ_2 to the approximated radius of the Gaussian rv $\hat{\mathbf{Z}}$ ($\sqrt{n\epsilon_2^n(\alpha^2\sigma_Z^2 + D)}$), i.e.,

$$\tilde{\rho}_2 = \frac{|\nu(\Lambda_2)|^{\frac{1}{n}}}{V_{\mathcal{B}}(1)^{\frac{1}{n}}\sqrt{n\epsilon_2^n(\alpha^2\sigma_Z^2 + D)}} = \frac{|\nu(\Lambda_2)|^{\frac{1}{n}}t_n}{\sqrt{(2\pi e)\epsilon_2^n(\alpha^2\sigma_Z^2 + D)}}; \tag{B-17}$$

and let $\tilde{\rho}_3$ denote the ratio of the radius of Λ_3 to the approximated radius of the Gaussian rv $\hat{\mathbf{X}}_1$ ($\sqrt{n\epsilon_2^n(\alpha^2\sigma_{X_1}^2 + D)}$), i.e.,

$$\tilde{\rho}_3 = \frac{|\nu(\Lambda_3)|^{\frac{1}{n}}t_n}{\sqrt{(2\pi e)\epsilon_2^n(\alpha^2\sigma_{X_1}^2 + D)}}, \tag{B-18}$$

where $V_{\mathcal{B}}(1)$ is the volume of the ball in $\mathbb{R}^n$ of radius 1 and, hence, from (B-6) $t_n = \sqrt{\frac{2\pi e}{n}} \left(\frac{\Gamma(n/2+1)}{\pi^{n/2}}\right)^{1/n} \to 1$ (see, e.g. [9]).

Using (B-14), it follows from (B-12) that $|\nu(\Lambda_2)| = e^{\frac{n}{2}(\log(2\pi e)P + o_n(1))}$ and $|\nu(\Lambda_3)| = e^{\frac{n}{2}(\log(2\pi e)Q + o_n(1))}$. Consequently, for all n sufficiently large, we get from (B-17) and (B-18) that

$$\tilde{\rho}_2 > \rho_2 = \sqrt{\frac{P}{r^2(\alpha^2\sigma_Z^2 + D)}} > 1 \tag{B-19}$$

and, from (B-18) and (B-18), that

$$\tilde{\rho}_3 > \rho_3 = \sqrt{\frac{Q}{r^2(\alpha^2\sigma_{X_1}^2 + D)}} > 1. \tag{B-20}$$

Let $\hat{\hat{\mathbf{Z}}}$ and $\hat{\hat{\mathbf{X}}}_1$ be n i.i.d. repetitions of Gaussian rvs with mean 0 and variances $r^2(\alpha^2\sigma_Z^2 + D)$ and $r^2(\alpha^2\sigma_{X_1}^2 + D)$, respectively. From Lemma B.5, (B-19), (B-20) and the fact that Λ_2, Λ_3 are exponentially good for AWGN channel coding and achieving the Poltyrev exponent evaluated at ρ_2, ρ_3, respectively, for all n sufficiently large,

$$Pr\{\alpha\mathbf{Z} - \mathbf{U} \notin \nu(\Lambda_2)\} \leq Pr\{\hat{\hat{\mathbf{Z}}} \notin \nu(\Lambda_2)\} \leq e^{-n(E_P(\rho_2) - o_n(1))};$$

$$Pr\{\alpha\mathbf{X}_1 - \mathbf{U} \notin \nu(\Lambda_3)\} \leq Pr\{\hat{\hat{\mathbf{X}}}_1 \notin \nu(\Lambda_3)\} \leq e^{-n(E_P(\rho_3) - o_n(1))},$$

thereby establishing (B-10) and (B-11). ∎

All assertions in Lemma 3.7 except for (3.55) follow from Lemma B.4 by noting from (3.44) and (3.50) that

$$R = \frac{1}{2}\log\frac{\alpha^2\sigma_{X_1}^2 + D}{D}, \qquad R_p = \frac{1}{2}\log\frac{\alpha^2\sigma_Z^2 + D}{D}.$$

To see (3.55), note that we have shown that $\log\frac{|\nu(\Lambda_1)|^{\frac{2}{n}}}{(2\pi e)D}$ tends to 0 in n. Consequently,

$$r_{\Lambda_1}^{eff}(n) = O(\frac{\sqrt{(2\pi e)D}}{V_{\mathcal{B}}(1)^{\frac{1}{n}}}) = O(\sqrt{nD}).$$

By the fact that Λ_1 is good for covering, $\frac{r^{cov}_{\Lambda_1}(n)}{r^{eff}_{\Lambda_1}(n)} \to 1$, $r^{cov}_{\Lambda_1}(n) = O(\sqrt{nD})$.

B.3 Proof of (3.57)

Denoting $\mathcal{A} = \{(\alpha\mathbf{X}_1 - \mathbf{E}) \bmod \Lambda_3 = \alpha\mathbf{X}_1 - \mathbf{E}\}$, we have that $Pr\{\mathcal{A}\} = 1 - o_n(1)$ by (3.56). Then

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{X}_1 - \hat{\mathbf{X}}_1\|^2\right] &= \mathbb{E}\left[\|\mathbf{X}_1 - c((\alpha\mathbf{X}_1 - \mathbf{E}) - Q_{\Lambda_3}(\alpha\mathbf{X}_1 - \mathbf{E}))\|^2\right] \\
&\leq \mathbb{E}\left[\|(1-c\alpha)\mathbf{X}_1 + c\mathbf{E}\|^2\right] + \\
&\quad \mathbb{E}\left[\|cQ_{\Lambda_3}(\alpha\mathbf{X}_1 - \mathbf{E}))\|^2\right] \\
&= \mathbb{E}\left[\|(1-c\alpha)\mathbf{X}_1 + c\mathbf{E}\|^2\right] + \\
&\quad \mathbb{E}\left[\|cQ_{\Lambda_3}(\alpha\mathbf{X}_1 - \mathbf{E}))\|^2\ 1_{(\mathcal{A}^c)}\right] \\
&\leq \mathbb{E}\left[\|(1-c\alpha)\mathbf{X}_1 + c\mathbf{E}\|^2\right] + \\
&\quad \sqrt{\mathbb{E}\left[\|cQ_{\Lambda_3}(\alpha\mathbf{X}_1 - \mathbf{E})\|^4\right] Pr(1_{(\mathcal{A}^c)})} \\
&= n\left[(1-c\alpha)^2\sigma^2_{X_1} + c^2 D\right] + \\
&\quad \sqrt{\mathbb{E}\left[\|cQ_{\Lambda_3}(\alpha\mathbf{X}_1 - \mathbf{E})\|^4\right]}\sqrt{Pr(1_{(\mathcal{A}^c)})},
\end{aligned}$$

where the second equality is by that fact that in $\mathcal{A}$, $Q_{\Lambda_3}(\alpha\mathbf{X}_1 - \mathbf{E}) = \mathbf{0}$; the two inequalities above are by the triangle inequality and the Cauchy-Schwarz inequality, respectively. Next,

$$\sqrt{\mathbb{E}\left[\|cQ_{\Lambda_3}(\alpha\mathbf{X}_1 - \mathbf{E})\|^4\right]}$$

