On Entropy Maximization, Opportunistic Selection and LLR Generation for Algebraic Group Secret-Key Generation

Thesis submitted by

Rohit Joshi
2019BSY7504

under the guidance of

Dr. Harshan Jagadeesh,

Indian Institute of Technology Delhi

for the award of the degree of

Masters of Science
(by Research)

Bharti School of Telecommunication
Technology & Management
Indian Institute of Technology Delhi
New Delhi, India
July 2021

Certificate

This is to certify that the thesis titled **" On Entropy Maximization, Opportunistic Selection and LLR Generation for Algebraic Group Secret-Key Generation "** being submitted by **Mr. Rohit Joshi** to the **Bharti School of Telecommunication Technology and Management, Indian Institute of Technology, Delhi,** for the award of the degree of **Masters of Science (by Research)** is the record of the bona-fide research work done by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted either in part or in full to any other university or institute for the award of any degree or diploma.

Prof. Harshan Jagadeesh
Department of Electrical Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi 110016
India

Place: New Delhi
Date: 29th September 2021

Acknowledgements

First and foremost, I wish to express my deepest gratitude to my respectable supervisor Prof. Harshan Jagadeesh for his support, guidance, and immense knowledge during the tenure of my MS(R) program at IIT Delhi. His constant encouragement to think independently has been a great source of motivation. I am really blessed to have such a supporting and experienced supervisor.

I would also like to extend my thanks to my research committee members, Prof. Saif K. Mohammed, Prof. Seshan Srirangarajan, and Prof. Deepak Patil, for their valuable comments and suggestions during my research work. Thanks are also due to the anonymous reviewers whose constructive suggestions and criticisms have significantly improved the quality of my work.

I take this opportunity to thank the funding agency, Department of Telecommunications, Ministry of Communications, New Delhi, India and I am really honoured to be a part of the project, "Indigenous 5G Test Bed". Special thanks to Mr. Gautam Tiwari, Mrs. Savita Sharma, Mr. Dilip Jana and Mrs. Manrika Lochan for all the administrative support they had provided me during my stay at IIT Delhi.

My time spent at IIT Delhi was made memorable by the friend who were always there to help and support me. My special heartfelt thanks to Amogh Vithalkar, Jagdeep Singh, Syed Mohd Afzal, Anjali Gupta, Vishal Srivastava, Manish Bansal, Shashank Shukla and others. I extend my heartful thanks to my collaborators-cum-friends, Aakash Radhakrishanan, Suraj Sajeev, and Major Sriram SV for the lengthy technical discussions and the enjoyable moments shared together. Special thanks to all the senior members in the research group, Vivek Chaudhary, Jaya Goel and Rusnik Mangang for helping me out with the research study.

Above all, I owe my deepest gratitude to Papa, Mummy, Chacha, Chachi, Rohan, Anushka and Agrima. All my success and achievements go to their credit. The one person that I want to mention the most is my late grandfather, who actually introduced me to this beautiful domain of engineering at a very early age. He is the one reason I wanted to do well in the field of study and research, and I thank him for all that he did for me during his lifetime and continues to do so from the heavenly abode.

- Rohit Joshi

Abstract

This thesis addresses the design of consensus algorithm and post-consensus enhancements in the context of dynamic key generation. The first part of the thesis is on multi-level quantization and the second part is on enhancements in terms of key-rate and error-rate.

The broadcast nature of the wireless system makes them vulnerable to several security attacks which is not the case with the wired systems. The wireless interface is open both to legitimate and unauthorised users which is why it is very crucial to design these systems carefully. To achieve the security, there are multiple techniques in use, generally it is classified as: (i) crypto-primitive based and (ii) physical-layer based methods. In the former scheme, security is ensured by relying on computationally exhaustive mathematical operations which are easy to decode at the user with the key whereas complex to solve otherwise, whereas, in the latter scheme, physical layer resources are exploited in such a way that the unauthorised user does not get access to the data. There is one more special category which is physical-layer based key generation, here the users exploit the common source of randomness (CSR) such as wireless channel coefficients between them to derive a key useful for encryption and decryption operations. Towards achieving high-rate keys, consensus algorithms using multi-level quantization seek to generate more than one bit per sample by carefully designing the boundaries over the support-set of observations at individual nodes. However, we have observed that the secret-keys delivered by such multi-level quantization schemes do not exhibit maximum entropy, owing to mismatch in the distribution of the observations at individual nodes and the distribution of the observations in consensus.

In this work, we revisit the design of multi-level quantization schemes by using the joint probability distribution of the observations at the participating nodes. In particular, we formulate the quantizer design problem as a constrained optimization problem of maximizing the key-rate with strict constraints on the entropy and the mismatch-rate of the generated keys. Subsequently, we propose an iterative algorithm, referred to as the EM-EM algorithm, which synthesizes a multi-level quantizer, for any given number of levels, satisfying the constraints. Unlike the existing multi-level quantization schemes, the EM-EM algorithm provides secret-keys with maximum entropy, which otherwise would not be possible using marginal distributions. Finally, since our solution is algorithmic in nature, it is applicable on a wide range of joint distributions, and also on data sets, akin to the celebrated Max-Lloyd algorithm in the parallel world of multi-level quantization on univariate distributions.

Although two-user key generation is well understood, group secret-key (GSK) generation, wherein more than two nodes in a network generate secret-keys, still poses open problems. Recently, Manish Rao et al., have proposed the Algebraic Symmetrically Quantized GSK (A-SQGSK) protocol for a network of three nodes wherein the nodes share quantized versions of the channel realizations over algebraic rings, and then harvest a GSK. On the CSR provided by A-SQGSK for the case of three-user network, formulated EM-EM framework is applied. Given that the proposed multi-level consensus algorithm is closely coupled with the protocol used to exchange the CSR, we lay out design rules to jointly choose the parameters of the A-SQGSK protocol and the EM-EM algorithm as a function of underlying signal-to-noise-ratio and the required mismatch rate on the generated GSK.

Although A-SQGSK protocol guarantees confidentiality of common randomness to an eavesdropper, we observe that the key-rate of the protocol is poor since only one channel in the network is used to harvest GSK. Identifying these, we propose an opportunistic selection method wherein more than one wireless channel is used to harvest GSKs without compromising the confidentiality feature, thereby resulting in remarkable improvements in the key-rate. Furthermore, we also propose a log-likelihood ratio (LLR) generation method for the common randomness observed at various nodes, so that the soft-values are applied to execute LDPC codes based reconciliation to reduce the bit mismatches among the nodes.

Contents

List of Figures

List of Tables

Abbreviations

AES	Advanced Encryption Standard
A-SQGSK	Algebraic-Symmetrically Quantized GSK
AWGN	Additive White Gaussian Noise
BER	Bit Error Rate
CSR	Common Source of Randomness
D2D	Device to Device
DES	Data Encryption Standard
DSSS	Direct-Sequence Spread Spectrum
EM-EM	Entropy Maximization Error Minimization
FHSS	Frequency-Hopping Spread Spectrum
LDPC	Low-Density Parity Check
LLR	Log-Likelihood Ratio
M2M	Machine to Machine
MAP	Maximum A posteriori Probability
MP	Message Passing
PDF	Probability Density Function
PMF	Probability Mass Function
PSK	Phase-Shift-keying
QAM	Quadrature-Amplitude-Modulation
RSA	Rivest, Shamir, & Adleman (Public Key Encryption Technology)
SER	Symbol Error Rate
SNR	Signal-to-Noise Ratio
TDD	Time-Division Duplex
UE	User Equipment
USIM	User Service Identity Module
VANET	Vehicular Ad-hoc Network

Chapter 1

Introduction

Given the broadcast nature of wireless communication, it is well known that messages transmitted to an intended receiver can also be heard by eavesdroppers in the vicinity, thereby compromising the much needed *confidentiality* feature. Physical layer security can be used in order to preserve the confidentiality, where the physical-layer of a communication system is exploited to hide the messages in the wireless channel using techniques listed in Fig. 1.1 (right part). Another standard technique to circumvent this problem is to employ crypto-primitives at the higher-layer between the transmitter and the receiver, e.g., symmetric-key encryption or public-key encryption methods [AC93] as shown in Fig. 1.1 (left part). With symmetric-key techniques being favoured for application in low-cost wireless devices, the communicating parties need to posses a pre-shared secret-key to execute the crypto-primitives. Traditional wireless security makes use of prior known cryptography keys to establish a secure link between two nodes, where either the two devices are not computationally capable enough of deriving the keys using key exchange techniques or in scenarios where maintaining a public key database everywhere to support asymmetric encryption is challenging. Cellular communication is one such scenario, where keys are already stored in the user service identity module (USIM) and core network to support encryption/ decryption between the user and network. In the new era of cellular communication, involvement of Device to Device (D2D), Vehicle Ad-hoc network (VANETs) and Machine to machine (M2M) will require the use of some new technique to secure the link as no prior common key is pre-shared between any two nodes to communicate securely. Similarly in a wireless Ad-hoc network, establishing the keys between low end processing sensors nodes is challenging.

While a plethora of crypto-techniques are well known for key-exchange mechanisms, the concept of physical-layer key generation techniques has also received traction in the wireless community as the communicating nodes can harvest shared-keys just by witnessing the randomness in their channel realizations [MTM$^+$08, YMR$^+$10, JPC$^+$09, PCB13, GK11, YRS06, LDS12, SH11, HHL17, RSW11, CSD09, LLB13, YDS11, ZWCM10, KWTP13, LLN$^+$19]. Thus, this mode of key generation can provide additional layer of security other than the higher-layer crypto-primitives.

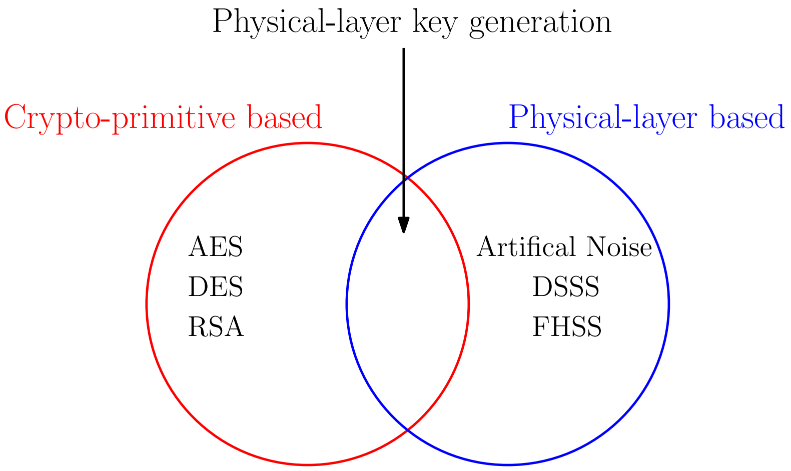


Figure 1.1: Crypto-primitve and physical-layer security

1.1 Background on Physical-Layer Key Generation

In the simple case of two-users key generation, consider two legitimate nodes, node-1 and node-2, which intend to secure their communication by harvesting an identical key using the shared wireless channel as the Common Source of Randomness (CSR). The two nodes first assist each other in observing the CSR by broadcasting a known pilot symbol to one another in a time-division duplex (TDD) fashion as depicted in Fig. 1.2.

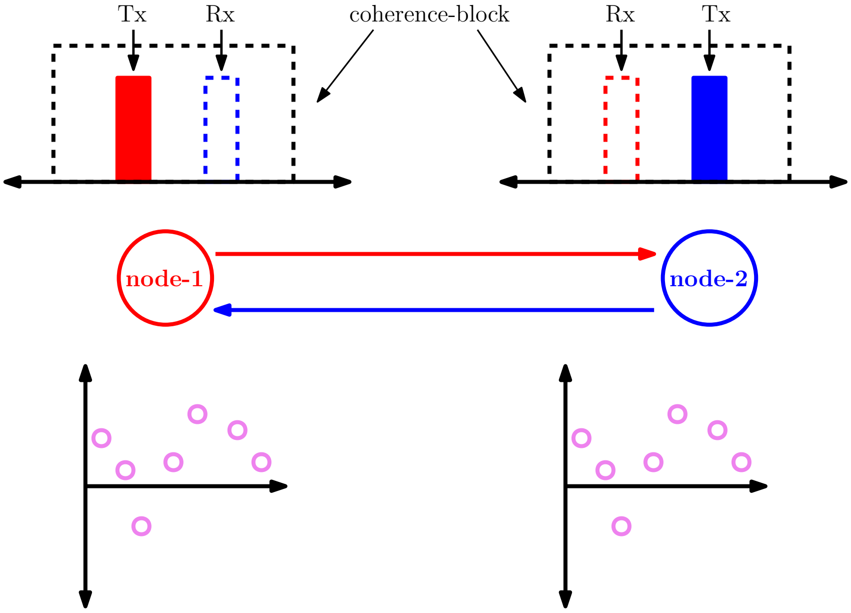


Figure 1.2: Figure depicts physical-layer based key generation technique where the upper part demonstrates the probing phase in TDD fashion and lower plots capture the channel estimates over different coherence blocks.

As a consequence of *Reciprocity Theorem*, if the role of transmit and receive antenna are interchanged functionally, the instantaneous characteristics of the channel remain unchanged. Further, if the two nodes probe back to each other within the same coherence-time, upon estimation, they will observe correlated channel estimates while fading characteristics ensure randomness in the correlated channel estimates. Overall, the CSR observed by node-1 and node-2 qualifies to generate a key with two desired properties- (i) common to both the users and (ii) random across the bits, given, a carefully designed consensus algorithm is used.

1.2 Mathur et. al. Consensus Algorithm

A straightforward technique to harvest shared secret-keys is to apply two-level quantization on the samples at node-1 and node-2, as proposed in [MTM+08, JPC+09]. Guard-bands are placed around the boundary in order to minimise the error-rate as shown in Fig. 1.3. Both the nodes use the indices which are out of the guard-band at both the sides to harvest a key. In order to do that, the two nodes first parse through their samples and form a list of all the indices which are out of the guard-band and then node-1 shares this list with node-2 without revealing the polarity of the indices. Upon receiving the list, node-2 finalises the indices which are common to both the users and shares it with node-1.

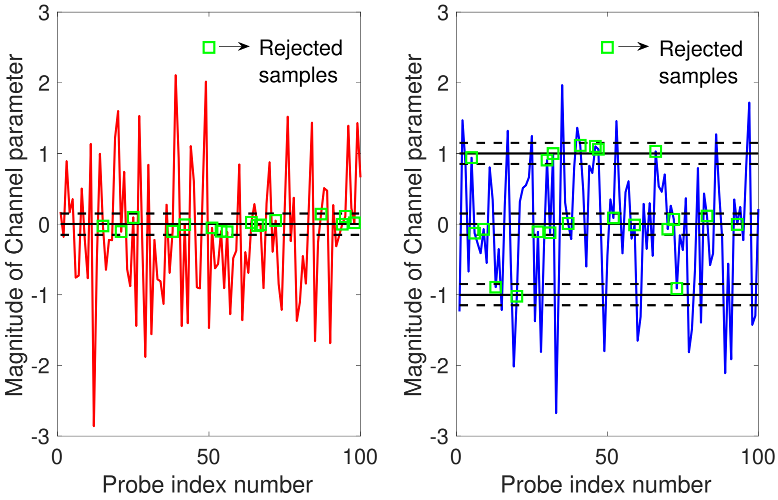


Figure 1.3: Figure depicts two-level (left) and four-level (right) quantizer.

Although the idea of two-level quantizer is effective, its limitation is its inability to generate more than one bit per sample when the channel offers significant randomness. A natural way to increase the number of bits per sample is to apply multi-level quantization on each sample, where the number of levels must be chosen depending on the common randomness offered by the wireless channel. We explore multi-level quantization method which ensures that the entropy is maximised and the error-rate is upper-bounded by some negligible number.

1.3 Key Generation Hierarchy

In order to ensure that the final key synthesised at legitimate nodes is random and identical, the hierarchy shown in Figure 1.4 is used. Advantage distillation phase includes channel probing and public discussion based consensus among the nodes, as already discussed. Upon advantage distillation, the nodes observe a correlated sequence, post-processing is required in order to make the sequences identical. Information reconciliation makes use of coding theory tools to achieve this. To compensate for the leakage while public discussion and information reconciliation phase, privacy amplification is followed, which uses hash functions. The end-product key can be used to fulfil the higher layer crypto-primitive operations.

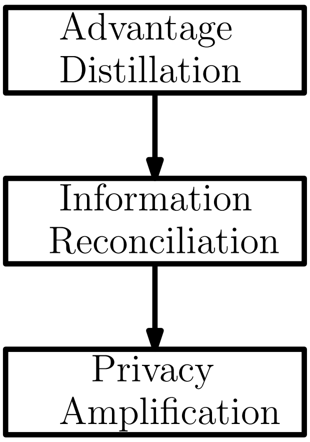
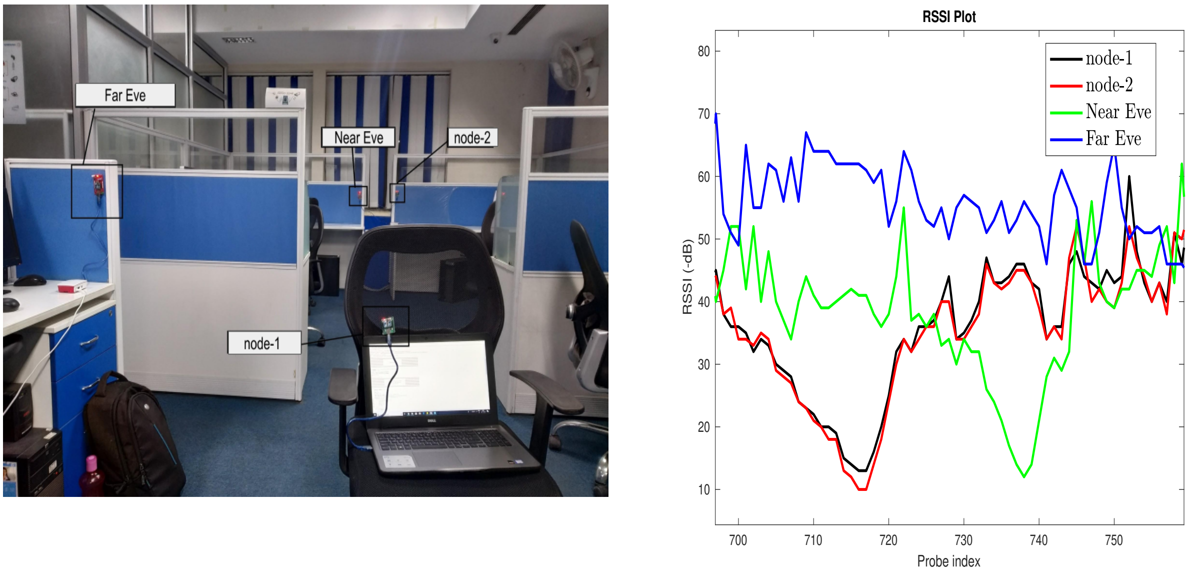


Figure 1.4: Dynamic key generation hierarchy

1.4 Eavesdropper's Perspective

Consider a passive eavesdropper, Eve, which intends to observe the same CSR as legitimate node. It can observe the channel which is shared between legitimate nodes, only if it is spatially located within the coherence distance, i.e. $\frac{\lambda}{2}$, to one of the nodes, where λ is the wavelength of the operating carrier signal. In context to 5G and next-generation networks, mmWave is prominent and hence, the only way Eve can observe the same coherent space is by implanting the antenna at one of the nodes, which falls under physical tampering and hardware security.

A small experiment conducted in the lab at IIT Delhi is shown in Fig. 1.5a. Two eavesdroppers are considered, near-Eve and far-Eve, the reference is with respect to node-2, so that, both the eavesdroppers estimate the channel with node-1. Xbee devices are used, which operates at 2.4 GHz bands. Eve fails to observe the CSR witnessed by the nodes whereas the channel between the two legitimate nodes is highly correlated as shown in Fig. 1.5b.



(a) Lab-setup consisting of node-1, node-2, far Eve and near Eve.

(b) Plot of received signal strength vs probe index for node-1, node-2, far Eve and near Eve.

Figure 1.5: Figure depicts a scenario where far and near Eve intends to observe the CSR.

1.5 Curious Case of Three-Users

While a number of contributions have been reported under physical-layer two-user key generation, its generalization to a network comprising more than two nodes have also been studied, under the framework of Group Secret-Key (GSK) generation techniques [YR07, XCD$^+$16, WZN12, LHH19, LYW$^+$14, TLQ15, SMDF11, HCH17, RH18, RH19]. In such a framework, more than two nodes generate a common secret-key by observing the temporal variation of their wireless channels so that these secret-keys can be used to keep their group messages confidential when implementing broadcast and relaying strategies among the group members. Typical applications of GSK generation for broadcast, relaying and multi-cast communication include Device-to-Device communication in ad hoc networks, e.g., vehicular networks and mobile networks.

Unlike the case of two-user key generation where the nodes inherently observe the CSR at probing phase itself, in a network consisting of multiple nodes, we highlight that not every node can inherently observe it.

Broadly, physical-layer GSK generation can be classified into two types: (i) Pairwise GSK generation, wherein a central authority (which is one of the nodes in the network) generates a secret-key by applying two-user key generation algorithm with one of its neighbours, and then shares it with the other nodes in the network in a confidential manner [YR07, XCD$^+$16, LHH19], and (ii) Group consensus based GSK generation, wherein the central authority assists multiple nodes in the network to witness a CSR so that all the nodes can synthesize a group secret-key using a group consensus algorithm [RH19, JJR21, WZN12, LYW$^+$14].

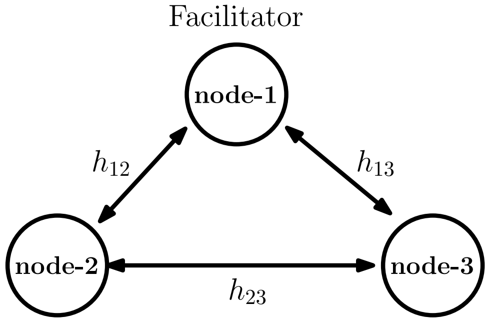


Figure 1.6: Three-user network model

It is noted that the former scheme trusts the central authority in the process of key generation and key distribution whereas in the latter scheme all the nodes share the responsibility of synthesising the key through a group consensus algorithm on the observed CSR. While the former class of methods piggyback on the simplicity of two-user key generation protocols, such schemes expose the generated digital key to threats to a possible insider attack in the wireless network. In contrast, using the latter class of methods, it has been shown that manipulating the CSR by an insider is power-inefficient, and may also be detected by the neighboring nodes provided the detection algorithms are carefully designed [HCH17]. In this paper, we are interested in the latter class of GSK generation protocols for the above mentioned reasons.

We use the three-node network shown in Fig. 1.6 to illustrate the challenges involved in securely sharing a CSR among the nodes in the network. From Fig. 1.6, it is clear that the available channels in the network are $h_{12}, h_{13},$ and h_{23}, where h_{jk} represents the complex baseband channel from node-j to node-k. In order to observe these channels systemically, all the nodes follow **Phase-1** to **Phase-3** of the protocol depicted in Fig. 1.7. In the first phase, node-1 transmits a pilot whereas other two nodes estimate the channel, likewise in **Phase-2** of the protocol, node-2 transmits and others estimate, similarly, **Phase-3** follows next.

Towards executing physical-layer GSK generation, the three nodes have to pick at least one of these channels as the CSR. Suppose that the channel h_{12} is the chosen CSR. With that choice, although node-1 and node-2 can witness this CSR by transmitting pilot symbol turn-by-turn within the coherence-time, node-3 cannot learn this channel during the pilot transmission phase, and therefore either node-1 or node-2 must act as *facilitator* to help node-3 in observing h_{12}. Assuming node-1 plays the role of facilitator, upon observing the channel realizations h_{12} and h_{13} in the pilot transmission phase, it must broadcast a function of h_{12} and h_{13}, denoted by $g(h_{12}, h_{13}) \in \mathbb{C}$ (as depicted in **Phase-4** Fig. 1.7, and Fig. 1.8) such that h_{12} must be recovered by node-3, whereas an eavesdropper in the vicinity must not be able to recover h_{12} from the broadcast signal. However, from a practical viewpoint, since

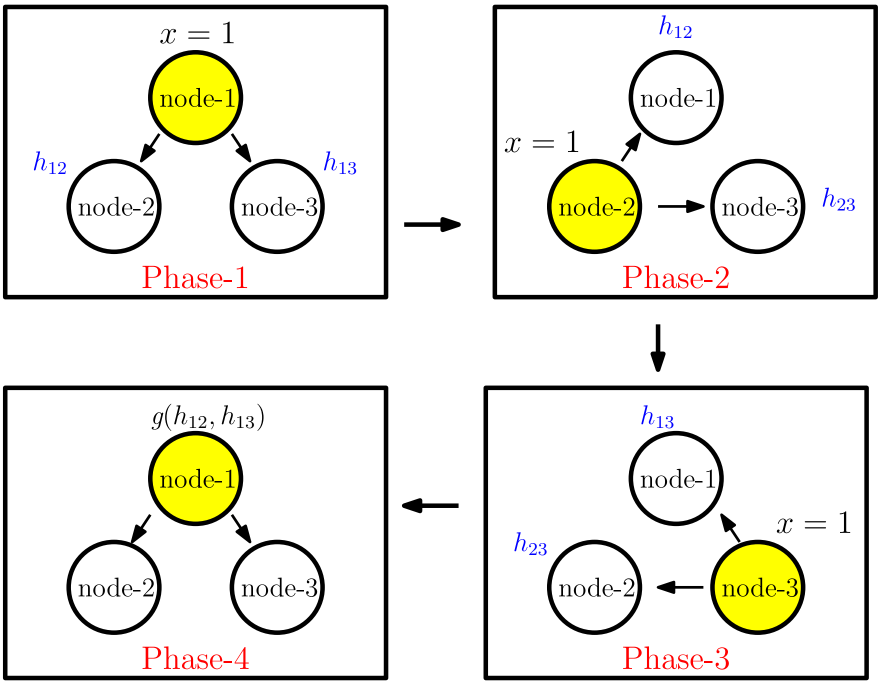


Figure 1.7: Systematic protocol to observe the reciprocal channels (**Phase-1** to **Phase-3**) and CSR sharing (**Phase-4**).

radio devices are designed to transmit baseband symbols from finite constellations, such as Quadrature-Amplitude-Modulation (QAM), Phase-Shift-Keying (PSK), etc, the broadcast signal $g(h_{12}, h_{13})$ cannot be transmitted as an arbitrary complex number in an unquantized fashion. Although the facilitator can quantize $g(h_{12}, h_{13})$ to binary sequences of large block-lengths and then map those bits to complex constellations to reduce the quantization error, such a strategy may not be applicable when the CSR has to be shared within a short coherence-block.

