Quantum Networks: State Transmission and Network Operation

by

Wenhan Dai

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2020

Author .....
Department of Aeronautics and Astronautics
January 22, 2020

Certified by .....
Moe Z. Win
Professor, Aeronautics and Astronautics
Thesis Supervisor

Certified by .....
Sanjoy Mitter
Professor, Electrical Engineering and Computer Science

Certified by .....
Marco Chiani
Professor, Electronic and Information Engineering "G. Marconi,"
University of Bologna

Accepted by .....
Sertac Karaman
Chair, Graduate Committee

Quantum Networks: State Transmission and Network Operation

by

Wenhan Dai

Submitted to the Department of Aeronautics and Astronautics
on January 22, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Aeronautics and Astronautics

Abstract

Quantum information science is believed to create the next technological revolution. As key ingredients of quantum information science, quantum networks enable various technologies such as secure communication, distributed quantum sensing, quantum cloud computing, and next-generation positioning, navigation, and timing. The main task of quantum networks is to enable quantum communication among different nodes in the network. This includes the topics such as the transmission of quantum states involving multiple parties, the processing of quantum information at end nodes, and the distribution of entanglement among remote nodes. Since quantum communication has its own peculiar properties that have no classical counterparts, the protocols and strategies designed for classical communication networks are not well-suited for quantum ones. This calls for new concepts, paradigms, and methodologies tailored for quantum networks. To that end, this thesis studies the design and operation of quantum networks, with focus on the following three topics: state transmission, queueing delay, and remote entanglement distribution.

The first part develops protocols to broadcast quantum states from a transmitter to N different receivers. The protocols exhibit resource tradeoffs between multiparty entanglement, broadcast classical bits (bcbits), and broadcast quantum bits (bqubits), where the latter two are new types of resources put forth in this thesis. We prove that to send 1 bqubit to N receivers using shared entanglement, $O(\log N)$ bcbits are both necessary and sufficient. We also show that the protocols can be implemented using poly(N) basic gates composed of single-qubit gates and CNOT gates.

The second part introduces a tractable model for analyzing the queuing delay of quantum data, referred to as quantum queuing delay (QQD). The model employs a dynamic programming formalism and accounts for practical aspects such as the finite memory size. Using this model, we develop a cognitive-memory-based policy for memory management and show that this policy can decrease the average queuing delay exponentially with respect to memory size.

The third part offers a design of remote entanglement distribution (RED) protocols that maximize the entanglement distribution rate (EDR). We introduce the concept

of enodes, representing the entangled quantum bit (qubit) pairs in the network. This concept enables us to design the optimal RED protocols based on the solutions of some linear programming problems. Moreover, we investigate RED in a homogeneous repeater chain, which is a building block for many quantum networks. In particular, we determine the maximum EDR for homogeneous repeater chains in a closed form.

The contributions of this work provide guidelines for the design and implementation of quantum networks.

Thesis Supervisor: Moe Z. Win
Title: Professor, Aeronautics and Astronautics

Committee Member: Sanjoy Mitter
Title: Professor, Electrical Engineering and Computer Science

Committee Member: Marco Chiani
Title: Professor, Electronic and Information Engineering "G. Marconi," University of Bologna

Acknowledgments

First, I am grateful to my advisor, Prof. Moe Win, for his guidance, support, and caring throughout my years at MIT. He gave me the opportunity and freedom to explore research topics. The lessons I learned from him will benefit me in my future career. I also thank my thesis committee members Prof. Marco Chiani and Prof. Sanjoy Mitter for their kind support and insightful discussions, and my thesis readers Prof. Andrea Conti and Dr. Jack Winters for their careful reading and helpful advice.

I am also thankful to my colleagues for their support and help, either technical or non-technical. I am especially grateful to Tianyi Peng, Liangzhong Ruan, Yuan Shen, Florian Meyer, Stefano Guerrini, Zhenyu Liu, Zehao Yu, Junting Chen, Tianheng Wang, Stefania Bartoletti, Igor Kadota, and Prof. Xiaodi Wu.

Moreover, I would like to thank the LIDS staff members Richard, Rachel, Jennifer, Brian, Lynne, and Gracie, for their kind help all these years. I am especially thankful to Beth Marois at the Department of Aeronautics and Astronautics of MIT, for always been so kind and helpful during my Ph.D. program.

I would also like to acknowledge support from the National Science Foundation, the Office of Naval Research, the National Institute of Standards and Technology, and the MIT Institute for Soldier Nanotechnologies.

Finally and most importantly, I cannot express enough my gratitude to my wife, Xiao Han, my parents, Xiuling Xu and Guanghui Dai, my parents-in-law, Yanhua Zheng and Zhizhong Han, for their unconditional love, encouragement, and support. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

Abbreviations

bcbit	broadcast cbit
bqubit	broadcast qubit
cbit	classical bit
DAG	directed acyclic graph
ebit	bit of entanglement
EDR	entanglement distribution rate
POVM	positive-operator valued measure
QQD	quantum queuing delay
RSP	remote state preparation
SDP	semidefinite programming

Chapter 1

Introduction

In this chapter, we provide background and preliminaries on quantum networks, motivate the research on quantum broadcasting, quantum queuing delay, as well as remote entanglement distribution, and review related research results. We then summarize the contributions of the thesis and provide the thesis organization.

1.1 Background and Preliminaries

Quantum information science has the potential to create the next technological revolution [1–3], enabling quantum communication [4–6], quantum sensing [7–9], quantum computing [10–14], as well as next-generation positioning, navigation, and timing [15–17].

The breakthroughs in the physical layer in quantum information science bring myriad novel research topics, most of which are interdisciplinary. The research on quantum networks is such an example. Quantum networks are the key ingredients in quantum information science as they can support quantum communication, quantum computing, and quantum metrology [18]. The main task of quantum networks is to enable quantum communication among different nodes that may be far apart. While physical layer implementations of quantum networks are advancing rapidly [19–22], little is known about the design and analysis of upper layer transmission and operation strategies that are responsible for sending quantum information reliably

and efficiently. The lack of a systematic framework for transmission and operation strategies limits the power of quantum networks. This critical issue motivates the research in this thesis, including the topics such as the transmission of quantum states involving multiple parties, the distribution of entanglement among remote nodes, and the processing of quantum information at end nodes.

Quantum networks have both similarities and differences compared to their classical counterparts. Regarding similarities, quantum networks employ many concepts that originate from classical communication networks such as protocol stack. On the other hand, quantum networks have their own peculiar properties that do not exist in classical networks. For example, a distinctive property of quantum networks comes from quantum entanglement that represents the non-local interconnection among quantum objects. Therefore, the protocols and strategies designed for classical communication networks are not well-suited for quantum ones. This calls for new concepts, paradigms, and methodologies tailored for quantum networks.

Before moving on to the specific discussions of the topics, we present some preliminaries that are the foundations of our work.

Quantum Teleportation

The quantum teleportation protocol aims to transmit quantum states to a receiver using entanglement and classical communication [23]. In quantum teleportation, a transmitter performs Bell measurements on the quantum bit (qubit) to be transmitted, denoted as $|\psi\rangle$, as well as its half of the entangled qubit pair; the measurement result is then sent to the receiver, which will perform operations on its half of the entangled qubit pair accordingly; the resulting qubit at the receiver then becomes $|\psi\rangle$. This protocol permits the quantum node to transmit 1 qubit at the cost of sending 2 classical bits (cbits) and consuming 1 entangled qubit pair.

Entanglement Swapping

A critical operation in the use of building quantum networks is entanglement swapping [24]. Entanglement swapping can be seen as a special case of teleportation [23] and is explained as follows. Suppose there are three nodes i, j, and k. Node i has one qubit, node j two, and node k one. Node i's qubit and node j's first qubit are maximally entangled, and so are node j's second and node k's qubit. Node j teleports the qubit entangled with the one in node i to node k. Then node i's qubit is entangled with node k's even though these two nodes have never directly interacted with each other.

Remote State Preparation

Another important technique for transmitting quantum states is remote state preparation (RSP) [25–29]. In contrast to teleportation, the transmitter is given the knowledge about the quantum states to be sent. Such knowledge makes it possible to transmit quantum states with fewer resources. In [30], it is shown that using backward communication and entanglement distillation, a pure quantum state in a two-dimensional Hilbert space can be transmitted by consuming 1 cbit and 3.79 bit of entanglements (ebits) asymptotically. In [31], a transmitter performs measurements (depending on the quantum state to be transmitted) on its half of the entanglement; the measurement result is then sent to the receiver, which will perform operations on its half of the entanglement accordingly; the resulting qubit at the receiver then becomes the desired quantum state. This protocol consumes 1 cbit and 1 ebit asymptotically per qubit.

1.2 Motivation and Related Research

In this section, we present the motivation and review relevant studies about quantum broadcasting, quantum queuing delay (QQD), and remote entanglement distribution (RED).

Quantum Broadcasting

The protocols for sending quantum states exploiting classical communication and quantum correlation, such as teleportation [23] and RSP [25, 26, 30, 31], exhibit fundamental tradeoffs among different types of resources. For example, a qubit can be sent via teleportation using 1 maximally entangled pair (or equivalently, 1 ebit) and 2 cbits [23], leading to the following schematic tradeoff:

$$1 \text{ ebit} + 2 \text{ cbits} \rightarrow 1 \text{ qubit}.$$

It is natural to ask if it is possible to generalize this tradeoff to a network setting. One class of the network settings that attracts research interest is quantum broadcasting [32], in which a transmitter aims at sending a quantum state identically to a collection of receivers using classical and quantum resources [33–35].[1] Such a broadcast setting calls for the new types of communication resources. In a network with 1 transmitter and N receivers, we define two types of resources: a broadcast cbit (bcbit) as the transmitter's ability to send 1 bit of classical information identically to N different receivers; and a broadcast qubit (bqubit) as the transmitter's ability to send a pure quantum state in a two-dimensional Hilbert space identically and exactly with high probability to N different receivers.[2] Note that bcbits have been used widely as a metric for applications such as radio and television [36], whereas bqubits can be used as a metric for applications such as quantum metrology [37] and quantum digital signatures [38]. A typical quantum broadcasting protocol consumes $f_1(N)$ ebits[3] and $f_2(N)$ bcbits, leading to the following schematic tradeoff:

$$f_1(N) \text{ ebits} + f_2(N) \text{ bcbits} \rightarrow 1 \text{ bqubit}. \tag{1.1}$$

[1] The notion of *identical* means that the reduced density matrices at each receiver are the same.

[2] The notion of *exact* means that the output state at each receiver is the same as the desired state.

[3] Since there is no commonly agreed measure of multiparty entanglement, this thesis considers the amount of ebits that are shared between two parties, namely, the transmitter and the receivers, as in [35].

There are some studies that investigated the problem of sending a quantum state identically to multiple receivers [32–35]. However, the protocol [33–35] results in the fidelity between the desired state and the state at each receiver to be at most $(2N+1)/(3N)$, thus unsuitable for sending bqubits. The protocols in [32] are valid only for $N \leqslant 3$ unless an infinite number of bcbits is utilized. To the best of the author's knowledge, there are no existing quantum broadcasting protocols for general N.

The difficulty of quantum broadcasting lies mainly in the fact that only distributed measurements and operations among spatially separated receivers are allowed. As a consequence, certain operations involving multiple receivers are not permitted in the design of protocols. The fundamental question related to quantum broadcasting is how to develop distributed measurements and operations that have capabilities comparable to centralized ones. The answer to this question enables the design of efficient quantum broadcasting protocols and provides fundamental tradeoff among different types of resources.

The goal is to develop protocols for sending an arbitrary pure quantum state $|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ from a transmitter, Alice, to N different receivers in two scenarios: 1) Alice knows the state $|\psi\rangle$ (non-oblivious with a known state) and 2) Alice has the state $|\psi\rangle^{\otimes N}$ (oblivious with multiple copies).[4] We aim at minimizing the resources, i.e., $f_1(N)$ and $f_2(N)$ in (1.1), for sending 1 bqubit.

Note that sending a bqubit to N receivers can be trivially achieved by sequentially performing teleportation, resulting in the schematic tradeoff $O(N)$ ebits + $O(N)$ bcbits $\rightarrow$ 1 bqubit. With this approach, the last receiver experiences the latency of $O(N)$, which may be undesirable for many applications, especially those involving large N. Note that the state $|\psi\rangle^{\otimes N}$ to be sent is in a Hilbert space $\mathcal{H}$ with

[4]The quantum no-cloning theorem prevents Alice from sending an unknown pure state to $N(\geqslant 2)$ receivers if Alice has no knowledge about this state (referred to as obliviousness) and has only one copy of the state $|\psi\rangle$. Relaxing the former leads to scenario 1 and relaxing the latter leads to scenario 2.

dimension $N+1$ instead of 2^N. Specifically,

$$|\psi\rangle^{\otimes N} = \sum_{j=0}^{N} \sqrt{\binom{N}{j}} \alpha^j \beta^{N-j} \, |j\rangle_{\mathrm{s}} \,, \tag{1.2}$$

where $|j\rangle_{\mathrm{s}}$ denotes the completely symmetric state of N qubits with j of them in state $|0\rangle$ and $(N-j)$ of them in state $|1\rangle$, i.e.,

$$|j\rangle_{\mathrm{s}} = \binom{N}{j}^{-1/2} \sum_{z_k \in \{0,1\}: \sum_{k=1}^{N} z_k = N-j} \otimes_{k=1}^{N} |z_k\rangle \,. \tag{1.3}$$

This gives us credence to develop a protocol that consumes $O(\log N)$ bcbits. In fact, the structure (1.2) has been widely used in processing a collection of identically prepared qubits [39,40]. Moreover, note that the set $\{(\alpha|0\rangle+\beta|1\rangle)^{\otimes N} : |\alpha|^2+|\beta|^2 = 1\}$ indeed spans the Hilbert space with dimension $N+1$, which suggests that a lower bound for the number of consumed bcbits is $\Omega(\log N)$.

Quantum Queuing Delay

Queuing delay of quantum data, referred to as QQD, is one of the critical issues in transmission of information across quantum networks. Compared to its classical counterpart, queuing delay is even more important for quantum networks: the quantum states interact with the environment and will lose a significant amount of information if not delivered on time [41–43]. The difficulties of analyzing QQD are two-fold. First, there is arguably no mathematical model that characterizes the queuing node and the queuing process in quantum networks. Such a model has to tally with the physical realizations (e.g., the quantum channels and quantum operations) and practical constraints (e.g., quantum memory size and quantum state lifetime). Second, quantum communication has its own peculiar properties that have no classical counterparts. For example, quantum communication may exploit quantum entanglement [44–47], a phenomenon representing non-local interconnection among quantum objects. These peculiar properties make the strategies designed for classical networks ill-suited for

quantum networks.

Existing work related to QQD can be divided into two groups: queuing delay in classical networks and operation design in quantum networks. There are myriad studies on queuing delay in classical networks [48–50]. Although some of the concepts such as queue length, Little's theorem, blocking probability, and stability [51–53] may be borrowed for studying quantum networks, the many methods tailored for the classical queuing theory do not apply directly to QQD. On the other hand, there are only a few studies on the operation designs for quantum networks, proposing ad-hoc protocols and verifying their performance via simulations. In [54], a decentralized entanglement routing protocol is proposed to find the shortest path in a quantum network using local knowledge of quantum nodes. In [55], optimized entanglement routing protocols are developed based on dynamic programming. In [56], link layer protocols are proposed for quantum networks and their performance is evaluated via simulations. These studies either maximize the quantum network throughput instead of queuing delay or provide heuristic protocols without performance guarantee.

The fundamental questions related to QQD are

- how to develop a tractable quantum queuing model that is consistent with the physical realizations and practical constraints; and
- how to characterize and exploit the properties specific to quantum nature for developing efficient policies that minimize QQD.

The answers to these questions will enable us to minimize QQD, which can further unleash the potential of quantum networks.

The goal of this part in the thesis are to build a mathematical model for characterizing QQD and to determine policies for controlling such a delay with performance guarantee. An essential part in this model is teleportation [23], a celebrated technique for sending quantum information from one node to another using entanglement and classical communication. The new technical idea in this part of the thesis is the introduction of dynamic programming in the modeling of QQD. Our view is that quantum nature can bring new phenomena in the area of queuing delay. We believe

that QQD can be significantly reduced by establishing entanglements and storing them in the entanglement memory before, instead of waiting until, quantum data arrive. Specifically, entanglements can be seen as resources that are reserved in the memory of a node, and quantum data in the queue can be delivered via teleportation using these resources instead of waiting for the transmission time.

Remote Entanglement Distribution

Many of quantum-enabled technologies rely on distributing quantum entanglement [57–59]. For example, distributing entanglement enables quantum teleportation [23, 60,61] and remote state preparation [30,31,62], sending quantum information without having to move physical particles.

The main difficulty of distributing entanglement at two distant nodes in quantum networks lies in the significant decay of communication capacity with the length of the channel. For example, the capacity of a lossy channel decays exponentially with the distance of optical fibers, thereby hindering distributing entanglement between two nodes that are far apart.

To address this issue, researchers introduced quantum repeaters in the design of quantum networks. A quantum repeater is a node with quantum memory and Bell-state measurement capability. With quantum repeaters, entanglement can be distributed between distant nodes without physically sending entangled qubits through the entire network [57–59]. The benefits of using quantum repeaters have been demonstrated in several studies [63, 64]. For example, inserting quantum repeaters between two nodes connected by optical fibers can improve the channel capacity; such capacity is determined by the maximum distance of the quantum channels divided by the quantum repeaters.

Several protocols have been proposed for entanglement distribution in quantum networks [54,63–68]. A key metric to evaluate these protocols is the entanglement distribution rate (EDR), i.e., the average amount of entanglement distributed between two specified nodes, referred to as source and sink, per time slot. Most protocols cannot provide the maximum EDR except for [63,64], in which optimal entanglement

routing protocols are proposed for basic decoherence channel models. However, the results in [63,64] rely on the assumption of perfect quantum repeaters, i.e., the entanglement swapping can be performed with success probability of one.[5] This assumption is unlikely to hold in the foreseeable future. In [68], the scenario with imperfect quantum repeaters is considered with the constraint that entanglement swapping is successfully and simultaneously performed at all the nodes along a path connecting the source and the sink, and this leads to suboptimal protocols since the order of entanglement swapping is not optimized. Little is known about the protocols that schedule entanglement swapping optimally with imperfect quantum repeaters.