$$
\begin{aligned}
&= c^2\sqrt{\mathbb{E}\left[\|(\alpha\mathbf{X}_1-\mathbf{E})-(\alpha\mathbf{X}_1-\mathbf{E}) \bmod \Lambda_3\|^4\right]} \\
&\leq c^2\left[\mathbb{E}\left[\|(\alpha\mathbf{X}_1-\mathbf{E})\|^4\right]^{\frac{1}{4}}+\mathbb{E}\left[\|(\alpha\mathbf{X}_1-\mathbf{E}) \bmod \Lambda_3\|^4\right]^{\frac{1}{4}}\right]^2 \\
&\leq c^2\left[\mathbb{E}\left[\|(\alpha\mathbf{X}_1-\mathbf{E})\|^4\right]^{\frac{1}{4}}+\mathbb{E}\left[\|(\alpha\mathbf{X}_1-\mathbf{E})\|^4\right]^{\frac{1}{4}}\right]^2 \\
&\leq 4c^2\left[\mathbb{E}\left[\|(\alpha\mathbf{X}_1-\mathbf{E})\|^4\right]^{\frac{1}{4}}\right]^2 \\
&\leq 4c^2\left[\mathbb{E}\left[\|\alpha\mathbf{X}_1\|^4\right]^{\frac{1}{4}}+\mathbb{E}\left[\|\mathbf{E}\|^4\right]^{\frac{1}{4}}\right]^2 \\
&= \left[O(n^2)^{\frac{1}{4}}+O(n^2)^{\frac{1}{4}}\right]^2 = O(n),
\end{aligned}
$$

where the first and the last inequalities are by Minkowski's inequality; the second inequality is from

$$\|(\alpha\mathbf{X}_1-\mathbf{E}) \bmod \Lambda_3\| \leq \|(\alpha\mathbf{X}_1-\mathbf{E})\|;$$

the last equality is a consequence of the components of $\mathbf{X}_1$ being i.i.d. Gaussian rvs and $\mathbf{E}$ taking values in $\nu(\Lambda_1)$ that is covered by a ball of radius $O(\sqrt{n})$ in (3.55). Consequently, we have that

$$\mathbb{E}\left[\|\mathbf{X}_1-\hat{\mathbf{X}}_1\|^2\right] \leq n(\left[(1-c\alpha)^2\sigma_{X_1}^2+c^2D\right]+o_n(1)).$$

Appendix C

Appendix for Chapter 4

C.1 Proof of Lemma 4.1

We prove a slightly stronger result that there exists an LC whose null space comprises only the all-zero and the all-one strings (corresponding to the edges in T being labelled all zero or all one) which clearly enables every terminal in V to recover all the edges of T. We prove the claim by induction. When $|T| = 2$, say, with $T = \{e_1 = (v_1, v_2), e_2 = (v_2, v_3)\}$, then $e_1 + e_2$ mod 2 constitutes an LC whose null space is $\{(00), (11)\}$. Next, suppose the claim is true for all trees with $k-1$ edges, $k \geq 3$. Given a tree with k edges, pick an end vertex v_{k+1} of the tree (a vertex with degree one), and let v_k be the sole vertex connecting to v_{k+1}. Then $G = (V, T' \bigcup \{(v_k, v_{k+1})\})$, and $G' = (V \backslash \{v_{k+1}\}, T')$ is a subtree of G. By the induction hypothesis, there exists an LC for G', say, $F(T')$ of length $k-2$ (bits) and whose null space is $\{\mathbf{0}^{k-1}, \mathbf{1}^{k-1}\}$. Let v_{k-1} be another vertex connecting to v_k and let $e_{k-1} = (v_{k-1}, v_k)$ and $e_k = (v_k, v_{k+1})$. Then, consider $\{F(T'), e_{k-1} + e_k\}$ as an LC of G of length $k-1$. It is now clear that the null space of this LC is $\{\mathbf{0}^k, \mathbf{1}^k\}$.

C.2 Proof of Claim in (The Proof of) Theorem 4.8

(a) Let $G_A = (A, E_A)$ denote a subgraph of G in A, where $E_A \subset E$ consists only of those edges in E whose both end vertices lie in A. Clearly,

$$\begin{aligned} |E| - INT_G(A) &\geq \mu(A, G) \geq \mu(A, G_A) \\ &= |E_A| - INT_{G_A}(A), \\ &\qquad \text{by Corollary 3 with } \mathcal{M} = A \\ &= |E| - (d_m + INT_{G_A}(A)) \,. \end{aligned}$$

Thus, it suffices to show that

$$d_m + INT_{G_A}(A) \leq INT_G(A). \tag{C-1}$$

Consider first the case where $(I_1^*, \ldots, I_{m-1}^*, 0)$ attains $INT_G(A)$. Without loss of generality, let $\{1, \ldots, a\},\ a \leq m-1$, be the set of vertices in A connecting to m. For any $v \in \{1, \ldots, a\}$, since $\{v, m\} \not\supseteq A$, we have that $I_v^* + I_m^* = I_v^* \geq e_{vm}$ (see (4.10)). Consequently, since $d_m = \sum_{u\,=\,1}^{a} e_{um}$, we see that $(I_1^* - e_{1m}, \ldots, I_a^* - e_{am}, I_{a+1}^*, \ldots, I_{m-1}^*)$, with components summing to $INT_G(A) - d_m$ is feasible for $INT_{G_A}(A)$. Thus, $INT_G(A) - d_m \geq INT_{G_A}(A)$, establishing (C-1). A nearly identical argument would show that (C-1) holds too for the case when at most vertex 1 is connected to m, and is omitted.

(b) Consider any $G^{uv} = (\mathcal{M}, E^{uv})$ as in the second paragraph of the proof of Theorem 4.8, and let $(I_1^{**}, \ldots, I_m^{**})$ attain $INT_{G^{uv}}(A)$. Then, $(I_1^{**}, \ldots, I_{m-1}^{**}, I_m^{**} + 1)$ is feasible for $INT_G(A)$, so that

$$INT_G(A) \leq INT_{G^{uv}}(A) + 1. \tag{C-2}$$

Without loss of generality, let $\{1,\ldots,a\}$ be as in the proof of Claim (a). To prove Claim (b), it suffices to show for $u=1$ that there exists $v\in\{2,\ldots,a\}$ such that $(I_1^*,\ldots,I_{m-1}^*,I_m^*-1)$ is feasible for $INT_{G^{1v}}(A)$ if $0<I_m^*\le\lfloor\frac{d_m}{2}\rfloor$. This would mean that

$$INT_G(A)-1\ge INT_{G^{1v}}(A). \tag{C-3}$$

which, together with the observation that $|E|-1=|E^{1v}|$, establishes Claim (b). To this end, referring to (4.10), for $B\subseteq\mathcal{M}$, set

$$e_G(B)\triangleq\sum_{1\le i<j\le m,\ i\in B,\ j\in B}e_{ij},\quad e_G(\emptyset)\triangleq 0, \tag{C-4}$$

and let

$$\mathcal{B}=\left\{\begin{array}{l} B,\ \emptyset\neq B\subset\mathcal{M},\ B\not\supseteq A,\\ \sum_{i\in B}I_i^*=e_G(B)\end{array}\right\}. \tag{C-5}$$

We make the following

Claim (d): For $u=1$, there exists $v\in\{2,\ldots,a\}$ connecting to m with the properties that

a) for $B\in\mathcal{B}$ such that $1\notin B,\ m\in B$, it holds that $v\in B$;

b) for $B\in\mathcal{B}$ such that $1\in B,\ m\notin B$, it holds that $v\notin B$.

Then, with the choice of v as in the Claim (d), a simple check of all the possibilities for B (in $\mathcal{B}$ or in $\mathcal{B}^c$) that are feasible in (4.10), shows that $(I_1^*,\ldots,I_{m-1}^*,I_m^*-1)$ is feasible for $INT_{G^{1v}}(A)$, thereby establishing (C-3) (and hence Claim (b)).

It only remains to establish Claim (d). We first state the following facts with accompanying proofs.

Fact 1: For $B_1,B_2\subset\mathcal{M},\ e_G(B_1)+e_G(B_2)\le e_G(B_1\cup B_2)+e_G(B_1\cap B_2)$.

This holds by observing that $e_G(B_1 \cup B_2) + e_G(B_1 \cap B_2) - e_G(B_1) - e_G(B_2) = \sum_{1 \le i < j \le m,\ i \in B_1 \setminus B_2,\ j \in B_2 \setminus B_1 \text{ or } i \in B_2 \setminus B_1,\ j \in B_1 \setminus B_2} e_{ij} \ge 0$.