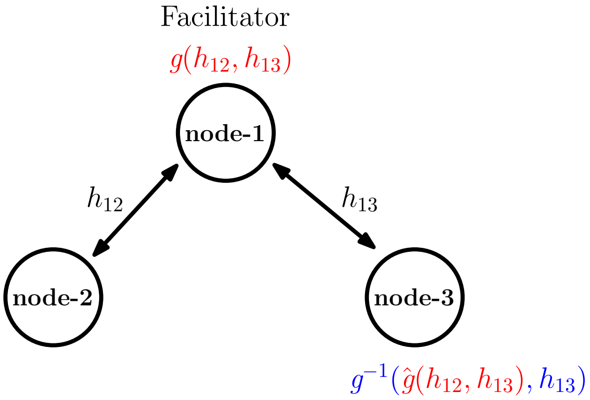


Figure 1.8: Depiction of CSR Sharing by the facilitator to node-3, $g^{-1}()$ is the inverse operation of $g()$ and $\hat{g}(h_{12}, h_{13})$ is the estimate of $g(h_{12}, h_{13})$.

In this work, we consider Algebraic- Symmetrically Quantized GSK (A-SQGSK) protocol which is the only protocol that preserves the confidentiality in the class of group consensus based GSK generation [RH19, JJR21]. Once, all the three nodes have the CSR available with them using A-SQGSK, they employ the extension of consensus algorithm discussed in

Section 1.2 and harvest a key. Although A-SQGSK protocol preserves the confidentiality but no consensus algorithm is presented yet which is tightly tied to it. We explore multi-level quantization aspects as well as some enhancements to it in terms of key-rate and error-rate.

1.6 Motivation and Contributions

1.6.1 Part- I: Multi-Level Quantization

For the case of two-user key generation, we identify that quantizers that are designed using marginal distributions (at individual nodes) cannot reap maximum benefits from wireless channels. This is because the generated secret-key does not exhibit maximum entropy since the distribution of the observations in consensus is different from their marginal distribution, especially at low and medium signal-to-noise ratio (SNR) values. Incorporating the above observations, in this part, we make the following two-fold contributions:

[**C1.1**:] We explore whether multi-level quantization schemes can be designed to generate secret-keys with maximum entropy provided the joint probability distribution function (PDF) of the CSR is available. Given the number of levels for quantization and the knowledge of the joint PDF, we formulate a constrained optimization problem to maximize the symbol-rate of the keys with strict constraints on the entropy and the mismatch-rate of the generated keys. Towards solving the optimization problem, we propose an iterative algorithm, referred to as the Entropy Maximization Error Minimization (EM-EM) algorithm, which carefully introduces guard bands in $\mathbb{R}^2$, to satisfy the underlying constraints. Unlike existing approaches on multi-level quantization, the EM-EM algorithm guarantees secret-keys with entropy of b-bits per sample in consensus when using a 2^b-level quantization. To emphasize the novelty, in Table 1.1, we point out the differences between our work over existing contributions. Our algorithm can be applied on a wide range of joint distributions, and thus can serve as a software package to design multi-level quantizers for key generation [JJR21, HJR20].

[**C1.2**:] On the CSR provided by A-SQGSK for the case of three-user network, formulated EM-EM framework is applied. As presented EM-EM framework is applicable only on two-dimensional joint PDF, in context to A-SQGSK protocol for multiple nodes, the three-dimensional PDF can not be applied directly to EM-EM. A relaxed criterion is presented in order to make EM-EM amenable to A-SQGSK. Given that the proposed multi-level consensus algorithm is closely coupled with the protocol used to exchange the CSR, we lay out design rules to jointly choose the parameters of the A-SQGSK protocol and the EM-EM algorithm as a function of underlying signal-to-noise-ratio and the required mismatch rate on the generated GSK.

Table 1.1: Novelty of the Proposed EM-EM Algorithm with respect to Existing Contributions

Reference	**Contributions on consensus algorithms**	**Limitation with respect to our work**
[Max60]	Max-Lloyd quantizer	Minimizes the distortion between input and quantized version. Not relevant to key generation.
[YRS06]	Marginal distribution based equiprobable quantizer	Does not guarantee maximum entropy since the distribution of samples in consensus is different from marginal distribution
[MTM+08]	Two-level quantizer and consensus algorithm	Two-level quantizer using the mean of data set and consecutive excursion length. Multiple levels of quantization not explored.
[SH11]	Optimal guard band quantizer	Placement of guard bands to bound symbol-error-rate. Maximum entropy not guaranteed.
[HHL17]	Vector quantization based multi-level quantizer	Handles temporally-correlated channel realizations. Specific to Gaussian distributions. In contrast, EM-EM algorithm is applicable on any joint distribution.

1.6.2 Part- II: Key-Rate Enhancement and Reconciliation

Although A-SQGSK protocol provides zero leakage, we point out that its key-rate is low because the protocol uses only one of the channels in the network as the CSR. Furthermore, in the context of two-user key generation, it is well known that reconciliation algorithms, e.g., using low-density parity check (LDPC) codes, can be used to arrive at zero mismatch rate among the keys at the nodes. Towards employing such reconciliation methods in A-SQGSK, we observe that aspects of generating log-likelihood ratios (LLRs) of the secret bits using the CSR samples have not been addressed hitherto. Identifying the above mentioned limitations, we make the following contributions in this part:

[**C2.1**:] For the framework of A-SQGSK protocol, we propose an opportunistic CSR selection strategy wherein the randomness offered by two channels are utilized to synthesize a GSK without compromising the confidentiality feature. We show that our approach provides higher key-rate than the protocol proposed in [JJR21, HJR20, RH19].

[**C2.2**:] Furthermore, when both the channels are amenable to key generation, we emphasize that using both would lead to compromise in the confidentiality, and as a result, we present a likelihood based strategy to choose one of them so as to minimize the mismatch rate among the nodes [JH21a, JH21b].

[**C2.3**:] Finally, for the A-SQGSK protocol, we also propose an LLR generation scheme, using which all the nodes employ a reconciliation algorithm. We also use LDPC based reconciliation to validate the effectiveness of our LLR generation scheme. Although LLR generation on CSR samples for two-user key generation is well known, we highlight that the statistics of the underlying noise is different at different set of nodes in the A-SQGSK protocol [JJR21]. Therefore, the proposed LLR generation scheme is a non-trivial contribution in GSK generation [JH21a, JH21b].

1.7 Thesis Organization

In the rest of this thesis, Chapter 2 and Chapter 3 cover EM-EM framework for two-user and three-user network while Chapter 4 covers our study on key-rate enhancement and reconciliation techniques. In Chapter 2, we propose an iterative algorithm to solve the presented optimization problem in order to maximise the symbol-rate and entropy of the key synthesised. We also present the pseudo-algorithm of different blocks which make the EM-EM algorithm. In Chapter 3, we extend the idea the of presented EM-EM algorithm for the case of three-user network, specifically, we highlight that the proposed algorithm is only applicable on a two-dimensional PDF and hence we present a relaxed design criterion for this network. Chapter 4 deals with the idea of opportunistic selection for key-rate improvement and LLR generation for minimising the error-rate for the network of three-users. Finally, a concluding summary and directions for further research are provided in Chapter 5.

1.8 Notations

We use $\mathbb{Z}$, $\mathbb{Z}[i]$, $\mathbb{N}$ and $\mathbb{C}$ to denote the set of all integers, Gaussian integers, natural numbers and complex numbers, respectively, where $i = \sqrt{-1}$. A circularly symmetric complex Gaussian random variable with mean 0 and variance σ^2 is represented as $x \sim \mathcal{CN}(0, \sigma^2)$. The set $\{0, 1, 2, \ldots, p-1\}$, for some integer $p > 1$, is denoted by $\mathbb{Z}_p$. The term $I(x; y)$ denotes the mutual information between two random variables x and y, and the term $H(x)$ denotes the entropy of a discrete random variable x. Given two sets $\mathcal{S}_1 \subset \mathbb{C}$ and $\mathcal{S}_2 \subset \mathbb{C}$, the term $\mathcal{S}_1 \bigoplus \mathcal{S}_2 = \{s_1 + s_2 \mid s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2\}$ denotes their direct sum. We refer to the additive white Gaussian noise using its acronym AWGN. The notation $|\mathcal{S}|$ represents the number of elements in the set $\mathcal{S}$. We use $\text{Prob}(\cdot)$ to represent the regular probability operator. We use the notation $[n]$ to represent the set of integers $\{1, 2, \ldots, n\}$. Given a two-dimensional probability density function $P(x, y)$ of continuous random variables X and Y, the probability that the pair lie in a given range is denoted by $\int_{a_j^-}^{a_j^+} \int_{a_k^-}^{a_k^+} P(x, y) dx dy$. In the special case of discrete random variables X and Y, the integral will collapse to summation of mass points in the given interval as $\sum_{a_j^-}^{a_j^+} \sum_{a_k^-}^{a_k^+} P(x, y)$, where $P(x, y)$ denotes the joint probability mass function $P(X = x, Y = y)$.

Part-I: Multi-Level Quantization Aspects of Two-User and A-SQGSK based Three-User Key Generation

Chapter 2

EM-EM Framework

2.1 Introduction

Dynamic key generation using wireless channel realizations as the CSR is a popular method to harvest shared secret-keys between multiple nodes in a wireless network [Mau93, AC93, MW03, MTM$^+$08]. While the main objective of this framework is to identify wireless channels that offer significant randomness, the subsequent goal is to design efficient consensus algorithms to harvest shared secret-keys with high-rate, measured in bits per second. One of the early contributions on consensus algorithms was from [MTM$^+$08], which proposed a two-level quantizer to derive secret bits in an unauthenticated channel as discussed in Section 1.2. Subsequently, [JPC$^+$09, YMR$^+$10] have proposed enhancements over the idea in [MTM$^+$08] to maximize the entropy of the generated key with two-level quantizer. Further, [YRS06, ZWCM10] have addressed scalar multi-level quantization schemes to generate more than one bit per sample, whereas [SH11] has proposed methods to reduce the mismatch-rate between the generated keys. Recently, [HHL17] has also explored vector quantization methods to achieve consensus on channels with correlated variations over time.

A common thread tying existing multi-level consensus algorithms is that the quantization boundaries are designed using the probability distribution on the observations at individual nodes, henceforth referred to as marginal distribution of the CSR. We identify that quantizers that are designed using marginal distributions cannot reap maximum benefits from wireless channels. This is because the generated secret-key does not exhibit maximum entropy since the distribution of the observations in consensus is not equal to the marginal distribution, especially at low and medium SNR values. As a result, the generated secret-key has to be further processed to maximize its entropy, which is turn is an overhead on the key generation process. Identifying these limitations, in this work, we explore whether multi-level quantization schemes can be designed to generate secret-keys with maximum entropy provided the joint PDF of the CSR is available. Potential applications of such a framework include device-to-device communications, wherein user-equipments (UEs) that intend to harvest secret-keys seek the assistance of their parent base-station to design a multi-level quantizer by providing their observations in a secure manner.

2.2 System Model and Problem Statement

We consider a wireless channel model between two nodes, namely: node-1 and node-2, which intend to harvest shared secret-keys by observing the channel realizations between them. We assume that the wireless channel is frequency-flat and remain quasi-static for a period of atleast two channel uses. Henceforth, we refer to this quasi-static block as coherence-block. Upon transmission of a probing signal $x = \sqrt{P}$ from node-1, node-2 receives a real baseband symbol[1] given by

$$y_B(l) = \sqrt{P}h(l) + n_B(l), \tag{2.1}$$

where P is the transmission power of the probing signals, $h(l) \in \mathbb{R}$ is the channel witnessed by node-2 during the coherence-block index l, $n_B(l)$ is the additive white Gaussian noise (AWGN), distributed as $\mathcal{N}(0, \sigma^2)$. Similarly, upon transmission of a probing signal $x = \sqrt{P}$ from node-2, node-1 receives a real baseband symbol given by

$$y_A(l) = \sqrt{P}h(l) + n_A(l), \tag{2.2}$$

where $n_A(l)$ is the AWGN, distributed as $\mathcal{N}(0, \sigma^2)$. We assume that the wireless channels from node-1-to-node-2 and node-2-to-node-1 exhibits perfect reciprocity, and also assume that $h(l)$ is a continuous random variable with PDF $P(h(l))$ over the support $(-\infty, +\infty)$. The average signal-to-noise-ratio (SNR) of the channel is defined as $\text{SNR} = \frac{P}{\sigma^2}$.

A straightforward technique to harvest shared secret-keys is to apply two-level quantization on the samples $\{y_A(l) \mid l = 1, 2, \ldots L\}$ at node-1 and $\{y_B(l) \mid l = 1, 2, \ldots L\}$ at node-2, as proposed in [MTM⁺08, JPC⁺09, YMR⁺10]. Here, L represents the number of probing signals exchanged between node-1 and node-2. Although this idea is effective, its limitation is its inability to generate more than one bit per sample when the channel offers significant randomness. A natural way to increase the number of bits per sample is to apply multi-level quantization on each sample, where the number of levels must be chosen depending on the common randomness offered by the wireless channel. To generate b bits per sample, for $b \in \mathbb{N}$, we formally define a 2^b-level quantizer as follows.

Definition 1. *A 2^b-level quantizer, denoted by $\mathcal{Q}_b \subset \mathbb{R}^2$, is defined by a set of 2^b pairs of real numbers, given by $\mathcal{Q}_b = \{(a_j^-, a_j^+) \mid j = 1, 2, \ldots, 2^b\}$, satisfying the following constraints:*

- *$a_j^- < a_j^+$ for each $1 \leq j \leq 2^b$,*
- *$a_j^+ \leq a_{j+1}^-$ for $1 \leq j \leq 2^b - 1$, and*
- *$a_1^- = -\infty$ and $a_{2^b}^+ = \infty$.*

[1] We consider real baseband models only for the narrative purpose. However, complex baseband channels with independent and identically distributed in-phase and quadrature components can also be reduced to this model.

Henceforth, we refer to the region (a_j^-, a_j^+) by a fixed representative in that region, denoted by $a_j \in (a_j^-, a_j^+)$. We call this set of representatives $\{a_j \mid 1 \leq j \leq 2^b\}$ as the finite constellation $\mathcal{C} \subset \mathbb{R}$ of size 2^b. We use $\mathcal{G}_{j,j+1} \triangleq (a_j^+, a_{j+1}^-)$ as the guard band separating the j-th and the $(j+1)$-th region, for $1 \leq j \leq 2^b - 1$. We also use $\Delta_j \triangleq a_{j+1}^- - a_j^+$ to represent the width of $\mathcal{G}_{j,j+1}$.

Definition 2. *A real number $y \in \mathbb{R}$ is quantized to $2^b + 1$ discrete values, denoted by $\bar{\mathcal{C}} = \mathcal{C} \cup \{X\}$, based on the following rule*

$$\mathcal{Q}_b(y) = \begin{cases} a_j, & \text{if } y \in (a_j^-, a_j^+) \\ X, & \text{if } y \in (a_j^+, a_{j+1}^-) \text{ for } 1 \leq j \leq 2^b - 1, \end{cases} \tag{2.3}$$

where a_j is the chosen representative of the region (a_j^-, a_j^+), and the symbol X is used to represent the samples lying in any of the guard bands.

Based on Definition 1 and Definition 2, the notation $\mathcal{Q}_b$ is used to represent a quantizer, whereas the notation $\mathcal{Q}_b(\cdot)$ is used to represent evaluation of the quantizer on a given real number. In the next section, we discuss a consensus algorithm using a quantizer $\mathcal{Q}_b$.

2.2.1 Consensus Phase

node-1 and node-2 agree upon a quantizer $\mathcal{Q}_b$, as presented in Definition 1. They collect a sufficiently large number of observations, denoted by $\mathcal{Y}_A = \{y_A(l) \mid l = 1, 2, \ldots, L\}$ and $\mathcal{Y}_B = \{y_B(l) \mid l = 1, 2, \ldots, L\}$, respectively, over L coherence-blocks. To achieve consensus, the two nodes execute the following protocol:

- node-1 obtains the set $\{\mathcal{Q}_b(y_A(l)) \mid y_A(l) \in \mathcal{Y}_A\}$, and then shares the index values $\mathcal{I}_A = \{l \in [L] \mid \mathcal{Q}_b(y_A(l)) \neq X\}$ to node-2, where $[L] = \{1, 2, \ldots, L\}$.
- node-2 obtains the set $\{\mathcal{Q}_b(y_B(l)) \mid y_B(l) \in \mathcal{Y}_B\}$, and then computes the corresponding set of index values $\mathcal{I}_B = \{l \in [L] \mid \mathcal{Q}_b(y_B(l)) \neq X\}$. Subsequently, node-2 shares $\mathcal{I}_C = \mathcal{I}_B \cap \mathcal{I}_A$ with node-1, where $\mathcal{I}_C$ denotes the set of index values in consensus between node-1 and node-2. We use N to denote the length of $\mathcal{I}_C$, i.e., $N = |\mathcal{I}_C|$.
- Using $\mathcal{I}_C$, node-1 and node-2 generate the following sequences $\mathcal{K}_A = \{\mathcal{Q}_b(y_A(l)) \mid l \in \mathcal{I}_C\}$, and $\mathcal{K}_B = \{\mathcal{Q}_b(y_B(l)) \mid l \in \mathcal{I}_C\}$. Note that both $\mathcal{K}_A$ and $\mathcal{K}_B$ are N-length sequences over the alphabet $\mathcal{C}$.

In the rest of this thesis, we drop the reference to the sample index l, and refer to the observations at node-1 and node-2 as two correlated random variables y_A and y_B with an underlying joint PDF $P(y_A, y_B)$.

2.2.2 Design Criteria on $\mathcal{Q}_b$

Based on the above consensus algorithm, the following properties are desired on $\mathcal{K}_A$ and $\mathcal{K}_B$:

1. The symbol-rate, given by $\frac{N}{L}$, which captures the fraction of samples in consensus between node-1 and node-2, is maximum.
2. The entropy of the two sequences must be maximum, i.e., $H(\mathcal{K}_A) = Nb$ and $H(\mathcal{K}_B) = Nb$, where $H(\mathcal{K}_A)$ and $H(\mathcal{K}_B)$, respectively denote the joint entropy of N random variables over $\mathcal{C}$.
3. The fraction of pair-wise disagreements between the two sequences must be negligible, i.e., $\frac{1}{N}d_H(\mathcal{K}_A, \mathcal{K}_B) \leq \beta$, where β is a small number of our choice, and $d_H(\mathcal{K}_A, \mathcal{K}_B)$ denotes the Hamming distance between $\mathcal{K}_A$ and $\mathcal{K}_B$.

It is apparent that 1, 2 and 3 are desired on the keys synthesised by the quantizer $\mathcal{Q}_b$, but there is a direct trade-off involved in optimising 1 and 3. Hence, we design a quantizer $\mathcal{Q}_b$ based on $P(y_A, y_B)$ such that the fraction of pair-wise disagreements between the two synthesised sequences are upper bounded by β, and symbol-rate is conditionally maximised on this upper bound on the error-rate. To capture the criterion of symbol-rate, the consensus probability, denoted by $p_c(\mathcal{Q}_b)$, is given by,

$$\begin{aligned} p_c(\mathcal{Q}_b) &= \text{Prob}(\mathcal{Q}_b(y_A) \in \mathcal{C}, \mathcal{Q}_b(y_B), \in \mathcal{C}) \\ &= \sum_{j=1}^{2^b}\sum_{k=1}^{2^b}\int_{a_j^-}^{a_j^+}\int_{a_k^-}^{a_k^+} P(y_A, y_B) dy_A dy_B. \end{aligned} \tag{2.4}$$

Out of the L real samples that undergo consensus, the average number of samples in agreement after the consensus phase is p_cL. Therefore, the symbol-rate of the quantizer $\mathcal{Q}_b$ is

$$\frac{N}{L} = p_c(\mathcal{Q}_b). \tag{2.5}$$

It is clear that $\mathcal{Q}_b(y_A) \in \mathcal{K}_A$ if and only if $\mathcal{Q}_b(y_A) \in \mathcal{C}$ and $\mathcal{Q}_b(y_B) \in \mathcal{C}$. As a result, the entropy of $\mathcal{Q}_b(y_A) \in \mathcal{K}_A$ is

$$H(\mathcal{Q}_b(y_A) \mid \mathcal{Q}_b(y_B), \mathcal{Q}_b(y_A) \in \mathcal{C}) = -\sum_{j=1}^{2^b} g_j \log_2 g_j, \tag{2.6}$$

where

$$\begin{aligned} g_j &= \text{Prob}(\mathcal{Q}_b(y_A) = a_j \mid \mathcal{Q}_b(y_A), \mathcal{Q}_b(y_B) \in \mathcal{C}) \\ &= \frac{\sum_{k=1}^{2^b} \int_{a_j^-}^{a_j^+} \int_{a_k^-}^{a_k^+} P(y_A, y_B) dy_A dy_B}{\sum_{j=1}^{2^b} \sum_{k=1}^{2^b} \int_{a_j^-}^{a_j^+} \int_{a_k^-}^{a_k^+} P(y_A, y_B) dy_A dy_B}. \end{aligned} \tag{2.7}$$

Since the samples after consensus are expected to be random, it is desired to have $H(\mathcal{Q}_b(y_A) \mid \mathcal{Q}_b(y_B), \mathcal{Q}_b(y_A) \in \mathcal{C}) = b$, when $\mathcal{C}$ comprises of 2^b levels.

Two samples, $\mathcal{Q}_b(y_A)$ and $\mathcal{Q}_b(y_B)$ that are already in consensus, i.e., $\mathcal{Q}_b(y_A), \mathcal{Q}_b(y_B) \in \mathcal{C}$, are said to be in error if $\mathcal{Q}_b(y_A) \neq \mathcal{Q}_b(y_B)$. Formally, using the joint PDF, the symbol error rate (SER) among the samples in consensus is given by

$$\begin{aligned} SER(\mathcal{Q}_b) &= \text{Prob}(\mathcal{Q}_b(y_A) \neq \mathcal{Q}_b(y_B) \mid \mathcal{Q}_b(y_A), \mathcal{Q}_b(y_B) \in \mathcal{C}) \\ &= \frac{p_{c,m}(\mathcal{Q}_b)}{p_c(\mathcal{Q}_b)}, \end{aligned} \tag{2.8}$$

where

$$p_{c,m}(\mathcal{Q}_b) = \sum_{j=1}^{2^b} \sum_{k \neq j}^{2^b} \int_{a_j^-}^{a_j^+} \int_{a_k^-}^{a_k^+} P(y_A, y_B) dy_A dy_B \tag{2.9}$$

and $p_c(\mathcal{Q}_b)$ is given in (2.4). The quantizer $\mathcal{Q}_b$ must be designed such that $SER(\mathcal{Q}_b)$ is upper bounded by a negligible number, say $\beta > 0$. In practice, the choice of β depends on the error-correction capability of the channel codes which are subsequently used to correct the residual errors in the secret-keys. Since $SER(\mathcal{Q}_b)$ is inversely proportional to $p_c(\mathcal{Q}_b)$, we take the approach of maximizing $p_c(\mathcal{Q}_b)$ for a given upper bound on $p_{c,m}(\mathcal{Q}_b)$.

Keeping in view of the expressions in (2.5), (2.6) and (2.8), the proposed objective function on the design of quantizer is formally given in Problem 1.