Among different quantum networks, we are particularly interested in the homogeneous repeater chain, a network that connects two distant nodes with a chain of identical and equally spaced repeaters. This is because the investigation of homogeneous repeater chains can shed light on quantum networks with general structures and lay the foundation for the study of more complicated network structures. Since the introduction of homogeneous repeater chains in [69], many protocols and EDR analyses have been provided [19, 70–77]. These protocols and analyses can be categorized into two groups. The first group of work considers quantum memories that can store qubits only for a short time. Such consideration accounts for the limitation of current quantum-hardware capabilities [75–77]. For example in [75], entanglement-based quantum key distribution rate is determined, and different factors such as fiber loss, detector dark counts, and detector inefficiency are accounted for. However, if the quantum memory cannot store qubits for a sufficiently long time, entanglement swapping operations need to succeed simultaneously at each quantum repeater. This makes the EDR decrease exponentially with respect to the length of a repeater chain. The second group of work considers quantum memories that can store qubits for a sufficiently long time. Within this group, some studies aim at experimentation of entanglement distribution using atomic ensembles and linear optics [19,70–73]; some studies aim at the design and analysis of entanglement distribution protocols [55, 69, 78, 79].

[5] In the rest of this thesis, "perfect/imperfect repeaters" and "perfect/imperfect entanglement swapping" are used interchangeably when clear in the context.

Most existing protocols, e.g. in [55], assume that two neighboring quantum repeaters stop generating entanglement once an entangled qubit pair (EQP) is generated between them, and this assumption may decrease the EDR. Despite the extensive entanglement distribution protocols, it still remains unclear whether the designed protocols can achieve the maximum EDR, even in the scenario that considers simple photon-loss quantum channels and probabilistic entanglement swapping.

The consideration of imperfect quantum repeaters creates a new research topic: how to design protocols that schedule entanglement swapping to maximize the EDR in a quantum network. These protocols are referred to as RED protocols in this thesis. Note that RED protocols differ from entanglement routing protocols [54, 64, 65, 68]: entanglement routing protocols find paths between two nodes and perform entanglement swapping at each node sequentially along the path, whereas RED protocols not only find the paths, but also determine the sequence of entanglement swapping. In this way, entanglement routing can be viewed as a special case of RED. To the best of the authors' knowledge, how to develop the optimal RED protocol with imperfect quantum repeaters remains unknown.

The fundamental questions related to RED in quantum networks are

- how entanglement swapping affects the EDR in quantum networks; and
- how to exploit the structure of the homogeneous repeater chains to maximize the EDR.

The answers to these questions will enable the distribution of entanglement over long distances with noisy intermediate-scale quantum (NISQ) technologies [2] and take an essential step for the development of quantum networks.

The goal of this part of the thesis is to develop the optimal RED protocols for quantum networks. We introduce the concept of enodes, representing the entangled qubit pairs in the network. Entanglement swapping can be viewed as an operation that exchanges entangled qubit pairs among different enodes. The introduction of enodes allows us to employ techniques from linear programming and classical networks to design the RED protocols in quantum networks.

1.3 Contributions

In this thesis, we develop protocols/policies for state transmission and network operation for quantum networks and quantify their performances. The main contributions are summarized in four areas: quantum broadcasting, multiparty RSP, QQD, and RED.

Quantum Broadcasting

In the first part of the thesis, we develop quantum broadcasting protocols and establish the resource tradeoff among entanglement, bcbits, and bqubits. The main contributions are summarized as follows.

- We develop a protocol for sending 1 bqubit in the two scenarios: 1) non-oblivious with a known state and 2) oblivious with multiple copies. This protocol leads to the result that $f_1(N) = \log N$ and $f_2(N) = O(\log N)$;
- We prove that to send 1 bqubit to N receivers using shared entanglement, $O(\log N)$ bcbits are both necessary and sufficient;
- We show that the protocols can be implemented using poly(N) gates composed of single-qubit gates and CNOT gates, demonstrating their practicality in the near future. Furthermore, we designed a gate-level quantum circuit for a proposed protocol and evaluated its performance using the IBM quantum simulator [80].

Quantum Queuing Delay

In the second part of the thesis, we establish a mathematical formalism for characterizing QQD and propose policies to significantly reduce QQD. The main contributions are summarized as follows.

- we establish a tractable model for characterizing QQD using dynamic programming;

- we determine the average queuing delay for the scenario with one receiver and show that such delay can decrease exponentially with respect to the memory size of a node;
- we develop a cognitive-memory-based policy for minimizing the average queue length in a general scenario and show that it is optimal for the scenario with two receivers;
- we derive an upper bound for the average queue length corresponding to the proposed cognitive-memory-based policy. This bound implies that the average queuing delay can decrease exponentially with respect to the memory size of a node.

Remote Entanglement Distribution

In the third part of the thesis, we establish a framework of designing RED protocols for quantum networks. We transform the design of the optimal RED protocols into linear programming problems. The main contributions are summarized as follows.

- We determine the maximum achievable EDR for quantum networks;
- We determine the structural properties of the graph corresponding to the optimal solution of the linear programming problem;
- We develop the optimal RED protocols for quantum networks based on the solution of the linear programming problem;
- We determine the maximum EDR for homogeneous repeater chains in a closed form.

1.4 Organization and Notations

Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters,

respectively. For example, a random variable and its realization are denoted by x and x; a random vector and its realization are denoted by $\mathbf{x}$ and $\boldsymbol{x}$; a random matrix and its realization are denoted by $\mathbf{X}$ and $\boldsymbol{X}$, respectively. Sets and random sets are denoted by upright sans serif and calligraphic font, respectively. For example, a random set and its realization are denoted by X and $\mathcal{X}$, respectively. The m-by-n matrix of zeros (resp. ones) is denoted by $\mathbf{0}_{m\times n}$ (resp. $\mathbf{1}_{m\times n}$); when $n = 1$, the m-dimensional vector of zeros (resp. ones) is simply denoted by $\mathbf{0}_m$ (resp. $\mathbf{1}_m$). The m-by-m identity matrix is denoted by $\boldsymbol{I}_m$: the subscript is removed when the dimension of the matrix is clear from the context. The cardinality of a set $\mathcal{S}$ is denoted by $\mathrm{Card}(\mathcal{S})$. The set $\{m, m+1, \ldots, n\}$ is denoted by $\mathcal{K}_{m:n}$. The sets of integers, even integers, odd integers, and positive integers are denoted by $\mathbb{Z}$, $\mathbb{Z}_{\mathrm{e}}$, $\mathbb{Z}_{\mathrm{o}}$, $\mathbb{N}_+$, respectively. The cardinality of a set $\mathcal{S}$ is denoted by $\mathrm{Card}(\mathcal{S})$. The set $\{m, m+1, \ldots, n\}$ is denoted by $\mathcal{K}_{m:n}$. For an edge set $\mathcal{E}$, the function $1_{\mathcal{E}}(i,j)$ is an indicator function defined to be 1 if $(i,j) \in \mathcal{E}$ and 0 otherwise. Throughout this thesis, the state of a quantum system and its corresponding density matrix will be used interchangeably; $\boldsymbol{\Xi}(|\psi\rangle)$ denotes the density matrix corresponding to the pure state $|\psi\rangle$ and the part $(|\psi\rangle)$ is omitted if clear in the context; the notation $\boldsymbol{\Xi}_{x,y}$ denotes a maximally entangled qubit pair between two systems x and y, each corresponding to one qubit; the notation $\boldsymbol{\Xi}_{i:j}$ denotes a maximally entangled qubit pair between two systems, where one of the systems is in node i and the other is in node j. The Hermitian adjoint is denoted by †; the transpose and the complex conjugate with respect to basis $|0\rangle$ and $|1\rangle$ are denoted by $^{\mathrm{T}}$ and *, respectively; exp denotes the exponential function with the natural basis.

The rest of the thesis is organized as follows. In Chapter 2, we present the design of quantum broadcasting protocols and establish the resource tradeoff among entanglement, bcbits, and bqubits. In Chapter 3, we introduce a formalism to analyze QQD and develop policies for managing entanglement memory. In Chapter 4, we establish a framework of designing RED protocols for quantum networks. The thesis is concluded with remarks in Chapter 5.

Chapter 2

Quantum Broadcasting

In this chapter, we will explore the design of quantum broadcasting protocols. We will consider two scenarios: 1) non-oblivious with a known state and 2) oblivious with multiple copies. We will establish the resource tradeoff among entanglement, broadcast cbits (bcbits), and broadcast qubits (bqubits).

2.1 A Non-oblivious Quantum Broadcasting Protocol with a Known State

Consider Scenario 1 in which Alice knows the state $|\psi\rangle$. Since the state $|\psi\rangle^{\otimes N}$ is a linear combination of states $|j\rangle_{\mathrm{s}}$, a natural form for the shared entangled state is

$$|\Phi\rangle = \frac{1}{\sqrt{N+1}} \sum_{j=0}^{N} |j\rangle_{\mathrm{s}}^{A} |j\rangle_{\mathrm{s}}^{B}, \tag{2.1}$$

where $|j\rangle_{\mathrm{s}}^{A}$ and $|j\rangle_{\mathrm{s}}^{B}$ denote bases of Alice's system and the receivers' system, respectively. Each qubit in $|j\rangle_{\mathrm{s}}^{B}$ corresponds to each of the N different receivers. The goal is to send an arbitrary pure state $|\psi\rangle \in \mathcal{H}_2$ to each of N receivers, where $\mathcal{H}_2$ is a two-dimensional Hilbert space. The non-oblivious quantum broadcasting protocol with a known state is described as follows.

1) Alice performs the positive-operator valued measure (POVM) on her part of

the entangled state $|\Phi\rangle$. The POVM is given by

$$\boldsymbol{H}_k = \frac{N+1}{K(1+\epsilon)} \boldsymbol{V}_{\mathrm{s},k}\, \boldsymbol{\Xi}\left(|\psi\rangle^{*\otimes N}\right) \boldsymbol{V}_{\mathrm{s},k}^{\dagger}, \;\; k \in \mathcal{K}_{1:K}$$
$$\boldsymbol{H}_{\text{failure}} = \mathbb{1}_{\mathrm{s}} - \sum_{k=1}^{K} \boldsymbol{H}_k$$

where $\epsilon \in (0,1)$ is a pre-determined constant, $\boldsymbol{V}_{\mathrm{s},k} = \boldsymbol{V}_k^{\otimes N}$ with $\boldsymbol{V}_k$ being unitary operators on $\mathcal{H}_2$ such that

$$\sum_{k=1}^{K} \boldsymbol{H}_k \preccurlyeq \mathbb{1}_{\mathrm{s}} \tag{2.2}$$

in which $\mathbb{1}_{\mathrm{s}}$ denotes the identity operator on the Hilbert space $\mathcal{H}$ with basis $|j\rangle_{\mathrm{s}}$.

2) Alice broadcasts the classical message (representing the outcome of the POVM) to N different receivers. If outcome k is received, each receiver applies $\boldsymbol{V}_k^{\mathrm{T}}$ to their part of the state. If outcome "failure" is received, the protocol fails.

In the protocol above, the outcome k of the POVM performed by Alice indicates that the state at each receiver is

$$\Big(\langle\psi|^{*\otimes N}\, \boldsymbol{V}_{\mathrm{s},k}^{\dagger}\Big)^{A} \Big(\sum_{j=0}^{N} |j\rangle_{\mathrm{s}}^{A}\, |j\rangle_{\mathrm{s}}^{B}\Big) = \sum_{j=0}^{N} \Big(\big(\langle\psi|^{*}\, \boldsymbol{V}_k^{\dagger}\big)^{\otimes N}\, |j\rangle_{\mathrm{s}}\Big)^{A} |j\rangle_{\mathrm{s}}^{B} = \big(\boldsymbol{V}_k^{*}\, |\psi\rangle\,\big)^{\otimes N}.$$

If each receiver applies $\boldsymbol{V}_k^{\mathrm{T}}$ to its state $\boldsymbol{V}_k^{*}\,|\psi\rangle$, the pure state $|\psi\rangle$ is obtained at each receiver.

The specific structure of $\boldsymbol{V}_{\mathrm{s},k}$ as a tensor product of N matrices is inspired by the need to accommodate the non-locality imposed by spatially separated receivers. The validity of the protocol above requires the existence of $\boldsymbol{V}_{\mathrm{s},k}$'s that satisfy (2.2) for *every* pure state $|\psi\rangle$. The following theorem shows such existence for sufficiently large K. The proof is in Appendix A.1.

Theorem 2.1. For $\epsilon > 0$, there exist unitary operators $\boldsymbol{V}_{\mathrm{s},k} = (\boldsymbol{V}_k)^{\otimes N}$, $k \in \mathcal{K}_{1:K}$,

such that for every pure state $|\psi\rangle \in \mathcal{H}_2$,

$$\frac{N+1}{K(1+\epsilon)} \sum_{k=1}^{K} \boldsymbol{V}_{\mathrm{s},k}\, \boldsymbol{\Xi}\left(|\psi\rangle^{*\otimes N}\right) \boldsymbol{V}_{\mathrm{s},k}^{\dagger} \preccurlyeq \mathbb{1}_{\mathrm{s}} \tag{2.3}$$

where

$$K \leqslant \left\lceil \frac{32(N+1)^2}{\epsilon^2} \ln \frac{20(N+1)}{\epsilon} + \frac{64(N+1)}{\epsilon^2} \ln \frac{20N(N+1)}{\epsilon} \right) \right\rceil. \tag{2.4}$$

Note that the outcome "failure" occurs with probability

$$\begin{aligned} \mathrm{tr}\big\{ \boldsymbol{H}_{\mathrm{failure}} \otimes \mathbb{1}_{\mathrm{s}}\, \boldsymbol{\Xi}\left(|\Phi\rangle\right) \big\} &= 1 - \sum_{k=1}^{K} \mathrm{tr}\big\{ \boldsymbol{H}_k \otimes \mathbb{1}_{\mathrm{s}}\, \boldsymbol{\Xi}\left(|\Phi\rangle\right) \big\} \\ &= 1 - K \frac{N+1}{K(1+\epsilon)(N+1)} \leqslant \epsilon. \end{aligned}$$

Moreover, the POVM is constructed according to the state $|\psi\rangle$ and thus, this protocol is non-oblivious. Regarding the resources, this protocol requires $\log K = 2\log N + o(\log N)$ bcbits according to (2.4), thus reducing the number of consumed bcbits from $O(N)$ to $O(\log N)$ compared to sequential transmission of quantum states.

2.2 An Oblivious Quantum Broadcasting Protocol with Multiple Copies

Consider Scenario 2 in which Alice possesses N qubits $|\psi\rangle^{\otimes N}$ (denoted by $\tilde{A}$), but does not have knowledge of the state $|\psi\rangle$. As in the previous protocol, the shared entanglement between Alice and receivers is given by (2.1). The oblivious quantum broadcasting protocol with multiple copies is described as follows.

1) Alice performs the POVM on the system consisting of $\tilde{A}$ and her part of the

entangled state $|\Phi\rangle$. The POVM is given by

$$\boldsymbol{M}_i = \frac{(N+1)^2}{I(1+\epsilon)} \boldsymbol{\Xi}\left(|\xi_i\rangle\right), \quad i \in \mathcal{K}_{1:I}$$
$$\boldsymbol{M}_{\text{failure}} = \mathbb{1}_{\text{s}}^{\tilde{A}} \otimes \mathbb{1}_{\text{s}}^{A} - \sum_{i=1}^{I} \boldsymbol{M}_i$$

where $\epsilon \in (0,1)$ is a pre-determined constant,

$$|\xi_i\rangle = \frac{1}{\sqrt{N+1}} (\boldsymbol{U}_i)^{\otimes N} \otimes \mathbb{1}^{\otimes N} \sum_{j=0}^{N} |j\rangle_{\text{s}}^{\tilde{A}} |j\rangle_{\text{s}}^{A}$$

in which $\mathbb{1}$ denotes the identity operator on $\mathcal{H}_2$ and $\boldsymbol{U}_i$ denotes a unitary operator on $\mathcal{H}_2$ such that

$$\sum_{i=1}^{I} \boldsymbol{M}_i \preccurlyeq \mathbb{1}_{\text{s}} \otimes \mathbb{1}_{\text{s}}. \tag{2.5}$$

2) Alice broadcasts the classical message (representing the outcome of the POVM) to N different receivers. If outcome i is received, each receiver applies $\boldsymbol{U}_i$ to their part of the state. If outcome "failure" is received, the protocol fails.

In the protocol above, the outcome i of the POVM performed by Alice indicates that the state at receivers is

$$\Big(\sum_{j=0}^{N} \big({}_{\text{s}}\langle j| \, \boldsymbol{U}_i^{\dagger \otimes N} \big)^{\widetilde{A}} \, {}_{\text{s}}\langle j|^{A} \Big) (|\psi\rangle^{\otimes N})^{\widetilde{A}} \Big(\sum_{j=0}^{N} |j\rangle_{\text{s}}^{A} \, |j\rangle_{\text{s}}^{B} \Big)$$
$$= \sum_{j=0}^{N} \Big({}_{\text{s}}\langle j| \, \big((\boldsymbol{U}_i)^{\dagger} |\psi\rangle \big)^{\otimes N} \Big)^{\widetilde{A}} |j\rangle_{\text{s}}^{B} = \big(\boldsymbol{U}_i^{\dagger} |\psi\rangle \big)^{\otimes N}.$$

If each receiver applies $\boldsymbol{U}_i$ to its state $\boldsymbol{U}_i^{\dagger} |\psi\rangle$, the pure state $|\psi\rangle$ is obtained at each receiver.

The validity of the protocol above requires the existence of $|\xi_i\rangle$'s that satisfies (2.5). The following theorem shows such existence for sufficiently large I. The proof is in Appendix A.2.

Theorem 2.2. For $\epsilon > 0$, there exist unit vectors $|\xi_i\rangle = \frac{1}{\sqrt{N+1}}(\boldsymbol{U}_i)^{\otimes N} \otimes \mathbb{1}^{\otimes N} \sum_{j=0}^{N} |j\rangle_{\mathrm{s}}^{\tilde{A}} |j\rangle_{\mathrm{s}}^{A}$, $i \in \mathcal{K}_{1:I}$ such that

$$\frac{(N+1)^2}{I} \sum_{i=1}^{I} \boldsymbol{\Xi}\left(|\xi_i\rangle\right) \preccurlyeq (1+\epsilon)\mathbb{1}_{\mathrm{s}} \otimes \mathbb{1}_{\mathrm{s}} \tag{2.6}$$

where

$$I \leqslant \left\lceil \frac{32(N+1)^4}{\epsilon^2} \ln\left(\frac{10(N+1)^2}{\epsilon}\right) \right\rceil. \tag{2.7}$$

Analysis similar to that in the previous section shows that the outcome "failure" occurs with probability no greater than ϵ. Moreover, the POVM does not rely on the state $|\psi\rangle$ and thus, this protocol is oblivious. Regarding the resources, this protocol requires $\log I = 4\log N + o(\log N)$ bcbits according to (2.7), thus reducing the number of consumed bcbits from $O(N)$ to $O(\log N)$ compared to sequential transmission of quantum states.