Fact 2: For $B_1, B_2 \in \mathcal{B}$ with $B_1 \cup B_2 \not\supseteq A$, it holds that $B_1 \cup B_2$ and $B_1 \cap B_2$ are both in $\mathcal{B}$. To see this, note first that

$$\begin{aligned}
\sum_{i \in B_1 \cup B_2} I_i^* &= \sum_{i \in B_1} I_i^* + \sum_{i \in B_2} I_i^* - \sum_{i \in B_1 \cap B_2} I_i^* \\
&= e_G(B_1) + e_G(B_2) - \sum_{i \in B_1 \cap B_2} I_i^* \\
&\le e_G(B_1) + e_G(B_2) - e_G(B_1 \cap B_2) \\
&\le e_G(B_1 \cup B_2), \quad \text{by Fact 1.}
\end{aligned}$$

Also, $\sum_{B_1 \cup B_2} I_i^* \ge e_G(B_1 \cup B_2)$, since $B_1 \cup B_2 \not\supseteq A$ is feasible in (4.10). The fact follows.

Fact 3: For $B \subseteq \mathcal{M}$, let $D_m(B)$ denote the total number of edges connecting m to all the vertices in $B \cap A$. Then, for $B \in \mathcal{B}$, if $m \in B$ then $D_m(B) \ge I_m^*$, and if $m \notin B$ then $D_m(B) \le I_m^*$. To see this, consider first the case $m \in B \in \mathcal{B}$. As $\{m\} \notin \mathcal{B}$ (since $I_m^* > 0$), we have $B \cap A \neq \emptyset$. Since $B \in \mathcal{B}$, $\sum_{i \in B} I_i^* = e_G(B \cap A) + D_m(B)$. Also, since $B \cap A \neq \emptyset$ is feasible in (4.10), $\sum_{i \in B \cap A} I_i^* \ge e_G(B \cap A)$. Subtracting the latter from the former gives $I_m^* \le D_m(B)$. The second assertion of the fact is proved similarly.

Fact 4: The intersection of all Bs in $\mathcal{B}$ satisfying $1 \notin B,\ m \in B$, when nonempty, is also in $\mathcal{B}$. The union of all Bs in $\mathcal{B}$ satisfying $1 \in B,\ m \notin B$, when nonempty, is also in $\mathcal{B}$.

The first assertion in Fact 4 is obtained by observing that the union of all Bs in $\mathcal{B}$ with $1 \notin B,\ m \in B$, does not contain A, and by a repeated use of Fact 2. The

second assertion would follow similarly by Fact 2 if the union of all Bs in $\mathcal{B}$ with $1 \in B,\ m \notin B$, is strictly contained in A. Suppose not; then this union is exactly A. The ensuing contradiction can be seen, for instance, with B_1, B_2 as above with $B_1 \cup B_2 = A$. Then

$$\begin{aligned}
d_m &= D_m(A) = D_m(B_1 \cup B_2) \\
&= D_m((B_1 \backslash B_2) \cup (B_1 \cap B_2) \cup (B_2 \backslash B_1)) \\
&= D_m(B_1 \backslash B_2) + D_m(B_1 \cap B_2) + D_m(B_2 \backslash B_1) \\
&= D_m(B_1) + D_m(B_2) - D_m(B_1 \cap B_2) \\
&\leq I_m^* + I_m^* - 1, \text{ by Fact 3 and } 1 \in B_1 \cap B_2 \\
&\leq 2\lfloor \frac{d_m}{2} \rfloor - 1, \text{ by the assumption } I_m^* \leq \lfloor \frac{d_m}{2} \rfloor \\
&< d_m,
\end{aligned}$$

a contradiction.

Finally, to prove Claim (d), let B' (resp. B'') represent the intersection (resp. union), when nonempty, in Fact 4. It suffices now to show that there exists $v \in B' \cap A$ (when $B' \neq \emptyset$) such that $v \notin B''$ and v connects to m; this follows from

$$\begin{aligned}
D_m(B' \backslash B'') &= D_m(B') - D_m(B' \cap B'') \\
&= D_m(B') - (D_m(B'') - D_m(B'' \backslash B')) \\
&\geq I_m^* - (I_m^* - 1), \text{ by Fact 3 and } 1 \in B'' \backslash B' \\
&= 1.
\end{aligned}$$

Then, any B as in Claim (d)(a) must contain B' and hence the v above. On the other hand, any B as in Claim (d)(b) must be contained in B'' and so cannot contain

the v above. The cases $B' = \emptyset$ or $B'' = \emptyset$ are handled trivially.

(c) Let $G^{uv} = (\mathcal{M}, E^{uv})$ and suppose that $T_1 \sqcup \ldots \sqcup T_k \subseteq E^{uv}$ attain $\mu(A, G^{uv})$. If $E^{uv} \backslash \{\sqcup_{i=1}^{k} T_i\}$ contains at least one edge connecting (u, v), then $\{T_1, \ldots, T_k\}$ is also a Steiner tree packing of $G = (\mathcal{M}, E)$, so that $\mu(A, G) \geq \mu(A, G^{uv})$. Else, let T_1, say, be the Steiner tree that contains an edge connecting u, v that emerged by splitting off (u, m) and (v, m) of $G = (\mathcal{M}, E)$. Then, $\{T_1 \backslash \{(u, v)\}\} \cup \{(u, m), (v, m)\}$ is A-connected and hence contains a Steiner tree T_1' for A in $G = (\mathcal{M}, E)$ that corresponds to T_1; clearly, again $\mu(A, G) \geq \mu(A, G^{uv})$.

C.3 Strong SK for the PIN Model

We shall be concerned in this appendix with a variant of the PIN model in Chapter 4. Suppose that terminals $1, \ldots, m,\ m \geq 2$, observe n independent and identically distributed (i.i.d.) repetitions of the rvs $X_1, \ldots, X_m$, denoted by $X_1^n, \ldots, X_m^n$, where $X_i^n = (X_{i,1}, \ldots, X_{i,n}),\ i \in \mathcal{M} = \{1, \ldots, m\}$. Each rv $X_i,\ i \in \mathcal{M}$, is of the form $X_i = (Y_{ij},\ j \in \mathcal{M} \backslash \{i\})$ with $m-1$ components, and the "reciprocal pairs" of rvs $\{(Y_{ij}, Y_{ji}),\ 1 \leq i < j \leq m\}$ are mutually independent.[1] See Figure 2. Thus, every pair of terminals in $\mathcal{M}$ is associated with a corresponding pair of rvs that are independent of pairs of rvs associated with all the other pairs of terminals. All the rvs are assumed to take their values in finite sets.

The overall goal is to generate a strong SK for a given set $A \subseteq \mathcal{M}$ of terminals at the largest rate possible, with the remaining terminals (if any) cooperating in

[1] *Unlike in Chapter 4, we do not require that* $Y_{ji} = Y_{ji},\ 1 \leq i < j \leq m$.

secrecy generation.