Problem 1. *Solve*

$$\arg\max_{\mathcal{Q}_b} \quad p_c(\mathcal{Q}_b)$$

such that

$$H(\mathcal{Q}_b(y_A) \mid \mathcal{Q}_b(y_B), \mathcal{Q}_b(y_A) \in \mathcal{C}) = b, \tag{2.10}$$

$$p_{c,m}(\mathcal{Q}_b) \leq \eta, \tag{2.11}$$

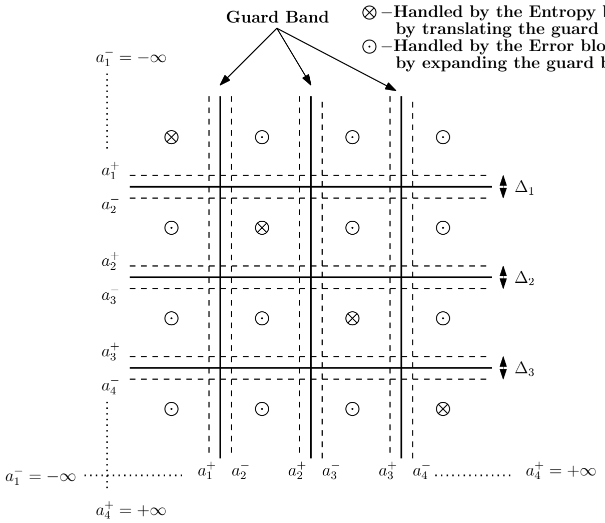


Figure 2.1: Depiction of the proposed two-dimensional approach to solve multi-level quantization for key-generation. Unlike existing approaches, joint PDF is exploited to identify the placements and the widths of the guard bands.

where $\eta > 0$ is a given negligible number.

The constrained optimization in Problem 1 must be solved for a given set of inputs $\{P(y_A, y_B), \eta, b, \sigma^2\}$. With $p_c(\mathfrak{Q}_b)$ denoting the symbol-rate offered by the quantizer, the SER offered by it is upper bounded by $\frac{\eta}{p_c(\mathfrak{Q}_b)}$. Therefore, one way to obtain a quantizer satisfying the upper bound $SER(\mathfrak{Q}_b) \leq \beta$, for some $\beta > 0$, is to solve Problem 1 for various values of $\eta > 0$, and then choose the one which satisfies $\frac{\eta}{p_c(\mathfrak{Q}_b)} \leq \beta$.

In the next section, we provide an iterative algorithm to obtain a quantizer that satisfies the constraints in (2.10)-(2.11) for a given $\eta > 0$.

2.3 EM-EM Algorithm

Towards solving Problem 1, we present an iterative algorithm, referred to as the EM-EM algorithm, which carefully introduces guard bands in $\mathbb{R}^2$, as shown in Fig. 2.1, to satisfy the underlying constraints. As shown in Fig. 2.2, our algorithm comprises of four blocks, namely: (i) the initialization block, which feeds an initial set of boundaries $\{(a_j^-, a_j^+) \mid \forall j\}$ for a given $b \in \mathbb{N}$, (ii) the entropy block, which handles the constraint in (2.10), (iii) the error block, which addresses the constraint in (2.11), and finally (iv) the refining block, which corrects the suboptimality of the entropy block in achieving equality constraint on conditional entropy.

Given the inputs $\{P(y_A, y_B), \eta, b, \sigma^2\}$, our approach is to solve Problem 1 by assuming $\sigma^2 = 0$, i.e., when $y_A(l) = y_B(l) = h(l)$ in (2.1) and (2.2), and then use the corresponding quantizer as the initial set of boundaries. Since $\sigma^2 = 0$, the initial boundaries $\{(a_j^-, a_j^+) \mid \forall j\}$ will be such that $\Delta_j = 0$ for each j. As a result, the constraint on $p_{c,m}(\mathcal{Q}_b)$ will not be satisfied when $\sigma^2 > 0$. To circumvent this problem, we feed these boundaries to the error block, which increases the width of the j-th guard band, for $1 \leq j \leq 2^b - 1$, as

$$(a_j^+, a_{j+1}^-) \leftarrow (a_j^+ - \theta_j, a_{j+1}^- + \theta_j),$$

for some $\theta_j \geq 0$, in order to satisfy the constraint $p_{c,m}(\mathcal{Q}_b) \leq \eta$. Here, the notation $\leftarrow$ is used to represent the update operator on the guard bands. Subsequently, since the conditional entropy might have been disturbed, the updated boundaries from the error block are fed to the entropy block, which translates the j-th guard band, for $1 \leq j \leq 2^b - 1$, as

$$(a_j^+, a_{j+1}^-) \leftarrow (a_j^+ + \phi_j, a_{j+1}^- + \phi_j),$$

for some $\phi_j \in \mathbb{R}$, to satisfy the constraint on conditional entropy. This way, iterations between the error block and the entropy block continue until the constraints on the conditional entropy and $p_{c,m}(\mathcal{Q}_b)$ are met. At the end of the algorithm, using the final set of boundaries $\{(a_j^-, a_j^+) \mid \forall j\}$, we compute $p_c(\mathcal{Q}_b)$ using $P(y_A, y_B)$. With that the SER achieved by the EM-EM algorithm is upper bounded by $\frac{\eta}{p_c(\mathcal{Q}_b)}$.

Definition 3. *Using the EM-EM algorithm, the pair (b, β), for a given $b \in \mathbb{N}$ and $\beta > 0$, is said to be feasible if there exists an $\eta \leq \beta$ such that $\frac{\eta}{p_c(\mathcal{Q}_b)} \leq \beta$.*

In the rest of this section, we explain the functionality of each block by providing the rationale behind its design.

2.3.1 Initialization Block

For a given $b \in \mathbb{N}$, we obtain a quantizer $\{(a_j^-, a_j^+) \mid \forall j\}$, which is optimized to $\sigma^2 = 0$, i.e., when y_A and y_B are identical. For this extreme case, the boundaries are obtained by equating

$$P_j = \int_{a_j^-}^{a_j^+} P(y_A) dy_A \tag{2.12}$$

to $\frac{1}{2^b}$, for $1 \leq j \leq 2^b$, where $a_1^- = -\infty$, $a_{2^b}^+ = +\infty$ and $P(y_A)$ is the PDF of y_A. A pseudocode description to solve (2.12) is given in Algorithm 1. It is straightforward to observe that $\Delta_j = 0$, for each j, since $\sigma^2 = 0$.

Algorithm 1. *Initialization Block: The case when $\sigma^2 = 0$*

Input: *$P(y_B)$, b, and step-size $\theta > 0$*

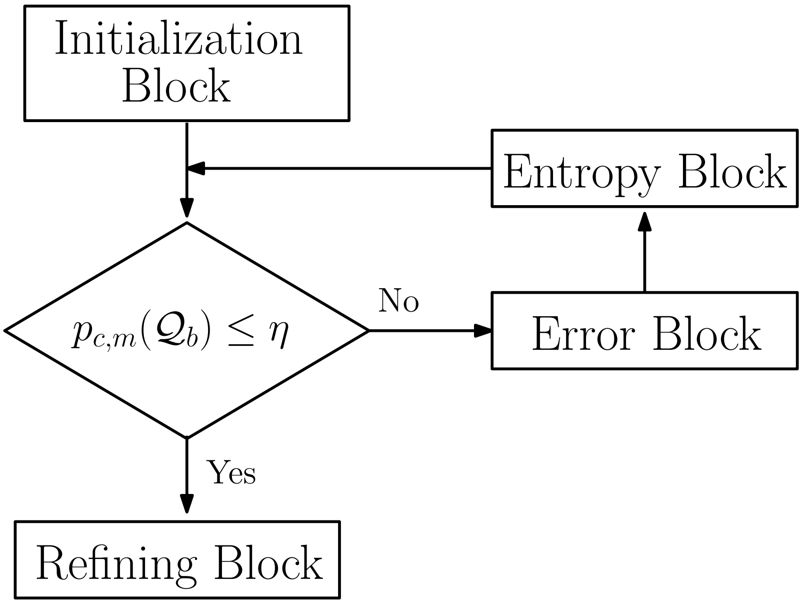


Figure 2.2: Depiction of the proposed EM-EM algorithm to generate a multi-level quantizer $\mathcal{Q}_b$, which is matched to the joint PDF $P(y_A, y_B)$. The inputs to the algorithm are $\{P(y_A, y_B), \eta, b, \sigma^2\}$, where 2^b is the number of levels, σ^2 is the variance of the additive noise, and η is the constraint on $p_{c,m}(\mathcal{Q}_b)$.

```
Output: $\{(a_j^-, a_j^+) \mid \forall j\}$
Initialize $a_1^- = -\infty$ and $a_1^+ = -\infty$
for $j = 1 \rightarrow 2^b - 1$ do
    Compute $P_j$ using (2.12)
    while $P_j < \frac{1}{2^b}$ do
        $a_j^+ \leftarrow a_j^+ + \theta$
        update $P_j$ using (2.12)
    end while
    $a_{j+1}^- = a_j^+$; $a_{j+1}^+ = a_j^+$
end for
$a_{2^b}^+ = +\infty$
```

2.3.2 Error Block

The objective of this block is to increase the widths of guard bands to satisfy the constraint

$$\sum_{j=1}^{2^b} \sum_{k \neq j}^{2^b} \int_{a_j^-}^{a_j^+} \int_{a_k^-}^{a_k^+} P(y_A, y_B) dy_A dy_B \leq \eta. \tag{2.13}$$

$$\delta_i = \int_{a_i^-}^{a_i^+} \int_{a_{i+1}^-}^{a_{i+1}^+} P(y_A, y_B) dy_A dy_B + \int_{a_{i+1}^-}^{a_{i+1}^+} \int_{a_i^-}^{a_i^+} P(y_A, y_B) dy_A dy_B \leq \frac{\eta}{2^b - 1} \tag{2.14}$$

Out of the $2^{2b} - 2^b$ terms in (2.13), the dominant terms are $\int_{a_j^-}^{a_j^+} \int_{a_k^-}^{a_k^+} P(y_A, y_B) dy_A dy_B$ such that $|k - j| = 1$. Therefore, instead of addressing the constraint in (2.13), the error block satisfies the constraint on δ_i given in (2.14), for each $1 \leq i \leq 2^b - 1$. Using $P(y_A, y_B)$, we compute the set $\{\delta_i \mid i = 1, 2, \ldots, 2^b - 1\}$. Starting from $i = 1$ to $2^b - 1$, the error block increases the width of $\mathcal{G}_{i,i+1}$ until the constraint on (2.14) is satisfied, as shown in Algorithm 2. Proposition 1 provides guarantee that increasing the width of $\mathcal{G}_{i,i+1}$ reduces δ_i. If δ_i, for some i, already satisfies the constraint, then $\mathcal{G}_{i,i+1}$ remains unchanged. This way, the error block increases the width of each guard band prudently based on $P(y_A, y_B)$.

Proposition 1. *Increasing the width of $\mathcal{G}_{i,i+1}$ reduces δ_i.*

Proof. This result follows from the definition of probability distribution function. □

Algorithm 2. *Error Block*

Input: $\{(a_j^-, a_j^+) \mid \forall j\}$, *$P(y_B, y_C)$, b, η, and step-size* $\theta > 0$

Output: $\{(a_j^-, a_j^+) \mid \forall j\}$

```
for $i = 1 \rightarrow 2^b - 1$ do
    Compute $\delta_i$ using (2.14)
    while $\delta_i > \frac{\eta}{(2^b - 1)}$ do
        $a_i^+ \leftarrow a_i^+ - \theta$; $a_{i+1}^- \leftarrow a_{i+1}^- + \theta$
        update $\delta_i$ using (2.14)
    end while
end for
```

2.3.3 Entropy Block

The role of the entropy block is to maximize the conditional entropy in (2.6). Based on the expressions of $\{g_j \mid 1 \leq j \leq 2^b\}$ given in (2.7), the entropy block translates the guard bands locally such that $g_j = \frac{1}{2^b}$, for each j. Unlike the error block, this block does not increase the widths of the guard bands; instead it translates them either to left or right to maximize the conditional entropy. Among the 2^b terms in the numerator of each g_j, terms of the form

$$\int_{a_j^-}^{a_j^+} \int_{a_k^-}^{a_k^+} P(y_A, y_B) dy_A dy_B, \text{ for } j \neq k,$$

are already driven to negligible values by the error block. As a result, the entropy block neglects such terms, and considers an approximation on g_j, denoted by $\tilde{g}_j$, as

$$\tilde{g}_j = \frac{\alpha_j}{\alpha_1 + \alpha_2 + ... + \alpha_{2^b}}, \tag{2.15}$$

where

$$\alpha_j = \int_{a_j^-}^{a_j^+} \int_{a_j^-}^{a_j^+} P(y_A, y_B) dy_A dy_B. \tag{2.16}$$

Using the boundaries received from the error block, $\tilde{g}_j$ is computed as in (2.15) sequentially from $j = 1$ to 2^b. For a given j, if $\tilde{g}_j$ is less than $\frac{1}{2^b}$, then the corresponding guard band is translated to right until $\tilde{g}_j = \frac{1}{2^b}$, as shown in Algorithm 3. On the other hand, if $\tilde{g}_j$ is more than $\frac{1}{2^b}$, then the corresponding guard band is translated to left by an appropriate amount until the equality $\tilde{g}_j = \frac{1}{2^b}$ is met. The following proposition shows that the direction of translation depends on whether $\tilde{g}_j$ is more or less than $\frac{1}{2^b}$.

Proposition 2. *If $\tilde{g}_j$ is less than $\frac{1}{2^b}$, then shifting the guard band to right increases $\tilde{g}_j$. Similarly, if $\tilde{g}_j$ is more than $\frac{1}{2^b}$, then shifting the guard band to left decreases $\tilde{g}_j$.*

Proof. We provide a proof to show that translating the guard band to right increases the corresponding value of $\tilde{g}_j$. The result for the other direction can be proved in a similar manner. Before translating the guard band $\mathcal{G}_{j,j+1} = (a_j^+, a_{j+1}^-)$, let $\tilde{g}_j$ be computed as in (2.15) using the initial set of values given by $\{\alpha_j \mid \forall j\}$. If this guard band is translated as $(a_j^+ + \gamma, a_{j+1}^- + \gamma)$, for some $\gamma > 0$, then based on the joint PDF, it is straightforward to observe that α_j increases to $\alpha_j + \gamma_{\alpha_j}$, for some $\gamma_{\alpha_j} > 0$, and α_{j+1} decreases to $\alpha_{j+1} - \gamma_{\alpha_{j+1}}$, for some $\gamma_{\alpha_{j+1}} > 0$, and the rest of the terms $\{\alpha_k, \mid k \neq j, k \neq j+1\}$ remain unchanged. As a result, the updated version of $\tilde{g}_j$ is of the form

$$\tilde{g}_j = \frac{\alpha_j + \gamma_{\alpha_j}}{(\sum_{k=1}^{2^b} \alpha_k) + \gamma_{\alpha_j} - \gamma_{\alpha_{j+1}}}. \tag{2.17}$$

Since $\tilde{g}_j$ is always less than one, it is straightforward to prove that the updated value in (2.17) will be more than $\tilde{g}_j$ for any $\gamma_{\alpha_j} \geq 0$, $\gamma_{\alpha_{j+1}} \geq 0$. This completes the proof. □

Algorithm 3. *Entropy Block: Maximizing conditional entropy*

```
Input: $\{(a_j^-, a_j^+) \mid \forall j\}$, $P(y_B, y_C)$, b, and step-size $\theta > 0$
Output: $\{(a_j^-, a_j^+) \mid \forall j\}$
for $j = 1 \rightarrow (2^b - 1)$ do
    Compute $\tilde{g}_j$ using (2.15)
    if $\tilde{g}_j < \frac{1}{2^b}$ then
        while $\tilde{g}_j < \frac{1}{2^b}$ do
```

```
            $a_j^+ \leftarrow a_j^+ + \theta$; $a_{j+1}^- \leftarrow a_{j+1}^- + \theta$
            update $\tilde{g}_j$ using (2.15)
        end while
    else $\tilde{g}_j > \frac{1}{2^b}$
        while $\tilde{g}_j > \frac{1}{2^b}$ do
            $a_j^+ \leftarrow a_j^+ - \theta$; $a_{j+1}^- \leftarrow a_{j+1}^- - \theta$
            update $\tilde{g}_j$ using (2.15)
        end while
    end if
end for
```

2.3.4 Refining Block

Notice that the entropy block forces each $\tilde{g}_j$ to take $\frac{1}{2^b}$ from $j = 1$ to 2^b in a sequential manner, and as a result, the overall entropy $-\sum_j \tilde{g}_j \log_2(\tilde{g}_j)$ may not be b after $\tilde{g}_{2^b}$ is updated. This is because the process of forcing $\tilde{g}_{j+1}$ to $\frac{1}{2^b}$ disturbs $\sum_j \alpha_j$, which in turn changes $\tilde{g}_j$, which was optimized in the preceding step. To correct this suboptimal behavior of the entropy block, the refining block expands the guard bands to ensure $\tilde{g}_j = \alpha_{min}$, where $\alpha_{min} = \min\{\alpha_j \mid j = 1, 2, \ldots, 2^b\}$. This way, the equality constraint on entropy is met, and moreover the constraint on $p_{c,m}(\mathcal{Q}_b)$ is not violated. A pseudocode description of the refining block is given in Algorithm 4.

Algorithm 4. *Refining Block: Refining* $\tilde{g}_j = \frac{1}{2^b}$

Input: $\{(a_j^-, a_j^+) \mid \forall j\}$, $P(y_B, y_C)$, *and step-size* $\theta > 0$

Output: $\{(a_j^-, a_j^+) \mid \forall j\}$

```
Compute $\{\alpha_1, \alpha_2, \ldots, \alpha_{2^b}\}$ using (2.16)
$\alpha_{min} = \min_{1 \le j \le 2^b} \alpha_j$
for $j = 1 \rightarrow 2^b$ do
    while $\alpha_j > \alpha_{min}$ do
        $a_j^- \leftarrow a_j^- + \theta$; $a_j^+ \leftarrow a_j^+ - \theta$
        update $\alpha_j$ using (2.16)
    end while
end for
```

2.4 Simulation Results

In this section, simulation results are presented to showcase the performance of the EM-EM algorithm on CSR in (2.1) and (2.2), wherein $h(l)$ is distributed as $\mathcal{N}(0,1)$. On the left-side of Fig. 2.3, we present the key-rate of the EM-EM algorithm, defined as $b * p_c(\mathcal{Q}_b)$, when $b = 1, 2,$ and 3, against various values of SNR $= \frac{P}{\sigma^2}$. To generate the results, the constraint $SER(\mathcal{Q}_b) \leq 10^{-3}$ is satisfied at each SNR by feeding an appropriate value of η to the EM-EM algorithm. The left-side plots in Fig. 2.3 show that b must be chosen based on SNR in order to fully exploit the shared randomness. We highlight that the secret-keys generated for the above values of b exhibit maximum entropy at each SNR.

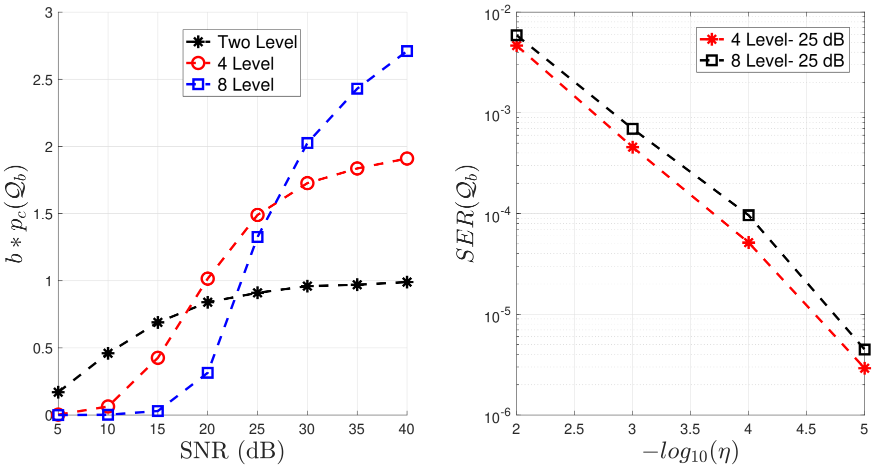


Figure 2.3: Left-side: Average bits per sample offered by the EM-EM algorithm with the constraint $SER(\mathcal{Q}_b) \leq 10^{-3}$. Right-side: $SER(\mathcal{Q}_b)$ values achieved by the EM-EM algorithm based on η.

The right-side plots of Fig. 2.3 show that lower values of $SER(\mathcal{Q}_b)$ can also be achieved by the EM-EM algorithm by choosing appropriate values of η. In Fig. 2.4, we compare the proposed EM-EM algorithm with the following baselines: (i) Multi-level quantizers which are optimized to maximize the entropy using the marginal distribution, (ii) Max-Lloyd algorithm, which provides 2^b points in $\mathbb{R}$ by optimizing the average quantization error using the marginal distribution, and (iii) Uniform quantizer, wherein the 2^b quantization levels are uniformly spread in $\mathbb{R}$, independent of the marginal distribution. With each baseline, the widths of the guards bands are increased until the constraint $SER(\mathcal{Q}_b) \leq 10^{-3}$ is satisfied. The plots confirm that none of the baselines achieves entropy of b bits. However, at high SNR, the entropy achieved by optimizing marginal distributions is close to that of EM-EM algorithm owing to high probability of consensus. Interestingly, the uniform quantizer achieves higher symbol-rate than the EM-EM algorithm because of significant mass points near the origin. However, the corresponding entropy is low, thereby resulting in lower key-rate.

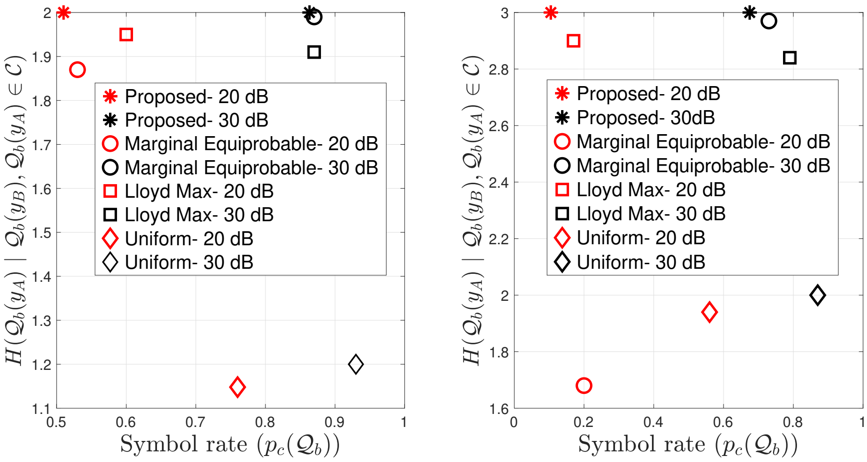


Figure 2.4: Scatter plots of pairs $(H(\mathcal{Q}_b(y_A) \mid \mathcal{Q}_b(y_A) \in \mathcal{C}, \mathcal{Q}_b(y_B) \in \mathcal{C}), p_c(\mathcal{Q}_b))$ for various multi-level quantizers with $b = 2$ and 3, and SNR $= 20$ and 30 dB, to achieve $SER(\mathcal{Q}_b) \leq 10^{-3}$. The plots confirm that the EM-EM algorithm guarantees maximum entropy.

2.5 Discussion

The quantizers which are designed based on the marginal distribution are not optimised to maximise the conditional entropy post consensus and hence the key harvested is defective. We have revisited the design of multi-level quantization schemes for key generation, and have shown that the knowledge of joint PDF can be exploited to generate secret-keys with maximum entropy and high symbol-rate. An iterative algorithm referred to as EM-EM was presented in order to maximise the entropy conditioned on the samples which are in consensus. EM-EM primarily comprises of two important sub-routines, entropy and error block, where the former takes care of conditional entropy whereas error block ensures that the error is upper-bounded by some negligible number. Presented algorithm can be applied on a wide range of joint distributions, and thus can serve as a software package to design multi-level quantizers for key generation.

Chapter 3

EM-EM for A-SQGSK

3.1 Introduction

Physical-layer key generation between two radio devices is well studied starting from theory that focuses on fundamental limits [YMR$^+$10], to testbed developments that showcase proof-of-concepts [PCB13]. A wide range of contributions exist in this topic, wherein the specific choice of CSR [GK11, LDS12, RSW11, CSD09, LLB13], used to generate the keys depend on wireless platforms such as OFDM [YDS11], multiple-input multiple-output (MIMO) systems [ZWCM10], and fibre optical networks [KWTP13], to name a few.

As discussed in Section 1.5, we are interested in the class of GSK generation protocols [WZN12, LYW$^+$14, TLQ15], wherein multiple nodes in the network first exchange a CSR, and then generate a GSK after executing a group consensus algorithm. To serve the purpose, a facilitator is required whose role is to assist the nodes in the network which can not observe the CSR inherently. The facilitator has to ensure that the CSR is shared with the respective node with no leakage to an eavesdropper in the vicinity. In context to GSK generation, A-SQGSK [RH19] is the only protocol for CSR sharing, which achieves practicality and confidentiality, and this feature is attributed to the use of algebraic rings. After all the three nodes have the CSR with them, they need to design a quantizer such that entropy is maximised and error-rate is bounded by some negligible number. With respect to the consensus phase, none of the existing work [MTM$^+$08, YMR$^+$10, JPC$^+$09, YRS06, LDS12, HHL17] have addressed the objective of maximising the entropy in the GSK generation and no systematic consensus algorithms that are customized to the distribution of the CSR have been presented hitherto. Hence, we present consensus algorithms among the three nodes to synthesize a GSK when using the shared CSR.

In this chapter, we extend the idea of EM-EM algorithm for a network of three users. We highlight that the EM-EM is only applicable on two-dimensional joint PDF and hence can not be applied directly to the three-dimensional PDF of the CSR seen at the three nodes. First, we will see the motivation for using A-SQGSK for GSK generation and then we will discuss the extension of EM-EM to it. In a nutshell, we propose relaxed design criterion for EM-EM algorithm which is applicable to this network.