2.3 Lower Bound for the bcbits

We next show that both in Scenario 1 and Scenario 2, the number of consumed bcbits for sending an arbitrary pure state $|\psi\rangle$ to each of the N different receivers is lower bounded by $\Omega(\log N)$.

Theorem 2.3. For any protocol that can send N copies of an arbitrary pure state $|\psi\rangle \in \mathcal{H}_2$ exactly with success probability no less than $1-\epsilon$ to a collection of receivers (the receiver is notified whether the protocol is successful), it is necessary for the transmitter to send at least $(1-\epsilon)\log(N+1)$ bcbits.

Proof. We prove Theorem 2.3 by contradiction. We consider a scenarios where all the receivers are colocated. Suppose there exists a Protocol H so that the transmitter can send N copies of an arbitrary pure state $|\psi\rangle$ exactly with a success probability no less than $1-\epsilon$ using only $(1-\epsilon)\log(N+1) - \Delta$ cbits for some $\Delta > 0$. Then for an arbitrary pure state $|\psi\rangle$ at the transmitter (either in Scenario 1 or Scenario 2),

Protocol H can prepare a state with the following density matrix at receivers:

$$\boldsymbol{W}(|\psi\rangle) := (1-\epsilon)\,|1\rangle\langle 1| \otimes \boldsymbol{\Xi}\left(|\psi\rangle^{\otimes N}\right) + \epsilon\,|0\rangle\langle 0| \otimes \left(\frac{1}{N+1}\mathbb{1}_{\mathrm{s}}\right)$$

where $|1\rangle\langle 1|$ and $|0\rangle\langle 0|$ indicate whether the protocol is successful or not. Note that if for certain states, the success probability is strictly less than ϵ, we can enforce the receivers to prepare the state $\mathbb{1}_{\mathrm{s}}/(N+1)$ with an appropriate probability even in the success case so that $\boldsymbol{W}(|\psi\rangle)$ can be prepared.

We choose pure states $|\psi_m\rangle$, $m \in \mathcal{K}_{1:M}$, such that

$$\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\Xi}\left(|\psi_m\rangle^{\otimes N}\right) \in \left[\frac{1-\epsilon_1}{N+1}\mathbb{1}_{\mathrm{s}}, \frac{1+\epsilon_1}{N+1}\mathbb{1}_{\mathrm{s}}\right] \tag{2.8}$$

where $\epsilon_1 \in (0,1)$ is a given constant and the closed interval refers to the operator order. This is valid for sufficiently large M because

$$\int_{|\psi\rangle\in\mathcal{H}_2} \boldsymbol{\Xi}\left(|\psi\rangle^{\otimes N}\right) d\,|\psi\rangle = \frac{1}{N+1}\mathbb{1}_{\mathrm{s}}$$

as shown by Lemma A.4 in the appendix.

We consider a transmitter Charlie and a receiver David, and Charlie would like to send classical information to David using Protocol H. Consider letter states with density matrices

$$\boldsymbol{W}_m = \boldsymbol{W}(|\psi_m\rangle), \quad m \in \mathcal{K}_{1:M}$$

with *a priori* probabilities $p_m = 1/M$. For fixed $\epsilon_2, \delta > 0$, there exist a code (whose codewords are strings of L letter states) and a decoder with a probability of error no greater than ϵ_2 such that the classical information carried per letter state is at least $\chi_W - \delta$ [81], where χ_W denotes the Holevo information:

$$\chi_W := H\Big(\sum_{m=1}^{M}\frac{1}{M}\boldsymbol{W}_m\Big) - \sum_{m=1}^{M}\frac{1}{M}H(\boldsymbol{W}_m)$$

in which $H(\cdot)$ denotes the von Neumann entropy. For a codeword $(c_1 c_2 \ldots c_L)$, Charlie uses Protocol H to send letter states with density matrices $\boldsymbol{W}_{c_l}$, $l \in \mathcal{K}_{1:L}$, to David. Due to the causality, the information obtained by David is no greater than that used in Protocol H, i.e.,

$$(1-\epsilon_2)L \cdot (\chi_W - \delta) \leqslant L \cdot \big((1-\epsilon)\log(N+1) - \Delta\big)$$

implying that

$$\chi_W \leqslant \frac{(1-\epsilon)\log(N+1) - \Delta}{1-\epsilon_2} + \delta \tag{2.9}$$

for arbitrary small ϵ_1, ϵ_2 and δ.

Note that

$$\begin{aligned}
\chi_W &= H\Big(\sum_{m=1}^{M} \frac{1}{M} \boldsymbol{W}_m\Big) - h(\epsilon) - \epsilon \log\left(N+1\right) \\
&= (1-\epsilon) H\left(\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\Xi}\left(|\psi_m\rangle^{\otimes N}\right)\right) \\
&\geqslant -(1-\epsilon)\log\left(\frac{1+\epsilon_1}{N+1}\right)
\end{aligned} \tag{2.10}$$

where $h(x) = -x\log x - (1-x)\log(1-x)$, and the inequality can be obtained by bounding the eigenvalues of $\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{\Xi}\left(|\psi_m\rangle^{\otimes N}\right)$ based on (2.8). Combining (2.9) and (2.10), we have

$$\begin{aligned}
&(1-\epsilon)\log(N+1) - (1-\epsilon)\log(1+\epsilon_1) \\
&\qquad \leqslant \frac{(1-\epsilon)\log(N+1) - \Delta}{1-\epsilon_2} + \delta.
\end{aligned}$$

Letting $\epsilon_1 \to 0$, $\epsilon_2 \to 0$, and $\delta \to 0$, the inequality above becomes $\Delta \leqslant 0$, which provides the desired contradiction. □

Theorem 2.3 implies that it is impossible to send an arbitrary pure state $|\psi\rangle$ to each of the N receivers by using fewer than $(1-\epsilon)\log\left(N+1\right)$ bcbits.

2.4 Efficient Implementation of Protocols

The proposed two protocols have success probability of at least $1-\epsilon$ for an arbitrary pure state $|\psi\rangle \in \mathcal{H}_2$. It is possible to increase the success probability by optimizing the POVM performed by Alice.[1] We next describe an optimization method for the non-oblivious protocol through an example; a similar method can be adopted for the oblivious one.

1) For a given pure state $|\psi\rangle$, Alice obtains the optimal solution $\mathring{\lambda}_k$ of the following semidefinite programming (SDP) problem

$$\begin{aligned} \mathscr{P}: \ \underset{\lambda_k, k\in\mathcal{K}_{1:K}}{\text{maximize}} \ & \sum_{k=1}^{K} \lambda_k \\ \text{subject to } & \frac{N+1}{K(1+\epsilon)} \sum_{k=1}^{K} \lambda_k \boldsymbol{V}_{\mathrm{s},k} \left(\boldsymbol{\Xi}\left(|\psi\rangle\right)^*\right)^{\otimes N} \boldsymbol{V}_{\mathrm{s},k}^\dagger \preccurlyeq \mathbb{1}_{\mathrm{s}} \\ & \lambda_k \geqslant 0, \ k \in \mathcal{K}_{1:K}. \end{aligned}$$

2) Alice performs the POVM on her part of the entangled state $|\Phi\rangle$. The POVM is given by

$$\begin{aligned} \mathring{\boldsymbol{H}}_k &= \mathring{\lambda}_k \frac{N+1}{K(1+\epsilon)} \boldsymbol{V}_{\mathrm{s},k} \left(\boldsymbol{\Xi}\left(|\psi\rangle\right)^*\right)^{\otimes N} \left(\boldsymbol{V}_{\mathrm{s},k}\right)^\dagger, \quad k \in \mathcal{K}_{1:K} \\ \mathring{\boldsymbol{H}}_{\text{failure}} &= \mathbb{1}_{\mathrm{s}} - \sum_{k=1}^{K} \mathring{\boldsymbol{H}}_k. \end{aligned}$$

3) Alice broadcasts the classical message (representing the outcome of the POVM) to N different receivers. If outcome k is received, each receiver applies $\boldsymbol{V}_k^{\mathrm{T}}$ to their part of the state. If outcome "failure" is received, the protocol fails.

The solutions of $\mathscr{P}$ do not affect the receivers' operations, thus the transmitter is not required to consume additional classical resources for sending these solutions to the receivers. The success probability achieved by this protocol is given by

[1] Using optimization techniques for improving performance has been widely used in quantum communication [82].

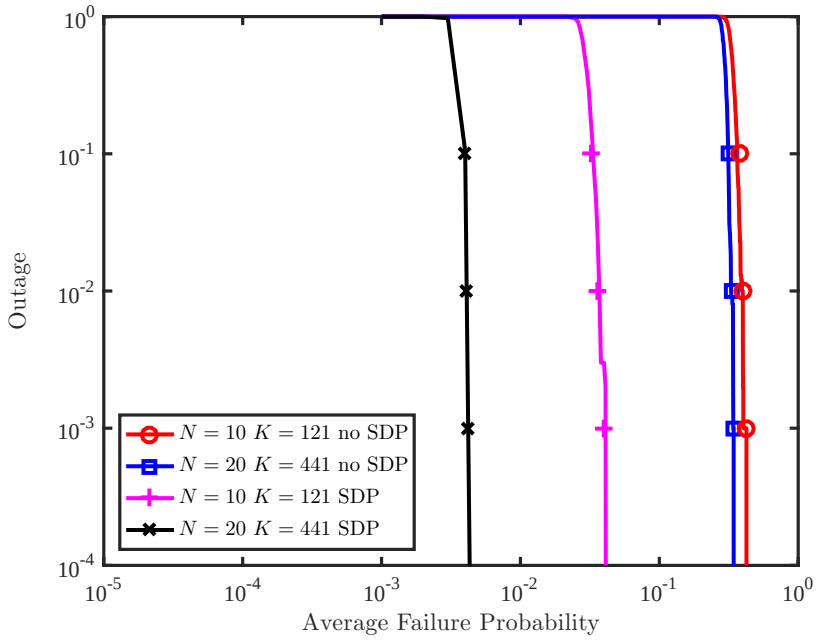


Figure 2-1: The CDF of the failure probability with and without SDP. K is set to be $(N+1)^2$.

$\sum_{k=1}^{K} \mathring{\lambda}_k / \big(K(1+\epsilon)\big)$, which is at least $1/(1+\epsilon)$ as $\lambda_k = 1$ is a feasible solution for $\mathscr{P}$. The optimization over λ_k can significantly improve the success probability in practice. To study such improvement, we consider the cases: $N = 10$ and $N = 20$. In particular, $K = (N+1)^2$ unitary operators are drawn from the Haar measure. For an instantiation of K unitary operators, the average failure probability of a protocol is obtained by averaging over pure states from $\mathcal{H}_2$ uniformly. The outage is defined as the empirical probability that the average failure probability is greater or equal than the abscissa. For $N = 20$, in 99% of cases, the protocol without SDP optimization has a failure probability less than or equal to 0.276, whereas the protocol with SDP optimization has a failure probability of 0.004. This corresponds to an reduction by a factor of around 70.

Note that the problem $\mathscr{P}$ has K variables, where the semidefinite constraints involve a matrix of size $N+1$ by $N+1$. Therefore, solving $\mathscr{P}$ requires computational complexity of only poly(N) [83]. Moreover, sparsity analysis similar to [84] shows that there exists an optimal solution of $\mathscr{P}$ with at most $(N+1)(N+2)/2$ non-zero

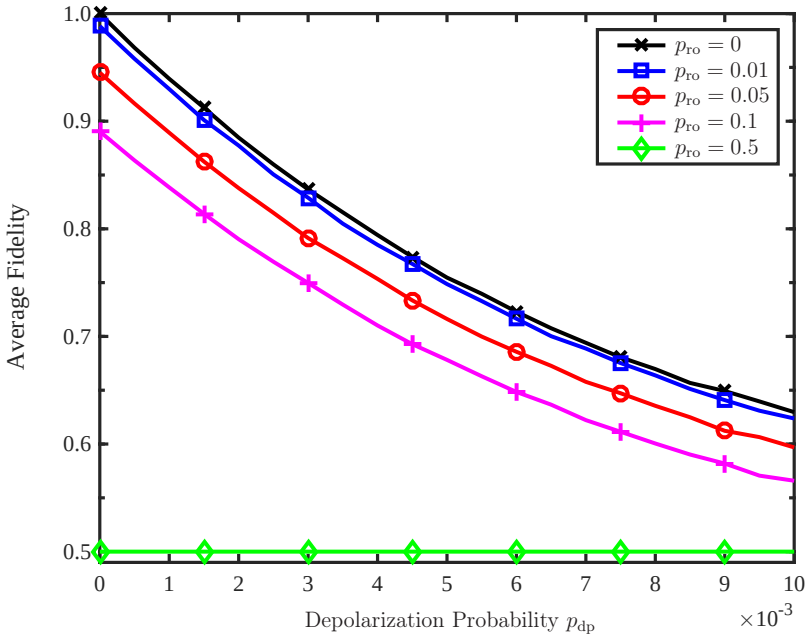


Figure 2-2: The state fidelity under different noise levels.

elements. This observation significantly decreases the complexity of circuits when implementing the POVM $\mathring{\boldsymbol{H}}_k$.

We next consider the implementation of the developed protocols. Note that in the optimized non-oblivious protocol, the success probability and the output will remain the same if the entanglement (2.1) is replaced by

$$|\Phi\rangle_{\rm b} = \frac{1}{\sqrt{N+1}} \sum_{j=0}^{N} |j\rangle_{\rm b}^{A} |j\rangle_{\rm s}^{B}$$

where $|j\rangle_{\rm b}$ is the binary representation of j with $\lceil \log N + 1 \rceil$ quantum bits (qubits) (e.g., for $N = 6$, $j = 5$, $|j\rangle_{\rm b} = |101\rangle$) and the POVMs $\mathring{\boldsymbol{H}}_k$ and $\mathring{\boldsymbol{H}}_{\rm failure}$ (with bases $|j\rangle_{\rm s}$) are replaced by their isometries $\mathring{\boldsymbol{H}}_{{\rm b},k}$ and $\mathring{\boldsymbol{H}}_{\rm b,failure}$ (with bases $|j\rangle_{\rm b}$), respectively. Similarly, in the optimized oblivious protocol, the success probability and the output will be the same if the entanglement (2.1) is replaced by $|\Phi\rangle_{\rm b}$, $\mathring{\boldsymbol{M}}_i$ and $\mathring{\boldsymbol{M}}_{\rm failure}$ (with bases $|j\rangle_{\rm s}|j\rangle_{\rm s}$) are replaced with by their isometries $\mathring{\boldsymbol{M}}_{{\rm b},i}$ and $\mathring{\boldsymbol{M}}_{\rm b,failure}$ (with bases $|j\rangle_{\rm b}|j\rangle_{\rm b}$), and the given state $|\psi\rangle^{\otimes N}$ (with bases $|j\rangle_{\rm s}$) is replaced by its isometry $|\psi\rangle_{\rm b}$ (with bases $|j\rangle_{\rm b}$).

The key steps in the implementation of the proposed protocols are (a) building the entangled state $|\Phi\rangle_{\rm b}$; (b) implementing the POVM $\mathring{\boldsymbol{H}}_{{\rm b},k}$, $\mathring{\boldsymbol{H}}_{\rm b,failure}$, $\mathring{\boldsymbol{M}}_{{\rm b},k}$, and $\mathring{\boldsymbol{M}}_{\rm b,failure}$; and (c) transforming the state $|\psi\rangle^{\otimes N}$ to the state $|\psi\rangle_{\rm b}$. The next theorem shows that these three tasks can be efficiently performed with basic gates.[2]

Theorem 2.4. (a) The entangled state $|\Phi\rangle_{\rm b}$ can be built with poly(N) basic gates; (b) the POVM $\mathring{\boldsymbol{H}}_{{\rm b},k}$, $\mathring{\boldsymbol{H}}_{\rm b,failure}$, $\mathring{\boldsymbol{M}}_{{\rm b},k}$, and $\mathring{\boldsymbol{M}}_{\rm b,failure}$ can be implemented with poly(N) basic gates and standard basis measurements; and (c) the state $|\psi\rangle^{\otimes N}$ can be transformed to the state $|\psi\rangle_{\rm b}$ with poly(N) basic gates.

Proof. See Appendix A.4. □

Based on Theorem 2.4, we have implemented quantum broadcasting protocol for both scenarios (non-oblivious with a known state and oblivious with multiple copies) using basic gates on the IBM quantum simulator [80]. To investigate the effects brought by practical imperfect operations, we follow the error model in [85]: the CNOT gates depolarizes with probability $p_{\rm dp}$, i.e., $\boldsymbol{\Xi} \to (1-p_{\rm dp})\boldsymbol{U}_{\rm CNOT}\boldsymbol{\Xi}\boldsymbol{U}_{\rm CNOT}^{\dagger} + \frac{p_{\rm dp}}{4}\boldsymbol{I}$, where $\boldsymbol{\Xi}$ is a density matrix, $\boldsymbol{U}_{\rm CNOT}$ represents the CNOT operator, and $\frac{1}{4}\boldsymbol{I}$ is the matrix representing the maximally mixed two-qubit state. We also assume the measurement readout error is $p_{\rm ro}$, i.e., the binary measurement result flips with probability $p_{\rm ro}$. As a representative case, we consider $N=3$ and $K=5$. Fig. 2-2 shows the average fidelity $\mathbb{E}\{\langle\psi|\,\varphi\,|\psi\rangle\}$ as a function of $p_{\rm dp}$, where φ denotes the state obtained at a single receiver when the transmission succeeds. When $p_{\rm ro}=0.5$, the resulting state is a maximally mixed state, hence the state fidelity is 0.5, regardless of the depolarization probability as expected. In practice, $p_{\rm ro}$ can be as low as 0.01 [86], at which case the fidelity of the received state is affected only slightly by the readout error. For a small gate error, the protocol can achieve a desirable performance (e.g., when $p_{\rm dp}\approx 0.001, p_{\rm ro}\approx 0.01$, the average fidelity achieves around 0.93). The current technology reaches two-qubit gate fidelity around 0.994 [87], which would correspond to $p_{\rm dp}=0.008$. As the technology is rapidly improving, the quantum broadcasting protocol becomes close to a reality.

[2]We refer to single-qubit gates and CNOTs as basic gates.

Chapter 3

Quantum Queuing Delay

In this chapter, we will introduce a formalism to analyze quantum queuing delay (QQD) and develop policies for managing entanglement memory.