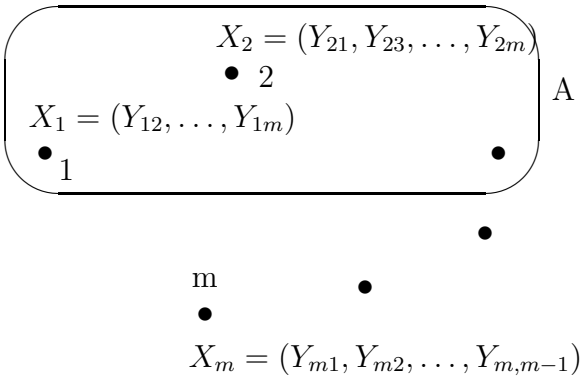

Figure 2: The PIN Model for Appendix C.3

C.3.1 Results

All references below to a "PIN model" are to the model in this appendix. Our main results are the following. First, we obtain, upon particularizing the results of [15], a (single-letter) expression for $C(A)$ for a PIN model, in terms of a linear combination of mutual information terms that involve only pairs of "reciprocal" rvs $\{(Y_{ij}, Y_{ji}),\ 1 \leq i \neq j \leq m\}$. Second, stemming from this observation, a connection is drawn between SK generation for the PIN model and the combinatorial problem of maximal packing of Steiner trees in an associated multigraph. Specifically, we show that the maximum rate of Steiner tree packing in the multigraph is always a lower bound for SK capacity. Third, for the case $|A| = 2$ (when the Steiner tree becomes a path connecting the two vertices in A) and for the case $A = \mathcal{M}$ (when the Steiner tree becomes a spanning tree), the previous lower bound is shown to be tight. This is done by means of an explicit algorithm, based on maximal path

packing and maximal spanning tree packing, respectively, that forms an SK out of independent SKs for pairs of terminals. In fact, the maximum rate of the SK thereby generated equals the previously known upper bound for SK capacity [15] mentioned above.

C.3.1.1 Strong SK Capacity

In order to state our result, we recall the notation of Section 3.1.

Proposition C.1: For a PIN model, the SK capacity for a set of terminals $A \subseteq \mathcal{M}$, with $|A| \geq 2$, is

$$C(A) \;=\; \min_{\lambda \in \Lambda(A)} \left[\sum_{1 \leq i < j \leq m} \left(\sum_{\substack{B \in \mathcal{B}(A): \\ i \in B,\, j \in B^c}} \lambda_B \right) I(Y_{ij} \wedge Y_{ji}) \right]. \tag{C-6}$$

Remarks: (i) It is of interest in (C-6) that the SK capacity for a PIN model depends on the joint probability distribution of the underlying rvs only through a linear combination of the pairwise reciprocal mutual information terms.

(ii) We note from [15, Theorem 3] that additional independent randomization at the terminals in $\mathcal{M}$, enabled by giving them access to the mutually independent rvs $M_1, \ldots, M_m$, respectively, that are independent also of $(X_1^n, \ldots, X_m^n)$, does not serve to enhance SK capacity. Heuristically speaking, the mentioned independence of the randomization forces any additional "common randomness" among the terminals in A to be acquired only through public communication, which is observed fully by the eavesdropper. On the other hand, randomization can serve to enhance secrecy generation for certain models (cf. e.g., [51])

Proof: The proof entails an application of the formula for SK capacity in [15, 16] to the PIN model. For $B \in \mathcal{B}(A)$, denote $X_B = (X_i,\ i \in B)$. From ([16, Theorem 3.1],

$$C(A) \ = \ H\left(X_1, \ldots, X_m\right) - \max_{\lambda \in \Lambda(A)} \sum_{B \in \mathcal{B}(A)} \lambda_B H\left(X_B | X_{B^c}\right). \tag{C-7}$$

For the PIN model, since $X_i = (Y_{ij},\ j \ \in \ \mathcal{M}\backslash\{i\})$, we observe in (C-7) that

$$\begin{aligned} H(X_1, \ldots, X_m) \ &= \ H\left(\{(Y_{ij}, Y_{ji})\}_{1 \le i < j \le m}\right) \\ &= \sum_{1 \le i < j \le m} H(Y_{ij}, Y_{ji}) \end{aligned} \tag{C-8}$$

and

$$\begin{aligned} H(X_B | X_{B^c}) \ &= \ H(X_{\mathcal{M}}) - H(X_{B^c}) \\ &= \sum_{1 \le i < j \le m} H(Y_{ij}, Y_{ji}) - \sum_{\substack{1 \le i < j \le m, \\ i \in B^c,\, j \in B^c}} H(Y_{ij}, Y_{ji}) \\ &\quad - \sum_{i \in B^c,\, j \in B} H(Y_{ij}) \\ &= \sum_{\substack{1 \le i < j \le m, \\ i \in B,\, j \in B}} H(Y_{ij}, Y_{ji}) + \sum_{i \in B,\, j \in B^c} H(Y_{ij} | Y_{ji}). \end{aligned} \tag{C-9}$$

A straightforward manipulation of (C-7), using (C-8), (C-9), gives

$$\begin{aligned} C(A) = \min_{\lambda \in \Lambda(A)} \sum_{1 \le i < j \le m} \Bigg[& H\left(Y_{ij}, Y_{ji}\right) - \left(\sum_{\substack{B \in \mathcal{B}(A): \\ i \in B,\, j \in B}} \lambda_B \right) H\left(Y_{ij}, Y_{ji}\right) \\ & - \left(\sum_{\substack{B \in \mathcal{B}(A): \\ i \in B,\, j \in B^c}} \lambda_B \right) H\left(Y_{ij} | Y_{ji}\right) - \left(\sum_{\substack{B \in \mathcal{B}(A): \\ i \in B^c,\, j \in B}} \lambda_B \right) H\left(Y_{ji} | Y_{ij}\right) \Bigg]. \end{aligned}$$

Since by (2),

$$\sum_{\substack{B\in\mathcal{B}(A):\\ i\in B,\, j\in B}} \lambda_B = 1 - \sum_{\substack{B\in\mathcal{B}(A):\\ i\in B,\, j\in B^c}} \lambda_B = 1 - \sum_{\substack{B\in\mathcal{B}(A):\\ i\in B^c,\, j\in B}} \lambda_B,$$

we get

$$C(A) \;=\; \min_{\lambda\in\Lambda(A)} \left[\sum_{1\le i<j\le m} \left(\sum_{\substack{B\in\mathcal{B}(A):\\ i\in B,\, j\in B^c}} \lambda_B \right) \left(\begin{array}{l} H(Y_{ij}, Y_{ji}) \\ -H(Y_{ij}|Y_{ji}) \\ -H(Y_{ji}|Y_{ij}) \end{array} \right) \right],$$

thereby completing the proof. ■

An upper bound had been established for SK capacity for a general multi-terminal source model [15, Example 4]. This bound was expressed in terms of the (Kullback-Leibler) divergence between the joint distribution of the rvs defining the underlying correlated sources and the product of the (marginal) distributions associated with appropriate partitions of these rvs, thereby measuring the minimum mutual dependence among the latter. The bound was particularized to the PIN model in [57], and is restated below in a slightly different form that will be used subsequently.

Let $\mathcal{P}$ be a partition of $\mathcal{M} = \{1, \ldots, m\}$, and denote the number of atoms of $\mathcal{P}$ by $|\mathcal{P}|$.

Lemma C.2 [57]: The SK capacity $C(A),\ A \subseteq \mathcal{M}$, for the PIN model is bounded above according to

$$C(A) \;\leq\; C^{ub}(A) \triangleq \min_{\mathcal{P}} \left(\frac{1}{|\mathcal{P}|-1}\right) \left[\sum_{\substack{1\leq i<j\leq m \\ (i,j) \text{ crosses } \mathcal{P}}} I(Y_{ij} \wedge Y_{ji}) \right], \qquad \text{(C-10)}$$

where for a fixed $\mathcal{P}$, a pair of indices (i,j) crosses $\mathcal{P}$ if i and j are in different atoms of $\mathcal{P}$. The minimization in the right side of (C-10) is over all partitions $\mathcal{P}$ of $\mathcal{M}$ for which every atom of $\mathcal{P}$ intersects A.

C.3.1.2 SK Capacity and Steiner Tree Packing

There exists a natural connection between SK generation for the PIN model and the combinatorial problem of Steiner tree packing in an associated multigraph.

We note that when $|A| = 2$, a Steiner tree for A always contains a path connecting the two vertices in A. Clearly, it suffices to take $\mu(A,G)$ to be the maximum number of edge disjoint paths connecting the two terminals in A.