3.2 System Model for GSK Generation

As shown in Fig. 3.1, we consider a three-user wireless network comprising node-1, node-2 and node-3 along with an eavesdropper, denoted by Eve. The wireless channel between any two nodes in this network model is assumed to be frequency-flat and remain quasi-static for a block of four channel-uses. Specifically, we use a complex Gaussian random variable, denoted by $h_{jk} \sim \mathcal{CN}(0, 1)$, to represent the channel between node-j and node-k, for $j \neq k$. We assume that all the channels $\{h_{jk} \mid j \neq k\}$ exhibit pair-wise reciprocity within the coherence-block, i.e., $h_{jk} = h_{kj}$, and moreover, every pair of channels in $\{h_{jk}\}$ are statistically independent. We also assume that the AWGN witnessed at all the nodes are distributed as $\mathcal{CN}(0, \sigma^2)$.

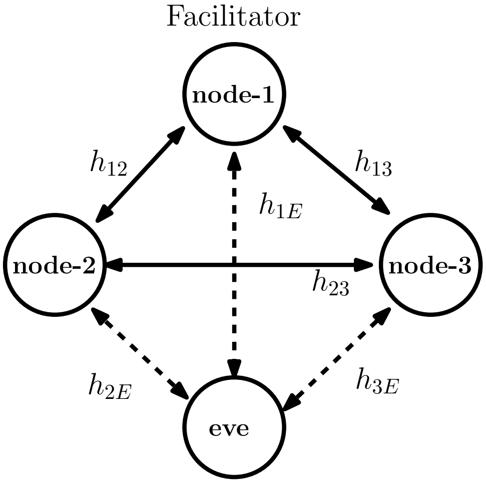


Figure 3.1: Network model with three wireless nodes, which intend to generate a GSK in the presence of the passive eavesdropper, referred to as Eve. All the channels in the network are assumed to be statistically independent. First, the three nodes share a common source of randomness, and then synthesize a group secret-key using a group consensus algorithm.

In this network model, the three nodes are interested in observing a subset of the channels $\{h_{12}, h_{13}, h_{23}\}$ so as to synthesize a GSK. Henceforth, throughout the paper, we refer to this subset of channels as the CSR. Towards witnessing the chosen CSR, the three nodes follow the conventional approach of broadcasting pilot symbols turn-by-turn within each coherence-block. As a result, every node can learn the corresponding channels upon receiving the pilot symbols. Since pilot transmission does not help the three nodes to witness the CSR, one of the nodes, referred to as the facilitator, broadcasts a linear combination of its observed channels in order to assist all the nodes to acquire the CSR within the coherence-block. Gathering the CSR observed over several coherence-blocks, the three nodes subsequently apply an appropriate key generation algorithm to generate a GSK. To highlight the role of each coherence-block, we denote the channels using the coherence-block index l as $\{h_{12}(l), h_{13}(l), h_{23}(l)\}$ for $l = 1, 2, \ldots, L$, where L is the total number of coherence-blocks used to witness the CSR. In the following section, we present a GSK protocol wherein the three nodes choose $\{h_{12}(l)\}$ for $1 \leq l \leq L$ as the CSR. Since the objective of this work is to study the effects of quantization when generating a GSK, we choose $\{h_{12}(l) \mid 1 \leq l \leq L\}$ to be CSR of interest throughout the paper.

3.2.1 GSK Generation with No Quantization at the Facilitator

We present a detailed description of a GSK protocol [HCH17] to exchange a CSR among the three nodes in the network shown in Fig. 3.1. We describe the four phases of the protocol for a given coherence-block $l \in \{1, 2, \ldots, L\}$:

Phase 1: node-1 transmits a pilot symbol $x = 1$, which is used by node-2 and node-3 to estimate the channels $h_{12}(l)$ and $h_{13}(l)$, respectively, as

$$\theta_2^{(1)}(l) = h_{12}(l) + e_2^{(1)}(l) \text{ and } \theta_3^{(1)}(l) = h_{13}(l) + e_3^{(1)}(l), \tag{3.1}$$

where $e_2^{(1)}(l) \sim \mathcal{CN}(0, \gamma)$ and $e_3^{(1)}(l) \sim \mathcal{CN}(0, \gamma)$ denote the channel estimation errors at node-2 and node-3, respectively. The superscripts denote the phase number in each coherence-block.

Phase 2: node-2 transmits a pilot symbol $x = 1$, which is used by node-1 and node-3 to estimate the channels $h_{12}(l)$ and $h_{23}(l)$, respectively, as

$$\theta_1^{(2)}(l) = h_{12}(l) + e_1^{(2)}(l) \text{ and } \theta_3^{(2)}(l) = h_{23}(l) + e_3^{(2)}(l), \tag{3.2}$$

where $e_1^{(2)}(l) \sim \mathcal{CN}(0, \gamma)$ and $e_3^{(2)}(l) \sim \mathcal{CN}(0, \gamma)$ are the corresponding estimation errors.

Phase 3: node-3 transmits a pilot symbol $x = 1$, which is used by node-2 and node-1 to estimate the channels $h_{23}(l)$ and $h_{13}(l)$, respectively, as

$$\theta_2^{(3)}(l) = h_{23}(l) + e_2^{(3)}(l) \text{ and } \theta_1^{(3)}(l) = h_{13}(l) + e_1^{(3)}(l), \tag{3.3}$$

where $e_2^{(3)}(l) \sim \mathcal{CN}(0, \gamma)$ and $e_1^{(3)}(l) \sim \mathcal{CN}(0, \gamma)$ are the corresponding estimation errors. We assume that all the nodes employ the same channel estimation algorithm, and as a result, we use γ as the variance of the estimation error at all the nodes.

Phase 4: By the end of Phase 3, node-1 and node-2 have noisy versions of the CSR $\{h_{12}(l)\}$, but not node-3. Therefore, to fill the gap, in the last phase, node-1 (which acts as the facilitator) transmits the sum $\theta_1^{(2)}(l) + \theta_1^{(3)}(l)$, using which node-3 receives $\theta_3^{(4)}(l) = h_{13}(l)\left(\theta_1^{(2)}(l) + \theta_1^{(3)}(l)\right) + n_3^{(4)}(l)$, where $n_3^{(4)}(l)$ denotes the additive noise at node-3 distributed as $\mathcal{CN}(0, \sigma^2)$. Using $\theta_3^{(1)}(l)$ and $\theta_3^{(4)}(l)$, node-3 learns a noisy version of $h_{12}(l)$ as

$$\bar{\theta}_3^{(4)}(l) = \left(\left(\theta_3^{(1)}(l)\right)^{-1} \theta_3^{(4)}(l)\right) - \theta_3^{(1)}(l).$$

Thus, by the end of Phase 4, the three nodes witness noisy versions of the CSR $\{h_{12}(l)\}$.

Observe that all the nodes witness noisy version of the CSR, wherein the noise levels depend on the node. Specifically, node-1 and node-2 observe $\{h_{12}(l)\}$, which are perturbed by estimation errors. However, node-3 observes a noisy version of $h_{12}(l)$, which is perturbed by both estimation error and the recovery noise during Phase 4.

In terms of leakage, an external eavesdropper receives the following symbols in the four phases: $y_E^{(1)}(l) = h_{1E}(l) + n_E^{(1)}(l)$, $y_E^{(2)}(l) = h_{2E}(l) + n_E^{(2)}(l)$, $y_E^{(3)}(l) = h_{3E}(l) + n_E^{(3)}(l)$, $y_E^{(4)}(l) = h_{1E}(l)(\theta_1^{(2)}(l) + \theta_1^{(3)}(l)) + n_E^{(4)}(l)$, where $h_{jE}(l)$ is the complex channel between node-j for $1 \leq j \leq 3$ and the eavesdropper, and $n_E^{(k)}(l)$ is the AWGN at the eavesdropper in Phase k for $k = 1, 2, 3, 4$. Note that the eavesdropper cannot learn the channel realizations $\{h_{12}(l)\}$ during the first three phases by the virtue of its physical location (with the assumption that $h_{1E}(l), h_{2E}(l), h_{3E}(l)$ are statistically independent of $h_{12}(l)$). However, in Phase 4, it is straightforward to verify that the CSR is not confidential since the mutual information between the sum $\theta_1^{(2)}(l) + \theta_1^{(3)}(l)$ and $\theta_1^{(2)}(l)$ is not zero. Overall, in addition to the asymmetry in the noise levels of the CSR at different nodes, this protocol also leaks the CSR to an eavesdropper.

3.2.2 GSK Generation with Quantization at the Facilitator

In this section, we discuss some practical aspects of GSK generation protocols. In Section 3.2.1, the first three phases involve broadcast of pilot symbols, wherein the receiver nodes estimate the corresponding channels using an appropriate channel estimation algorithm. However, in Phase 4, the sum of the two channel realizations, i.e., $\theta_1^{(2)}(l) + \theta_1^{(3)}(l)$, is transmitted by node-1. Observe that the in-phase and the quadrature components of $\theta_1^{(2)}(l) + \theta_1^{(3)}(l)$ can be irrational, and as a result, there will be loss of precision when the radio devices are implemented with limited hardware. Furthermore, most practical radios are designed to transmit baseband signals from finite constellations such as PSK, QAM etc. Due to constraints of short coherence-blocks, node-1 would need to transmit a quantized version of the sum $\theta_1^{(2)}(l) + \theta_1^{(3)}(l)$, given by $\varphi(\theta_1^{(2)}(l) + \theta_1^{(3)}(l)) = \theta_1^{(2)}(l) + \theta_1^{(3)}(l) + z_{sum}(l)$, where $\varphi(\cdot)$ is an appropriate quantization algorithm that directly quantizes the channel estimates to points in a complex constellation, denoted by $\mathcal{A}$, and $z_{sum}(l)$ is the corresponding quantization noise. We refer to this form of the GSK protocol as Asymmetrically Quantized GSK (AQGSK) [RH18, RH19]. After transmitting the quantized version, the received symbols at node-3 is given by $\theta_3^4(l) = h_{13}(l)\left(\theta_1^{(2)}(l) + \theta_1^{(3)}(l) + z_{sum}(l)\right) + n_3^{(4)}(l)$.

Using the above received symbols, the channel estimates recovered at node-3 are corrupted by the quantization noise in addition to the recovery noise in Phase 4. Thus, with quantization at node-1, the common randomness across the three nodes are affected by different levels of noise. Although more practical than the GSK protocol in Section 3.2.1, this method suffers from disparity in the effective noise levels at the three nodes, and importantly, the transmitted symbol from the facilitator $\theta_1^{(2)}(l) + \theta_1^{(3)}(l) + z_{sum}(l)$ continues to leak the CSR $\theta_1^{(2)}(l)$ at an external eavesdropper. Identifying these disadvantages of the AQGSK protocol, we explore a GSK protocol that enables the facilitator (node-1) to transmit symbols from a finite constellation, and yet provide zero-leakage to an external eavesdropper.

3.3 Algebraic SQGSK (A-SQGSK) Protocol

Unlike the ideas discussed in Section 3.2, the main idea to ensure zero-leakage to an external eavesdropper is to avoid quantizing $\theta_1^{(2)}(l) + \theta_1^{(3)}(l)$, which is the sum of the noisy channel realizations observed by node-1. In contrast, [RH19, JJR21] propose to quantize the noisy channel realizations $\theta_1^{(2)}(l)$ and $\theta_1^{(3)}(l)$ separately at node-1, then appropriately transform them to points in an algebraic ring, and then compute their linear combination over the algebraic ring. Henceforth, throughout the thesis, we refer to this scheme as the A-SQGSK protocol [RH19, JJR21]. In the following section, we first present the ingredients required to describe the A-SQGSK protocol.

3.3.1 Ingredients

The A-SQGSK protocol requires three complex constellations for the following purposes: (i) to quantize the channel realizations at all the three nodes, (ii) to execute the algebraic operations at node-1, and (iii) to broadcast a function of the channel realizations at node-1 (the facilitator). An example for the three constellations is provided in Fig. 3.2, along with the description of their functionality in Table 3.1. In the proposed A-SQGSK protocol, the facilitator quantizes the complex channel realizations to points in a complex constellation $\mathcal{A} \subset \mathbb{C}$, of size 2^m, given by $\mathcal{A} = \mathcal{A}_I \bigoplus i\mathcal{A}_Q$, where $i = \sqrt{-1}$, such that $\mathcal{A}_I = \mathcal{A}_Q$ and $|\mathcal{A}_I| = 2^{\frac{m}{2}}$. Using $\varphi : \mathbb{C} \rightarrow \mathcal{A}$ to denote the quantization operator, we assume that $\varphi(\cdot)$ works independently on the in-phase and the quadrature components of the input. For instance, with $\beta \sim \mathcal{CN}(0, \Sigma)$, we have

$$\varphi(\beta) = \arg min_{a \in \mathcal{A}} |\beta - a|^2 \in \mathcal{A}. \tag{3.4}$$

We choose the constellation $\mathcal{A}$ such that $\varphi(\beta)$ is uniformly distributed over the support $\mathcal{A}$ when $\beta \sim \mathcal{CN}(0, \Sigma)$. Given that $\mathcal{A}_I = \mathcal{A}_Q$, and the in-phase and the quadrature components of β are independent and identically distributed, it suffices to choose $\mathcal{A}_I \subset \mathbb{R}$ such that real$(\varphi(\beta))$ and imag$(\varphi(\beta))$ are uniformly distributed over the support $\mathcal{A}_I$ and $\mathcal{A}_Q$, respectively.

Table 3.1: Functionality of $\mathcal{A}$, $\bar{\mathcal{A}}$, and $\mathcal{A}'$

Functionality	**Constellation**
Quantization at all the nodes	$\mathcal{A}$ is used. Both $\bar{\mathcal{A}}$ and $\mathcal{A}'$ do not provide uniform distribution.
Algebraic operations at the facilitator	$\mathcal{A}'$ is used. Both $\bar{\mathcal{A}}$ and $\mathcal{A}$ have no algebraic structure.
Broadcasting at the facilitator	$\mathcal{A}'$ would be energy inefficient, $\mathcal{A}$ would lead to more errors due to smaller minimum distance. Therefore, $\bar{\mathcal{A}}$ is used.

We also require a regular square quadrature amplitude modulation (QAM) constellation $\bar{\mathcal{A}} \subset \mathbb{C}$, of size 2^m, given by $\bar{\mathcal{A}} = \bar{\mathcal{A}}_I \bigoplus i\bar{\mathcal{A}}_Q$, such that $\bar{\mathcal{A}}_I = \bar{\mathcal{A}}_Q = \{-2^{\frac{m}{2}} + 1, -2^{\frac{m}{2}} + 3, \ldots, 2^{\frac{m}{2}} - 3, 2^{\frac{m}{2}} - 1\}$, where $i = \sqrt{-1}$, and m is even. Assuming that the numbers of $\mathcal{A}_I$ are arranged in the ascending order, for $\nu \in \mathcal{A}$, we define a one-to-one mapping, denoted by $\psi : \mathcal{A} \to \bar{\mathcal{A}}$ as

$$\psi(\nu) = \bar{\mathcal{A}}_I(\mathcal{I}(\text{real}(\nu))) + i\bar{\mathcal{A}}_Q(\mathcal{I}(\text{imag}(\nu))), \tag{3.5}$$

where $\mathcal{I}(\cdot) \in [2^{\frac{m}{2}}]$ provides the position of the argument in the ordered set $\mathcal{A}_I$, and $\bar{\mathcal{A}}_I(t)$, for $t \in [2^{\frac{m}{2}}]$, provides the t-th element in the ordered set $\bar{\mathcal{A}}_I$. We require a complex constellation $\mathcal{A}' = \{0, 1, \ldots, 2^{\frac{m}{2}} - 1\} \bigoplus \{0, i, \ldots, i(2^{\frac{m}{2}} - 1)\}$, which forms an algebraic ring $\mathbb{Z}_{2^{\frac{m}{2}}}[i]$, defined over regular addition and multiplication, however, with modulo $2^{\frac{m}{2}}$ operation on both the in-phase and the quadrature components. A regular 2^m-QAM constellation $\bar{\mathcal{A}}$ can be written as a scaled and shifted version of $\mathcal{A}'$. In particular, for $\alpha \in \bar{\mathcal{A}}$, the one-one transformation from $\bar{\mathcal{A}}$ to $\mathcal{A}'$, represented by $\phi : \bar{\mathcal{A}} \to \mathcal{A}'$, is

$$\phi(\alpha) = \frac{\alpha + 2^{\frac{m}{2}} - 1 + i(2^{\frac{m}{2}} - 1)}{2}. \tag{3.6}$$

Using the mappings $\phi(\cdot)$ and $\psi(\cdot)$, it is straightforward to note that there is a one-to-one correspondence between the constellations $\mathcal{A}$, $\bar{\mathcal{A}}$, and $\mathcal{A}'$. Henceforth, throughout the paper, the composite mapping $\phi(\psi(\cdot))$ from $\mathcal{A}$ to $\mathcal{A}'$ is denoted by $\Theta(\cdot)$, and its inverse from $\mathcal{A}'$ to $\mathcal{A}$ is denoted by $\Theta^{-1}(\cdot)$.

3.3.2 A-SQGSK Protocol

The three nodes agree upon the discrete constellations $\mathcal{A}, \bar{\mathcal{A}}, \mathcal{A}' \subset \mathbb{C}$ of size 2^m, for some integer m, where m is even. Similar to the GSK protocol in Section 3.2.1, the A-SQGSK

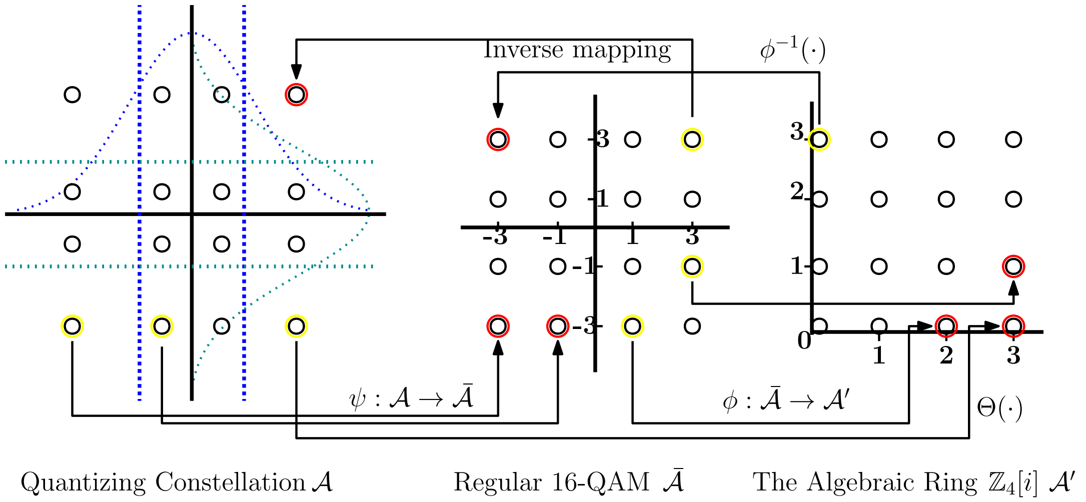

Inverse mapping
$\phi^{-1}(\cdot)$
3
1
-3
-1
1
3
-1
-3
3
2
1
0
1
2
3
$\psi : \mathcal{A} \rightarrow \bar{\mathcal{A}}$
$\phi : \bar{\mathcal{A}} \rightarrow \mathcal{A}'$
$\Theta(\cdot)$
Quantizing Constellation $\mathcal{A}$
Regular 16-QAM $\bar{\mathcal{A}}$
The Algebraic Ring $\mathbb{Z}_4[i]$ $\mathcal{A}'$

Figure 3.2: Example for the three constellations used in the A-SQGSK protocol with $m = 4$: Constellation $\mathcal{A}$ is used to quantize the channel realizations such that the quantized values are uniformly distributed. Constellation $\mathcal{A}'$ is the ring $\mathbb{Z}_{2^{\frac{m}{2}}}[i]$ used to compute linear combination of the quantized samples at the facilitator, whereas $\bar{\mathcal{A}}$ is used by the facilitator for broadcasting the linear combination of the CSR samples to node-2 and node-3.

protocol also comprises four phases (as shown in Fig. 3.3), which are described below:

Phase 1: node-1 broadcasts a pilot symbol $x = 1$ using which node-2 and node-3 receive $y_2^{(1)}(l) = h_{12}(l)x + n_2^{(1)}(l)$, and $y_3^{(1)}(l) = h_{13}(l)x + n_3^{(1)}(l)$, respectively, where $n_2^{(1)}(l)$ and $n_3^{(1)}(l)$ are the AWGN distributed as $\mathcal{CN}(0, \sigma^2)$. Using the received symbols, node-2 and node-3 estimate the channels $h_{12}(l)$ and $h_{13}(l)$, respectively, as $h_{12}(l) + e_2^{(1)}(l)$ and $h_{13}(l) + e_3^{(1)}(l)$, where $e_2^{(1)}(l) \sim \mathcal{CN}(0, \gamma)$ and $e_3^{(1)}(l) \sim \mathcal{CN}(0, \gamma)$ denote the channel estimation errors at node-2 and node-3, respectively. Further, these estimates are quantized to points in $\mathcal{A}$ as

$$\begin{aligned} \theta_2^{(1)}(l) &= \varphi(h_{12}(l) + e_2^{(1)}(l)) \in \mathcal{A}, \\ \theta_3^{(1)}(l) &= \varphi(h_{13}(l) + e_3^{(1)}(l)) \in \mathcal{A}, \end{aligned}$$

where $\varphi(\cdot)$ is as given in (3.4).

Phase 2: Similar to **Phase 1**, node-2 transmits a pilot symbol $x = 1$, which is used by node-1 and node-3 to estimate the channels $h_{12}(l)$ and $h_{23}(l)$, respectively, as $h_{12}(l) + e_1^{(2)}(l)$ and $h_{23}(l) + e_3^{(2)}(l)$. Subsequently, the estimates are quantized as

$$\begin{aligned} \theta_1^{(2)}(l) &= \varphi(h_{12}(l) + e_1^{(2)}(l)) \in \mathcal{A}, \\ \theta_3^{(2)}(l) &= \varphi(h_{23}(l) + e_3^{(2)}(l)) \in \mathcal{A}, \end{aligned}$$

where $e_1^{(2)}(l) \sim \mathcal{CN}(0, \gamma)$ and $e_3^{(2)}(l) \sim \mathcal{CN}(0, \gamma)$ are the corresponding estimation errors.

Phase 3: Similar to **Phase 1** and **Phase 2**, node-3 transmits a pilot symbol $x = 1$, which

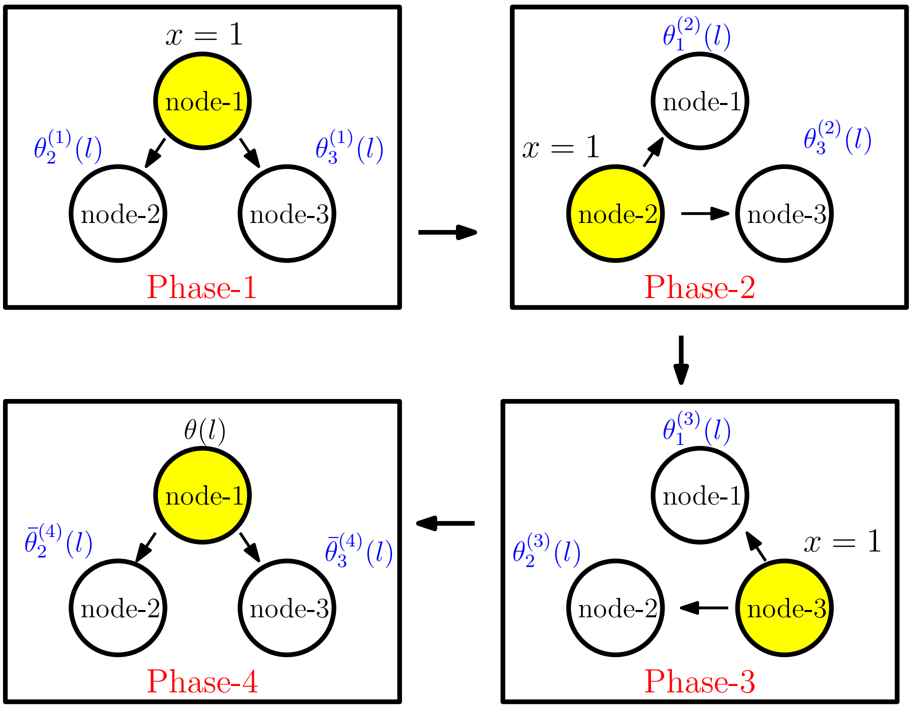


Figure 3.3: Depiction of the four phases of the A-SQGSK protocol. In the first three phases, each node transmits a pilot symbol, whereas in the last phase, node-1, which acts as the facilitator, broadcasts a linear combination of its channels suitably quantized over the algebraic ring $\mathbb{Z}_{2^{\frac{m}{2}}}[i]$.