3.1 System Model

This section presents the system model and introduces the problem of minimizing queueing delay in quantum networks.

3.1.1 Quantum Node

We consider a quantum system composed of a quantum node and a collection of receivers as illustrated in Fig. 3-1. A quantum node is composed of an entanglement-generating platform, an entanglement memory, and N_{r} quantum data queues, where N_{r} is the number of receivers. Quantum data are quantum states (e.g., the spin of an electron and the polarization of a photon), and they arrive at the quantum node according to a stochastic process.[1] Each quantum datum is associated with a destined receiver and is quantified by a quantum bit (qubit). Next we describe each component

[1]These quantum data are either generated by the quantum node itself or are sent to the quantum node by a collection of transmitters, which are not illustrated in Fig. 3-1 explicitly. In the latter case, the quantum data are sent to the node from transmitters via either direct transmission or teleportation; then the quantum node needs to move the data to the corresponding data queues.

in the quantum node as well as the protocol for transmitting quantum data to the corresponding receiver.

Quantum Data Queue: Each quantum data queue is associated with a receiver. Quantum data are stored in the quantum data queue after arrival. These data are assumed to be stored perfectly for infinite time in the quantum data queue.

Entanglement-generating Platform: In practice, there are multiple ways of generating entanglement, i.e., entangled qubit pairs, between two nodes [88–90]. We focus on abstract models and leave more detailed physics realizations for later. We consider a platform for making attempts to generate entangled qubit pairs between the node and a receiver. Such attempts may fail due to imperfection in practical operations. Note that the platform does not have the storing capability and qubits are assumed to be collapsed in the next time slot if stored at the platform.

Entanglement Memory: If the attempt of entanglement generation succeeds, an entangled qubit pair will be shared between the quantum node and a receiver, with one qubit at the quantum node and the other at the receiver. The quantum node will use this entangled qubit pair for teleportation immediately or move half of the entangled qubit pair (i.e., the qubit at the quantum node) to the entanglement memory for future use [91]. Such half of the entangled qubit pair will be referred to as entangled qubit in the rest of the thesis. The entangled qubit pairs are assumed to be stored perfectly for infinite time in the entanglement memory.

Teleportation Protocol: The quantum node employs the teleportation protocol to transmit quantum data to a receiver [23].[2] In teleportation, the quantum node performs Bell measurements on the qubit to be transmitted, denoted as $|\psi\rangle$, as well as its half of the entangled qubit pair; the measurement result is then sent to the receiver, which will perform operations on its half of the entangled qubit pair accordingly; the resulting qubit at the receiver then becomes $|\psi\rangle$.

The teleportation protocol permits the quantum node to transmit 1 qubit at

[2]The reason for employing teleportation rather direct transmission using quantum channels are two folds. First, it is less challenging to establish entanglement and perform teleportation using near-future quantum technologies compared to direct transmission, which requires entanglement involving many qubits for quantum error correction coding. Second, as will be shown in this thesis, using teleportation results in much less QQD compared to direct transmission.

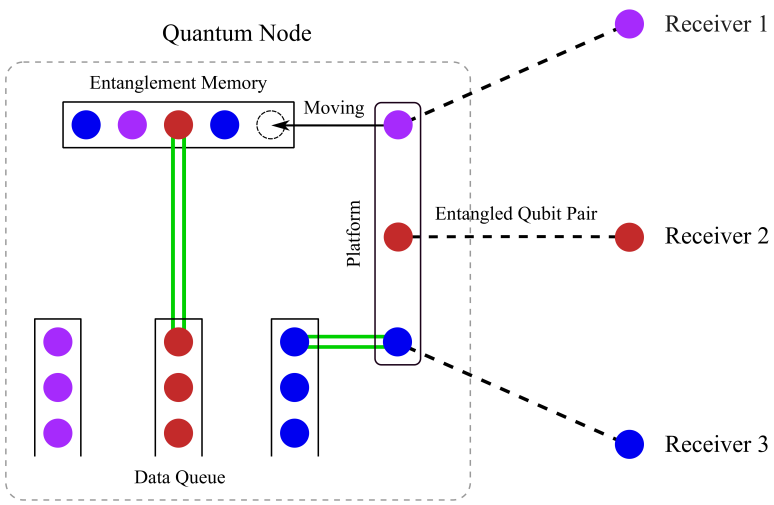


Figure 3-1: An illustration of the quantum node and receivers. Quantum nodes are composed of quantum data queues, platforms, and entanglement memories. Purple, red, and blue circles represent qubits corresponding to Receiver 1, 2, 3, respectively. Hollow circles represent vacant spot in the entanglement memory. Double green lines represent teleportation. Dashed lines represent entangled qubit pairs. Black arrows represent the operation of moving qubits from the platform to the entanglement memory.

the cost of sending 2 classical bits (cbits) and consuming 1 entangled qubit pair. In this thesis, we consider that classical communication resources are free since in the foreseeable future, the communication capability of classical information will be significantly larger than that of quantum information. Therefore, the delay brought by the classical communication is negligible compared to QQD, and is omitted in this thesis. The bottleneck of quantum information transmission then lies in the entanglement generation rate and in the storage capability.

3.1.2 Dynamic Programming Formalism

We formulate the quantum data transmission at a quantum node as a dynamic programming problem. Consider a dynamic system with discrete time slots t_n ($n = 0, 1, 2, \dots$). Let $s_i^{(n)}$ denote the number of qubits associated with Receiver i in the system at the beginning of time slot t_n. In particular, $s_i^{(n)} \geqslant 0$ implies there are $s_i^{(n)}$ qubits in the ith data queue; $s_i^{(n)} < 0$ implies there are $-s_i^{(n)}$ entangled qubits

associated with Receiver i in the memory.

Each time slot is divided into the following three phases.

1. Data arrival and entanglement generation: $a_i^{(n)}$ qubits of quantum data associated with Receiver i arrive at the node; meanwhile, the node makes an attempt to generate entanglement in the platform with every receiver; $b_i^{(n)}$ entangled qubit pairs associated with Receiver i are successfully generated.

2. Teleportation: the node adopts teleportation to transmit quantum data by consuming entangled qubit pairs. The entangled qubits at the node are in the platform and the entanglement memory. The entangled qubits in the platform are given priority for serving as resources in teleportation.

3. Entanglement storage: the remaining entangled qubits in the platform are moved to the entanglement memory.

Note that in the third phase, if the entanglement memory cannot accommodate all the entangled qubits in the platform, the node has to discard some entangled qubits and move the rest to the memory. The policy for discarding entangled qubits in such a scenario is referred to as *entanglement memory management*. Moreover, we consider that there is a physical limit Q_{t} for the number of quantum data corresponding to a receiver, meaning that after the teleportation phase, if the number of quantum data in the queue is greater than Q_{t}, the node has to discard the newly arrived quantum data. This gives $s_i^{(n)} \leqslant Q_{\mathrm{t}}$, $\forall i, n$.[3]

The number of entangled qubit pairs that are successfully generated, i.e., $b_i^{(n)}$, are known to the node before the control decision $u^{(n)}$ is made. If this assumption does not hold (e.g., if the node and the receiver are far apart), one can create a slightly different model, but the insights obtained in this thesis would still be valid.

With the background introduced above, the mathematical model for quantum

[3]If the discarded quantum data are required to be recovered, one can either encode the quantum data with error correction coding techniques, or require the retransmission of the quantum data. In the latter case, the transmitters need to have knowledge about the quantum data so that the no-cloning theorem is not violated.

data transmission at a quantum node has the form

$$\boldsymbol{x}^{(n+1)} = f(\boldsymbol{x}^{(n)}, u^{(n)}, \boldsymbol{w}^{(n)}), \quad n = 0, 1, 2, \ldots \tag{3.1}$$

where

- $\boldsymbol{x}^{(n)}$: the state of the system, consisting of the number of qubits at the beginning of time slot t_n, as well as the difference between the number of quantum data arriving at the quantum node and the number of entangled qubit pairs successfully generated during the first phase of the time slot, i.e.,

  $$\boldsymbol{x}^{(n)} = \left[\left(\boldsymbol{s}^{(n)}\right)^{\mathrm{T}} \left(\boldsymbol{c}^{(n)}\right)^{\mathrm{T}} \right]^{\mathrm{T}}$$

  in which

  $$\boldsymbol{s}^{(n)} = \left[s_1^{(n)} \; s_2^{(n)} \; \ldots \; s_{N_\mathrm{r}}^{(n)} \right]^{\mathrm{T}}$$
  $$\boldsymbol{c}^{(n)} = \left[c_1^{(n)} \; c_2^{(n)} \; \ldots \; c_{N_\mathrm{r}}^{(n)} \right]^{\mathrm{T}}$$

  and $c_i^{(n)} = a_i^{(n)} - b_i^{(n)}$.[4]

- $u^{(n)}$: the control policy at time slot n, i.e., the policy for entanglement memory management. In particular, $u^{(n)}$ is a function that maps $\boldsymbol{x}^{(n)}$ into $\boldsymbol{s}^{(n+1)}$ with the constraint that

  $$\sum_{i=1}^{N_\mathrm{r}} \max\{0, -s_i^{(n+1)}\} \leqslant M$$

  where M denotes the capacity of the entanglement memory. Moreover, due to the upper bound for the number of quantum data, we require that

  $$s_i^{(n+1)} = Q_\mathrm{t} \quad \text{if } s_i^{(n)} + c_i^{(n)} > Q_\mathrm{t}. \tag{3.2}$$

[4]Note that due to the teleportation in the second phase, the difference between $a_i^{(n)}$ and $b_i^{(n)}$ is sufficient for entanglement memory management.

- $\boldsymbol{w}^{(n)}$: the instantiation of the random vector that represents all the uncertainty in the system at time slot t_{n+1}, i.e., $\boldsymbol{w}^{(n)} = \boldsymbol{c}^{(n+1)}$. The distribution of $\mathbf{w}^{(n)}$ is assumed known *a priori*. In particular, consider that $\mathsf{a}_i^{(n)}$ are independent and identically distributed (i.i.d.) Bernoulli random variables with parameter $\mathbb{P}\{\mathsf{a}_i^{(n)} = 1\} = p_i$ (independent over i and n) and that $\mathsf{b}_i^{(n)}$ are also i.i.d. Bernoulli random variables with parameter $\mathbb{P}\{\mathsf{b}_i^{(n)} = 1\} = q_i$.[5] The distribution of $\mathsf{c}_i^{(n)}$ is found to be

$$\mathbb{P}\{\mathsf{c}_i^{(n)} = k\} = \begin{cases} (1-p_i)q_i, & \text{if } k = -1 \\ p_i q_i + (1-p_i)(1-q_i) & \text{if } k = 0 \\ p_i(1-q_i) & \text{if } k = +1 \\ 0 & \text{otherwise.} \end{cases}$$

  In the sequel, we assume that $0 < p_i, q_i < 1$, $i \in \mathcal{K}_{1:N_\mathrm{r}}$.

- f: a function that describes the system. Note that $u^{(n)}(\boldsymbol{x}^{(n)}) = \boldsymbol{s}^{(n+1)}$ and $\boldsymbol{w}^{(n)} = \boldsymbol{c}^{(n+1)}$; therefore, f can be expressed as

$$f(\boldsymbol{x}^{(n)}, u^{(n)}, \boldsymbol{w}^{(n)}) = \left[\left(u^{(n)}(\boldsymbol{x}^{(n)})\right)^{\mathrm{T}} \left(\boldsymbol{w}^{(n)}\right)^{\mathrm{T}} \right]^{\mathrm{T}}.$$

This model is illustrated in Fig. 3-2. When $n = 0$, $s_i^{(0)} = 0$, $i \in \mathcal{K}_{1:N_\mathrm{r}}$, $a_1^{(0)} = 0$, $a_2^{(0)} = 1$, $a_3^{(0)} = 1$, $b_1^{(0)} = 1$, $b_2^{(0)} = 0$, $b_3^{(0)} = 1$. Consequently, $c_1^{(0)} = -1$, $c_2^{(0)} = 1$, $c_3^{(0)} = 0$. The entangled qubit corresponding to Receiver 1 is moved to the memory, which gives $s_1^{(1)} = -1$. When $n = 1$, $s_1^{(1)} = -1$, $s_2^{(1)} = 1$, $s_3^{(1)} = 0$, $a_1^{(1)} = 0$, $a_2^{(1)} = 1$, $a_3^{(1)} = 0$, $b_1^{(1)} = 1$, $b_2^{(1)} = 0$, $b_3^{(1)} = 1$. Consequently, $c_1^{(1)} = -1$, $c_2^{(1)} = 1$, $c_3^{(1)} = -1$. Note that in this case, there are three entangled qubits in the memory and the platform after the teleportation phase, and this is beyond the capacity of the memory. The control policy in this case chooses to discard the entangled qubit in the platform

[5] As no large-scale quantum networks have been implemented, it is not clear how to model the distribution of the quantum data arrival. In this thesis, motivated by classical networks, we consider that $\mathsf{a}_i^{(n)}$ follows the Bernoulli distribution. The justification for $\mathsf{b}_i^{(n)}$ following a Bernoulli distribution will be shown in the next section.

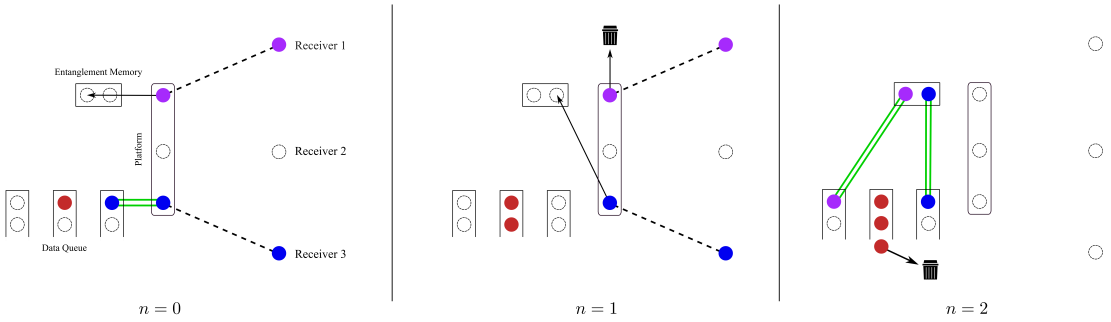


Figure 3-2: An illustration of dynamic programming formalism for $N_{\mathrm{r}} = 3$, $M = 2$, $Q_{\mathrm{t}} = 2$. Purple, red, and blue circles represent qubits corresponding to Receiver 1, 2, and 3, respectively. Hallow circles represent vacant spot in the receivers, platforms, data queues, and the entanglement memory. Double green lines represent teleportation. Dashed lines represent entangled qubit pairs. Black arrows represent the operation of moving qubits from the platform to the entanglement memory and the operation of discarding qubits.

corresponding to Receiver 1, which gives $s_1^{(2)} = -1$ and $s_3^{(2)} = -1$. When $n = 2$, $s_1^{(2)} = -1$, $s_2^{(2)} = 2$, $s_3^{(2)} = -1$, $a_i^{(2)} = 1$, $b_i^{(2)} = 0$, $i \in \mathcal{K}_{1:N_{\mathrm{r}}}$. Consequently, $c_i^{(2)} = 1$, $i \in \mathcal{K}_{1:N_{\mathrm{r}}}$. After the teleportation phase, there are three qubits in the second data queue, which is beyond the capacity of the data queue. The node has to discard one qubit, which gives $s_2^{(3)} = 2$.

The goal is to minimize the average delay of the quantum queuing system. Due to the relationship between average delay and average queue length shown by Little's law [92], we aim at designing the control policy $u^{(n)}$ for minimizing the expected average queue length.[6] In particular, let $J_{\mathrm{L}}(\boldsymbol{x}^{(1)})$ denote the expected average queue length starting in the state $\boldsymbol{x}^{(1)}$, i.e.,

$$J_{\mathrm{L}}(\boldsymbol{x}^{(1)}) := \limsup_{N\to\infty} \mathbb{E}\left\{ \frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{N_{\mathrm{r}}} \max\left\{0, s_i^{(n)}\right\} \middle| \boldsymbol{x}^{(1)} \right\}. \tag{3.3}$$

[6]Little's law shows that the average queue length is the multiplication of the average queuing delay and the effective arrival rate of quantum data. Later we will see that the blocking probability is almost zero for the developed policies and hence the effective arrival rate is almost a constant.

3.1.3 Practical Implementation

We now consider one of the practical methods of entanglement generation and justify the Bernoulli distribution assumption of $\mathsf{b}_i^{(n)}$ in the previous section. As mentioned in [56, 93], Nitrogen-Vacancy (NV) platforms are available for entanglement generation. There are three practical factors that need to be considered:

- Time for one entanglement attempt: consider that the distance between the quantum node and a receiver is about 25 kilometers (a typical distance between two neighboring European cities) [56]. The time for an attempt to generate an entangled qubit pair between the node and the receiver is about 145 μs, including photon emission, electron readout, and the communication delay.

- Success probability for one entanglement attempt: according to [56], the success probability for one entanglement attempt is about $\alpha \times 10^{-3}$, where α is a parameter such that the fidelity of the entanglement is $1 - \alpha$.[7] For example, if the fidelity of the entanglement is required to be 0.9, the success probability is about 10^{-4}.

- Times for teleportation and entanglement storage: the operation for teleportation in the transmitter's side is essentially a Bell-state measurement, which takes about 100 μs using an NV platform [93]; the time for moving qubits to the memory or the queue is about 1040 μs.

Note that the times for teleportation and entanglement storage are akin to the overhead in the classical communication networks, whereas the time for the entanglement attempt is akin to communication time. Moreover, the total time for teleportation and entanglement storage is much larger than that for one entanglement attempt. To improve entanglement generation efficiency, multiple attempts of entanglement generation can be made in the first phase of a time slot. Suppose there are 500 attempts in the first phase. Then the duration of the phase is 7.25×10^4 μs,

[7] Fidelity is a real number in $[0, 1]$ that characterizes the quality of entanglement. Higher fidelity implies higher quality.

which is much larger than the time of the second and third phase (about 1.1×10^3 μs). With the mild assumption that the results of these attempts are i.i.d., the number of generated entanglements follows a binomial distribution. If the success probability for one attempt is set to be 10^{-4}, corresponding to a fidelity value of 0.9 as shown above, then

$$\mathbb{P}\{\text{number of the generated entanglement} = k\} \begin{cases} = 0.9512 & \text{if } k = 0 \\ = 0.0476 & \text{if } k = 1 \\ \approx 0 & \text{if } k > 1 \end{cases}$$

which can be approximated as a Bernoulli distribution as the probability of generating more than one entangled qubit pair is lower than 1.2×10^{-3}. If the required fidelity value is higher than 0.9, $\mathsf{b}_i^{(n)}$ can be approximated only better as a Bernoulli random variable. This justifies the assumption that $\mathsf{b}_i^{(n)}$ follows a Bernoulli distribution.