Next, assume without any loss of generality in the PIN model that all pairwise reciprocal mutual information values $I(Y_{ij} \wedge Y_{ji}),\ 1 \leq i \neq j \leq m$, are rational numbers. Let $\mathcal{N}$ denote the collection of positive integers n such that the number of edges between any pair of vertices i, j is equal to $nI(Y_{ij} \wedge Y_{ji})$ is integer-valued for all $1 \leq i \neq j \leq m$; clearly, the elements of $\mathcal{N}$ form an arithmetic progression. For a PIN model, consider a sequence of associated multigraphs $\{G^{(n)} = \left(\mathcal{M}, E^{(n)}\right),\ n \in \mathcal{N}\}$, where $E^{(n)},\ n \in \mathcal{N}$, is such that $e_{ij} = nI(Y_{ij} \wedge Y_{ji})$. We term $\sup_{n\in\mathcal{N}} \frac{1}{n}\mu(A, G^{(n)})$ as the *maximum rate of Steiner tree packing* in the multigraph $G = (\mathcal{M}, E)$. The connection between SK generation for the PIN model and Steiner tree packing is

formalized below.

Theorem C.3: For a PIN model,

(i) the SK capacity satisfies

$$C(A) \geq \sup_{n \in \mathcal{N}} \frac{1}{n} \mu(A, G^{(n)}) \tag{C-11}$$

for every $A \subseteq \mathcal{M}$;

(ii) when $|A| = 2$, the SK capacity is

$$C(A) \;\; = \;\; \sup_{n \in \mathcal{N}} \frac{1}{n} \mu(A, G^{(n)}) = C^{ub}(A). \tag{C-12}$$

Remarks: (i) The inequality in (C-11) can be strict, as shown by the example after Theorem 4.5. See also the remark following Theorem C.4 for a heuristic explanation.

(ii) An exact determination of $\mu(A, G)$ is known to be NP-hard [8]. A nontrivial upper bound for $\mu(A, G)$, similar in form to (C-10), is known [28, Paragraph 5 of Section 1]. This bound can be extended to yield an upper bound for $\sup_{n \in \mathcal{N}} \frac{1}{n} \mu(A, G^{(n)})$ which, in general, is inferior to that provided by $C(A)$ in (C-11).

Proof: (i) The proof consists of two main steps. In the first step, fix an $\epsilon > 0$ that is smaller than every positive $I(Y_{ij} \wedge Y_{ji})$, $1 \leq i < j \leq m$. Each pair of terminals i, j with $I(Y_{ij} \wedge Y_{ji}) > 0$, generates a (pairwise) SK $K_{ij} = K_{ij}^{(n)}$ of size $\lfloor n(I(Y_{ij} \wedge Y_{ji}) - \epsilon) \rfloor$ bits, using public communication $F_{ij} = F_{ij}^{(n)}$, and satisfying

$$s(K_{ij}; F_{ij}) = o_n(1); \tag{C-13}$$

the existence of such an SK follows from [40]. The SK achievability scheme in [40]

consists of a "weak" SK generated by Slepian-Wolf data compression, followed by "privacy amplification" to extract a "strong" SK. Note by the definition of the PIN model that $\{(K_{ij}, F_{ij})\}_{1\leq i<j\leq m}$ are mutually independent.

In the second step, consider the sequence of multigraphs $\left\{G_\epsilon^{(n)} = (\mathcal{M}, \widetilde{E^{(n)}})\right\}_{n=1}^{\infty}$, where $\widetilde{E^{(n)}}$ is such that the number of edges between any pair of vertices i, j equals $\lfloor n(I(Y_{ij} \wedge Y_{ji}) - \epsilon)\rfloor$. We next show that every Steiner tree in a Steiner tree packing of $G_\epsilon^{(n)}$ yields one shared bit for the terminals in A that is independent of the communication in that Steiner tree. Specifically, for edges (i, j) and (i, j'), $j \neq j'$, with common vertex i in the Steiner tree, vertex i broadcasts to vertices j, j' the binary sum of two independent SK bits – one with j and the other with j' – obtained from the first step. This enables i, j, j' to share any one of these two bits, with the attribute that the shared bit is independent of the binary sum. This method of propagation ([15, Proof of Theorem 5]) enables all the vertices in A, which are connected in the Steiner tree, to share one bit that is independent of all the broadcast binary sums from this tree. Therefore, the maximum number of such shared bits for the terminals in A that can be generated by this procedure equals $\mu(A, G_\epsilon^{(n)})$. Denote these shared bits (of size $\mu(A, G_\epsilon^{(n)})$) and the communication messages generated by the mechanism in this second step by $K = K^{(n)}(\{K_{ij}\}_{1\leq i<j\leq m})$ and $F = F^{(n)}(\{K_{ij}\}_{1\leq i<j\leq m})$, respectively.

We claim that K constitutes an SK for A. Specifically, it remains to show that K satisfies the secrecy condition (2.1) with respect to the overall communication in steps 1 and 2. To this end, we denote by $K_R^{(n)}(\{K_{ij}\}_{1\leq i<j\leq m})$ all the pairwise SK bits generated in the first step, that are residual from the maximal Steiner tree packing

of $G_\epsilon^{(n)}$ used to generate K by means of F. Clearly,

$$\{K_{ij}\}_{1\le i<j\le m} = (K, F, K_R). \tag{C-14}$$

Moreover, since the total number of edges in any Steiner tree equals the sum of unity (i.e., the shared bit of K) and the number of bits of public communication for that shared bit, we have

$$|\widetilde{E^{(n)}}| = \log|\mathcal{K}| + \log|\mathcal{F}| + \log|\mathcal{K}_R|, \tag{C-15}$$

where $\mathcal{K}$, $\mathcal{F}$ and $\mathcal{K}_R$ denote the respective ranges of K, F and K_R. Note that $\log|\mathcal{K}| = \mu(A, G_\epsilon^{(n)})$. Then,

$$\begin{aligned}
s(K;\{F_{ij}\}_{1\le i<j\le m}, F) &= \log|\mathcal{K}| - H(K|\{F_{ij}\}_{1\le i<j\le m}, F) \\
&\le \log|\mathcal{K}| - H(K|\{F_{ij}\}_{1\le i<j\le m}, F, K_R) \\
&= \log|\mathcal{K}| - H(K, F, K_R|\{F_{ij}\}_{1\le i<j\le m}) \\
&\quad + H(F, K_R|\{F_{ij}\}_{1\le i<j\le m}) \\
&= \log|\mathcal{K}| - H(\{K_{ij}\}_{1\le i<j\le m}|\{F_{ij}\}_{1\le i<j\le m}) \\
&\quad + H(F, K_R|\{F_{ij}\}_{1\le i<j\le m}), \quad \text{by (C-14)} \\
&\le \log|\mathcal{K}| + s(\{K_{ij}\}_{1\le i<j\le m}; \{F_{ij}\}_{1\le i<j\le m}) \\
&\quad - |\widetilde{E^{(n)}}| + H(F, K_R) \\
&\le s(\{K_{ij}\}_{1\le i<j\le m}; \{F_{ij}\}_{1\le i<j\le m}), \quad \text{by (C-15)} \\
&= \sum_{1\le i<j\le m} s(K_{ij}; F_{ij}), \\
&= \frac{m(m-1)}{2} o_n(1),
\end{aligned}$$

where the second-to-last equality is by the fact that $\{(K_{ij}, F_{ij})\}_{1\le i<j\le m}$ are mutually independent, and the last equality is by (C-13). The maximum rate of the SK thus generated is equal to $\lim_{n\to\infty} \frac{1}{n}\mu(A, G_\epsilon^{(n)})$ which, since $\epsilon > 0$ was arbitrary, equals $\sup_{n\in\mathcal{N}} \frac{1}{n}\ \mu(A, G^{(n)})$.