.

is used by node-1 and node-2 to obtain quantized version of estimates in $\mathcal{A}$ as

$$\begin{aligned}\theta_1^{(3)}(l) &= \varphi(h_{13}(l) + e_1^{(3)}(l)) \in \mathcal{A},\\ \theta_2^{(3)}(l) &= \varphi(h_{23}(l) + e_2^{(3)}(l)) \in \mathcal{A},\end{aligned}$$

where $e_2^{(3)}(l) \sim \mathcal{CN}(0,\gamma)$ and $e_1^{(3)}(l) \sim \mathcal{CN}(0,\gamma)$ are the corresponding estimation errors. All the three nodes employ the same channel estimation algorithm, and as a result, γ is identical at the three nodes.

Phase 4: By the end of Phase 3, node-1 and node-2 have quantized versions of the estimates of the channel $h_{12}(l)$, whereas node-3 does not have access to $h_{12}(l)$. Therefore, to fill the gap, in the last phase, node-1 applies the composite transformation $\Theta(\cdot)$ on $\theta_1^{(2)}(l)$ and $\theta_1^{(3)}(l)$ to obtain $\Theta(\theta_1^{(2)}(l)) \in \mathcal{A}'$ and $\Theta(\theta_1^{(3)}(l)) \in \mathcal{A}'$, respectively. Subsequently, node-1 computes $\theta_{sum}(l) = \Theta(\theta_1^{(2)}(l)) \oplus \Theta(\theta_1^{(3)}(l)) \in \mathcal{A}'$, where $\oplus$ denotes addition over the ring $\mathbb{Z}_{2^{\frac{m}{2}}}[i]$, and then it broadcasts $\theta(l) \triangleq \phi^{-1}(\theta_{sum}(l)) \in \bar{\mathcal{A}}$ to node-2 and node-3. Here $\phi^{-1}(\cdot)$ denotes the inverse of ϕ, defined in (3.6). With this, the received symbol at node-3 is given by $\theta_3^{(4)}(l) = \frac{1}{\sqrt{E_{avg}}} h_{13}(l)\theta(l) + n_3^{(4)}(l)$, where $\sqrt{E_{avg}}$ is the scalar used to normalize the transmit power in Phase 4 such that $\mathbb{E}[|\theta(l)|^2] = 1$. Since node-3 has the knowledge of both $h_{13}(l) + e_3^{(1)}(l)$ and its quantized version, it obtains a *maximum a posteriori probability* (MAP) estimate $\hat{\theta}_3(l) \in \bar{\mathcal{A}}$ of $\theta(l)$. Using the above estimate, node-3 obtains an estimate of

$$\left\{\varphi\left(h_{12}(l)+e_1^{(2)}(l)\right)\right\},\left\{\varphi\left(h_{12}(l)+e_2^{(1)}(l)\right)\right\}, \text{ and } \left\{\bar{\theta}_3^{(4)}(l)\right\} \tag{3.8}$$

the quantized version of $h_{12}(l)$ as

$$\bar{\theta}_3^{(4)}(l)=\Theta^{-1}\left(\phi(\hat{\theta}_3(l))\ominus\Theta(h_{13}(l)+e_3^{(1)}(l))\right)\in\mathcal{A}, \tag{3.7}$$

where the subtraction operator $\ominus$ is over the ring $\mathbb{Z}_{2^{\frac{m}{2}}}[i]$. Thus, the CSR seen by node-1, node-2, node-3 are respectively given in (3.8). Unlike the protocol in Section 3.2.1, the CSR witnessed at the three nodes belong to $\mathcal{A}$. In [JJR21, Theorem 1] it is proved that A-SQGSK provides no leakage at an eavesdropper, i.e, $I(C_1^{h_{12}};\phi^{-1}(c_{sum}))=0$.

3.4 Group Consensus Algorithm for A-SQGSK Protocol

By the end of the A-SQGSK protocol, the CSR at the three nodes, as given in (3.8), belong to the complex constellation $\mathcal{A}$, for some $m>1$. Out of the three observations in (3.8), the first two are a result of direct quantization of channel realizations during the first three phases, whereas as the third one is a consequence of MAP decoding and successive cancellation at node-3 during Phase 4. We note that the in-phase and the quadrature components of a complex CSR sample in (3.8) are statistically independent at each node owing to circularly symmetric complex channel and also perfect knowledge of the channel during the phase of MAP decoding and successive interference cancellation at node-3. Since the in-phase and the quadrature components of the CSR belong to $\mathcal{A}_I$, the three nodes can agree upon a consensus algorithm to identify the location of the real samples that lie at the same level. A straightforward technique to harvest shared secret-keys is to apply two-level quantization on the in-phase and the quadrature components of the samples given in (3.8), as proposed in [YMR$^+$10, JPC$^+$09]. Although this idea is effective, its limitation is its inability to generate more than one bit per sample when the CSR offers significant randomness. A natural way to increase the number of bits per sample is to apply multi-level quantization on each sample, where the number of levels must be chosen depending on the CSR at the end of the A-SQGSK protocol. In order to generate b bits per real sample, for $b\in\mathbb{N}$, we formally define a 2^b-level quantizer in Section 2.2. Although the quantizer is applicable on any real number, in this paper, we use this quantizer to apply on real samples in $\mathcal{A}_I$. In the next section, we discuss a group consensus algorithm among the three nodes using an appropriately designed quantizer $\mathcal{Q}_b$.

3.4.1 Consensus Phase for Group-Key Generation

To generate a GSK, node-1, node-2 and node-3 agree upon a quantizer $\mathcal{Q}_b$, as presented in Definition 1. Furthermore, they collect a sufficiently large number of CSR observations, denoted by

$$\mathcal{Y}_A^c = \left\{\varphi\left(h_{12}(l) + e_1^{(2)}(l)\right) \mid l = 1, 2, \ldots, L\right\},$$

$$\mathcal{Y}_B^c = \{\varphi\left(h_{12}(l) + e_2^{(1)}(l)\right) \mid l = 1, 2, \ldots, L\},$$

and

$$\mathcal{Y}_C^c = \{\bar{\theta}_3^{(4)}(l) \mid l = 1, 2, \ldots, L\},$$

over L coherence-blocks. After unfolding the in-phase and the quadrature components of the CSR, node-1, node-2 and node-3, respectively gather the sets of real samples $\mathcal{Y}_A$, $\mathcal{Y}_B$ and $\mathcal{Y}_C$, each of size $2L$. To achieve consensus, the three nodes execute the following protocol using the excursion length $e \geq 1$:[1]

- node-2 obtains the set $\bar{Y}_B = \{\mathcal{Q}_b(y_B(r)) \mid y_B(r) \in \mathcal{Y}_B\}$, and then shares the index values $\mathcal{I}_B = \{r \in [2L] \mid \mathcal{Q}_b(y_B(r)) = \mathcal{Q}_b(y_B(r+1)) = \ldots = \mathcal{Q}_b(y_B(r+e-1)) = a_j,$ for some $a_j \in \mathcal{C}\}$ to node-1.
- node-1 obtains the set $\bar{Y}_A = \{\mathcal{Q}_b(y_A(r)) \mid y_A(r) \in \mathcal{Y}_A\}$, and then computes the corresponding set of index values $\mathcal{I}_A = \{r \in [2L] \mid \mathcal{Q}_b(y_A(r)) = \mathcal{Q}_b(y_A(r+1)) = \ldots = \mathcal{Q}_b(y_A(r+e-1)) = a_j,$ for $a_j \in \mathcal{C}\}$. Subsequently, node-1 shares $\mathcal{I}_{BA} \triangleq \mathcal{I}_B \cap \mathcal{I}_A$ with node-3, where $\mathcal{I}_{BA}$ denotes the set of index values in consensus between node-2 and node-1.
- node-3 obtains the set $\bar{Y}_C = \{\mathcal{Q}_c(y_C(r)) \mid y_C(r) \in \mathcal{Y}_C\}$, and then computes the corresponding set of index values $\mathcal{I}_C = \{r \in [2L] \mid \mathcal{Q}_b(y_C(r)) = \mathcal{Q}_b(y_C(r+1)) = \ldots = \mathcal{Q}_b(y_C(r+e-1)) = a_j,$ for $a_j \in \mathcal{C}\}$. Subsequently, node-3 shares $\mathcal{I}_{CBA} \triangleq \mathcal{I}_C \cap \mathcal{I}_{BA}$ with node-1 and node-2, where $\mathcal{I}_{CBA}$ denotes the set of index values in consensus between node-1, node-2 and node-3. We use N_{group} to denote the length of $\mathcal{I}_{CBA}$, i.e., $N_{group} = |\mathcal{I}_{CBA}|$.

Using $\mathcal{I}_{CBA}$, node-1, node-2 and node-3 generate the following sequences $\mathcal{K}_A = \{\mathcal{Q}_b(y_A(r)) \mid r \in \mathcal{I}_{CBA}\}$, $\mathcal{K}_B = \{\mathcal{Q}_b(y_B(r)) \mid r \in \mathcal{I}_{CBA}\}$, and $\mathcal{K}_C = \{\mathcal{Q}_b(y_C(r)) \mid r \in \mathcal{I}_{CBA}\}$. Note that $\mathcal{K}_A$, $\mathcal{K}_B$, and $\mathcal{K}_C$ are N_{group}-length sequences over the alphabet $\mathcal{C}$.

[1] This protocol for group consensus is a generalization of the protocol proposed for pair-wise key generation in [YMR+10]

3.4.2 Relaxed Design Criteria on $\mathcal{Q}_b$

Based on the above consensus algorithm, the following properties are desired on $\mathcal{K}_A$, $\mathcal{K}_B$, and $\mathcal{K}_C$:

1. The symbol-rate, given by $\frac{N_{group}}{2L}$, which captures the fraction of samples in consensus among the three nodes, is maximum.
2. The entropy of the three sequences must be maximum, i.e., $H(\mathcal{K}_A) = N_{group}b$, $H(\mathcal{K}_B) = N_{group}b$, and $H(\mathcal{K}_C) = N_{group}b$ where $H(\mathcal{K}_A)$, $H(\mathcal{K}_B)$ and $H(\mathcal{K}_C)$, respectively denote the joint entropy of N_{group} random variables over $\mathcal{C}$.
3. The fraction of pair-wise disagreements between any two sequences must be negligible, i.e.,
$$\frac{1}{N_{group}} d_H(\mathcal{K}_E, \mathcal{K}_F) \leq \beta,$$
for $E, F \in \{A, B, C\}$, where $d_H(\cdot,\cdot)$ denotes the Hamming distance operator, and β is a negligible number of our choice.

In the rest of this work, we drop the reference to the sample index r, and refer to the real numbers at node-1, node-2 and node-3 as three correlated random variables y_A, y_B and y_C with an underlying joint probability distribution function $P(y_A, y_B, y_C)$. The above listed criteria can be met provided we use EM-EM based on $P(y_A, y_B, y_C)$. However, given that this three-dimensional distribution is intractable to handle and does not comply with EM-EM, we propose relaxed design criteria on $\mathcal{Q}_b$ which takes into account the two-dimensional joint distribution $P(y_B, y_C)$ instead of $P(y_A, y_B, y_C)$. We choose the pair-wise distribution $P(y_B, y_C)$ over $P(y_A, y_B)$ and $P(y_A, y_C)$ since y_C is more distorted with respect to y_B than y_A because of the combination of quantization noise as well as the recovery noise. Note that while the quantizer design is based on the CSR between the worst-pair of nodes in the network, the same quantizer will be used by all the three nodes during the group consensus phase of Section 3.4.1.

By focusing on the CSR available at node-2 and node-3, we design $\mathcal{Q}_b$ using EM-EM framework, assuming that only node-2 and node-3 are participating in the key-generation process using their CSR $\mathcal{Y}_B$ and $\mathcal{Y}_C$ as presented in Section 2.3. However, when the CSR samples are discrete, as in the case of A-SQGSK, the formulation presented in equations (2.4) - (2.16) continues to hold after replacing the integrals by summation operations.

3.4.3 On Achieving the Desired SER using the EM-EM Algorithm

After the refining block presented in Algorithm 4, the EM-EM algorithm guarantees that the entropy of the key is maximized for a given $\eta > 0$. However, at this point, the desired

SER may not be achieved, i.e., $SER(\mathcal{Q}_b) > \beta$. Further decreasing η decreases both the numerator and the denominator of (2.8), and as a result, lower values of $SER(\mathcal{Q}_b)$ may not be guaranteed by decreasing η. In such cases, the pair (b, β) is not feasible as defined below:

Definition 4. *When using a quantizer $\mathcal{Q}_b$ with excursion length $e = 1$, the pair (b, β), for a given $b \in \mathbb{N}$ and $\beta > 0$, is said to be feasible if there exists an $\eta \leq \beta$ such that $SER(\mathcal{Q}_b) \leq \beta$.*

Based on Definition 4, when (b, β) is not feasible, we propose to design $\mathcal{Q}_b$ using the EM-EM algorithm by feeding an $\eta > 0$ such that $SER(\mathcal{Q}_b)$ is minimized, i.e.,

$$\eta^* = \arg min_{\eta} \ SER(\mathcal{Q}_b). \tag{3.9}$$

Subsequently, we use the corresponding quantizer $\mathcal{Q}_b$ (which is designed with η^*) in the consensus algorithm with excursion length $e > 1$. The minimum value of e for which the mismatch rate of β is achieved will be used in the consensus phase. The following result proves that if the consensus algorithm is employed with $\mathcal{Q}_b$ (which is designed using the EM-EM algorithm) and excursion length $e > 1$, then the entropy of the synthesized key continues to be maximum.

Proposition 3. *When using the quantizer $\mathcal{Q}_b$ from the EM-EM algorithm along with excursion length $e > 1$ in the consensus algorithm, the entropy of the synthesized key continues to be b bits per sample.*

Proof. The EM-EM algorithm generates a quantizer $\mathcal{Q}_b$ that guarantees maximum entropy on the generated key when $e = 1$, and this feature is contributed by the refining block of the algorithm. When using the consensus algorithm with $\mathcal{Q}_b$ and $e > 1$, let $\bar{Y}_B^e$ and $\bar{Y}_C^e$ denote e-successive samples of $\bar{Y}_B$ and $\bar{Y}_C$, respectively, and let $P(\bar{Y}_B^e, \bar{Y}_C^e)$ denote the corresponding joint probability mass function of $\bar{Y}_B^e$ and $\bar{Y}_C^e$. Note that the support of $\bar{Y}_B^e$ and $\bar{Y}_C^e$ are e-fold cross products of the set $\bar{\mathcal{C}}$ given by $\bar{\mathcal{C}}^e = \underbrace{\bar{\mathcal{C}} \times \bar{\mathcal{C}} \times \ldots \times \bar{\mathcal{C}}}_{e \text{ times}}$ with size $(2^b + 1)^e$, where $\bar{\mathcal{C}} = \mathcal{C} \cup X$. As per the consensus algorithm in Section **??**, the consensus probability is given by $\sum_j \sum_k \text{Prob}(\bar{Y}_B^e = \bar{a}_j, \bar{Y}_C^e = \bar{a}_k)$, where $\bar{a}_j = [a_j \ a_j \ \ldots \ a_j]$ and $\bar{a}_k = [a_k \ a_k \ \ldots \ a_k]$ such that $a_j, a_k \in \mathcal{C}$. Let us define the set $\bar{\mathcal{C}}_X^e = \{v \in \bar{\mathcal{C}}^e \mid v(t) = X \text{ for some } 1 \leq t \leq e\}$. With that, the conditional entropy with excursion length $e > 1$ is given by $H\left(\bar{Y}_B^e | \bar{Y}_B^e \notin \bar{\mathcal{C}}_X^e, \bar{Y}_C^e \notin \bar{\mathcal{C}}_X^e\right) = -\sum_{j=1}^{2^b} t_j \log_2(t_j)$, where

$$\begin{aligned} t_j &= \frac{\sum_k \text{Prob}(\bar{Y}_B^e = \bar{a}_j, \bar{Y}_C^e = \bar{a}_k)}{\sum_j \sum_k \text{Prob}(\bar{Y}_B^e = \bar{a}_j, \bar{Y}_C = \bar{a}_k)} \\ &= \frac{\sum_k \left(\text{Prob}(\bar{Y}_B = a_j, \bar{Y}_C = a_k)\right)^e}{\sum_j \sum_k \left(\text{Prob}(\bar{Y}_B = a_j, \bar{Y}_C = a_k)\right)^e}, \end{aligned}$$

where the second equality is applicable because of statistical independence across the e

samples. If the parameter γ of the EM-EM algorithm is appropriately chosen, then the cross-terms $\text{Prob}(\bar{Y}_B = a_j, \bar{Y}_C = a_k)$, for $j \neq k$, are negligible. As a result, we can approximate t_j as

$$t_j \approx \frac{\left(\text{Prob}(\bar{Y}_B = a_j, \bar{Y}_C = a_j)\right)^e}{\sum_{k=1}^{2^b} \left(\text{Prob}(\bar{Y}_B = a_k, \bar{Y}_C = a_k)\right)^e}.$$

Since the refining block of the EM-EM algorithm ensures identical values of $\text{Prob}(\bar{Y}_B = a_j, \bar{Y}_C = a_j)$ for each $1 \leq j \leq 2^b$, we note that $t_j \approx \frac{1}{2^b}$, and therefore the conditional entropy of the symbols in consensus is maximum. □

3.4.4 Complexity of the EM-EM Algorithm

Given a joint probability distribution function on two correlated random variables, the EM-EM algorithm generates a two-dimensional discrete probability mass function (PMF) on an alphabet of size 2^b so as to satisfy the conditions in Problem 1. Towards quantifying its complexity, the complexity of the error block, the entropy block, and the refining block, must be quantified. With respect to the error and the entropy blocks, we can quantify their complexity by counting the number of times the two-dimensional PMF is updated through the iterative process. Note that each time a guard band is expanded or shifted, the PMF has to be updated to check the condition given in line 3 of Algorithm 2 and Algorithm 3. If the two correlated random variables are continuous, then the number of PMF updates depends on the choice of the step-size θ within each block. In particular, with smaller step-size, the EM-EM algorithm provides higher precision in achieving the entropy of b bits per symbol, however, at the cost of a large number of PMF updates. On the other hand, with large step-size, while the number of PMF updates decreases, the EM-EM algorithm may generate secret-keys with lower key-rate as guard bands may not be prudently expanded in the error block.

When applied to the A-SQGSK protocol, the input to the EM-EM algorithm is the joint PMF on the CSR between node-2 and node-3. Note that the CSR is discrete, wherein the size of the support set is given by $2^{\frac{m}{2}}$. Furthermore, upon mapping the CSR samples in (3.8) to the ring $\mathbb{Z}_{2^{\frac{m}{2}}}[i]$ using $\Theta(\cdot)$, the minimum step-size that can be used in the EM-EM algorithm is $\theta = 1$. This implies that once the algorithm enters the error block, the maximum number of PMF updates is $\lceil \frac{2^{\frac{m}{2}} - 2^b}{2} \rceil$. However, when the algorithm is inside the entropy block the maximum number of PMF updates is $2^{\frac{m}{2}} - 2^b$. This difference in the PMF updates between the error block and the entropy block is attributed to the fact that the error block increments the width of a guard band on both sides, whereas the entropy block only shifts a guard band without increasing its width. Finally, after a number of iterations between these two blocks, the refining block also updates the PMF one more time so as to achieve the conditional entropy of b bits per sample (see line 4 of Algorithm 4). In

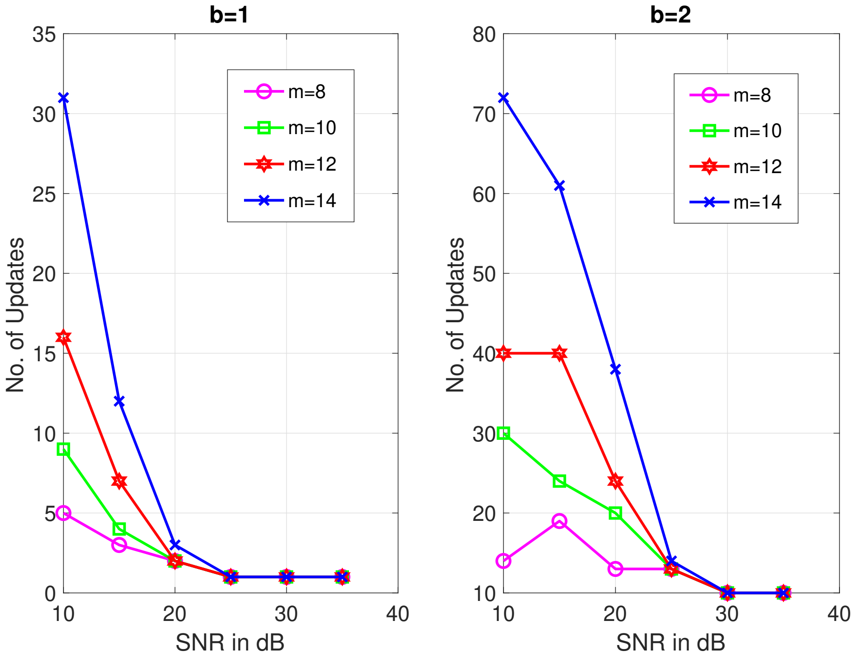


Figure 3.4: Complexity of the EM-EM algorithm as a function of SNR, and the size of the constellation used to generate the CSR samples. We use the number of PMF updates during the iterations within the EM-EM block as a measure of complexity.

order to present the total number of PMF updates through the iterative process, we count them when the CSR samples of the A-SQGSK protocol are fed to the EM-EM algorithm at various values of SNR and m. These plots are presented in Fig. 3.4 for both $b = 1$ and $b = 2$ in order to achieve the bound $p_{c,m}(\mathcal{Q}_b) \leq \eta$, for $\eta = 0.1$. As shown in each of the two cases, it is clear that with larger values of m, the number of updates is high since there is more room for the guard bands to shift and expand. Furthermore, as SNR increases, the number of updates decreases since the error block only needs to expand a guard band few times since the underlying noise is negligible to help achieve the bound $p_{c,m}(\mathcal{Q}_b) \leq \eta$. Overall, the plots in Fig. 3.4 show that the EM-EM algorithm can be deployed in practice owing to few PMF updates at moderate- and high-SNR values.

3.5 Simulation Results

In this section, we present simulation results on the performance of the A-SQGSK protocol in conjunction with the proposed EM-EM algorithm. In the first three phases of the A-SQGSK protocol, all the nodes use the received symbols as the noisy estimates of the channels, i.e., $\gamma = \sigma^2$. For a given value of $m \in \{2, 4, 6, 8, 10, 12, 14\}$, and the underlying signal-to-noise-ratio, defined by SNR $= \frac{1}{\sigma^2}$, we choose the complex constellation $\mathcal{A}$ to ensure that the outputs of $\varphi(\cdot)$ are uniformly distributed. Accordingly, we use 4-, 16-, 64-, 256−, 1024−, 4096− and

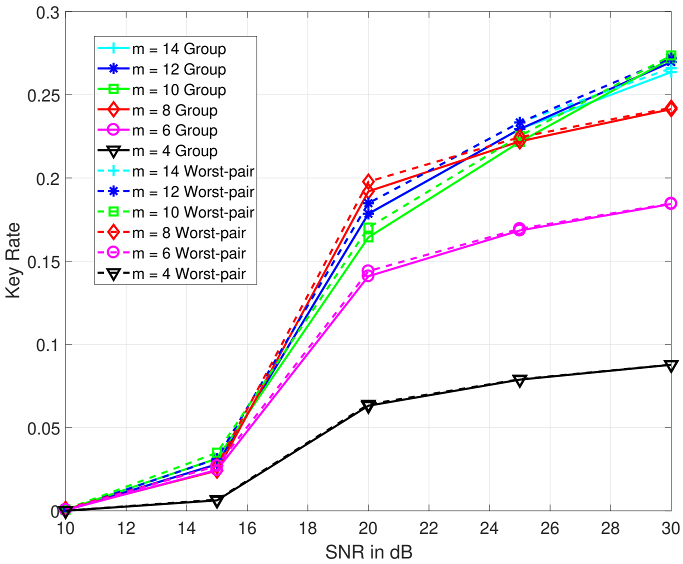


Figure 3.5: Key rate against various SNR values and various sizes of the constellation $\mathcal{A}$ to achieve entropy of $b = 1$ bit per sample and a mismatch rate (BER) of at most 10^{-2}.

16384− QAM constellations as $\bar{\mathcal{A}}$. We use $L = 10000$ coherence-blocks to generate the CSR samples, after which each node generates 20000 real samples, which correspond to the in-phase and the quadrature components of their CSR samples. Subsequently, we feed the joint probability distribution of the real samples at node-2 and node-3 as the input to the EM-EM algorithm by specifying the value of $b \geq 1$ (which is the number of bits generated per real sample) and the mismatch rate. As discussed in Section 3.4.3, if the bound on mismatch rate is not achieved with excursion length $e = 1$, then we design the quantizer $\mathcal{Q}_b$ with η^* (as given in (3.9)) and then use it in the consensus algorithm with $e > 1$. Finally, we employ the designed quantizer $\mathcal{Q}_b$ to achieve consensus on a GSK as per the protocol in Section 3.4.1. Throughout this section, mismatch rate is referred to as bit-error-rate (BER) and symbol-error-rate (SER) when $b = 1$ and $b > 1$, respectively. Out of the $2L$ real samples available for consensus, we define the key rate as the average number of secret bits generated per real sample.