3.2 One-Receiver Scenario

We consider a simple scenario, where there is only one receiver, i.e., $N_{\mathrm{r}} = 1$, to gain insights into the quantum queuing system.

Recall that the number of quantum data $\mathsf{a}_i^{(n)}$ and the number of entangled qubits $\mathsf{b}_i^{(n)}$ are i.i.d. Bernoulli random variables. In the presence of only one receiver, the subscript is dropped in this section, e.g.,

$$\mathbb{P}\{\mathsf{a}^{(n)} = 1\} = p, \quad \mathbb{P}\{\mathsf{b}^{(n)} = 1\} = q. \tag{3.4}$$

In the scenario with $N_{\mathrm{r}} = 1$, the control $u^{(n)}$ is trivial: in the third phase, the quantum node simply moves the remaining entangled qubits in the platform to the

entanglement memory as long as the memory does not reach its full capacity, i.e.,

$$u^{(n)}\big([\, s^{(n)}\ c^{(n)}\,]^{\mathrm{T}}\big) = \begin{cases} Q_{\mathrm{t}} & \text{if } s^{(n)} + c^{(n)} > Q_{\mathrm{t}} \\ -M & \text{if } s^{(n)} + c^{(n)} < -M \\ s^{(n)} + c^{(n)} & \text{otherwise.} \end{cases}$$

Note that if there are multiple receivers, then determining $u^{(n)}$ becomes challenging as it involves the allocation of memory among different receivers. We next evaluate the expected average queue length in the following proposition.

Proposition 3.1. For $N_{\mathrm{r}} = 1$, if $p, q \in (0, 1)$, the expected average queue length is

$$J_{\mathrm{L}}(\boldsymbol{x}^{(1)}) = \begin{cases} \dfrac{\alpha^M\big[\alpha - \alpha^{Q_{\mathrm{t}}+1} - (1-\alpha)Q_{\mathrm{t}}\alpha^{Q_{\mathrm{t}}+1}\big]}{(1-\alpha)(1-\alpha^{Q_{\mathrm{t}}+M+1})} & \text{if } p \neq q \\ \dfrac{Q_{\mathrm{t}}(Q_{\mathrm{t}}+1)}{2(Q_{\mathrm{t}}+M+1)} & \text{if } p = q \end{cases}$$

where

$$\alpha = \frac{p(1-q)}{q(1-p)}.$$

Proof. Consider the evolution of $\mathsf{s}^{(n)}$. The transition probability from the state $\mathsf{s}^{(n)}$

to the state $\mathsf{s}^{(n+1)}$ is

$$\mathbb{P}\Big\{\mathsf{s}^{(n+1)} = s^{(n+1)} \Big| \mathsf{s}^{(n)} = s^{(n)}\Big\}$$
$$= \begin{cases} p(1-q) & \text{if } s^{(n+1)} = s^{(n)} + 1 \text{ and} \\ & \qquad s^{(n+1)} \leqslant Q_{\mathrm{t}} \\ pq + (1-q)(1-p) & \text{if } s^{(n+1)} = s^{(n)} \text{ and} \\ & \qquad s^{(n)} \in \mathcal{K}_{-M+1:Q_{\mathrm{t}}-1} \\ q(1-p) & \text{if } s^{(n+1)} = s^{(n)} - 1 \text{ and} \\ & \qquad s^{(n+1)} \geqslant -M \\ 1-p+pq & \text{if } s^{(n+1)} = s^{(n)} = -M \\ 1-q+pq & \text{if } s^{(n+1)} = s^{(n)} = Q_{\mathrm{t}}. \end{cases}$$

Let π denote the stationary distribution of $\mathsf{s}^{(n)}$, i.e.,

$$\pi_i = \lim_{n\to\infty} \mathbb{P}\{\mathsf{s}^{(n)} = i | \mathsf{s}^{(1)} = s^{(1)}\}, \quad i \in \mathcal{K}_{-M:Q_{\mathrm{t}}}.$$

The stationary distribution π satisfies the following properties:

$$\begin{aligned} \pi_{-M} &= (1-p+pq)\pi_{-M} + q(1-p)\pi_{-M+1} \\ \pi_i &= (1-q)p\pi_{i-1} + [\,pq + (1-p)(1-q)\,]\pi_i + q(1-p)\pi_{i+1}, \quad i \in \mathcal{K}_{-M+1:Q_{\mathrm{t}}-1} \\ \pi_{Q_{\mathrm{t}}} &= (1-q)p\pi_{Q_{\mathrm{t}}-1} + (1-q+pq)\pi_{Q_{\mathrm{t}}}. \end{aligned}$$

Solving these equations gives

$$\pi_i = \begin{cases} \alpha^i \dfrac{\alpha^M(1-\alpha)}{1-\alpha^{Q_{\mathrm{t}}+M+1}}, & \text{if } p \neq q \\ \dfrac{1}{Q_{\mathrm{t}}+M+1} & \text{if } p = q \end{cases}$$

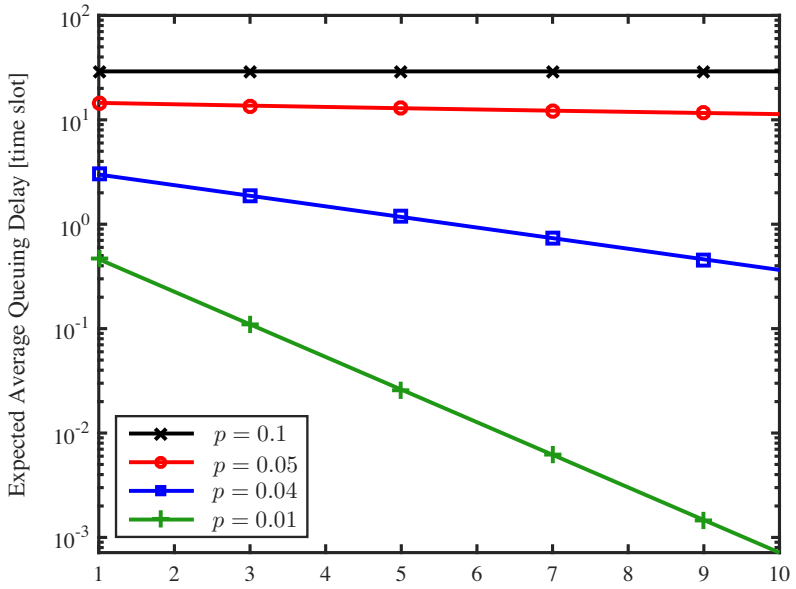


Figure 3-3: Expected average queue length as a function of the memory size M with $N_{\mathrm{r}} = 1$ and $Q_{\mathrm{t}} = 30$. The parameter q is fixed as 0.05, whereas different values of p are considered.

where $i \in \mathcal{K}_{-M:Q_{\mathrm{t}}}$. The expected average queue length is

$$J_{\mathrm{L}}(\boldsymbol{x}^{(1)}) = \sum_{i=1}^{Q_{\mathrm{t}}} i \cdot \pi_i$$

which is the desired result after some calculation. □

Fig. 3-3 shows the expected average queue length as a function of M with $N_{\mathrm{r}} = 1$ and $Q_{\mathrm{t}} = 30$. The parameter q is set as 0.05 for consistency with the analysis in Section 3.1.3. The observations obtained from Fig. 3-3 are as follows. These observations will also hold for other values of q.

- If $p < q$, i.e., $\alpha < 1$, the expected average queue length decreases exponentially as a function of M. This is because for $\alpha < 1$,

$$C_1 \alpha^M \leqslant J_{\mathrm{L}}(\boldsymbol{x}^{(1)}) \leqslant C_2 \alpha^M$$

where $C_1 = \left[\alpha - \alpha^{Q_{\mathrm{t}}+1} - (1-\alpha) Q_{\mathrm{t}} \alpha^{Q_{\mathrm{t}}+1}\right]/(1-\alpha)$ and $C_2 = C_1/(1-\alpha^{Q_{\mathrm{t}}+1})$, and they do not rely on M. This observation manifests the peculiar property of

quantum information transmission via teleportation: entanglement can be built before the quantum data arrive so that the delay can be significantly reduced. Note that the scenario with $\alpha < 1$ is common in practice since it is akin to the case that the arrival rate is lower than the processing rate in classical queuing problems.

- If $p = q$, the expected average queue length is approximately inversely proportional as a function of M. If $p > q$, the expected average queue length converges to Q_{t} (for sufficiently large Q_{t}). These two scenarios correspond to the uncommon case for which the arrival rate is no less than the processing rate.

One can also verify that the "blocking probability," i.e., the probability that the quantum data are dropped because the queue is full, is

$$\begin{aligned}
&\mathbb{P}\{\text{blocking of quantum data}\} \\
&\quad = (1-q) \lim_{n\to\infty} \mathbb{P}\{\mathsf{s}^{(n)} = Q_{\mathrm{t}} | \mathsf{s}^{(1)} = s^{(1)}\} \\
&\quad = \begin{cases} \dfrac{(1-q)\alpha^{Q_{\mathrm{t}}+M}(1-\alpha)}{1-\alpha^{Q_{\mathrm{t}}+M+1}} & \text{if } p \neq q \\ \dfrac{1-q}{Q_{\mathrm{t}}+M+1} & \text{if } p = q\,. \end{cases}
\end{aligned}$$

It is straightforward to see that this probability is a decreasing function of Q_{t}, whereas the expected average queue length is an increasing function of Q_{t}. One may optimize over Q_{t} to achieve a desirable tradeoff between the blocking probability and the expected average queue length.

Before finishing the analysis of single-receiver scenario, we would like to discuss the scenario where the discarded qubits are required for retransmission, referred to as *retransmission model*. In this model, the assumption that $\mathsf{a}^{(n)}$ are i.i.d. may not hold. However, one can show that the expected average queue length under the retransmission model is lower-bounded by $J_{\mathrm{L}}(\boldsymbol{x}^{(1)})$ shown in Proposition 3.1 and

upper-bounded by

$$\overline{J}_{\mathrm{L,retra}}(\boldsymbol{x}^{(1)}) = \begin{cases} \dfrac{\alpha^{M+1}}{1-\alpha} & \text{if } p \neq q \\ \infty & \text{if } p = q \end{cases}$$

where $\overline{J}_{\mathrm{L,retra}}(\boldsymbol{x}^{(1)})$ is obtained by letting Q_{t} go to infinity in $J_{\mathrm{L}}(\boldsymbol{x}^{(1)})$. Note that if $p < q$, both $J_{\mathrm{L}}(\boldsymbol{x}^{(1)})$ and $\overline{J}_{\mathrm{L,retra}}(\boldsymbol{x}^{(1)})$ decrease to zero exponentially with respect to the memory size M; hence, the expected average queue length under the retransmission model also decreases to zero exponentially with M.

3.3 Multiple-Receiver Scenario

In this section, we design the control policy $u^{(n)}$ for $N_{\mathrm{r}} \geqslant 2$. Recall that $u^{(n)}$ is a function that maps $\boldsymbol{x}^{(n)}$ into $\boldsymbol{s}^{(n+1)}$. We first consider the scenario where $N_{\mathrm{r}} = 2$, and extend the analysis to the scenario where $N_{\mathrm{r}} > 2$.

3.3.1 The Two-Receiver Scenario

In this section, $N_{\mathrm{r}} = 2$. Recall that $\boldsymbol{s}^{(n)}$ represents the numbers of qubits in the system at the beginning of the time slot t_n, and $\boldsymbol{c}^{(n)}$ represents the difference between the numbers of qubits arriving at the node and the numbers of entangled qubit pairs generated in the first phase of the time slot t_n. For certain values of $\boldsymbol{s}^{(n)}$ and $\boldsymbol{c}^{(n)}$, the optimal control $u^{(n)}$ is trivial: in the third phase, the quantum node simply moves the remaining entangled qubits in the platform to the entanglement memory as long as the memory does not reach its full capacity. Specifically, recall that $u^{(n)}\big([\,\boldsymbol{s}^{(n)}\ \boldsymbol{c}^{(n)}\,]^{\mathrm{T}}\big) = \boldsymbol{s}^{(n+1)}$ and

$$s_i^{(n+1)} = \min\{Q_{\mathrm{t}}, s_i^{(n)} + c_i^{(n)}\} \tag{3.5}$$

provided that

$$\sum_{i=1}^{2} \max\{0, -s_i^{(n)} - c_i^{(n)}\} \leqslant M.$$

In addition, one can verify that (3.5) also holds if $s_j^{(n)} + c_j^{(n)} \geqslant 0$, $j = 1$ or 2.

We now consider the scenarios where

$$\sum_{i=1}^{2} \max\{0, -s_i^{(n)} - c_i^{(n)}\} > M \tag{3.6}$$

$$s_i^{(n)} + c_i^{(n)} < 0 \quad i = 1, 2. \tag{3.7}$$

Note that in these scenarios, $c_i^{(n)} \leqslant 0$ $(i = 1, 2)$ since $\sum_{i=1}^{2} \max\{0, -s_i^{(n)}\} \leqslant M$. The next theorem shows that the optimal control is a threshold-based policy for these scenarios.

Theorem 3.1. For $N_\mathrm{r} = 2$, if (3.6) and (3.7) hold, there exists $T^* \in \{0, -1, -2, \ldots, -M\}$ such that the optimal control u^* for minimizing the expected average queue length gives

$$s_1^{(n+1)} = \begin{cases} s_1^{(n)} + c_1^{(n)} & \text{for } s_1^{(n)} + c_1^{(n)} \geqslant T^* \\ -M - \left(s_2^{(n)} + c_2^{(n)}\right) & \text{otherwise} \end{cases}$$

and

$$s_2^{(n+1)} = \begin{cases} -M - \left(s_1^{(n)} + c_1^{(n)}\right) & \text{for } s_1^{(n)} + c_1^{(n)} \geqslant T^* \\ s_2^{(n)} + c_2^{(n)} & \text{otherwise.} \end{cases}$$

Proof. See Appendix B.1 and therein Lemmas B.1-B.4. □

Remark 1. Given a threshold T, we develop the corresponding control policy for $N_\mathrm{r} = 2$, summarized in Algorithm 1. One way to interpret the entanglement memory management in Algorithm 1 is as follows. We assign $M_1 = T$ and $M_2 = M - T$ slots of memory budget to the entangled qubit pairs corresponding to the Receivers 1 and

Algorithm 1 Control for the Two-Receiver Scenario

Require: the current state $(s_1^{(n)}, s_2^{(n)}, c_1^{(n)}, c_2^{(n)})$, the threshold T, the memory size M, the upper bound of the queue length Q_{t}

Ensure: the qubit number in the system $s_1^{(n+1)}, s_2^{(n+1)}$ at the next time slot

if $s_i^{(n)} + c_i^{(n)} \geqslant 0$, $i \in \{1, 2\}$ or $\sum_{i=1}^{2} \max\{0, -s_i^{(n)} - c_i^{(n)}\} \leqslant M$ **then**

    // Control is trivial in these scenarios

    for $i = 1 : 2$ **do**

        $s_i^{(n+1)} = \min\{Q_{\mathrm{t}}, s_i^{(n)} + c_i^{(n)}\}$;

    end for

else

    // Entanglement memory management

$$s_1^{(n+1)} = \begin{cases} s_1^{(n)} + c_1^{(n)} & \text{if } s_1^{(n)} + c_1^{(n)} \geqslant -T \\ -M - \left(s_2^{(n)} + c_2^{(n)}\right) & \text{otherwise} \end{cases}$$

and

$$s_2^{(n+1)} = \begin{cases} -M - \left(s_1^{(n)} + c_1^{(n)}\right) & \text{if } s_1^{(n)} + c_1^{(n)} \geqslant -T \\ s_2^{(n)} + c_2^{(n)} & \text{otherwise.} \end{cases}$$

end if

2, respectively. When the budget of Receiver i is not used up, the entangled qubits corresponding the Receiver i have a higher priority of using the memory: the entangled qubits are moved to the memory, even at the cost of discarding the entangled qubits corresponding to the other receiver if the memory is full.

This structure of memory use is akin to the spectrum use in cognitive radio. Consider two virtual sets of entanglement memory slots, denoted as $\mathcal{R}_1$ and $\mathcal{R}_2$ with $\mathrm{Card}(\mathcal{R}_1) = M_1$ and $\mathrm{Card}(\mathcal{R}_2) = M_2$. The entangled qubits associated with Receiver i have a higher priority on the usage of the memory set $\mathcal{R}_i$, and we can exploit the memory set of the other Receiver j as long as $\mathcal{R}_j$ is not full. In other words, the entangled qubits associated with Receiver i are akin to the primary user regarding the usage of $\mathcal{R}_i$; and they are akin to the secondary user regarding the usage of $\mathcal{R}_j$. This idea is referred to as the cognitive memory and can be used for the policy design in general scenarios where $N_{\mathrm{r}} > 2$.

Remark 2. Algorithm 1 requires the value of the threshold T as input. It remains

unclear how to determine the optimal value threshold T^*, but in the next section we will provide an upper bound for the expected average queue length of Algorithm 1 as a function of T, which can guide the choice of T.

3.3.2 Harnessing Cognitive Memory

For the design of the control policy in the scenario with $N_\mathrm{r} > 2$, one can use the value iteration method or the policy iteration method, which is commonly used for solving the dynamic programming problem [94]. However, these methods require computing the values of certain functions corresponding to every state in the system in which the number of states increases in the order of $O(M^{N_\mathrm{r}})$. Therefore, these methods become computationally unfavorable for large N_r.

Alternatively, inspired by the idea of cognitive memory, we assign M_i slots of memory budget to the entangled qubits corresponding to Receiver i, $i \in \mathcal{K}_{1:N_\mathrm{r}}$, where

$$M_i \in \mathbb{N}_+, i \in N_\mathrm{r} \tag{3.8}$$

$$\sum_{i=1}^{N_\mathrm{r}} M_i \leqslant M. \tag{3.9}$$

Similarly to Algorithm 1, we design control policies for $N_\mathrm{r} > 2$, summarized in Algorithm 2. The control policy based on the output of Algorithm 2 is referred to as the cognitive-memory-based policy. In particular, the threshold-based policy for $N_\mathrm{r} = 2$ can be viewed as a special case of the cognitive-memory-based policy.