(ii) Suppose that $A = \{1, 2\}$, and note from the paragraph after Definition 3 that $\mu(A, G)$ is the maximum number of edge disjoint paths in G connecting terminals 1 and 2. It is clear that $\frac{1}{n}\mu(A, G^{(n)})$ is nondecreasing in $n \in \mathcal{N}$, by the definition of $G^{(n)}$. According to Menger's theorem [42, 5], given a multigraph $G = (\mathcal{M}, E)$, the maximum number of edge disjoint paths in G connecting terminals 1 and 2 is equal to

$$\min_{\substack{\emptyset\neq B\subset\mathcal{M}\\ 1\in B,\ 2\in B^c}} (\text{number of edges that cross } \{B, B^c\})\,.$$

Applying this to $G^{(n)}$ as above, we have that for $n \in \mathcal{N}$,

$$\frac{1}{n}\mu(A, G^{(n)}) \;=\; \frac{1}{n}\left[\min_{\substack{\emptyset\neq B\subset\mathcal{M}\\ 1\in B,\ 2\in B^c}} \left(\sum_{\substack{1\le i<j\le m:\\ (i,j) \text{ crosses } \{B,B^c\}}} nI(Y_{ij}\wedge Y_{ji})\right)\right].$$

It then follows that

$$\begin{aligned} C(A) &\ge \sup_{n\in\mathcal{N}} \frac{1}{n}\mu(A, G^{(n)}), \quad \text{by (C-11)} \\ &= \min_{\substack{\emptyset\neq B\subset\mathcal{M}\\ 1\in B,\ 2\in B^c}} \left(\sum_{\substack{1\le i<j\le m:\\ (i,j) \text{ crosses } \{B,B^c\}}} nI(Y_{ij}\wedge Y_{ji})\right) \\ &= C^{ub}(A), \quad \text{by (C-10)}. \end{aligned}$$

The last equality follows upon noting that when $|A| = 2$, the minimization in (C-10) is over only those partitions that contain two atoms, each of which includes terminal 1 and terminal 2, respectively. This proves (ii). ∎

C.3.2 SK Capacity and Spanning Tree Packing for $A = \mathcal{M}$

When all the terminals in $\mathcal{M}$ seek a shared SK, i.e., when $A = \mathcal{M}$, a Steiner tree for A is a spanning tree for $\mathcal{M}$. In this case, we show that the lower bound for SK capacity in Theorem C.3 (i) is, in fact, tight. Specifically, we show that the algorithm in the proof of Theorem C.3 yields an SK of maximum rate that coincides with the upper bound for $C(\mathcal{M})$ in Lemma C.2.

Theorem C.4: For a PIN model, the SK capacity $C(\mathcal{M})$ is

$$\begin{aligned} C(\mathcal{M}) &= \sup_{n \in \mathcal{N}} \frac{1}{n} \mu(\mathcal{M}, G^{(n)}) \\ &= C^{ub}(\mathcal{M}). \end{aligned} \tag{C-16}$$

Remark: When $A \subset \mathcal{M}$, Steiner tree packing may not attain SK capacity. In SK generation, a helper terminal in A^c helps link the user terminals in A in complex ways through various combinations of subsets of A. In general, an optimal such linkage need not be attained by Steiner tree packing. However, when $|A| = 2$, the two user terminals are either directly connected or are connected by a path through helpers in A^c; both can be accomplished by Steiner tree packing. When $A = \mathcal{M}$, the mentioned complexity of a helper is nonexistent.

Proof: The proof relies on a graph-theoretic result of Nash-Williams [43] and

Tutte [48], that gives a min max formula for the maximum size of spanning tree packing in a multigraph.

It is clear that $\frac{1}{n}\mu(\mathcal{M}, G^{(n)})$ is nondecreasing in $n \in \mathcal{N}$, by the definition of $G^{(n)}$. By [43, 48], given a multigraph $G = (\mathcal{M}, E)$, the maximum number of edge disjoint spanning trees that can be packed in G is equal to

$$\min_{\mathcal{P}} \left\lfloor \frac{1}{|\mathcal{P}| - 1} \left(\text{number of edges that cross } \mathcal{P}\right) \right\rfloor,$$

with the minimization being over all partitions $\mathcal{P}$ of $\mathcal{M}$. Applying this to $G^{(n)}$ as above, we have that for $n \in \mathcal{N}$,

$$\frac{1}{n}\mu(\mathcal{M}, G^{(n)}) = \frac{1}{n}\left[\min_{\mathcal{P}} \left\lfloor \frac{1}{|\mathcal{P}| - 1} \left(\sum_{\substack{1 \le i < j \le m: \\ (i,j) \text{ crosses } \mathcal{P}}} n I(Y_{ij} \wedge Y_{ji}) \right) \right\rfloor \right].$$

Denoting by D the quantity in $\left[\ \right]$ above, it follows that

$$\begin{aligned} C(\mathcal{M}) &\ge \sup_{n \in \mathcal{N}} \frac{1}{n}\mu(\mathcal{M}, G^{(n)}), \quad \text{by Theorem C.3} \\ &\ge \sup_{n \in \mathcal{N}} \{D - \frac{1}{n}\} \\ &\ge \min_{\mathcal{P}} \frac{1}{|\mathcal{P}| - 1} \left(\sum_{\substack{1 \le i < i \le m: \\ (i,j) \text{ crosses } \mathcal{P}}} I(Y_{ij} \wedge Y_{ji}) \right) \\ &= C^{ub}(\mathcal{M}), \quad \text{by (C-10)}. \end{aligned}$$

The assertion in (C-16) is now immediate. ∎

Lastly, the following observation is of independent interest. Given a combinatorial problem of finding the maximal packing of Steiner trees in a multigraph, we can always associate with it a problem of SK generation for an associated PIN

model. By Theorem C.3 (i), the SK capacity for the PIN model yields an upper bound for the maximum rate of edge disjoint Steiner trees that can be packed in the multigraph; the upper bound is tight both in the case of path packing by Theorem C.3 (ii) and in the case of spanning tree packing by Theorem C.4.

C.3.3 Discussion

Our proofs of Theorems 3.3 and 3.4 give rise to explicit polynomial-time schemes for forming a group-wide SK for the terminals in A from the collection of optimum and mutually independent SKs for pairs of terminals in $\mathcal{M}$ (namely the K_{ij}s in the proof of Theorem C.3). When $|A| = 2$ or $A = \mathcal{M}$, our schemes achieve SK capacity. Specifically, the schemes combine known polynomial-time algorithms for finding a maximal collection of edge-disjoint paths (resp. spanning trees) connecting the vertices in A when $|A| = 2$ (resp. $A = \mathcal{M}$) [19, 20, 23] with the technique for SK propagation in each tree as in the proof of Theorem C.3.

For a general multiterminal source model, the notions of wiretap secret key (WSK) [38, 1, 15] and private key (PK) [15] have also been proposed. Specifically, these notions involve an extra "wiretapped" terminal, say $m+1$, that observes n i.i.d. repetitions of a rv X_{m+1} with a given joint pmf with $(X_1, \ldots, X_m)$, and to which the eavesdropper has access. The key must now be concealed from the eavesdropper's observations of $X_{m+1}^n = (X_{m+1,1}, \ldots, X_{m+1,n})$ and the public communication. The notion of a WSK requires that terminal $m + 1$ not cooperate in key generation. The less restrictive notion of a PK allows cooperation by terminal $m + 1$ by way of

public communication. The corresponding capacities for the terminals in $A \subseteq \mathcal{M}$ are defined in the usual manner, and denoted by $C_W(A)$ and $C_P(A)$. We remark that in the context of a PIN model, terminal $m+1$ represents a compromised entity.