Using the EM-EM algorithm with $b = 1$, i.e., two-level quantization, the key rates of the A-SQGSK protocol are presented in Fig. 3.5 against SNR $\in \{10, 15, 20, 25, 30\}$ dB so as to achieve an upper bound on the mismatch rate of 10^{-2}. In this context, a bit is said to be in error if any two nodes disagree with the value at that location after executing the protocol in Section 3.4.1. At each SNR value, we capture the impact of the A-SQGSK protocol by employing different sizes of the discrete constellation $\mathcal{A}$. The plots in Fig. 3.5 show that the key rate increases with SNR, and this behaviour is attributed to more samples in consensus among the three nodes since the BER is upper bounded by 10^{-2}. In addition to the key rate of the GSK, we also plot the key rate obtained by executing the pair-wise consensus algorithm (given in Section 3.4.2) which uses the EM-EM algorithm using the

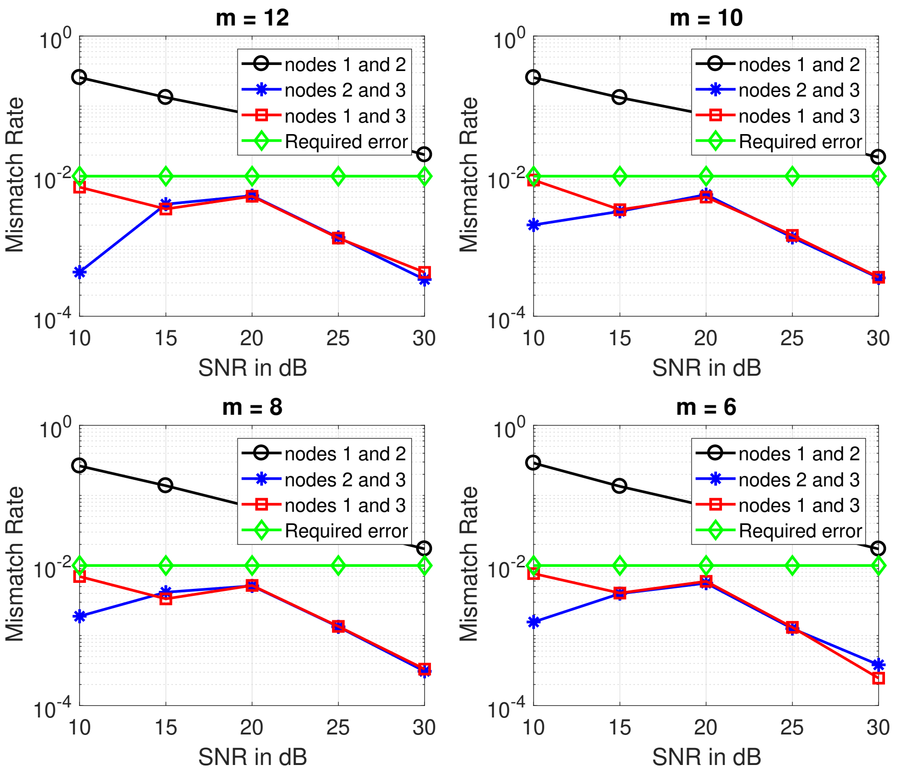


Figure 3.6: Comparing the mismatch rate offered by the group consensus algorithm when $\mathcal{Q}_b$ is designed based on the pair-wise CSR samples at (i) node-1 and node-2, (ii) node-1 and node-3, and (iii) node-2 and node-3.

CSR samples at node-2 and node-3. Since the CSR samples between node-2 and node-3 capture the worst-case scenario (due to the combined effect of quantization noise as well as the recovery noise in Phase 4 of the A-SQGSK protocol), the plots show that the key rate of the GSK is marginally lower than that of the pair-wise key. The plots also show that while the key rate increases with m, it saturates after a certain value of m, which is dependent on the underlying SNR. The intuition for this behaviour is that as the size of $\mathcal{A}$ increases, the recovery noise in phase 4 increases, and a result, the CSR witnessed between node-2 and node-3 are further degraded when compared to those at lower values of m. While this is the case with respect to recovery noise, larger value of m also provides finer granularity to expand and shift the guard bands in the EM-EM algorithm, thus providing more degrees of freedom to upper bound the BER within 10^{-2}. Overall, due to the conflicting behaviour between the fraction of decoding errors and the smoothness offered to the EM-EM algorithm, the benefits in terms of key rate are marginal after a certain value of m.

In Fig. 3.6, we capture the performance of group key generation when different pair-wise samples are considered to design $\mathcal{Q}_b$. We plot the mismatch rate of the group key offered by feeding the joint distribution of several pairs, along with the threshold of 10^{-2}, which is the intended mismatch rate fed to the EM-EM algorithm. The plots show that for all values of m, the joint distribution of CSR at node-1 and node-2 must not be used to

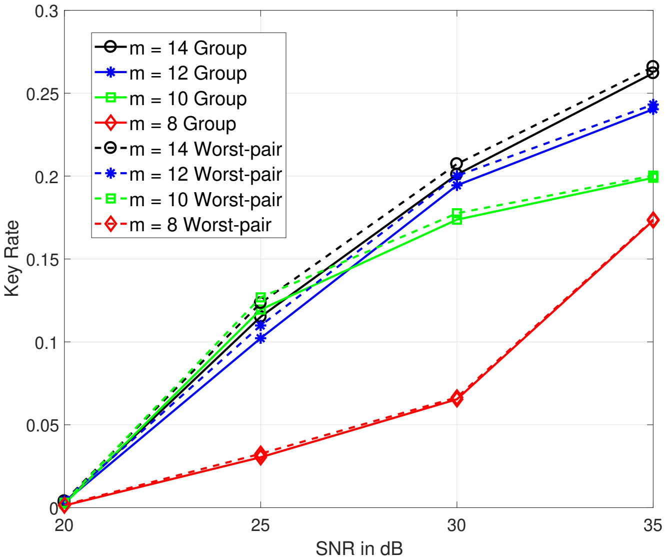


Figure 3.7: Similar to Fig. 3.5, key rate against various SNR values and various sizes of the constellation $\mathcal{A}$ to achieve entropy of $b = 2$ bits per sample and a mismatch rate (SER) of at most 10^{-2}.

design the quantizer; this is because those CSR samples are only perturbed by the effect of additive noise in the quantization process. As a result, at high SNR values, the impact of recovery noise in Phase 4 of the A-SQGSK protocol is neglected. Thus, while the quantizer design is made on good pair of CSR samples, the subsequently designed quantizer is used to achieve consensus on CSR samples that are poorer compared to that of node-1 and node-2. As a result, the overall mismatch rate achieved at the GSK level is more than the desired reliability level. The plots also show that instead of CSR samples at node-2 and node-3, we could also make use of the CSR samples at node-1 and node-3. This is because between the two sources of noise, the recovery noise in Phase 4 is more dominant, and therefore the resultant mismatch rate continues to lie below the desired reliability number. At lower SNR values, the quantization noise between node-1 and node-2 are also considered, and therefore the CSR samples at node-2 and node-3 offer better mismatch rate.

As a generalization of results presented in Fig. 3.5, in Fig. 3.7, we also present the key rate to synthesize a GSK with entropy of $b = 2$ bits per real sample. In this context, $b = 2$ implies that the quantizer $\mathcal{Q}_b$ divides the CSR samples (which take $2^{\frac{m}{2}}$ levels) at each user into 2^b zones in order to arrive at consensus. Similar to the case of $b = 1$, the quantizer design continues to maintain an upper bound on the mismatch rate of 10^{-2}. However, the mismatch rate corresponds to SER since the synthesized key is over the alphabet $\mathcal{C}$ containing 4 values. The plots show that inferences drawn by observing Fig. 3.5 continue to hold when $b = 2$. We have also verified that the entropy of the generated keys is 2 bits per real sample. At lower SNR values, the proposed EM-EM algorithm was unable to generate non-zero key rate satisfying entropy of 2 bits per sample and a mismatch rate of 10^{-2}.

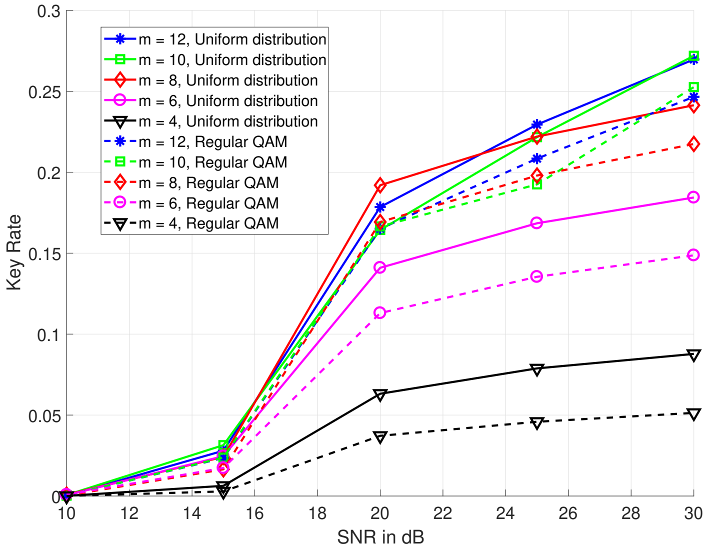


Figure 3.8: Comparison of key rates to achieve entropy of $b = 1$ bit per sample and a mismatch rate (BER) of at most 10^{-2} using the EM-EM algorithm in conjunction with the A-SQGSK protocol: (i) $\mathcal{A}$ is chosen to induce uniform distribution on the quantized values during the A-SQGSK protocol, and (ii) $\mathcal{A}$ is uniformly spaced regular QAM constellation with unit average energy.

Finally, we present a comparison between the key rate offered by the proposed combination of the A-SQGSK protocol and the EM-EM algorithm when the following two types of discrete constellations are considered for $\mathcal{A}$: (i) $\mathcal{A}$ is chosen such that the resulting CSR samples after the A-SQGSK protocol exhibit uniform distribution, and (ii) $\mathcal{A}$ is a regular square QAM constellation. To generate the simulation results under (ii), we use a regular square QAM normalized to unit average energy. The corresponding plots on key rate are presented in Fig. 3.8 for $b = 1$. The plots show that forcing uniform distribution on the CSR samples after A-SQGSK protocol outperforms regular-QAM since the latter method accumulates large number of samples around the mean value, and as a result, increasing the guard band drops significant number of samples when compared to the former case.

In summary, stitching together the advantages of the A-SQGSK protocol and the EM-EM algorithm, we recommend to choose m and b based on the underlying SNR values. From the viewpoint of designing $\mathcal{Q}_b$, we recommend the use of joint distribution of CSR samples at node-2 and node-3, which constitute the worst-pair of common randomness when using h_{12} to harvest GSKs.

3.6 Discussion

In GSK generation, as a consequence of the underlying three-dimensional PMF, EM-EM can not be applied in its original form in order to maximise the entropy. Hence, we have presented a relaxed criterion where the joint PMF of the CSR seen at worst pair of the nodes is fed to the EM-EM. It is shown that this approach continues to maximise the conditional entropy on the final key. Further, as the CSR associated with A-SQGSK is discrete in nature, we lay out some design rules for achieving the desired SER using the EM-EM.

Part-II: Opportunistic Selection and LLR Generation for A-SQGSK

Chapter 4

Key-Rate Enhancement and Reconciliation

4.1 Introduction

For a network of three-users, not every node has access to the CSR in the probing phase itself, so, one of the nodes is declared as the facilitator whose role is to secretly share the CSR with such blind nodes. We continue to use A-SQGSK [RH19, JJR21] protocol for the purpose of sharing the CSR in the network. A-SQGSK proposed in its original form utilises only one of the two available channels, i.e., h_{12} and h_{13}, in order to achieve the confidentiality. Hence, we explore a method to harvest the two available CSR in an opportunistic fashion for the same set-up. It is theoretically shown that in the case when both the channels are in consensus, only one of the two channels can be used as the CSR otherwise confidentiality will be compromised. Therefore, a selection criterion is presented which ensures that the error is minimised when selecting one of the two available channels as the CSR.

In order to facilitate reconciliation, LDPC codes are considered and Message Passing (MP) decoder is utilised for low-complexity decoding of the codes. In context of GSK generation, input to MP algorithm is the LLR associated with the CSR, hence, we present an LLR generation technique which is tightly tied to A-SQGSK protocol.

4.2 System Model and Background

A network comprising three nodes: node-1, node-2, and node-3 is considered, as shown in Fig. 4.1. The channel between node-j and node-k is denoted by h_{jk}, where $j \neq k$ and $h_{jk} \sim \mathcal{CN}(0,\, 1)$. Channel h_{jk} is assumed: (i) to be flat-fading and remains quasi-static for a block of at least 4 channel uses, (ii) it exhibits pairwise reciprocity within the coherence-block i.e., $h_{jk} = h_{kj}$ and (iii) $\{h_{jk}\}$ are statistically independent and identically distributed. Another assumption made is that all the nodes witness AWGN distributed as $\mathcal{CN}(0,\, \sigma^2)$, so that the average SNR is $\frac{1}{\sigma^2}$. To synthesize a GSK, a subset of $\{h_{12}, h_{13}, h_{23}\}$, referred to as the CSR, must be learned by all the nodes. While a noisy version of h_{12} can be estimated at node-2 and node-1 by probing pilot symbols within a coherence-block, it is clear that node-3 needs to learn h_{12} explicitly. To help this cause, [RH19, JJR21] proposed a protocol to share quantized version of h_{12} over an algebraic ring .First, we recall the A-SQGSK protocol, and then point out its limitations.

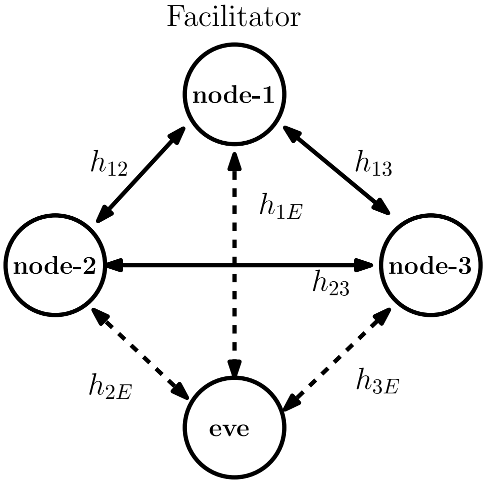


Figure 4.1: A network of three nodes along with an eavesdropper

4.2.1 A-SQGSK Protocol

To execute the A-SQGSK protocol, the three nodes must be equipped with three complex constellations $\mathcal{A}$, $\mathcal{A}'$ and $\bar{\mathcal{A}}$, as already exemplified in Fig. 3.2. One-to-one transformations $\psi : \mathcal{A} \rightarrow \bar{\mathcal{A}}$ is defined in (3.5) and $\phi : \bar{\mathcal{A}} \rightarrow \mathcal{A}'$ is defined in (3.6). The composite mapping $\phi(\psi(\cdot))$ from $\mathcal{A}$ to $\mathcal{A}'$ is denoted by $\Theta(\cdot)$, and its inverse from $\mathcal{A}'$ to $\mathcal{A}$ is denoted by $\Theta^{-1}(\cdot)$. The A-SQGSK protocol consists of four phases as discussed in Section 3.3.2. The protocol is revisited here again with some new notations for the sake of clarity.

Phase-1: node-1 broadcasts a pilot symbol $x = 1$ using which node-2 and node-3 receive $y_2^{(1)} = h_{12}x + n_2^{(1)}$ and $y_3^{(1)} = h_{13}x + n_3^{(1)}$, where the noise $n_2^{(1)}$ and $n_3^{(1)}$ are distributed as $\mathcal{CN}(0, \sigma^2)$ respectively. In this notation, the superscript denotes the phase number in each coherence- block and the subscript denotes the node index. These two nodes then estimate the channel as $h_{12} + e_2^{(1)}$ and $h_{13} + e_3^{(1)}$, where the estimation errors are distributed as $e_2^{(1)}, e_3^{(1)} \sim \mathcal{CN}(0, \gamma)$. Furthermore, these estimates are then quantized as

$$C_2^{h_{12}} = \varphi(h_{12} + e_2^{(1)}) \in \mathcal{A} \text{ and}$$

$$C_3^{h_{13}} = \varphi(h_{13} + e_3^{(1)}) \in \mathcal{A},$$

wherein in the superscript denotes the inherently observed channel and the subscript denotes the node index. Also, the quantization operator $\varphi(\beta)$ for $\beta \in \mathbb{C}$ is already defined by (3.4).
Phase-2: Similarly, node-2 broadcasts a pilot symbol $x = 1$, which is used by node-1 and node-3 to estimate the channel as $h_{12} + e_1^{(2)}$ and $h_{23} + e_3^{(2)}$, respectively, with similar noise

statistics as in **Phase-1**. Subsequently, the estimates are quantized as

$$C_1^{h_{12}} = \varphi(h_{12} + e_1^{(2)}) \in \mathcal{A} \ \text{ and}$$

$$C_3^{h_{23}} = \varphi(h_{23} + e_3^{(2)}) \in \mathcal{A}.$$

Phase-3: Similarly, node-3 transmits a pilot symbol $x = 1$, whereby node-1 and node-2 estimate the channel as $h_{13} + e_1^{(3)}$ and $h_{23} + e_2^{(3)}$, respectively. Subsequently, both the nodes obtain the quantized version of estimates as

$$C_1^{h_{13}} = \varphi(h_{13} + e_1^{(3)}) \in \mathcal{A} \ \text{ and}$$

$$C_2^{h_{23}} = \varphi(h_{23} + e_2^{(3)}) \in \mathcal{A}.$$

Phase-4: Assuming that the CSR is derived using h_{12}, node-3 does not have the access to it. To bridge the gap, node-1 applies the composite transformation $\Theta(\cdot)$ on $C_1^{h_{12}}$ and $C_1^{h_{13}}$ to obtain $\Theta(C_1^{h_{12}}) \in \mathcal{A}'$ and $\Theta(C_1^{h_{13}}) \in \mathcal{A}'$, respectively. Subsequently, node-1 computes $c_{sum} = \Theta(C_1^{h_{12}}) \oplus \Theta(C_1^{h_{13}}) \in \mathcal{A}'$, where $\oplus$ denotes addition over the ring $\mathbb{Z}_{2^{\frac{m}{2}}}[i]$, and then it broadcasts $\frac{1}{\sqrt{E_{avg}}}\phi^{-1}(c_{sum}) \in \bar{\mathcal{A}}$ to node-2 and node-3, where E_{avg} is the average energy of the QAM constellation and $\phi^{-1}(\cdot)$ denotes the inverse of ϕ, defined in (3.6).

With the knowledge of $h_{13} + e_3^{(1)}$, node-3 obtains the *maximum aposteriori probability* (MAP) estimate of $\phi^{-1}(c_{sum})$, denoted by $\hat{\theta}_3 \in \bar{\mathcal{A}}$. Using the above estimate, node-3 obtains the CSR as

$$C_3^{h_{12}} = \Theta^{-1}\left(\phi(\hat{\theta}_3) \ominus \Theta(\varphi(h_{13} + e_3^{(1)}))\right) \in \mathcal{A},$$

where $\ominus$ is the subtraction over the ring $\mathbb{Z}2_{\frac{m}{2}}[i]$. Similarly, node-2 will recover $C_2^{h_{13}}$, which is the decoded version of the quantized channel h_{13}. By the end of the A-SQGSK protocol, node-j has $\{C_j^{h_{12}}, C_j^{h_{13}}\}$ for $j = 1, 2, 3$. node-j unfolds the real and the imaginary components of $C_j^{h_{12}}$, $C_j^{h_{13}}$, and then uses the samples for key extraction. Henceforth, throughout this paper, we refer to a CSR from the unfolded set of h_{12} as $R_j^{h_{12}}$, and similarly, we refer to a CSR sample from the unfolded set of h_{13} as $R_j^{h_{13}}$.

4.2.2 Consensus Phase

In order to extract the secret-key, [RH19, JJR21] proposed to run the A-SQGSK protocol for a number of coherence-blocks, and then used $\{R_j^{h_{12}}\}$ as the CSR at node-j. Subsequently, a two-level consensus algorithm [MTM$^+$08], with guard bands $q_+ \geq 0$, and $q_- \leq 0$, was employed to synthesize secret bits by satisfying the rule $\mathcal{Q}(\alpha) = 1$ if $\alpha > q_+$, and $\mathcal{Q}(\alpha) = 0$ if $\alpha < q_-$, for any real sample α. A sample α is said to be out of consensus if $q_- \leq \alpha \leq q_+$. The guards bands were appropriately chosen to upper bound the mismatch rate (referred to as initial error rate), which is the fraction of bits that do not agree between any two nodes. To achieve consensus among the three nodes, [RH19, JJR21] proposed all the three nodes to parse through their quantized samples of $\{R_1^{h_{12}}, R_2^{h_{12}}, R_3^{h_{12}}\}$, and then create a list of all the CSR samples that are lying outside the guard bands. Subsequently, all the nodes mutually agree on common indices and then generate the secret-key using the samples on the common indices.

4.3 Opportunistic CSR Selection

One of the limitations of [RH19, JJR21] is that only the CSR $\{R_1^{h_{12}}, R_2^{h_{12}}, R_3^{h_{12}}\}$ was used to extract secret-keys. However, as depicted on the right-side of Fig. 4.2, it is clear that the three nodes can opportunistically make use of both $\{R_1^{h_{12}}, R_2^{h_{12}}, R_3^{h_{12}}\}$ and $\{R_1^{h_{13}}, R_2^{h_{13}}, R_3^{h_{13}}\}$ based on the coherence-block under consideration. In particular, the following possibilities arise on a given coherence-block: (i) Only one of the two sets of CSR samples is out of the guard band at all the three nodes, thereby contributing to the key. (ii) Both the CSR sets lie in the guard band of at least one of the nodes, thereby not contributing to the key for this coherence block. (iii) Both the CSR sets lie outside the guard band at all the nodes, and therefore, either of them is a good choice of CSR. In the first case, the nodes can use the CSR which is in consensus on a given coherence-block. As a result, there will be improvement in the key-rate in comparison with the A-SQGSK protocol [RH19, JJR21]. We prove that using either of the subset as the CSR preserves confidentiality.

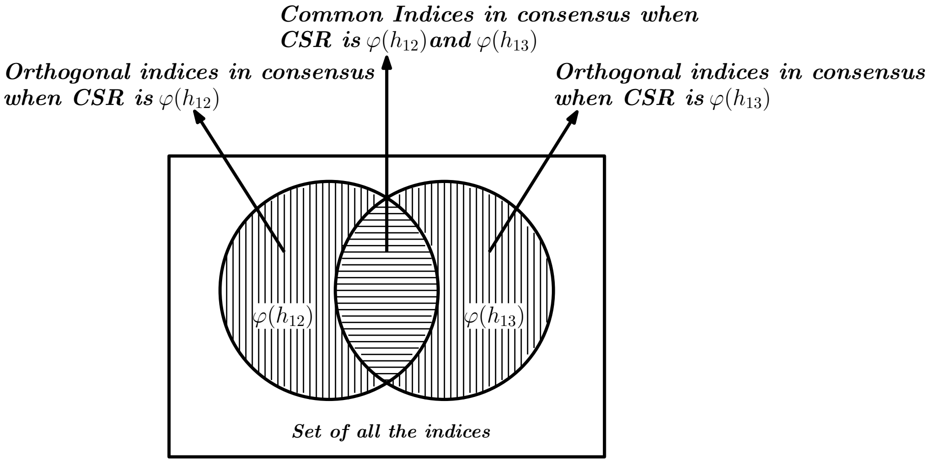

Common Indices in consensus when CSR is $\varphi(h_{12})$and $\varphi(h_{13})$
Orthogonal indices in consensus when CSR is $\varphi(h_{12})$
Orthogonal indices in consensus when CSR is $\varphi(h_{13})$
$\varphi(h_{12})$
$\varphi(h_{13})$
Set of all the indices

Figure 4.2: Venn diagram depicting two possibilities of channels in consensus - (i) indices belongs to only one type of CSR, either $\varphi(h_{12})$ or $\varphi(h_{13})$, and (ii) indices belong to both the CSR (shown in the intersection).