A subroutine in the line 12 of Algorithm 2 determines how to discard the entangled qubits that use up their memory budgets. Note that this subroutine can be designed heuristically. Later in Section 4.4, we will see that even a simple design of the subroutine can achieve near-optimal performance. The performance of Algorithm 2 relies on the subroutine. However, we can provide an upper bound for the expected average queue length regardless of the design of the subroutine.

Proposition 3.2. The expected average queue length achieved by Algorithm 2 is

Algorithm 2 Control for the N_{r}-Receiver Scenario

Require: the current state $s_i^{(n)}$, $c_i^{(n)}$, the budget M_i, the probabilities p_i and q_i, $i \in \mathcal{K}_{1:N_{\mathrm{r}}}$, the memory size M, and the upper bound of the queue length Q_{t}

Ensure: the qubit number in the system $s_i^{(n+1)}$, $i \in \mathcal{K}_{1:N_{\mathrm{r}}}$ at the next time slot

1: Let $\mathcal{D} = \varnothing$;
2: **if** $\sum_{i=1}^{N_{\mathrm{r}}} \max\{0, -s_i^{(n)} - c_i^{(n)}\} \leqslant M$ **then**
3: $\quad s_i^{(n+1)} = \min\{Q_{\mathrm{t}}, s_i^{(n)} + c_i^{(n)}\}$;
4: **else**
5: $\quad$ **for** $i = 1 : N_{\mathrm{r}}$ **do**
6: $\quad\quad$ **if** $s_i^{(n)} + c_i^{(n)} \geqslant -M_i$ **then**
7: $\quad\quad\quad s_i^{(n+1)} = \min\{Q_{\mathrm{t}}, s_i^{(n)} + c_i^{(n)}\}$;
8: $\quad\quad\quad \mathcal{D} \to \mathcal{D} \cup \{i\}$;
9: $\quad\quad$ **end if**
10: $\quad$ **end for**
11: $\quad$ **if** $\mathrm{Card}(\mathcal{D}) \neq N_{\mathrm{r}}$ **then**
12: $\quad\quad$ Call a subroutine to determine $s_j^{(n+1)}$, $j \in \mathcal{K}_{1:N_{\mathrm{r}}} \setminus \mathcal{D}$;
13: $\quad$ **end if**
14: **end if**

upper bounded by

$$J(M_1, M_2, \ldots, M_{N_{\mathrm{r}}}) = \sum_{i=1}^{N_{\mathrm{r}}} \left[(1 - \delta_{p_i,q_i}) \frac{\alpha_i^{M_i} \left[\alpha_i - \alpha_i^{Q_{\mathrm{t}}+1} - (1-\alpha_i) Q_{\mathrm{t}} \alpha_i^{Q_{\mathrm{t}}+1} \right]}{(1-\alpha_i)(1-\alpha_i^{Q_{\mathrm{t}}+M_i+1})} + \delta_{p_i,q_i} \frac{Q_{\mathrm{t}}(Q_{\mathrm{t}}+1)}{2(Q_{\mathrm{t}}+M_i)} \right] \tag{3.10}$$

where $\delta_{x,y}$ is the Kronecker delta function and

$$\alpha_i = \frac{p_i(1-q_i)}{q_i(1-p_i)},$$

Proof. The set $\mathcal{K}_{1:N_{\mathrm{r}}} \setminus \mathcal{D}$ represents the set of receivers whose corresponding entangled qubits use up the memory budget. The subroutine in the line 12 aims to determine $s_i^{(n+1)}$, $i \in \mathcal{K}_{1:N_{\mathrm{r}}} \setminus \mathcal{D}$. From the way that $\mathcal{D}$ is generated, we have that $s_i^{(n)} = -M_i$ and $c_i^{(n)} = -1$, $i \in \mathcal{K}_{1:N_{\mathrm{r}}} \setminus \mathcal{D}$. Consider a naive algorithm that discards the entangled

qubits corresponding to Receiver $i \in \mathcal{K}_{1:N_\mathrm{r}} \setminus \mathcal{D}$ in the platform, i.e.,

$$s_i^{(n+1)} = -M_i, \quad i \in \mathcal{K}_{1:N_\mathrm{r}} \setminus \mathcal{D}.$$

This algorithm has the worst performance among all possible designs of subroutines. Using this algorithm, the queuing system is decomposed into N_r independent subsystems. Each subsystem i has a memory size M_i and the corresponding expected average queue length is derived in Proposition 3.1. The total expected average queue length is then given by Proposition 3.2. □

Next, consider the choice of M_i for minimizing the expected average queue length. Since the expected average queue length depends on the design of subroutine, we use the upper bound $J(M_1, M_2, \ldots, M_{N_\mathrm{r}})$ as the objective and formulate an optimization problem as follows:

$$\begin{aligned} \mathscr{P}: \underset{\{M_i, i \in \mathcal{K}_{1:N_\mathrm{r}}\}}{\text{minimize}} \quad & J(M_1, M_2, \ldots, M_{N_\mathrm{r}}) \\ \text{subject to} \quad & \text{(3.8) and (3.9)}. \end{aligned}$$

Note that $\mathscr{P}$ is an integer programming problem and solving $\mathscr{P}$ may be computationally cumbersome.[8] Therefore, we relax the constraints (3.8) and (3.9) to $M_i \geqslant 0$, $i \in N_\mathrm{r}$, and $\sum_{i=1}^{N_\mathrm{r}} M_i \leqslant M$, respectively. The effect brought by such relaxation becomes negligible for large M. Moreover, we consider the scenarios where the following conditions hold:

$$p_i < q_i \quad i \in N_\mathrm{r} \tag{3.11}$$

$$\alpha^{Q_\mathrm{t}+M_i} \ll 1 \quad i \in N_\mathrm{r}. \tag{3.12}$$

The condition (3.11) is consistent with the discussion in Section 3.2 and the condition (3.12) is reasonable for achieving the low blocking probability shown in Section 3.2.

[8] Note that optimization techniques have been used to solve problems on quantum information science [82, 95–97].

With these two conditions, the objective function of $\mathscr{P}$ can be approximated by

$$\sum_{i=1}^{N_{\mathrm{r}}} \frac{\alpha_i^{M_i}\left[\alpha_i - \alpha_i^{Q_{\mathrm{t}}+1} - (1-\alpha_i)Q_{\mathrm{t}}\alpha_i^{Q_{\mathrm{t}}+1}\right]}{1-\alpha_i}.$$

Using this approximated term as the objective function and the relaxation of the constraints on M_i from integers to real numbers, we obtain the following relaxed optimization problem:

$$\begin{aligned} \mathscr{P}_{\mathrm{R}}: \underset{\{M_i, i\in\mathcal{K}_{1:N_{\mathrm{r}}}\}}{\text{minimize}} \quad & \sum_{i=1}^{N_{\mathrm{r}}} \lambda_i \alpha_i^{M_i} \\ \text{subject to} \quad & M_i \geqslant 0, i \in N_{\mathrm{r}} \\ & \sum_{i=1}^{N_{\mathrm{r}}} M_i \leqslant M \end{aligned}$$

where

$$\lambda_i = \frac{\alpha_i - \alpha_i^{Q_{\mathrm{t}}+1} - (1-\alpha_i)Q_{\mathrm{t}}\alpha_i^{Q_{\mathrm{t}}+1}}{1-\alpha_i}.$$

One can easily verify that $\lambda_i \geqslant 0$; moreover, the objective function is convex and the constraints are linear. Therefore, $\mathscr{P}_{\mathrm{R}}$ can be solved by standard convex programming. We next show that, by checking the Karush-Kuhn-Tucker conditions, we can obtain a closed-form solution for the problem $\mathscr{P}_{\mathrm{R}}$ under some mild assumptions.

Proposition 3.3. For

$$\begin{aligned} \mu := & -\exp\left\{\frac{1}{\sum_{i=1}^{N_{\mathrm{r}}}(1/\log\alpha_i)}\left[M + \sum_{i=1}^{N_{\mathrm{r}}}\frac{\log\left(-\lambda_i\log\alpha_i\right)}{\log\alpha_i}\right]\right\} \\ & \geqslant \lambda_i \log\alpha_i, \qquad i \in \mathcal{K}_{1:N_{\mathrm{r}}} \end{aligned}$$

the optimal solution for $\mathscr{P}_{\mathrm{R}}$ is

$$M_i^* = \frac{1}{\log\alpha_i}\log\frac{\mu}{\lambda_i\log\alpha_i}.$$

Remark 3. The condition in Proposition 3.3 holds for a large M. In such a scenario, the minimum of the objective function is

$$\mu \sum_{i=1}^{N_{\mathrm{r}}} \frac{1}{\log \alpha_i}$$

which decreases to zero exponentially as a function of M and the exponent is

$$\frac{1}{\sum_{i=1}^{N_{\mathrm{r}}} 1/\log \alpha_i}.$$

This shows that for $N_{\mathrm{r}} > 1$, the delay can be significantly decreased with entanglement built before the quantum data arrive. Such an observation is consistent with the one shown in Section 3.2, where $N_{\mathrm{r}} = 1$.

3.4 Numerical Results

This section illustrates the performance of the proposed policies through numerical results. The probabilities are set as $q_i = 0.05$, $i \in \mathcal{K}_{1:N_{\mathrm{r}}}$. Here q_i is chosen according to the analysis of the practical implementation in Section 3.1.3.

3.4.1 Two-Receiver Scenario

We first evaluate the expected average queue length in the scenario with two receivers. As an example, consider $p_1 = 0.5q_1$ and $p_2 = 0.6q_2$. We compare the following four policies:

- Value Iteration (Optimal): the policy obtained by value-iteration [98], which is also the optimal policy.

- Threshold-based I: the policy shown in Algorithm 1, with the optimal choice of the threshold T. The optimal choice is obtained by evaluating the performance of all choices of the threshold.

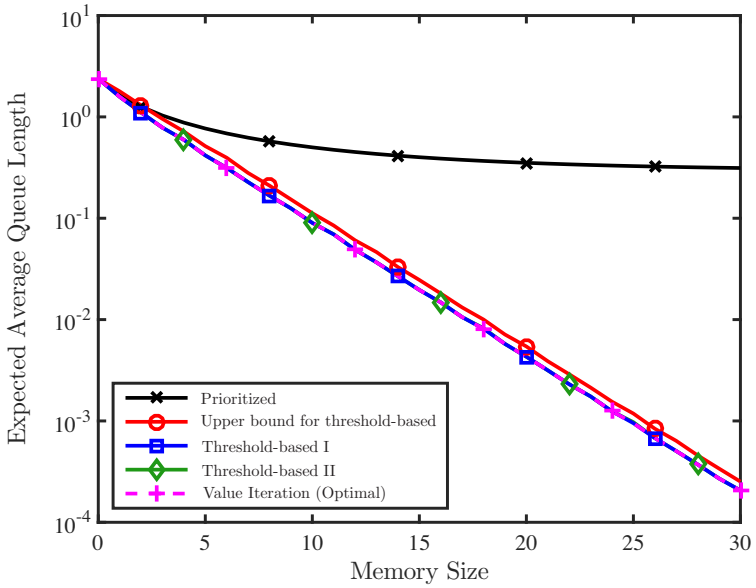


Figure 3-4: Expected average queue length as a function of the memory size M with $N_{\mathrm{r}} = 2$.

- Threshold-based II: the policy shown in Algorithm 1, with the choice of threshold T by solving the problem $\mathscr{P}$.
- Prioritized: the policy that gives priority to discarding the entangled qubits of the receiver with a lower arrival rate when the memory is full.

Note that the prioritized policy severs as a baseline. The upper bound (3.10) for the threshold-based policy will also be shown for evaluating the performance of the policies.

Fig. 3-4 shows the expected average queue length of the Value Iteration policy, the Threshold-based policies, and the Prioritized policy, as well as the upper bound for the performance of the threshold-based policy as a function of the memory size. First, the curves associated with the Value Iteration policy, the Threshold-based policies, and the upper bound demonstrate exponential decrease as a function of the memory size M. This observation is consistent with the analysis in Section 3.3.2 and further validates the insights into QQD: by building and storing entanglement before the quantum data arrive, the queuing delay can be significantly reduced and such reduction largely relies on the memory size. Second, Threshold-based Policy I

achieves the same performance as the optimal policy. This observation is consistent with Theorem 3.1, which shows that the Value Iteration policy is threshold-based. Third, the benefit of the Value Iteration and the Threshold-based policy is evident. For example, the expected average queue length is 0.4143 for the Prioritized policy when $M = 14$, whereas it is 0.0268 and 0.0268 for Threshold-based Policy I and Threshold-based Policy II, respectively. This corresponds to an average queue length reduction of 93.5%. Fourth, the curve associated with the upper bound is close to the one associated with the Value Iteration policy, showing that the upper bound is tight.

3.4.2 Multi-Receiver Scenario

We next evaluate the expected average queue length in the scenario with multiple receivers. In particular, we consider that $N_{\mathrm{r}} = 3$ and $p_1 = 0.3q_1, p_2 = 0.35q_2, p_3 = 0.4q_3$.[9] We compare the following three policies:

- Value Iteration: the policy obtained by value-iteration [98], which is also the optimal policy.
- Cognitive-memory-based: the policy shown in Algorithm 2, with the choice of M_i by solving the problem $\mathscr{P}_{\mathrm{R}}$. The subroutine in Algorithm 2 gives priority to discarding the entangled qubits of the receiver with a lower arrival rate.
- Prioritized: the policy that gives priority to discarding the entangled qubits of the receiver with a lower arrival rate when the memory is full.

Note that the prioritized policy severs as a baseline. The upper bound (3.10) will also be shown for evaluating the performance of the policies.

Fig. 3-5 shows the expected average queue length of the Value Iteration policy, the Cognitive-memory-based policy, and the Prioritized policy, as well as the upper bound for the performance of the Cognitive-memory-based policy as a function of

[9]For large N_{r}, the optimal policy is computationally cumbersome as mentioned in Section 3.3.2. Therefore, we choose a moderate value of N_{r} in this section.

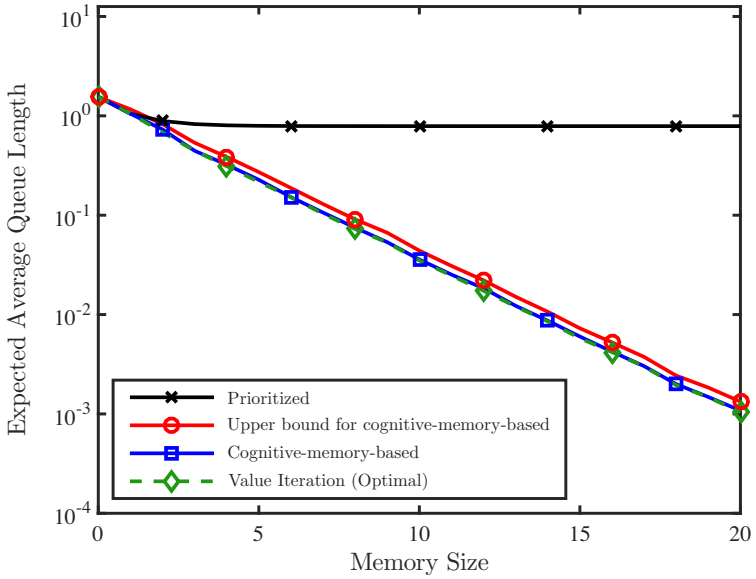


Figure 3-5: Expected average queue length as a function of the memory size M with $N_{\mathrm{r}} = 3$.

the memory size. First, the curves associated with the Value Iteration policy, the Cognitive-memory-based policy, and the upper bound demonstrate exponential decrease as a function of the memory size M. Again, this observation shows that the queuing delay can be significantly reduced and such reduction largely relies on the memory size. Second, the benefit of the Value Iteration and the Cognitive-memory-based policy is evident. For example, the expected average queue length is 0.7644 for the Prioritized policy when $M = 10$, whereas it is 0.0738 and 0.0815 for the Value Iteration policy and the Cognitive-memory-based policy, respectively. This corresponds to expected average queue length reductions of 90.4% and 89.3%, respectively. This observation is consistent with the two-receiver scenario and Proposition 3.2. Third, the curve associated with the upper bound is close to the one associated with the Value Iteration policy, showing that the upper bound is tight.

Next, we evaluate the blocking probability in the scenario with multiple receivers. Recall that the blocking probability is the probability that the quantum data are dropped because the queue is full. Fig. 3-6 shows the blocking probability of the Value Iteration policy, Cognitive-memory-based policy, and the Prioritized policy as

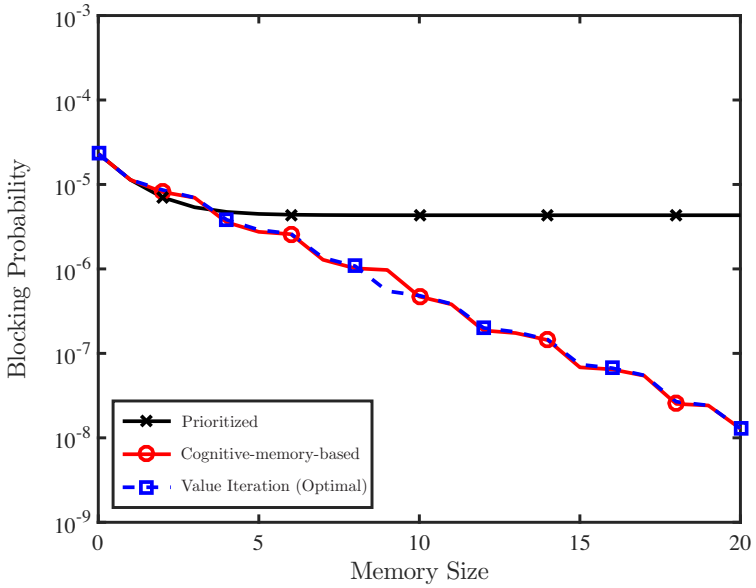


Figure 3-6: Blocking probability as a function of the memory size M with $N_{\mathrm{r}} = 3$.

a function of the memory size. The blocking probability converges to zero quickly a function of the memory size M for all the policies. This is consistent with the analysis in Section 3.2 and shows that almost none of the quantum data are dropped in the considered scenario. Together with Little's law, this shows that the expected average queuing delay is proportional to the expected average queue length, which justifies the use of expected average queue length as the objective function.

Chapter 4

Remote Entanglement Distribution

In this chapter, we will establish a framework of designing remote entanglement distribution (RED) protocols for quantum networks.