One model for the wiretapped rv X_{m+1} entails its consisting of $\begin{pmatrix} m \\ 2 \end{pmatrix}$ mutually independent components, one corresponding to each pair $(Y_{ij}, Y_{ji}),\ 1 \leq i < j \leq m$, of legitimate correlated signals. This model is unresolved even in the simplest case of $m = 2$ terminals [39, 1, 15, 24, 25]. Instead, we consider a different model which depicts the situation in which an erstwhile legitimate terminal $m+1$ becomes compromised. Specifically, the model now involves every legitimate terminal i in $\mathcal{M}$ observing n i.i.d. repetitions of the rv $(X_i, Y_{i,m+1})$, while terminal $m+1$ observes n i.i.d. repetitions of $X_{m+1} = (Y_{m+1,j},\ j \in \mathcal{M})$. We argue in the following proposition that the WSK and PK capacities for this PIN model are the same as the SK capacity of a reduced PIN model obtained by disregarding terminal $m+1$ and with each legitimate terminal i in $\mathcal{M}$ observing just X_i^n.

Proposition C.5: It holds that

$$C_W(A) = C_P(A) = C(A).$$

Proof: We shall prove that

$$C(A) \overset{(a)}{\leq} C_W(A) \overset{(b)}{\leq} C_P(A) \overset{(c)}{\leq} C(A).$$

The inequality (b) is by definition. Next, let $K = K(X_1^n, \ldots, X_m^n)$ be a SK for A achieved with communication $\mathbf{F} = \mathbf{F}(X_1^n, \ldots, X_m^n)$ for the reduced PIN model.

Then K is also a WSK since

$$\begin{aligned} s\left(K; \mathbf{F}, (Y_{m+1,j}^n,\ j \in \mathcal{M})\right) &= \log|K| - H\left(K|\mathbf{F}, (Y_{m+1,j}^n,\ j \in \mathcal{M})\right) \\ &= s(K; \mathbf{F}) + I(K \wedge (Y_{m+1,j}^n,\ j \in \mathcal{M})|\mathbf{F}) \\ &= o_n(1) \end{aligned}$$

since $I\left(K, \mathbf{F} \wedge (Y_{m+1,j}^n,\ j \in \mathcal{M})\right) = 0$, thereby establishing (a). In order to establish (c), we claim that every achievable PK rate is an achievable SK rate for the reduced PIN model upon using randomization at the terminals in $\mathcal{M}$; by Remark (ii) after Proposition C.1, (c) then follows. Since $(Y_{m+1,j}^n,\ j \in \mathcal{M})$ is independent of $(X_1^n, \ldots, X_m^n)$, any terminal in $\mathcal{M}$, say terminal 1, can simulate $(Y_{m+1,j}^n,\ j \in \mathcal{M})$ and broadcast it to all the terminals. Next, each terminal i in $\mathcal{M}$ can simulate $Y_{i,m+1}^n$ conditioned on $(Y_{m+1,j}^n,\ j \in \mathcal{M}) = (y_{m+1,j}^n,\ j \in \mathcal{M})$. This second step of randomization is feasible since $(X_1^n, \ldots, X_m^n), Y_{1,m+1}^n, \ldots, Y_{m,m+1}^n$ are conditionally mutually independent conditioned on $(Y_{m+1,j}^n,\ j \in \mathcal{M}) = (y_{m+1,j}^n,\ j \in \mathcal{M})$. Thus, each terminal i in $\mathcal{M}$ now has access to $(X_i^n, Y_{i,m+1}^n)$ while the eavesdropper observes $(Y_{m+1,j}^n,\ j \in \mathcal{M})$, so that the reduced PIN model for SK generation can be used to simulate a PIN model for PK generation with the given underlying joint pmf. Thus, any achievable rate of a PK for A in the *given* PIN model for PK generation is an achievable rate of a PK for A in the *simulated* model. Further, the latter PK is a fortiori an SK for A in the reduced PIN model with randomization permitted at the terminals in $\mathcal{M}$. This establishes (c). ∎

In the proof of achievability of SK capacity for the general multiterminal source model in [15], an SK of optimum rate was extracted from "omniscience," i.e., from

a reconstruction by the terminals in A of *all* the signals $(X_i^n,\ i \in \mathcal{M})$ observed by the terminals in $\mathcal{M}$. In contrast, the scheme in Theorem C.3 (ii) (resp. Theorem C.4) for achieving SK capacity for a PIN model with $|A| = 2$ (resp. $A = \mathcal{M}$) neither seeks nor attains omniscience; however, we note that omniscience can be attained by letting the terminals in $\mathcal{M}$ simply broadcast all the residual bits left over from a maximal path packing (resp. maximal spanning tree packing).

Bibliography

[1] R. Ahlswede and I. Csiszár, "Common Randomness in Information Theory and Cryptography–Part I: Secret Sharing," IEEE Trans. Inf. Theory **39**, pp. 1121-1132 (1993).

[2] R. Ahlswede and G. Dueck, "Identification in the Presence of Feed-back — A Discovery of New Capacity Formulas," IEEE Trans. Inf. Theory **35**, pp. 30-36 (1989).

[3] C.H. Bennett, G. Brassard, C. Crépeau and U.M. Maurer, "Generalized Privacy Amplification," IEEE Trans. Inf. Theory **41**, pp. 1915-1923 (1995).

[4] C. H. Bennett, G. Brassard and J. -M. Robert, "Privacy Amplification by Public Discussion," SIAM J. Computing **17**, pp. 210-229 (1988).

[5] A. Bondy and U. S. R. Murty, *Graph Theory, Series: Graduate Texts in Mathematics* **244** (Springer, 2008).

[6] C. Boyd and A. Mathuria, *Protocols for Authentication and Key Establishment* (Springer-Verlag, New York, NY, 2003).

[7] C. Cachin, *Entropy Measures and Unconditional Security in Cryptography*, Ph.D. Dissertation, ETH Zurich, Switzerland (1997).

[8] J. Cheriyan and M. Salavatipour, "Hardness and Approximation Results for Packing Steiner Trees," Algorithmica **45, 1**, pp. 21-43 (2006).

[9] J.H. Conway and N. J. A. Sloane, *Sphere Packing, Lattices and Groups,* (Springer Verlag, New York, NY, 1988).

[10] I. Csiszár, "Linear Codes for Sources and Source Networks: Error Exponents, Universal Coding," IEEE Trans. Inf. Theory **28**, pp. 585-592 (1982).

[11] I. Csiszár, "Almost Independence and Secrecy Capacity," Probl. Pered. Inf. (Special issue devoted to M.S. Pinsker) **32**, pp. 48-57 (1996).

[12] I. Csiszár and J. Körner, "Broadcast Channels with Confidential Messages," IEEE Trans. Inf. Theory **24**, pp. 339-348 (1978).

[13] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Academic Press, Inc., New York, NY, 1981).

[14] I. Csiszár and P. Narayan, "Common Randomness and Secret Key Generation with a Helper," IEEE Trans. Inf. Theory **46**, pp. 344-366 (2000).

[15] I. Csiszár and P. Narayan, "Secrecy Capacities for Multiple Terminals," IEEE Trans. Inf. Theory **50**, pp. 3047-3061 (2004).

[16] I. Csiszár and P. Narayan, "Secrecy Capacities for Multiterminal Channel Models," IEEE Trans. Inf. Theory (Special Issue on Information Theoretic Security) **54**, pp. 2437-2452 (2008).

[17] I. Csiszár and P. Narayan, "Secrecy Capacities for Multiple Input Multiple Output Channel Models," in Proc. of the IEEE Internat. Sympos. on Inf. Theory, pp. 2447-2451 (Seoul, Korea, June 28-July 3, 2009).

[18] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Springer-Verlag, New York, NY, 1998).

[19] E. A. Dinic, "An Algorithm for the Solution of the Problem of Maximal Flow in a Network with Power Estimation," Dokl. Akad. Nauk SSSR. **194**, pp. 754-757 (1970).