Proposition 4. *For a 2^m-QAM constellation, when the two CSR, $C_1^{h_{12}}$ and $C_1^{h_{13}}$, are identically distributed, we have $I(C_1^{h_{12}}; \phi^{-1}(c_{sum})) = 0$ and $I(C_1^{h_{13}}; \phi^{-1}(c_{sum})) = 0$, where $\phi^{-1}(c_{sum})$ is the symbol transmitted by node-*1.

$$I(C_1^{h_{12}}, C_1^{h_{13}}; \phi^{-1}(c_{sum})) = H(C_1^{h_{12}}, C_1^{h_{13}}) - H(C_1^{h_{12}}, C_1^{h_{13}}|\phi^{-1}(c_{sum})) \tag{4.1}$$

$$H(C_1^{h_{12}}, C_1^{h_{13}}) = H(C_1^{h_{12}}) + H(C_1^{h_{13}}|C_1^{h_{12}}) = H(C_1^{h_{12}}) + H(C_1^{h_{13}}) \tag{4.2}$$

$$H(C_1^{h_{12}}, C_1^{h_{13}}|\phi^{-1}(c_{sum})) = \sum_{j=1}^{2^m} H(C_1^{h_{12}}, C_1^{h_{13}}|\phi^{-1}(c_{sum}) = c_j)P(\phi^{-1}(c_{sum}) = c_j) \tag{4.3}$$

$$-\sum_{k=1}^{2^m}\sum_{l=1}^{2^m} P(C_1^{h_{12}} = b_k, C_1^{h_{13}} = a_l|\phi^{-1}(c_{sum}) = c_j)log_2(P(C_1^{h_{12}}, C_1^{h_{13}} = a_l|\phi^{-1}(c_{sum}) = c_j)) \tag{4.4}$$

$$P(C_1^{h_{12}} = b_k, C_1^{h_{12}} = \phi^{-1}(\Theta(a_l) \ominus \Theta(c_j))) = \begin{cases} P(C_1^{h_{12}} = b_k), & \text{if } b_k = \phi^{-1}(\Theta(a_l) \ominus \Theta(c_j)) \\ 0, & \text{otherwise} \end{cases} \tag{4.5}$$

$$H(C_1^{h_{12}}, C_1^{h_{13}}|\phi^{-1}(c_{sum}) = c_j) = -\sum_{k=1}^{2^m} P(C_1^{h_{12}} = b_k)log_2(P(C_1^{h_{12}} = b_k)) = H(C_1^{h_{12}}) \tag{4.6}$$

Proof. Following the similar lines of the proof in [JJR21, Theorem 1], it can be proved that $I(C_1^{h_{12}}; \phi^{-1}(c_{sum})) = 0$ and likewise, $I(C_1^{h_{13}}; \phi^{-1}(c_{sum})) = 0$. □

Using Proposition 4, it follows that choosing either of the two CSR sets for a coherence block will not compromise the confidentiality feature of the CSR. As seen in the third case, on a given coherence-block, when both the CSR samples are in consensus, we cannot use both to extract the keys as it does not ensure confidentiality of the CSR samples as we prove next.

Proposition 5. *For a 2^m-QAM constellation, when the two complex CSR, $C_1^{h_{12}}$ and $C_1^{h_{13}}$, are identically distributed, we have non-zero value of $I(C_1^{h_{12}}, C_1^{h_{13}}; \phi^{-1}(c_{sum}))$, where $\phi^{-1}(c_{sum})$ is the symbol transmitted by node-1.*

Proof. The expression for $I(C_1^{h_{12}}, C_1^{h_{13}}; \phi^{-1}(c_{sum}))$ is expanded in (4.1), where $H(C_1^{h_{12}}, C_1^{h_{13}})$ is given in (4.2) such that the second equality holds as $C_1^{h_{12}}$ and $C_1^{h_{13}}$ are statistically independent. Furthermore, $H(C_1^{h_{12}}, C_1^{h_{13}}|\phi^{-1}(c_{sum}))$ is given in (4.3). (4.4) to (4.6) show that the conditional entropy, $H(C_1^{h_{12}}, C_1^{h_{13}}|\phi^{-1}(c_{sum}))$, is equal to $H(C_1^{h_{12}})$ as follows. The first terms after summation in (4.3), $H(C_1^{h_{12}}, C_1^{h_{13}}|\phi^{-1}(c_{sum}) = c_j)$, is given in (4.4). The probability, $P(C_1^{h_{12}} = b_k, C_1^{h_{13}} = a_l|\phi^{-1}(c_{sum}) = c_j)$ in (4.4) is given as $P(C_1^{h_{12}} = b_k, C_1^{h_{12}} = \phi^{-1}(\Theta(a_l) \ominus \Theta(c_j)))$ in (4.5). Substituting (4.5) in (4.4) gives $H(C_1^{h_{12}}, C_1^{h_{13}}|\phi^{-1}(c_{sum}) = c_j)$ in (4.6) which is equal to $H(C_1^{h_{12}})$ and then substituting (4.6) in (4.3) and (4.3) in (4.1) gives $I(C_1^{h_{12}}, C_1^{h_{13}}; \phi^{-1}(C_{sum})) = H(C_1^{h_{13}})$. □

Assuming Eve can perfectly retrieve $\phi^{-1}(c_{sum})$, the leakage at Eve is non-zero using Proposition 5. Therefore, when $\{R_1^{h_{12}}, R_2^{h_{12}}, R_3^{h_{12}}\}$ and $\{R_1^{h_{13}}, R_2^{h_{13}}, R_3^{h_{13}}\}$ are in consensus on a given coherence-block, we present a method for selecting one of them such that the mismatch rate among the keys is minimized.

$$\mho(\mathcal{N}(\mu,\gamma),\bar{\mathcal{A}}_I) = \left\{\int_{-\infty}^{\bar{\mathcal{A}}_I(1)+\frac{d_{min}}{2}} P_\Theta(\theta)d\theta, \int_{\bar{\mathcal{A}}_I(2)-\frac{d_{min}}{2}}^{\bar{\mathcal{A}}_I(2)+\frac{d_{min}}{2}} P_\Theta(\theta)d\theta,\ldots,\int_{\bar{\mathcal{A}}_I(2^{\frac{m}{2}})-\frac{d_{min}}{2}}^{\infty} P_\Theta(\theta)d\theta\right\} \tag{4.7}$$

$$\text{Prob}\left(R_2^{h_{13}} \in \mathcal{S} | R_1^{h_{13}} \in \bar{\mathcal{S}}\right) = \sum_{x_u \in \mathcal{S}} \varrho^{x_u}\left(\varnothing^{\phi(R_1^{h_{12}})}\left(\mho\left(\mathcal{N}\left(\phi^{-1}\left(\Theta(R_1^{h_{12}}) \oplus \Theta(R_1^{h_{13}})\right), \frac{E_{avg}\sigma^2}{|h_{12}+e_1^{(2)}|^2}\right), \bar{\mathcal{A}}_I\right)\right)\right) \tag{4.8}$$

$$\text{Prob}\left(R_3^{h_{12}} \in \mathcal{S} | R_1^{h_{12}} \in \bar{\mathcal{S}}\right) = \sum_{x_u \in \mathcal{S}} \varrho^{x_u}\left(\varnothing^{\phi(R_1^{h_{13}})}\left(\mho\left(\mathcal{N}\left(\phi^{-1}\left(\Theta(R_1^{h_{12}}) \oplus \Theta(R_1^{h_{13}})\right), \frac{E_{avg}\sigma^2}{|h_{13}+e_1^{(3)}|^2}\right), \bar{\mathcal{A}}_I\right)\right)\right) \tag{4.9}$$

4.3.1 Likelihood Based CSR Selection Strategy

We present an optimal CSR selection strategy, wherein the facilitator first builds the likelihood functions on the CSR observed at node-2 and node-3, and then chooses the one that provides smaller probability of error. First, we define notations needed to explain the technique. For a Gaussian PDF $P_\Theta(\theta)$, denoted by $\mathcal{N}(\mu,\gamma)$ with mean μ and variance γ, the notation $\mho(\mathcal{N}(\mu,\gamma),\bar{\mathcal{A}}_I)$ represents the PMF induced on the discrete constellation $\bar{\mathcal{A}}_I$ when quantizing $P_\Theta(\theta)$ onto the points in $\bar{\mathcal{A}}_I$. In other words, $\mho(\mathcal{N}(\mu,\gamma),\bar{\mathcal{A}}_I)$ is given in (4.7), where d_{min} represents the minimum Euclidean distance of the constellation $\bar{\mathcal{A}}_I$, and $\bar{\mathcal{A}}_I(t)$ denotes the t-th component for $1 \leq t \leq 2^{\frac{m}{2}}$. We use $\varnothing^s(\mathbf{g})$ to denote circular shift of the elements of the vector $\mathbf{g}$ to the left by s units. For a given PMF $\mathcal{H}$ on $\bar{\mathcal{A}}_I$, the notation $\varrho^p(\mathcal{H})$ denotes the probability of the sample point $p \in \bar{\mathcal{A}}_I$. Let $\bar{\mathcal{A}}_I^-$ and $\bar{\mathcal{A}}_I^+$ denote the set of PAM points that lies on the negative and positive sides in $\bar{\mathcal{A}}_I$, respectively.

Theorem 1. *On a coherence-block when both $\{R_1^{h_{12}}, R_2^{h_{12}}, R_3^{h_{12}}\}$ and $\{R_1^{h_{13}}, R_2^{h_{13}}, R_3^{h_{13}}\}$ are in consensus, the CSR of interest must be $\{R_1^{h_{12}}, R_2^{h_{12}}, R_3^{h_{12}}\}$ if Prob $\left(R_2^{h_{13}} \in \mathcal{S} | R_1^{h_{13}} \in \bar{\mathcal{S}}\right) \geq$ Prob $\left(R_3^{h_{12}} \in \mathcal{S} | R_1^{h_{12}} \in \bar{\mathcal{S}}\right)$, or $\{R_1^{h_{13}}, R_2^{h_{13}}, R_3^{h_{13}}\}$ otherwise, where Prob $\left(R_2^{h_{13}} \in \mathcal{S} | R_1^{h_{13}} \in \bar{\mathcal{S}}\right)$ and Prob $\left(R_3^{h_{12}} \in \mathcal{S} | R_1^{h_{12}} \in \bar{\mathcal{S}}\right)$ are given in* (4.8) *and* (4.9)*, respectively. In this context, we have $\mathcal{S} = \bar{\mathcal{A}}_I^-$, when $\bar{\mathcal{S}} = \bar{\mathcal{A}}_I^+$. Similarly, we have $\mathcal{S} = \bar{\mathcal{A}}_I^+$, when $\bar{\mathcal{S}} = \bar{\mathcal{A}}_I^-$.*

Proof. Let us consider a coherence-block for which both $R_j^{h_{12}}$ and $R_j^{h_{13}}$ are in consensus for each $j \in \{1,2,3\}$. The corresponding quantized versions of the complex channels are $C_j^{h_{12}}$ and $C_j^{h_{13}}$. At sufficiently large SNR values, and an appropriate value of m, we have $C_1^{h_{12}} = C_2^{h_{12}}$ and $C_1^{h_{13}} = C_3^{h_{13}}$ with high probability. Using $R_1^{h_{12}}$ and $R_1^{h_{13}}$, node-1 generates the point that was broadcast to node-2 and node-3 as $\mu = \phi^{-1}\left(\Theta(R_1^{h_{12}}) \oplus \Theta(R_1^{h_{13}})\right)$. Using $h_{12}+e_1^{(2)}$ as an estimate of the channel seen between node-1 and node-2, node-1 builds an aposteriori PDF at node-2, and in this case it is Gaussian distributed given by $\mathcal{N}\left(\mu, \frac{E_{avg}\sigma^2}{|h_{12}+e_1^{(2)}|^2}\right)$. Furthermore, since node-2 decodes on the PAM constellation using MAP decoder, the corresponding aposteriori PMF on the PAM points is given by $\mathcal{H}_{R_2^{h_{13}}} = \mho\left(\mathcal{N}\left(\mu, \frac{E_{avg}\sigma^2}{|h_{12}+e_1^{(2)}|^2}\right), \bar{\mathcal{A}}\right)$, wherein the PMF $\mathcal{H}_{R_2^{h_{13}}}$ is listed on PAM points when enumerated in the increasing order. Finally, since node-2 obtains the CSR $R_2^{h_{13}}$ by performing a modulo subtraction on the algebraic ring, node-1 incorporates the corresponding changes in the PMF as $\bar{\mathcal{H}}_{R_2^{h_{13}}} = \varnothing^{\phi(R_1^{h_{12}})}(\mathcal{H}_{R_2^{h_{13}}})$. Thus, node-1 generates an aposteriori PMF on the CSR $R_2^{h_{13}}$ seen at node-2. Once the PMFs are generated, then the probability of error at node-2 is computed by summing over the mass points in the complementary region of the PAM constellation with respect to the CSR $R_1^{h_{13}}$. By mimicking similar operations at node-3, node-1 also generates $\bar{\mathcal{H}}_{R_3^{h_{12}}}$, which is an aposteriori PMF on the CSR $R_3^{h_{12}}$ seen at node-3, and then computes the probability of error at node-3. Finally, the CSR that provides lower probability of error is chosen for key generation. □

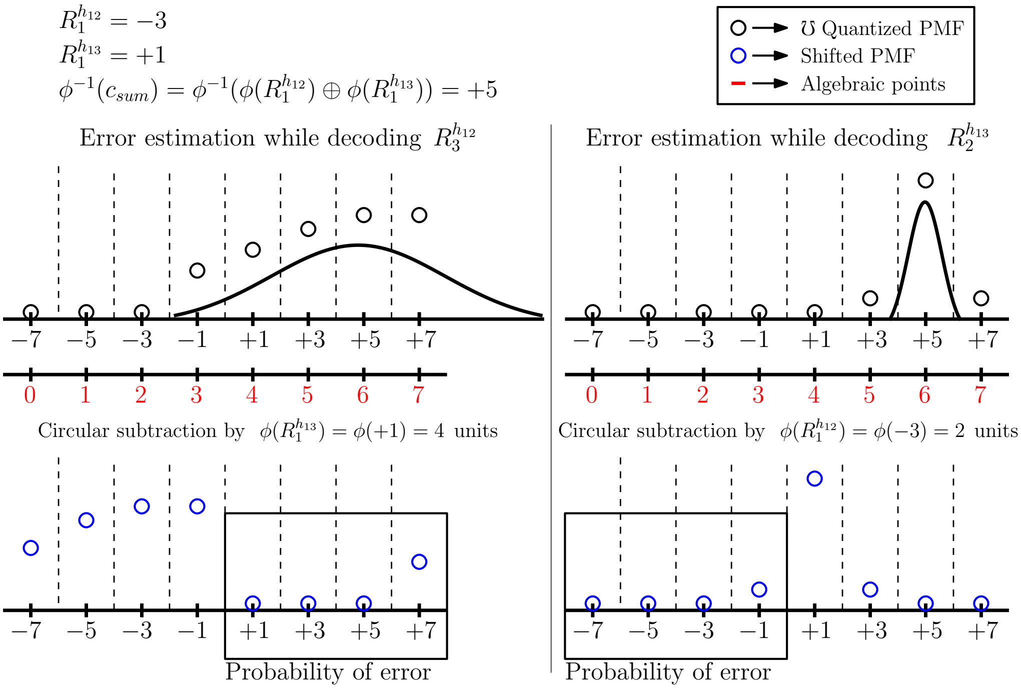

$R_1^{h_{12}} = -3$
$R_1^{h_{13}} = +1$
$\phi^{-1}(c_{sum}) = \phi^{-1}(\phi(R_1^{h_{12}}) \oplus \phi(R_1^{h_{13}})) = +5$
$\mho$ Quantized PMF
Shifted PMF
Algebraic points
Error estimation while decoding $R_3^{h_{12}}$
Error estimation while decoding $R_2^{h_{13}}$
−7 −5 −3 −1 +1 +3 +5 +7
0 1 2 3 4 5 6 7
Circular subtraction by $\phi(R_1^{h_{13}}) = \phi(+1) = 4$ units
Circular subtraction by $\phi(R_1^{h_{12}}) = \phi(-3) = 2$ units
Probability of error
Probability of error

Figure 4.3: Figure depicts an example for likelihood based CSR selection at node-1. On the left: an illustration of the computation of probability of error when the CSR is $R_1^{h_{12}}$. On the right: an illustration of the computation of probability of error when the CSR is $R_1^{h_{13}}$. For simplification, it is assumed that the CSR lies in the constellation $\bar{\mathcal{A}}$.

Fig. 4.3 depicts an example for the likelihood selection strategy at node-1 when both the CSR are in consensus. We highlight that node-1 is able to generate the aposteriori PMFs seen at node-2 and node-3 by using the channel realizations available in the first four phases of the A-SQGSK protocol. As a result, no additional communication-overheads are involved. Furthermore, this method is optimal at moderate values of m and mid-to-high SNR values since the quantized values of the channels used at node-1 would be the same used at node-2 and node-3 with high probability.

4.3.2 Consensus Algorithm for Opportunistic Selection

After executing the A-SQGSK protocol over L coherence-blocks, the three nodes have a sequence of samples. For $r \in \{2, 3\}$, a CSR sample is said to come from $\varphi(h_{1r})$ if it is obtained from either the real or the imaginary part of the quantized version of h_{1r} on any coherence-block. To achieve consensus, the three nodes use a generalized version of the consensus algorithm in [MTM$^+$08] as follows: node-2 obtains two sets of indices, which comprises index values of the CSR samples of $\varphi(h_{12})$ and $\varphi(h_{13})$ lying outside the guard band, and then shares it to node-1. Upon receiving the indices, node-1 computes the corresponding sets of indices in a similar fashion for the two sets of CSR samples, and then broadcasts the set of indices that are in consensus with node-2. node-3 computes the corresponding sets of indices lying outside the guard band, and then broadcasts the two sets of indices that are in consensus with both node-1 and node-2. Let $(\mathcal{R}_{\varphi(h_{12})}, \mathcal{R}_{\varphi(h_{13})})$ denote the two sets of indices in consensus among the three nodes. Then, all the three nodes generate the set $\mathcal{V} = \mathcal{R}_{\varphi(h_{12})} \cap \mathcal{R}_{\varphi(h_{13})}$, where $\mathcal{V}$ denotes the indices where both $\{R_1^{h_{12}}, R_2^{h_{12}}, R_3^{h_{12}}\}$ and $\{R_1^{h_{13}}, R_2^{h_{13}}, R_3^{h_{13}}\}$ are in consensus. With the likelihood based CSR selection strategy, for the indices in $\mathcal{V}$, node-1 calculates the probability of errors at node-2 and node-3 by locally generating the distributions at their side. Then it broadcasts the index of the chosen CSR to both node-2 and node-3. Finally, all the nodes use the CSR samples of $\varphi(h_{12})$ and $\varphi(h_{13})$ to extract a secret-key.

4.4 LLR Based Reconciliation with LDPC codes

We present an optimal LLR generation scheme for the CSR generated by the A-SQGSK protocol. Since node-1 observes the CSR samples from $\varphi(h_{12})$ and $\varphi(h_{13})$ through probing signals, we use the key generated at node-1 as the reference key, and then apply the LDPC reconciliation algorithm at node-2 and node-3. As the statistics of the underlying noise are different, the LLR generation scheme depends on whether the reconciliation is implemented at the (i) reciprocal node, which is the node that inherently observes the CSR through channel reciprocity, e.g., node-2 when the CSR is $\{R_1^{h_{12}}, R_2^{h_{12}}, R_3^{h_{12}}\}$, or the (ii) decoding node, which is the node that learns the unseen CSR through the process of decoding and subtraction over the ring, e.g., node-2 when the CSR is $\{R_1^{h_{13}}, R_2^{h_{13}}, R_3^{h_{13}}\}$. We discuss the LLR generation scheme at both these types of nodes.

$$\text{Prob}_b(R_2^{h_{13}}) = \sum_{x_u \in \mathcal{S}} \text{Prob}(\psi^{-1}(x_u)) \left(\varrho^{R_2^{h_{13}}} \left(\varnothing^{\phi(R_2^{h_{12}})} \left(\mho \left(\mathcal{N} \left(\phi^{-1}(\phi(x_u) \oplus \Theta(R_2^{h_{12}}) \right), \frac{E_{avg}\sigma^2}{|h_{12} + e_2^{(1)}|^2}), \bar{\mathcal{A}}_I \right) \right) \right) \right) \tag{4.10}$$

4.4.1 LLR Generation at the Reciprocal Node

Let $P(X, Y)$ denote the joint PMF between the CSR at node-1, denoted by X, and the CSR at the reciprocal node, denoted by Y. For all the $2^{\frac{m}{2}}$ points in $\bar{\mathcal{A}}_I$, the probability that the CSR sample at node-1 is quantized to bit 1 and bit 0 is given by $\text{Prob}_1(p) = \sum_{X \in \bar{\mathcal{A}}_I^+} P(X, Y = p)$ and $\text{Prob}_0(p) = \sum_{X \in \bar{\mathcal{A}}_I^-} P(X, Y = p)$, respectively, where $\bar{\mathcal{A}}_I^+$ and $\bar{\mathcal{A}}_I^-$ are the positive and negative points of $\bar{\mathcal{A}}_I$, respectively, and p is the CSR sample observed at the reciprocal node. Finally, the LLR is computed as $\log\left(\frac{\text{Prob}_0(p)}{\text{Prob}_1(p)}\right)$.

4.4.2 LLR Generation at the Decoding Node

For exposition, we explain the LLR generation scheme at node-2 when the CSR is $\{R_1^{13}, R_2^{13}, R_3^{13}\}$. As a result, using $R_2^{h_{13}}$, node-2 generates the LLR on the bit generated at node-1 using $R_1^{h_{13}}$. The corresponding quantized version of the channel with node-1, as seen by node-2, is $C_2^{h_{12}}$. Henceforth, we use $\hat{\bar{\mathcal{A}}}_I^+$ and $\hat{\bar{\mathcal{A}}}_I^-$ to represent the positive and negative points in $\bar{\mathcal{A}}_I$, respectively, that are out of guard bands upon quantization using $\mathcal{Q}(\cdot)$.

Theorem 2. *Using the CSR $R_2^{h_{13}}$ at node-2, the probability that node-1 quantizes its CSR $R_1^{h_{13}}$ to bit b, for $b \in \{0, 1\}$, is given in (4.10), where $\mathcal{S} = \hat{\bar{\mathcal{A}}}_I^-$ and $\mathcal{S} = \hat{\bar{\mathcal{A}}}_I^+$ when $b = 0$ and $b = 1$, respectively.*

Proof. node-2 intends to build an aposteriori PMF on its samples conditioned on the hypothesis that the CSR $h_1^{h_{13}} \in \hat{\bar{\mathcal{A}}}_I^+$. This way, node-2 generates the likelihood of CSR $R_1^{h_{13}}$ being bit 1 conditioned on its CSR $R_2^{h_{13}}$. Similarly, using all possible cases of $h_1^{h_{13}} \in \hat{\bar{\mathcal{A}}}_I^-$, it generates the likelihood of CSR $R_1^{h_{13}}$ being bit 0 conditioned on its CSR $R_2^{h_{13}}$. Henceforth, throughout this proof, we explain the steps for generating the likelihood of CSR $R_1^{h_{13}}$ being bit 1. Similar steps can be followed to obtain the likelihood of CSR $R_1^{h_{13}}$ being bit 0. Assuming CSR $R_1^{h_{13}} = \psi^{-1}(x_u), x_u \in \hat{\bar{\mathcal{A}}}_I^+$, wherein $R_1^{h_{13}}$ takes the ψ^{-1} of u-th element of $\hat{\bar{\mathcal{A}}}_I^+$, for $1 \leq u \leq |\hat{\bar{\mathcal{A}}}_I^+|$. node-2 hypothesizes the point broadcast by node-1 as $\mu_{x_u} = \phi^{-1}\Bigg(\phi(x_u) \oplus \Theta(R_2^{h_{12}})\Bigg)$. Note that it is possible to assume this since at mid-to-high SNR ranges, we have $R_1^{h_{12}} = R_2^{h_{12}}$. Using the estimate of the channel between node-1 and node-2, the instantaneous SNR at node-2 is $\frac{|h_{12}+e_2^{(1)}|^2}{E_{avg}\sigma^2}$. Therefore, the PDF of the effective noise as seen by node-2 is Gaussian distributed given by $\mathcal{N}\left(\mu_{x_u}, \frac{E_{avg}\sigma^2}{|h_{12}+e_2^{(1)}|^2}\right)$. Furthermore, since node-2 decodes on the PAM constellation using MAP decoder, the corresponding aposteriori PMF on the PAM points is given by $\mathcal{H}_{R_2^{h_{13}}} = \mho\left(\mathcal{N}\left(\mu_{x_u}, \frac{E_{avg}\sigma^2}{|h_{12}+e_2^{(1)}|^2}\right), \bar{\mathcal{A}}_I\right)$, wherein the PMF $\mathcal{H}_{R_2^{h_{13}}}$ is listed on PAM points when enumerated in the increasing order. Finally, since node-2 obtains the CSR $R_2^{h_{13}}$ by performing a modulo subtraction on the algebraic ring using $R_2^{h_{12}}$, node-1 incorporates the corresponding changes in the PMF as $\bar{\mathcal{H}}_{R_2^{h_{13}}} = \varnothing^{\phi(R_2^{h_{12}})}(\mathcal{H}_{R_2^{h_{13}}})$. Thus, node-2 generates an aposteriori PMF on the CSR $R_2^{h_{13}}$ under the hypothesis that $\psi^{-1}(x_u)$ was the CSR at node-1. Using the recovered CSR point $R_2^{h_{13}}$, node-2 evaluates the probability using the aposteriori PMF as $\varrho^{R_2^{h_{13}}}(\bar{\mathcal{H}}_{R_2^{h_{13}}})$. Overall, by considering all possible CSR points of $\hat{\bar{\mathcal{A}}}_I^+$, the probability that the CSR point at node-1 is quantized to bit 1 is given by $\text{Prob}_1(R_2^{h_{13}}) = \sum_{x_u \in \hat{\bar{\mathcal{A}}}_I^+} \varrho^{R_2^{h_{13}}}(\bar{\mathcal{H}}_{R_2^{h_{13}}})\text{Prob}(\psi^{-1}(x_u))$, where $\text{Prob}(\psi^{-1}(x_u))$ denotes the probability that the CSR $R_1^{h_{13}}$ takes the value $\psi^{-1}(x_u)$. Along the similar lines, the probability that the CSR point at node-1 is quantized to bit 0 is given by $\text{Prob}_0(R_2^{h_{13}}) = \sum_{x_u \in \hat{\bar{\mathcal{A}}}_I^-} \varrho^{R_2^{h_{13}}}(\bar{\mathcal{H}}_{R_2^{h_{13}}})\text{Prob}(\psi^{-1}(x_u))$. Finally, the LLR of the bit at node-1 is given by $\log\left(\frac{\text{Prob}_0(R_2^{h_{13}})}{\text{Prob}_1(R_2^{h_{13}})}\right)$. Fig. 4.4 depicts an example for LLR generation at node-2 when it observes $R_2^{h_{13}}$ as the decoded CSR. □