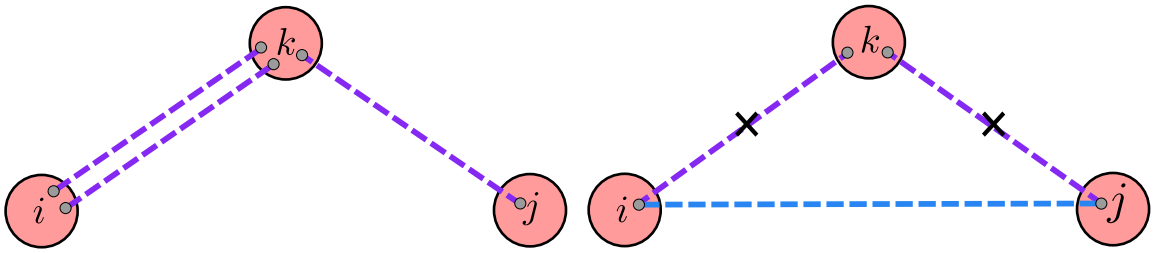
4.1 System Model

Consider a quantum network consisting of nodes equipped with quantum devices. Such a network can be abstracted by a graph consisting of nodes and edges. Let $\mathcal{N}$ and $\mathcal{E}$ denote the set of nodes and the set of edges, respectively. Each node in $\mathcal{N}$ has the capability of performing quantum measurements and storing quantum bits (qubits) for a sufficiently long time. Consequently, each node can serve as a quantum repeater.[1] Each edge $(i, j) \in \mathcal{E}$ represents a quantum channel, and this channel can be used to generate entanglement between nodes i and j. We aim to distribute entanglement at two remote nodes with the help of quantum repeaters for general networks. Note that we are particularly interested in homogeneous repeater chains, a special case of quantum networks with desired properties, and will discuss them in Section 4.3.

There are two essential operations in RED: entanglement generation and entanglement swapping, as illustrated in Fig. 4-1.

- Entanglement generation - For $(i, k) \in \mathcal{E}$, nodes i and k are connected by

[1] We use "repeater" and "node" interchangeably in the rest of the thesis.



(a) entanglement generation (b) entanglement swapping

Figure 4-1: An illustration of entanglement generation and entanglement swapping. In (a), nodes i and k are connected by a quantum channel, and so are nodes k and j. Purple lines represent generated EQPs. Entanglement can be generated between i and k. In (b), the node k performs entanglement swapping using $\boldsymbol{\Xi}_{i:k}$ and $\boldsymbol{\Xi}_{k:j}$ to distribute $\boldsymbol{\Xi}_{i:j}$. Blue lines represent distributed EQPs.

a quantum channel. Nodes i and k can attempt to generate[2] entangled qubit pairs (grey dots connected by purple dashed lines in Fig. 4-1(a)). Entanglement generation can be implemented, for example, by locally preparing an entangled qubit pair at one of the nodes (e.g., i) and sending one entangled qubit in this pair to the other node (e.g., k) [90]. The entangled qubits are stored in the nodes for a sufficiently long time. The attempt of entanglement generation may not always succeed. If the attempt succeeds, the density matrix of the entangled qubit pair is:

$$\boldsymbol{\Xi}_{i_a,k_b} := \frac{1}{2}(|00\rangle_{i_a,k_b} + |11\rangle_{i_a,k_b})(\langle 00|_{i_a,k_b} + \langle 11|_{i_a,k_b})$$

where i_a and k_b represent physical systems in node i and k, respectively; otherwise, no entangled qubit pair is obtained.

- Entanglement swapping - Entanglement swapping [24, 99] can be seen as a special case of teleportation [23]. Suppose there are three nodes i, k, and j. Node

[2]In this thesis, we use the term "generate/generation" to describe the operation of preparing an entangled qubit pair at one node and sending one of the qubits through a quantum channel; we use the term "distribute/distribution" to describe the operation of preparing entangled qubit pairs at two nodes that are not directly connected by a quantum channel via entanglement generation and entanglement swapping. Moreover, "distribution" in entanglement distribution should not be confused with that in probability distribution.

i has one qubit, node k two, and node j one. Node i's qubit and node k's first qubit are maximally entangled, and so are node k's second qubit and node j's qubit. Node k teleports the qubit entangled with the one in node i to node j. Then, the node i's qubit is entangled with node j's even though these two nodes have never directly interacted with each other. This operation is referred to as entanglement swapping. The attempt of entanglement swapping may not always succeed. If the attempt succeeds, one entangled qubit pair $\boldsymbol{\Xi}_{i:j}$ is distributed; otherwise, no entangled qubit pair is obtained even though the two EQPs are consumed.

We consider a time-slotted system in which slots are indexed by $\tau \in \mathbb{N}_+$. Each time slot is divided into two phases:

- Phase I: For any $(i,k) \in \mathcal{E}$, nodes i and k can make an attempt to generate an entangled qubit pair with success probability of $p_{i:k}$.[3]

- Phase II: For any $i,j,k \in \mathcal{N}$, node k can attempt to perform entanglement swapping with success probability q_k to distributed entanglement $\boldsymbol{\Xi}_{i:j}$ using $\boldsymbol{\Xi}_{i:k}$ and $\boldsymbol{\Xi}_{k:j}$.

Note that the model for entanglement generation includes the lossy optical channel, which is commonly used for quantum communication. The results presented in this thesis can be extended to a general quantum channel by replacing $p_{i:j}$ with the quantum channel capacity between i and j, $\forall (i,j) \in \mathcal{E}$.

We are interested in designing efficient RED protocols for scheduling entanglement swapping in the network so that a large number of entangled qubit pairs can be distributed between a source node s and a sink node t ($s,t \in \mathcal{N}$). The node s may be remote from the node t. Once an entangled qubit pair between s and t is distributed, it is stored and will not be used for entanglement swapping. For a given network described by $\mathcal{N}$, $\mathcal{E}$, and $\{p_{i:j}\}_{(i,j)\in\mathcal{E}}$, our goal is to maximize the entanglement

[3]Note that this probability does not rely on the order of i and k, i.e., $p_{i:k} = p_{k:i}$.

distribution rate (EDR) achieved by a protocol π, i.e.,

$$\lambda^{\pi} = \liminf_{T\to\infty} \frac{1}{T} \sum_{\tau=1}^{T} \mathbb{E}\{\mathsf{g}_{s:t}^{\pi}(\tau)\} \tag{4.1}$$

where $\mathsf{g}_{s:t}^{\pi}(\tau)$ denotes the number of EQPs generated and/or distributed between s and t distributed at time slot τ using an RED protocol π. The maximum EDR over all RED protocols is denoted by λ^*.

4.2 Imperfect Entanglement Swapping Operation

In this section, we determine the maximum EDR λ^* and design the optimal RED protocol.

4.2.1 Optimal EDR

Analyzing entanglement swapping directly on the original network is complicated because the entanglement swapping operation involve pairs of nodes rather than individual nodes. Such difficulty motivates us to consider a graph consisting of nodes that represent node pairs in $\mathcal{N}$. Specifically, we introduce a directed graph $\mathcal{G}$:

- A node in $\mathcal{G}$ corresponds to a pair of nodes in $\mathcal{N}$, denoted by $e_{i:k}$ for some $i, k \in \mathcal{N}$. To distinguish the nodes in $\mathcal{G}$ from those in $\mathcal{N}$, we refer to $e_{i:k}$, $i, k \in \mathcal{N}$ in $\mathcal{G}$ as *enodes*. We do not differentiate the order of the two nodes i and k in an enode, i.e., $e_{i:k} = e_{k:i}$. These enodes represent the entangled qubit pairs.

- Let a nonnegative number $f_{i:j}^{i:k}$ denote the *eflow* from $e_{i:k}$ to $e_{i:j}$. If $f_{i:j}^{i:k} > 0$, there is a directed edge from the enode $e_{i:k}$ to the enode $e_{i:j}$. The eflow $f_{i:j}^{i:k}$ represents the amount of EQPs $\boldsymbol{\Xi}_{i:k}$ used for distributing $\boldsymbol{\Xi}_{i:j}$ via entanglement swapping. Since the order of the two nodes in an enode is not differentiated, we have $f_{i:j}^{i:k} = f_{i:j}^{k:i} = f_{j:i}^{i:k} = f_{j:i}^{k:i}$. However, $f_{i:k}^{i:j}$ is not necessarily equal to $f_{i:j}^{i:k}$.

An illustration of the enode and eflow is depicted in Fig. 4-2. The next theorem provides an upper bound for the optimal EDR. This upper bound is based on an optimization problem, where the variables are the eflows.

Theorem 4.1. For a given graph with node set $\mathcal{N}$, edge set $\mathcal{E}$, entanglement generation success probability $\{p_{i:j} : (i,j) \in \mathcal{E}\}$, and entanglement swapping success probability $\{q_k : k \in \mathcal{N}\}$, the optimal EDR is upper-bounded by the optimal value of the following problem $\mathscr{P}$:

$$\begin{aligned}
\mathscr{P}: \quad & \underset{\{f_{i:j}^{i:k}: i,j,k \in \mathcal{N}\}}{\text{maximize}} \quad I(s,t) \\
& \text{subject to} \quad I(i,j) \geqslant \sum_{k \in \mathcal{N}\setminus\{i,j\}} (f_{i:k}^{i:j} + f_{k:j}^{i:j}), \qquad i,j \in \mathcal{N}, \{i,j\} \neq \{s,t\} \quad (4.2) \\
& \qquad\qquad f_{i:j}^{i:k} = f_{i:j}^{k:j} \geqslant 0, i,j,k \in \mathcal{N} \quad (4.3) \\
& \qquad\qquad f_{s:k}^{s:t} = f_{k:t}^{s:t} = 0, \quad k \in \mathcal{N} \quad (4.4)
\end{aligned}$$

where for $i,j \in \mathcal{N}$,

$$I(i,j) = 1_{\mathcal{E}}(i,j) p_{i:j} + \sum_{k \in \mathcal{N}\setminus\{i,j\}} q_k \frac{f_{i:j}^{i:k} + f_{i:j}^{k:j}}{2}.$$

Proof. See Appendix C.1. □

Remark 4. The objective function and the constrains of $\mathscr{P}$ can be interpreted as follows. The objective function $I(s,t)$ represents the amount of entanglement $\boldsymbol{\Xi}_{s:t}$ generated/distributed between the source node and the sink node. The constraint (4.2) represents the entanglement balance corresponding to an arbitrary enode $e_{i:j}$. In particular, the left hand side of (4.2) represents the amount of entanglement $\boldsymbol{\Xi}_{s:t}$ generated/distributed between node i and node j; the right hand side of (4.2) represents the amount of entanglement $\boldsymbol{\Xi}_{i:j}$ that is used for distributing other entanglement. The constraint (4.3) represents the symmetry in entanglement swapping. In particular, in entanglement swapping, the amount of entanglement $\boldsymbol{\Xi}_{i:k}$ and $\boldsymbol{\Xi}_{k:j}$ consumed to distribute $\boldsymbol{\Xi}_{i:j}$ needs to be the same. The constraint (4.4) represents

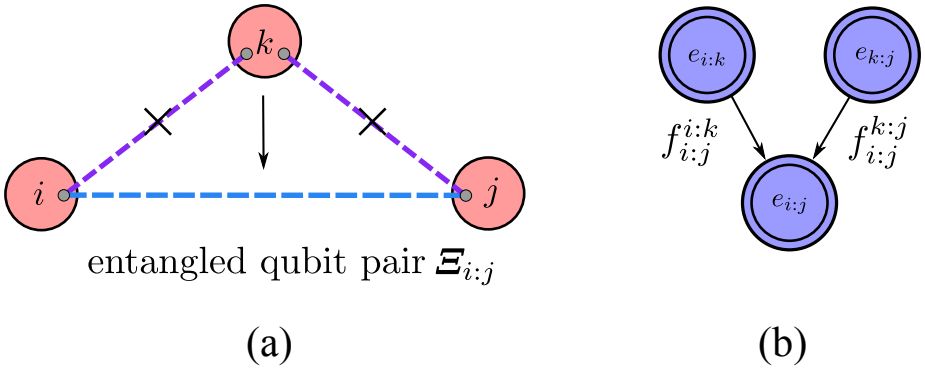


Figure 4-2: An illustration of enodes and eflows. In (a), the node j performs entanglement swapping using $\boldsymbol{\Xi}_{i:j}$ and $\boldsymbol{\Xi}_{k:j}$ to distribute $\boldsymbol{\Xi}_{i:k}$ with EDR λ. Correspondingly, in (b), the enodes $e_{i:k}$ and $e_{k:j}$ contribute eflows to $e_{i:j}$ with $f_{i:j}^{i:k} = f_{i:j}^{k:j} = \lambda$.

the requirement that the entanglement $\boldsymbol{\Xi}_{s:t}$ is not used for entanglement swapping.

Remark 5. The problem $\mathscr{P}$ is a linear programming problem. The computational complexity of $\mathscr{P}$ depends on the numbers of variables and constraints, and both numbers are poly(Card($\mathcal{N}$)) [100–103]. Therefore, the optimal solution of the problem $\mathscr{P}$ can be obtained efficiently by standard linear optimization algorithms. Note that optimization techniques have been used to solve problems on quantum information science [82, 95, 96, 104].

We next determine some structural properties of a graph that corresponds to an optimal solution of $\mathscr{P}$.

4.2.2 Structural Properties

Definition 1 (Directed acyclic graph)**.** A directed acyclic graph (DAG) is a directed graph with no directed cycle.

Proposition 4.1. There exists an optimal solution of $\mathscr{P}$ such that the graph $\mathcal{G}$ corresponding to this solution is a DAG.

Proof. See Appendix C.2. □

There are some standard concepts such as isolated node, parent, child, ancestor, and descendant for DAGs and we can employ them in the graph consisting of enodes.

Definition 2 (Efficiency)**.** A directed graph consisting of enodes is efficient if for every enode $e_{i:j}$ except for $e_{s:t}$, either $e_{i:j}$ is isolated or $e_{i:j}$ is an ancestor of $e_{s:t}$.

Proposition 4.2. There exists an optimal solution of $\mathscr{P}$ such that the graph corresponding to this solution is acyclic and efficient.

Proof. See Appendix C.3. □

The constraint (4.2) in the problem $\mathscr{P}$ can be interpreted as follows: the incoming flow into the enode $e_{i:j}$ is no less than the outgoing flow. If the equality does not hold, some of the entanglement will be wasted. To reduce such waste, we introduce *control variables* $\{u_{i:j} : (i,j) \in \mathcal{E}\}$ for limiting the number of generated EQPs, which induces the following problem $\mathscr{P}_{\mathrm{s}}$:

$$
\begin{aligned}
\mathscr{P}_{\mathrm{s}}: \underset{\substack{\{f_{i:j}^{i:k}: i,j,k\in\mathcal{N}\}\\ \{u_{i:j}:(i,j)\in\mathcal{E}\}}}{\text{maximize}} \quad & I(s,t) \\
\text{subject to} \quad & u_{i:j}p_{i:j}1_{\mathcal{E}}(i,j) + \sum_{k\in\mathcal{N}\setminus\{i,j\}} q_k \frac{f_{i:j}^{i:k} + f_{i:j}^{k:j}}{2} = \sum_{k\in\mathcal{N}\setminus\{i,j\}} (f_{i:k}^{i:j} + f_{k:j}^{i:j}), \\
& \qquad i,j \in \mathcal{N}, \{i,j\} \neq \{s,t\} && (4.5) \\
& f_{i:j}^{i:k} = f_{i:j}^{k:j} \geqslant 0, \qquad i,j,k \in \mathcal{N} && (4.6) \\
& f_{s:i}^{s:t} = f_{t:i}^{s:t} = 0, \qquad i \in \mathcal{N} && (4.7) \\
& 0 \leqslant u_{i:j} \leqslant 1, \qquad (i,j) \in \mathcal{E}. && (4.8)
\end{aligned}
$$

Evidently, if $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i,j) \in \mathcal{E}\}$ is an optimal solution of $\mathscr{P}_{\mathrm{s}}$, then $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ is a feasible solution of $\mathscr{P}$. The next theorem implies that $\mathscr{P}$ and $\mathscr{P}_{\mathrm{s}}$ are equivalent.

Theorem 4.2. The optimal value of $\mathscr{P}_{\mathrm{s}}$ is the same as that of $\mathscr{P}$. Moreover, there exists an optimal solution of $\mathscr{P}_{\mathrm{s}}$ such that the graph corresponding to this solution is acyclic and efficient.

Proof. See Appendix C.4. □

Theorems 4.1 and 4.2 have shown that the optimal value of $\mathscr{P}_{\rm s}$ is an upper bound for the EDR. Next we will show that this bound is tight by designing an RED protocol that achieves this bound. The RED protocol will employ the optimal solution of $\mathscr{P}_{\rm s}$.

4.2.3 Stationary Protocol

By Theorem 4.2, we can consider an optimal solution $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i,j) \in \mathcal{E}\}$ of $\mathscr{P}_{\rm s}$ such that the graph corresponding to this solution is acyclic and efficient. This solution will be used to develop the RED protocol. Note that topological ordering is possible for a DAG consisting of enodes, and we can find a linear ordering of all the enodes, such that if $f_{i:j}^{i:k} > 0$, the enode $e_{i:k}$ precedes the enode $e_{i:j}$ in the ordering.

For each enode $e_{i:k}$ that has a descendant $e_{s:t}$ in $\mathcal{G}$, we consider a set $\mathcal{M}_{i:k}$, consisting of the EQPs between i and k; we also consider a collection of sets, denoted as $\mathcal{F}_{i:j}^{i:k}$ and $\mathcal{F}_{j:k}^{i:k}$, $j \in \mathcal{N} \setminus \{i,k\}$, consisting of the EQPs between i and k that will be used for distributing $\boldsymbol{\Xi}_{i:j}$ and $\boldsymbol{\Xi}_{j:k}$, respectively. This stationary protocol, denoted by $\mathring{\pi}$, is described as follows:

- In Phase I,
  - For any $(i,k) \in \mathcal{E}$, nodes i and k attempt to generate an entangled qubit pair $\boldsymbol{\Xi}_{i:k}$ with the probability $\mathring{u}_{i:k}$.[4] If the attempt is successful, the entangled qubit pair $\boldsymbol{\Xi}_{i:k}$ is moved to the set $\mathcal{M}_{i:k}$.
  - For every enode $e_{i:k}$, if $e_{i:k} \neq e_{s:t}$, we move each EQP $\boldsymbol{\Xi}_{i:k}$ in $\mathcal{M}_{i:k}$ to the set $\mathcal{F}_{i:j}^{i:k}$ or the set $\mathcal{F}_{j:k}^{i:k}$, $j \in \mathcal{N} \setminus \{i,k\}$ randomly. In particular,

$$\mathbb{P}\{\text{move } \boldsymbol{\Xi}_{i:k} \text{ to } \mathcal{F}_{i:j}^{i:k}\} = \frac{\mathring{f}_{i:j}^{i:k}}{\sum_{l\in\mathcal{N}\setminus\{i,k\}}(\mathring{f}_{i:l}^{i:k} + \mathring{f}_{l:k}^{i:k})} \tag{4.9}$$

$$\mathbb{P}\{\text{move } \boldsymbol{\Xi}_{i:k} \text{ to } \mathcal{F}_{j:k}^{i:k}\} = \frac{\mathring{f}_{j:k}^{i:k}}{\sum_{l\in\mathcal{N}\setminus\{i,k\}}(\mathring{f}_{i:l}^{i:k} + \mathring{f}_{l:k}^{i:k})}. \tag{4.10}$$

[4]The control variable $\mathring{u}_{i:k}$ is the probability of attempting to generate entanglements. Note that only if they make the attempt *and* the attempt is successful, can an entangled qubit pair be generated between nodes i and k. Hence, an entangled qubit pair can be generated with probability $\mathring{u}_{i:k}p_{i:k}$.