[20] J. Edmonds and R.M. Karp, "Theoretical Improvements in Algorithmic Efficiency for Network Flow Problems," *Combinatorial Structures and Their Applications* (Gordon and Breach, New York, NY , 1970), pp. 93-96.

[21] U. Erez, S. Litsyn and R. Zamir, "Lattices which Are Good for (Almost) Everything," IEEE Trans. Inf. Theory **51**, pp. 3401-3416 (2005).

[22] U. Erez and R.Zamir, "Achieving $\frac{1}{2}\log(1+SNR)$ on the AWGN Channel With Lattice Encoding and Decoding," IEEE Trans. Inf. Theory **50**, pp. 2293-2314 (2004).

[23] H. N. Gabow and H. H. Westermann, "Forests, Frames and Games: Algorithms for Matroid Sums and Applications," Algorithmica **7**, pp. 465-497 (1992).

[24] A. Gohari and V. Anantharam, "Communication for Omniscience by a Neutral Observer and Information-theoretic Key Agreement of Multiple Terminals," in Proc. IEEE Internat. Sympos. on Inf. Theory, pp. 2056-2060 (Nice, France, 24-29 June, 2007).

[25] A. Gohari and V. Anantharam, "New Bounds on the Information-theoretic Key Agreement of Multiple Terminals," in Proc. of the IEEE Internat. Sympos. on Inf. Theory, pp. 742-746 (Toronto, Ontario, Canada, 6-11 July, 2008).

[26] M. Grötschel, A. Martin and R. Weismantel, "Packing Steiner Trees: A Cutting Plane Algorithm and Computational Results," Math. Programming **72**, pp. 125-145 (1996).

[27] G. H. Hardy and E. M Wright, *An Introduction to the Theory of Numbers, 5th ed.* (Oxford Univ. Press, Oxford, U.K., 1979).

[28] K. Jain, M. Mahdian, and M.R. Salavatipour, "Packing Steiner trees," in Proc. of the 14th ACM-SIAM Sympos. on Discrete Algorithms, pp. 266-274 (Baltimore, Maryland, USA, 12-14 January, 2003).

[29] J. Katz and M. Yung "Scalable Protocols for Authenticated Group Key Exchange," J. of Cryptology **20**, pp. 85-113 (2007).

[30] J. Körner and A. Orlitsky "Zero-Error Information Theory," IEEE Trans. Inf. Theory **44**, pp. 2207-2229 (1998).

[31] M. Kriesell, "Edge-disjoint Trees Containing Some Given Vertices in a Graph," Journal of Combinatorial Theory, Series B, **88**, pp. 53-65 (2003).

[32] Z. Li, B. Li and L. C. Lau, "On Achieving Maximum Multicast Throughput in Undirected Networks," IEEE Trans. Inf. Theory **52**, pp. 2467-2485 (2006).

[33] Y. Liang, H. V. Poor and S. Shamai (Shitz), "Information Theoretic Security," Found. and Trends in Commun. and Inf. Theory **5**, pp. 355-580 (2008).

[34] L. Lovász, "On Some Connectivity Properties of Eulerian Graphs," Acta Math. Akad. Sci. Hung. **28**, pp. 129-139 (1976).

[35] M. Manulis, "Security-Focused Survey on Group Key Exchange Protocols," HGI Network and Data Security Group Tech. Rep. 2006/03, (2006).

[36] M. Manulis, *Provably Secure Group Key Exchange*, Ph.D. Dissertation, Ruhr-Universität Bochum, Bochum, Germany (2007).

[37] M. Manulis, "Survey on Security Requirements and Models for Group Key Exchange," HGI Network and Data Security Group Tech. Rep. 2006/02, (2006) (Revised 2008).

[38] U. M. Maurer, "Provably Secure Key Distribution Based on Independent Channels," in Proc. of the IEEE Workshop Inf. Theory (Eindhoven, The Netherlands, June, 1990).

[39] U. M. Maurer, "Secret Key Agreement by Public Discussion from Common Information," IEEE Trans. Inf. Theory **39**, pp. 733-742 (1993).

[40] U. M. Maurer, "The Strong Secret Key Rate of Discrete Random Triples," *Communications and Cryptography: Two Sides of One Tapestry, R. E. Blahut et al., Ed.* (Kluwer, Norwell, MA, 1994), Chap. 26, pp. 271-285.

[41] U. Maurer and S. Wolf, "Information-theoretic Key Agreement: From Weak to Strong Secrecy for Free," *Lecture Notes in Computer Science* **1807** (Springer-Verlag, 2000), pp. 352-368.

[42] K. Menger, "Zur Allgemeinen Kurventheorie," Fund. Math. **10**, pp. 96-115 (1927).

[43] C. St.J. A. Nash-Williams, "Edge-Disjoint Spanning Trees of Finite Graphs," J. London Math. Soc. **36**, pp. 445-450 (1961).

[44] S. Nitinawarat and P. Narayan, "Perfect Secrecy, Perfect Omniscience and Steiner Tree Packing," IEEE Trans. Inf. Theory, to appear, December 2010.

[45] S. Nitinawarat, C. Ye, A. Barg, P. Narayan and A. Reznik, "Secret Key Generation for a Pairwise Independent Network Model," IEEE Trans. Inf. Theory, to appear, December 2010.

[46] G. Poltyrev,"On Coding without Restrictions for the AWGN Channel," IEEE Trans. Inf. Theory **40**, pp. 409-417 (1994).

[47] C. E. Shannon,"Communication Theory of Secrecy Systems," Bell Sys. Tech. J. **28** pp. 656-715 (1949).

[48] W. T. Tutte, "On the Problem of Decomposing a Graph into n Connected Factors," J. London Math. Soc. **36**, pp. 221-230 (1961).

[49] A. B. Wagner, S. Tavildar and P. Viswanath, "Rate Region of the Quadratic Gaussian Two-Encoder Source-Coding Problem," IEEE Trans. Inf. Theory **5**, pp. 1938-1961 (2008).

[50] Y. Wu, K. Jain and S.-Y. Kung, "A Unification of Network Coding and Tree-packing (Routing) Theorems," IEEE Trans. Inf. Theory **52**, pp. 2398-2409 (2006).

[51] A. D. Wyner, "The Wire-tap Channel," Bell Sys. Tech. J. **54**, pp. 1355-1387 (1975).

[52] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," IEEE Trans. Inf. Theory **22**, pp. 1-11 (1976).

[53] C. Ye, *Information Theoretic Generation of Multiple Secret Keys*, Ph.D. Dissertation, University of Maryland, MD (2005).

[54] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe and N. Mandayam, "Information Theoretic Secret Key Generation from Wireless Channels," IEEE Trans. Inf. Forensics and Security **5**, pp. 240-254 (2010).

[55] C. Ye and P. Narayan, "Secret Key and Private Key Constructions for Simple Multiterminal Source Models," in Proc. of the IEEE Internat. Sympos. on Inf. Theory, pp. 2133-2137 (Adelaide, Australia, 4-9 September 2005).

[56] C. Ye and P. Narayan, "Secret Key and Private Key Constructions for Simple Multiterminal Source Models," IEEE Trans. Inf. Theory, in review.

[57] C. Ye and A. Reznik, "Group Secret Key Generation Algorithms," in Proc. of the IEEE Internat. Sympos. on Inf. Theory, pp. 2896-2900 (Nice, France, 24-29 June 2007).

[58] R. Zamir and M. Feder, "On Universal Quantization by Randomized Uniform/ lattice Quantizer," IEEE Trans. Inf. Theory **38**, pp. 428-436 (1992).

[59] R. Zamir and M. Feder, "On Lattice Quantization Noise," IEEE Trans. Inf. Theory **42**, pp. 1152-1159 (1996).

[60] R. Zamir, S. Shamai and U. Erez, "Nested Linear/lattice Codes for Structured Multiterminal Binning," IEEE Trans. Inf. Theory **48**, pp. 1250-1276 (2002).