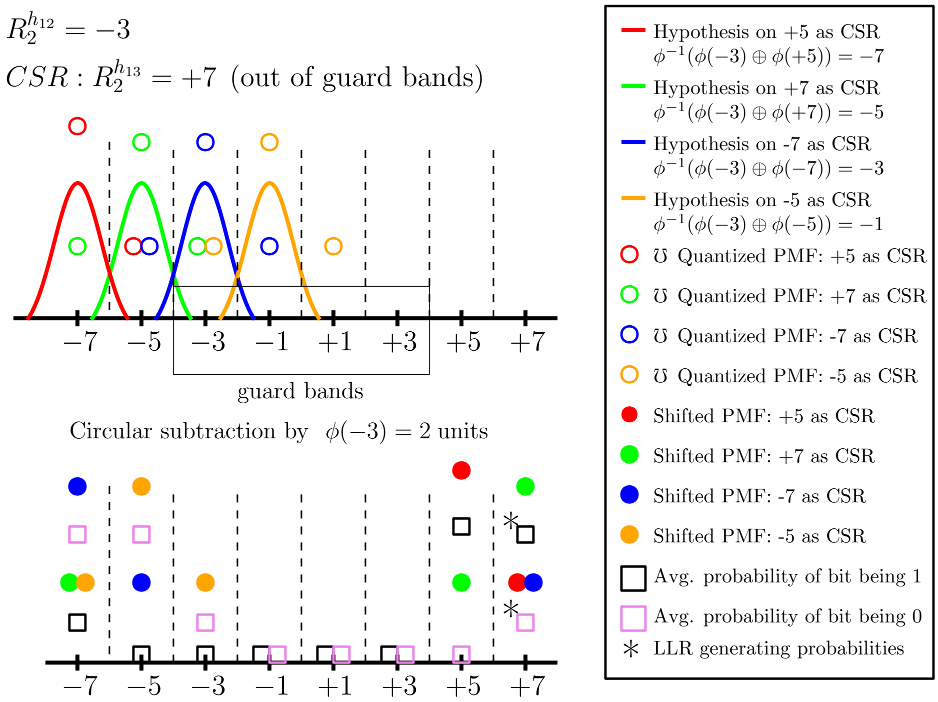

$R_2^{h_{12}} = -3$
$CSR : R_2^{h_{13}} = +7$ (out of guard bands)
−7 −5 −3 −1 +1 +3 +5 +7
guard bands
Circular subtraction by $\phi(-3) = 2$ units
−7 −5 −3 −1 +1 +3 +5 +7
Hypothesis on +5 as CSR
$\phi^{-1}(\phi(-3) \oplus \phi(+5)) = -7$
Hypothesis on +7 as CSR
$\phi^{-1}(\phi(-3) \oplus \phi(+7)) = -5$
Hypothesis on -7 as CSR
$\phi^{-1}(\phi(-3) \oplus \phi(-7)) = -3$
Hypothesis on -5 as CSR
$\phi^{-1}(\phi(-3) \oplus \phi(-5)) = -1$
$\mho$ Quantized PMF: +5 as CSR
$\mho$ Quantized PMF: +7 as CSR
$\mho$ Quantized PMF: -7 as CSR
$\mho$ Quantized PMF: -5 as CSR
Shifted PMF: +5 as CSR
Shifted PMF: +7 as CSR
Shifted PMF: -7 as CSR
Shifted PMF: -5 as CSR
Avg. probability of bit being 1
Avg. probability of bit being 0
∗ LLR generating probabilities

Figure 4.4: Figure depicts an example for LLR generation for $R_2^{h_{13}}$ at node-2. The top figure illustrates the process of generating aposteriori PMFs on the QAM symbols, whereas the bottom figure illustrates the circular shift operation incorporating subtraction over algebraic ring, and also the computation of LLR. For simplification, it is assumed that the CSR lies in the constellation $\bar{\mathcal{A}}$.

We highlight that no additional communication-overheads are involved in LLR generation. Furthermore, this method is also optimal at mid-to-high SNR and moderate values of m.

4.5 Simulation Results

4.5.1 Opportunistic Selection of CSR

To showcase the advantages of the opportunistic CSR selection, we present its key-rate along with that of the A-SQGSK protocol, wherein the CSR is fixed to $\varphi(h_{12})$. In this context, key-rate is defined as the average number of secret bits generated among the three nodes per CSR sample. The plots are presented in Fig. 4.5 for the cases when the two-level consensus algorithm delivers secret-keys with an initial error rate of 10^{-1} and 10^{-2}. In this context, initial error rate is defined as the upper bound on the desired mismatch rate among the nodes when choosing the guard bands for the quantizer $\mathcal{Q}(\cdot)$. With an initial error rate of 10^{-1}, the plots show that the benefits of the opportunistic method is marginal, and this observation is attributed to the fact that the number of samples from $\varphi(h_{12})$ and $\varphi(h_{13})$ that are jointly in consensus is large. However, when the initial error rate is 10^{-2}, the benefits are significant since the number of additional CSR samples coming out of $\varphi(h_{13})$ is large. We note that the above observations continue to hold good for different values of m, which captures the size of the constellation. In this work, we have also proposed a method to choose the CSR sample on those coherence-blocks whenever both $\varphi(h_{12})$ and $\varphi(h_{13})$ are in consensus.

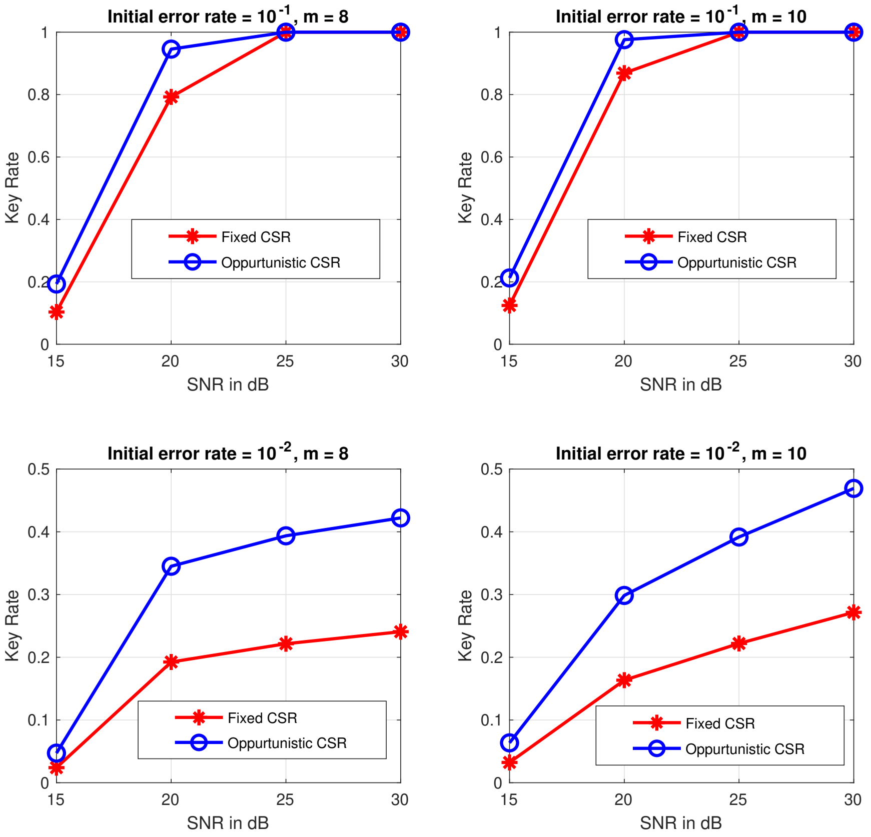


Figure 4.5: Key-rate improvement with opportunistic selection of CSR.

To showcase the efficacy of the CSR selection method, in Fig. 4.6, we plot the error rate offered by our scheme on the CSR samples when both $\varphi(h_{12})$ and $\varphi(h_{13})$ are in consensus. The plots show that the likelihood based CSR selection outperforms the channel-strength based CSR selection, wherein node-1 chooses the CSR that offers weaker channel-strength since the weaker channel degrades the SNR when recovering the other CSR.

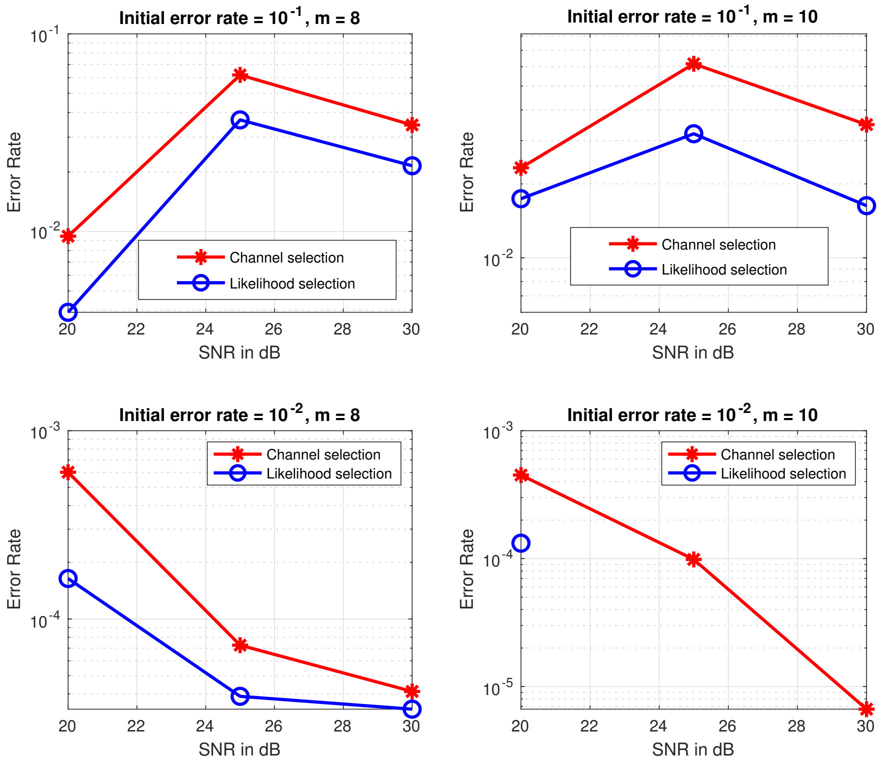


Figure 4.6: Mismatch rate on CSR selected from coherence-blocks when both $\varphi(h_{12})$ and $\varphi(h_{13})$ are in consensus. When the initial error rate is 10^{-2} and $m = 10$, an error rate of 0 is achieved at 25 and 30 dB.

4.5.2 LDPC Based Reconciliation for Algebraic-SQGSK

In the context of opportunistic A-SQGSK protocol, the CSR samples in consensus come from either $\varphi(h_{12})$ or $\varphi(h_{13})$. With respect to the CSR samples from $\varphi(h_{12})$, node-2 and node-3 generate the LLR values on the bits generated at node-1 by following the algorithm at the reciprocal node and the decoding node, respectively. Similarly, for the CSR samples from $\varphi(h_{13})$, node-3 and node-2 generate the LLR values on the bits generated at node-1 by following the algorithm at the reciprocal node and the decoding node, respectively. To execute LDPC based reconciliation, an (N, K) binary LDPC code characterized by the parity check matrix $\mathbf{H}$ of dimension $(N-K) \times N$ is used. With $\mathbf{x} \in \mathbb{F}_2^N$ denoting an N-length binary key generated at node-1, let $\mathbf{s}$ denote its syndrome vector $\mathbf{s} = \mathbf{H}\mathbf{x} \in \mathbb{F}_2^{(N-K)\times 1}$, wherein the multiplication operation is over the field $\mathbb{F}_2$. Subsequently, the syndrome vector

s is broadcast to node-2 and node-3, which in turn use it to reduce the mismatch rate by using a message-passing algorithm [LQDD20].

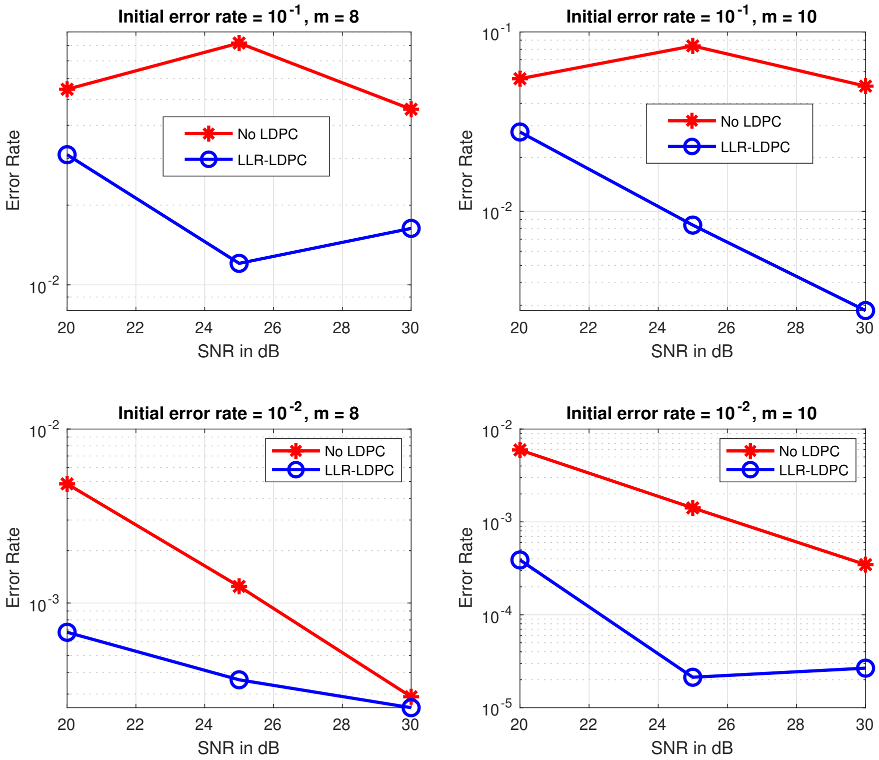


Figure 4.7: Improvement in the mismatch rate among the keys at the three nodes when LDPC based reconciliation is employed at node-2 and node-3.

In Fig. 4.7, we plot the performance of LDPC reconciliation when an $(N = 12, K = 9)$ LDPC code [Wol] is employed. To generate the simulation results, we use the CSR samples out of a two-level consensus algorithm with an initial error rate of 10^{-1} and 10^{-2}. Upon using LDPC reconciliation, we observe that the mismatch rate among the nodes reduces significantly. It is important to note that the benefits of the reconciliation algorithm is attributed to the LLR generation method at the reciprocal node and the decoding node. While we see significant improvements in the mismatch rate in the GSK, we believe that with the use of large block-length LDPC codes, our LLR generation method can ensure zero mismatch rate among the nodes. From the plots, we also remark that the error rate values at 25 dB and 30 dB are more than that at 20 dB when the initial error rate is 10^{-1}, and this is because the initial error rate is only used as an upper bound.

4.6 Discussion

State-of-the-art A-SQGSK protocol does not provide higher key-rate as it utilises only one of the channel as the CSR, therefore, we have explored a method to harvest the two available channels as the CSR in an opportunistic fashion. It is shown that in the case when both the channels are in consensus, only one of the two channels can be used as the CSR otherwise confidentiality will be compromised. Therefore, a selection criterion is presented which ensures that the error is minimised when selecting one of the two available channels as the CSR. In order to facilitate reconciliation, input to message passing algorithm is the LLR associated with the CSR provided by A-SQGSK protocol, hence, we present an LLR generation technique for two different set of nodes, i.e., reciprocal and decoding node.

Chapter 5

Conclusion

We have revisited the design of multi-level quantization schemes for key generation, and have shown that the knowledge of joint PDF can be exploited to generate secret-keys with maximum entropy and high symbol-rate. For the case of group key generation, we have laid out design rules to jointly choose the parameters of the A-SQGSK protocol and the EM-EM algorithm, particularly, a relaxed criterion is presented to make EM-EM amenable to A-SQGSK. An opportunistic CSR selection scheme was presented to achieve a higher key-rate than the state-of-the-art A-SQGSK scheme when synthesizing a GSK in a three-node network. Towards guaranteeing non-zero leakage of the CSR to an eavesdropper, we have shown that the proposed CSR selection strategy picks the CSR that minimizes the mismatch rate between the nodes. Finally, to facilitate information reconciliation on the proposed opportunistic CSR selection scheme, we have proposed a novel LLR generation scheme that exploits the underlying noise statistics at the nodes as well as the algebraic ring structure.

5.1 Direction for Further Research

The ideas and the techniques proposed in this thesis can be extended by generalizing the protocols to a wireless networks with more than three nodes. The extension of EM-EM framework, opportunistic selection and LLR generation to such a network would require the knowledge of inter-connection of nodes and a carefully designed CSR sharing protocol.

Bibliography

[AC93] R. Ahlswede and I. Csiszar. Common randomness in information theory and cryptography. i. secret sharing. *IEEE Transactions on Information Theory*, 39(4):1121–1132, 1993.

[CSD09] T. Chou, A. M. Sayeed, and S. C. Draper. Minimum energy per bit for secret key acquisition over multipath wireless channels. In *2009 IEEE International Symposium on Information Theory*, pages 2296–2300, 2009.

[GK11] S. Gollakota and D. Katabi. Physical layer wireless security made fast and channel independent. In *the Proceedings of 2011 IEEE INFOCOM*, pages 1125–1133, 2011.

[HCH17] J. Harshan, S. Chang, and Y. Hu. Insider-attacks on physical-layer group secret-key generation in wireless networks. In *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6, 2017.

[HHL17] Y. W. P. Hong, L. Huang, and H. Li. Vector quantization and clustered key mapping for channel-based secret key generation. *IEEE TIFS*, 12(5):1170–1181, 2017.

[HJR20] J. Harshan, Rohit Joshi, and Manish Rao. Group secret-key generation using algebraic rings in wireless networks. *arXiv 1805.00743 cs.IT*, 2020.

[JH21a] Rohit Joshi and J. Harshan. On opportunistic selection of common randomness and llr generation for algebraic group secret-key generation. In *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, pages 1–7, 2021.

[JH21b] Rohit Joshi and J. Harshan. On opportunistic selection of common randomness and llr generation for algebraic group secret-key generation. *arXiv 2103.02195 cs.IT*, 2021.

[JJR21] Harshan Jagadeesh, Rohit Joshi, and Manish Rao. Group secret-key generation using algebraic rings in wireless networks. *IEEE Transactions on Vehicular Technology*, 70(2):1538–1553, 2021.

[JPC$^+$09] S. Jana, S. Nandha Premnath, M. Clark, S. K. Kasera, N. Patwari, and S. V. Krishnamurthy. On the effectiveness of secret key extraction from wireless signal strength in real environments. In *the Proc. of the 15th ACM MobiCom '09*, pages 321–332, 2009.

[KWTP13] Konstantin Kravtsov, Zhenxing Wang, Wade Trappe, and Paul R. Prucnal. Physical layer secret key generation for fiber-optical networks. *Opt. Express*, 21(20):23756–23771, Oct 2013.

[LDS12] Y. Liu, S. C. Draper, and A. M. Sayeed. Exploiting channel diversity in secret key generation from multipath fading randomness. *IEEE TIFS*, 7(5):1484–1497, 2012.

[LHH19] G. Li, L. Hu, and A. Hu. Lightweight group secret key generation leveraging non-reconciled received signal strength in mobile wireless networks. In *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–6, 2019.

[LLB13] C. Ling, L. Luzzi, and M. R. Bloch. Secret key generation from gaussian sources using lattice hashing. In *2013 IEEE International Symposium on Information Theory*, pages 2621–2625, 2013.

[LLN^{+}19] K. Li, L. Lu, W. Ni, E. Tovar, and M. Guizani. Cooperative secret key generation for platoon-based vehicular communications. In *the Proc. of IEEE International Conference on Communications, 2019*, pages 1–6, 2019.

[LQDD20] G. Limei, R. Qi, J. Di, and H. Duan. Qkd iterative information reconciliation based on ldpc codes. *Int. Journal of Theoritical Physics*, pages 1717–1729, March 2020.

[LYW^{+}14] H. Liu, J. Yang, Y. Wang, Y. Chen, and C. E. Koksal. Group secret key generation via received signal strength: Protocols, achievable rates, and implementation. *IEEE Transactions on Mobile Computing*, 13(12):2820–2835, 2014.

[Mau93] U. M. Maurer. Secret key agreement by public discussion from common information. volume 39, pages 733–742, Piscataway, NJ, USA, May 1993. IEEE Press.

[Max60] J. Max. Quantizing for minimum distortion. *IRE Transactions on Information Theory*, 6(1):7–12, 1960.

[MTM^{+}08] Suhas Mathur, Wade Trappe, Narayan Mandayam, Chunxuan Ye, and Alex Reznik. Radio-telepathy: Extracting a secret key from an unauthenticated wireless channel. pages 128–139, 2008.

[MW03] U. Maurer and S. Wolf. Secret-key agreement over unauthenticated public channels - part ii: the simulatability condition. volume 49, pages 832–838, April 2003.

[PCB13] Alexandre J. Pierrot, Rémi A. Chou, and Matthieu R. Bloch. The Effect of Eavesdropper's Statistics in Experimental Wireless Secret-Key Generation. *arXiv e-prints*, December 2013.

[RH18] M. Rao and J. Harshan. Practical physical-layer group secret-key generation in three-user wireless networks. pages 342–346, 2018.

[RH19] M. Rao and J. Harshan. Low-latency exchange of common randomness for group-key generation. In *the Proc. of PIMRC*, pages 1–6, 2019.

[RSW11] K. Ren, H. Su, and Q. Wang. Secret key generation exploiting channel characteristics in wireless communications. *IEEE Wireless Communications*, 18(4):6–12, 2011.

[SH11] Y. E. H. Shehadeh and D. Hogrefe. An optimal guard-intervals based mechanism for key generation from multipath wireless channels. In *2011 4th IFIP International Conference on New Technologies, Mobility and Security*, pages 1–5, 2011.

[SMDF11] M. J. Siavoshani, S. Mishra, S. N. Diggavi, and C. Fragouli. Group secret key agreement over state-dependent wireless broadcast channels. In *the Proc. of 2011 IEEE International Symposium on Information Theory*, pages 1960–1964, 2011.

[TLQ15] C. D. Truyen Thai, J. Lee, and T. Q. S. Quek. Secret group key generation in physical layer for mesh topology. In *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, 2015.

[Wol] Jack Keil Wolf. *An Introduction to Error Correcting Codes.* available online at: `acsweb.ucsd.edu/~afazelic/ece154c/ErrorCorrection-JackWolf.pdf`.

[WZN12] Y. Wei, C. Zhu, and J. Ni. Group secret key generation algorithm from wireless signal strength. In *2012 Sixth International Conference on Internet Computing for Science and Engineering*, pages 239–245, 2012.

[XCD$^+$16] P. Xu, K. Cumanan, Z. Ding, X. Dai, and K. K. Leung. Group secret key generation in wireless networks: Algorithms and rate optimization. *IEEE Transactions on Information Forensics and Security*, 11(8):1831–1846, 2016.

[YDS11] Yanpei Liu, S. C. Draper, and A. M. Sayeed. Secret key generation through ofdm multipath channel. In *2011 45th Annual Conference on Information Sciences and Systems*, pages 1–6, 2011.

[YMR$^+$10] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. B. Mandayam. Information-theoretically secret key generation for fading wireless channels. *IEEE TIFS*, 5(2):240–254, 2010.

[YR07] C. Ye and A. Reznik. Group secret key generation algorithms. In *2007 IEEE International Symposium on Information Theory*, pages 2596–2600, 2007.

[YRS06] C. Ye, A. Reznik, and Y. Shah. Extracting secrecy from jointly gaussian random variables. In *2006 IEEE International Symposium on Information Theory*, pages 2593–2597, 2006.

[ZWCM10] K. Zeng, D. Wu, A. Chan, and P. Mohapatra. Exploiting multiple-antenna diversity for shared secret key generation in wireless networks. In *2010 Proceedings IEEE INFOCOM*, pages 1–9, 2010.

LIST OF PAPERS BASED ON THESIS

Papers in Refereed Journals

1. H. Jagadeesh, **R. Joshi** and M. Rao, "Group Secret-Key Generation Using Algebraic Rings in Wireless Networks," in IEEE Transactions on Vehicular Technology (IEEE TVT), vol. 70, no. 2, pp. 1538-1553, Feb. 2021.

Presentations in Conferences

1. **Rohit Joshi** and J. Harshan, "On Opportunistic Selection of Common Randomness and LLR Generation for Algebraic Group Secret-Key Generation," IEEE Vehicular Technology Conference (VTC2021-Spring), Helsinki, 2021.

2. S. Sriraam, S. Sajeev, **R. Joshi**, A. Vithalkar, M. Bansal and H. Jagadeesh, "Implementation of 5G Authentication and Key Agreement Protocol on Xbee Networks," (IEEE COMSNETS 2020).