- In Phase II, for every $i, j, k \in \mathcal{N}$, the node k performs entanglement swapping to distribute $\boldsymbol{\Xi}_{i:j}$ using $\boldsymbol{\Xi}_{i:k}$ in $\mathcal{F}_{i:j}^{i:k}$ and $\boldsymbol{\Xi}_{k:j}$ in $\mathcal{F}_{i:j}^{k:j}$ until the set $\mathcal{F}_{i:j}^{i:k}$ or $\mathcal{F}_{i:j}^{k:j}$ is empty. The distributed EQPs between i and j are then moved to $\mathcal{M}_{i:j}$.

Note that the target entangled qubit pair $\boldsymbol{\Xi}_{s:t}$ is in $\mathcal{M}_{s:t}$. We claim that the number of entangled qubit pairs that the enode $e_{i:j}$ has accumulated after sufficiently large time slot T in $\mathcal{G}$, denoted by $\mathsf{n}_{i:j}(T)$, satisfies

$$\lim_{T\to\infty} \frac{1}{T}\mathsf{n}_{i:j}(T) \overset{\text{a.s.}}{=} u_{i:j}p_{i:j}1_{\mathcal{E}}(i,j) + \sum_{k\in\mathcal{N}\setminus\{i,j\}} q_k \frac{\mathring{f}_{i:j}^{i:k} + \mathring{f}_{i:j}^{k:j}}{2}. \tag{4.11}$$

This claim can be rewritten as follows: for any $x \in \mathcal{K}_{1:\mathrm{Card}(\mathcal{G})}$, the equation (4.11) holds, where $e_{i:j}$ denotes the xth enode in the topological order of $\mathcal{G}$. This claim can be proved by strong mathematical induction on the position x of an enode in the topological order determined by $\mathcal{G}$ [105].

Base case: If $x = 1$, then $\mathring{f}_{i:j}^{i:k} = \mathring{f}_{i:j}^{k:j} = 0$ for $\forall k \in \mathcal{N}$, and by the strong law of large numbers, the number of the EQPs $\boldsymbol{\Xi}_{i:j}$ accumulated in the enode $e_{i:j}$ satisfies

$$\lim_{T\to\infty} \frac{1}{T}\mathsf{n}_{i:j}(T) \overset{\text{a.s.}}{=} u_{i:j}p_{i:j}1_{\mathcal{E}}(i,j)$$

which equals (4.11) with $\mathring{f}_{i:j}^{i:k} = \mathring{f}_{i:j}^{k:j} = 0$ for $\forall k \in \mathcal{N}$.

Induction step: Suppose equation (4.11) holds for $x = 1, 2, \ldots, r$. We will prove (4.11) holds for $x = r+1$. For $k \in \mathcal{N} \setminus \{i, j\}$, if $\mathring{f}_{i:j}^{i:k} > 0$, then the enode $e_{i:k}$ precedes $e_{i:j}$ in the topological ordering, showing that the position of $e_{i:k}$ is less than $r+1$. By the induction hypothesis,

$$\lim_{T\to\infty} \frac{1}{T}\mathsf{n}_{i:k}(T) \overset{\text{a.s.}}{=} u_{i:k}p_{i:k}1_{\mathcal{E}}(i,k) + \sum_{l\in\mathcal{N}\setminus\{i,k\}} q_l \frac{\mathring{f}_{i:k}^{i:l} + \mathring{f}_{i:k}^{l:k}}{2}. \tag{4.12}$$

Consequently, the number of the EQPs $\boldsymbol{\Xi}_{i:k}$ consumed for distributing $\boldsymbol{\Xi}_{i:j}$, denoted

by $\mathsf{d}_{i:j}^{i:k}(T)$, satisfies

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\mathsf{d}_{i:j}^{i:k}(T) &\overset{\text{a.s.}}{=} \lim_{T\to\infty}\frac{1}{T}\Big[\mathsf{n}_{i:k}(T)\,\mathbb{P}\{\text{move } \boldsymbol{\Xi}_{i:k} \text{ to } \mathcal{F}_{i:j}^{i:k}\}\Big]\\ &\overset{\text{a.s.}}{=} \mathring{f}_{i:j}^{i:k} \end{aligned} \tag{4.13}$$

where we have used (4.12) and (4.9) together with (4.5) to obtain (4.13). Similarly, we have

$$\lim_{T\to\infty}\frac{1}{T}\mathsf{d}_{i:j}^{k:j}(T) \overset{\text{a.s.}}{=} \mathring{f}_{i:j}^{k:j}. \tag{4.14}$$

Then

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\mathsf{n}_{i:j}(T) &\overset{\text{a.s.}}{=} u_{i:j}p_{i:j}1_{\mathcal{E}}(i,j)\\ &\quad+\sum_{k\in\mathcal{N}\setminus\{i,j\}} q_k \lim_{T\to\infty}\frac{1}{T}\min\{\mathsf{d}_{i:j}^{i:k}(T),\mathsf{d}_{i:j}^{k:j}(T)\}\\ &\overset{\text{a.s.}}{=} u_{i:j}p_{i:j}1_{\mathcal{E}}(i,j)+\sum_{k\in\mathcal{N}\setminus\{i,j\}} q_k\frac{\mathring{f}_{i:j}^{i:k}+\mathring{f}_{i:j}^{k:j}}{2}\end{aligned}$$

where the summand $T\,u_{i:j}p_{i:j}1_{\mathcal{E}}(i,j)$ comes from direct entanglement generation, and the rest of the summand comes from entanglement swapping due to (4.13) and (4.14). This proves that (4.11) holds for $x = r+1$.

Let $\mathring{I}(s,t)$ denote the optimal value of $\mathscr{P}_{\text{s}}$. Then

$$\begin{aligned}\lim_{T\to\infty}\mathbb{E}\Big\{\frac{1}{T}\mathsf{n}_{s:t}(T)\Big\} &= \mathbb{E}\Big\{\lim_{T\to\infty}\frac{1}{T}\mathsf{n}_{s:t}(T)\Big\}\\ &= p_{s:t}1_{\mathcal{E}}(s,t)+\sum_{k\in\mathcal{N}\setminus\{s,t\}} q_k\frac{\mathring{f}_{s:t}^{s:k}+\mathring{f}_{s:t}^{k:t}}{2}\\ &= \mathring{I}(s,t)\end{aligned}$$

where the first equality is due to the dominated convergence theorem [106]. Therefore,

the EDR achieved by the stationary protocol $\mathring{\pi}$ is

$$\lambda^{\mathring{\pi}} = \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}\{\mathsf{n}_{s:t}(T)\} = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\{\mathsf{n}_{s:t}(T)\} = \mathring{I}(s,t)$$

where the second equality is due to the existence of the limit. This shows that the protocol $\mathring{\pi}$ achieves the maximum EDR.

4.3 Homogeneous Repeater Chains

In this section, we consider a special quantum network, homogeneous repeater chains, and derive the solution for $\mathscr{P}_{\mathrm{s}}$ in a closed form.

4.3.1 Network Model

The model of homogeneous repeater chains is described in several related works, e.g., [78]. A repeater chain is connected by N quantum channels as illustrated in Fig. 4-3. Specifically, the nodes are labeled as 0, 1, 2, ..., N and the edge set $\mathcal{E} = \{(0,1),(1,2),(2,3),\ldots,(N-1,N)\}$. Each edge $(i-1,i)$ has associated probability $p_{i-1:i}$ describing the probability of success in generating entangled qubit pair $\boldsymbol{\Xi}_{i-1:i}$. The node 0 and the node N are the source and sink nodes, respectively. The enode $e_{s:t}$ is then $e_{0:N}$. In a homogeneous repeater chain, the quantum channels between neighboring nodes are the same, i.e., $p := p_{0:1} = p_{1:2} = p_{2:3} = \ldots = p_{N-1:N}$. Furthermore, the success probability of entanglement swapping in the chain is assumed to be the same for all nodes and denoted by q.

Section 4.2 shows that designing the optimal RED protocol is equivalent to solving the optimization problem $\mathscr{P}_{\mathrm{s}}$. Therefore, we aim at solving $\mathscr{P}_{\mathrm{s}}$ corresponding to homogeneous repeater chains. We have the following result:

Claim 1. For homogeneous repeater chains connected by N quantum channels, the maximal EDR is

$$\frac{(N-\xi(N))pq^{n+1}}{2(N-2^n)+(2^{n+1}-N-\xi(N))q}$$

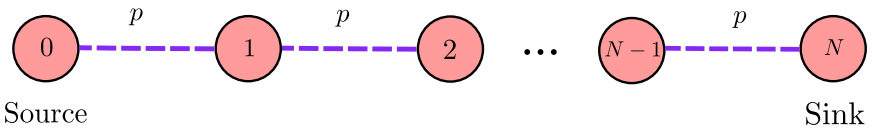


Figure 4-3: Illustration of repeater chains. The purple dashed lines represent the quantum channels between neighboring nodes.

where $n = \lceil \log_2 N \rceil - 1$, and $\xi(N)$ is a parity indicator function of N: $\xi(N) = 1$ if N is odd, otherwise $\xi(N) = 0$.

The proof of the claim and the design of the optimal protocol will be discussed in the next subsections.

4.3.2 Scenarios with an Even N

We consider the scenario where N is even, i.e., $N \in \mathbb{Z}_{\rm e}$. We first present a solution of $\mathscr{P}_{\rm s}$, denoted by $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i,j) \in \mathcal{E}\}$, and then prove that it is the optimal solution of $\mathscr{P}_{\rm s}$.

Let $n = \lceil \log_2 N \rceil - 1$. If $N = 2^{n+1}$, the solution $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i,j) \in \mathcal{E}\}$ of $\mathscr{P}_{\rm s}$ is

$$\begin{aligned} &\mathring{u}_{k:k+1} = 1, \quad k \in \mathcal{K}_{0:N-1} \\ &\mathring{f}_{2^k a:2^k a+2^k}^{2^k a:2^k a+2^{k-1}} = \mathring{f}_{2^k a:2^k a+2^k}^{2^k a+2^{k-1}:2^k a+2^k} = pq^{k-1}, \\ &\qquad a = 0,1,2,\ldots,2^{n+1-k}-1, \quad k = 1,2,\ldots,n+1. \end{aligned} \tag{4.15}$$

For example, if $a = 1$ and $k = 2$, then $2^k a = 4$, $2^k a + 2^{k-1} = 6$, $2^k a + 2^k = 8$, and consequently $\mathring{f}_{4:8}^{4:6} = \mathring{f}_{4:8}^{6:8} = pq$. An illustration of this solution for $N = 8$ is given in Fig. 4-4. This solution corresponds to a graph with the structure of a perfect binary tree. Specifically, the entanglement swapping is first performed between the enodes $e_{2a:2a+1}$ and $e_{2a+1:2a+2}$, and this gives the EQPs corresponding to the enode $e_{2a:2a+2}$, for $a = 0,1,2,\ldots,N/2-1$. Then the entanglement swapping is performed between $e_{4a:4a+2}$ and $e_{4a:4a+4}$, and this gives the EQPs corresponding to the enode $e_{4a:4a+4}$, for $a = 0,1,2,\ldots,N/4-1$. Such entanglement swapping can continue until the EQPs corresponding to the enode $e_{0:N}$ are obtained.

If $N < 2^{n+1}$, the optimal solution $\{\mathring{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i, j) \in \mathcal{E}\}$ can be obtained in six steps:

1. $\mathring{u}_{k:k+1} = 1$, $k \in \mathcal{K}_{0:N-1}$;

2. We consider the enode: $e_{0:1}$, $e_{1:2}$, ..., $e_{N-1:N}$ and divide them into $N/2$ pairs of enodes: $e_{0:1}$ and $e_{1:2}$, $e_{2:3}$ and $e_{3:4}$, ..., $e_{N-2:N-1}$ and $e_{N-1:N}$;

3. There are $\binom{N/2}{N-2^n}$ possible ways to choose $(N - 2^n)$ out of $N/2$ enode pairs. These choices are labeled as 1, 2, ..., $\binom{N/2}{N-2^n}$;

4. For the choice l, we have $(N - 2^n)$ enode pairs chosen. Entanglement swapping is performed between two enodes in each chosen pair, resulting in an enode that is the common child of these two enodes; now we have a chain consisting of 2^n quantum channels and we can relabel the nodes in the chain as $\overline{0}$, $\overline{1}$, ..., $\overline{2^n}$;

5. For the newly labeled chain, we determine the following eflow:

$$\mathring{f}_{\overline{2^k a}:\overline{2^k a+2^{k-1}}}^{\overline{2^k a}:\overline{2^k a+2^k}}(l) = \mathring{f}_{\overline{2^k a+2^{k-1}}:\overline{2^k a+2^k}}^{\overline{2^k a}:\overline{2^k a+2^k}}(l) = \frac{1}{\binom{N/2}{N-2^n}} \frac{N/2}{N - 2^n + q(2^n - N/2)} pq^k,$$
$$a = 0, 1, 2, \ldots, 2^{n-k} - 1, \quad k = 1, 2, \ldots, n+1$$

6. We repeat Steps 4 and 5 until all $\binom{N/2}{N-2^n}$ combinations are iterated. We add all the eflows to obtain the optimal solution.

In Step 3, we transfer the problem from the scenario with $N \in \mathbb{Z}_{\rm e}$ to the scenario where N is a power of two. An illustration of this step is in Fig. 4-5. The method of developing the optimal solution for $N = 6$ is illustrated in Fig. 4-6. In this case, $n = \lceil \log_2 6 \rceil - 1 = 2$. We divide six enodes into three pairs: $e_{0:1}$ and $e_{1:2}$, $e_{2:3}$ and $e_{3:4}$, as well as $e_{4:5}$ and $e_{5:6}$. There are three possible ways of choosing $N - 2^n = 2$ out of $N/2 = 3$ pairs. The choices are labeled as 1, 2, and 3, and are shown in three red rectangles. In each rectangle, we perform the Steps 4 and 5. The eflows in three rectangles can then be added to obtain the optimal solution of $\mathscr{P}_{\rm s}$.

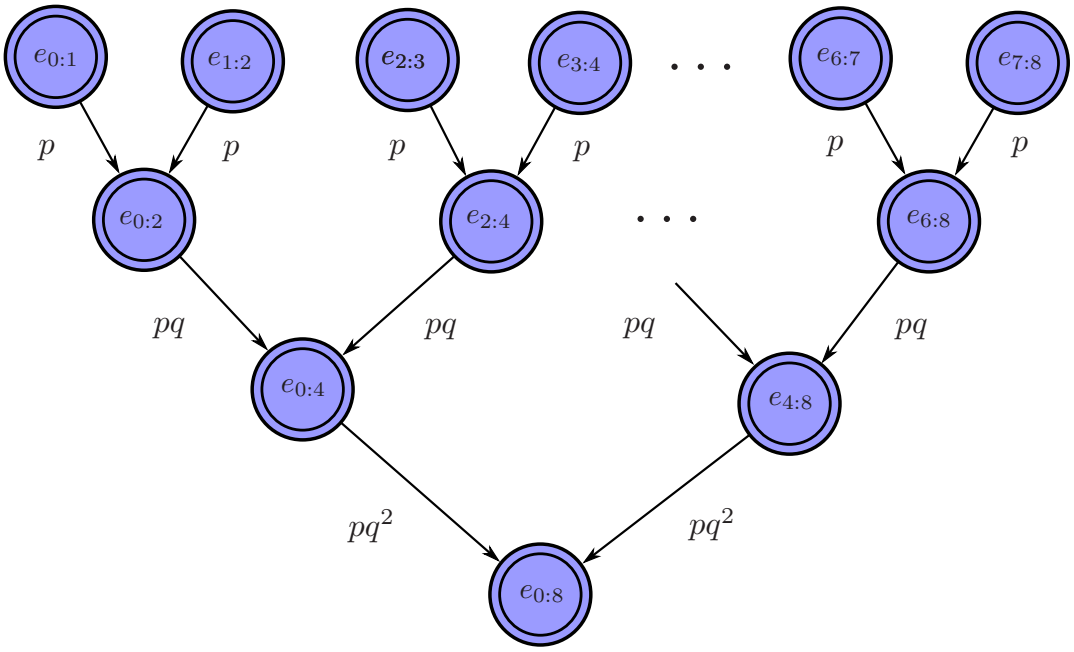


Figure 4-4: Illustration of the optimal RED protocol for $N = 8 = 2^3$. The quantities near the edges represent the eflow.

Remark 6. The feasibility of the $\{\mathring{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i, j) \in \mathcal{E}\}$ as a solution of $\mathscr{P}_{\rm s}$ can be verified by checking conditions (4.5)-(4.8). The value of the objective function corresponding to this solution is

$$\begin{aligned}\mathring{I}(s,t) &= \frac{1}{\binom{N/2}{N-2^n}} \frac{N/2}{N-2^n+q(2^n-N/2)} pq^{n+1} \cdot \binom{N/2}{N-2^n} \\ &= \frac{Np}{g(N)}\end{aligned}$$

where $g(\cdot)$ is a function defined as

$$g(L) = \frac{2(L-2^l)+q(2^{l+1}-L)}{q^{l+1}} \tag{4.16}$$

in which $l = \lceil \log_2 L \rceil - 1$.

We next determine an upper bound for the optimal value of $\mathscr{P}_{\rm s}$ for $N \in \mathbb{Z}_{\rm e}$. If this upper bound coincides with $\mathring{I}(s,t)$, then the optimality of $\{\mathring{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i, j) \in \mathcal{E}\}$ will be proved.

Note that the EQPs generated in Phase I correspond to the term $u_{i:j}p_{i:j}1_{\mathcal{E}}(i, j)$

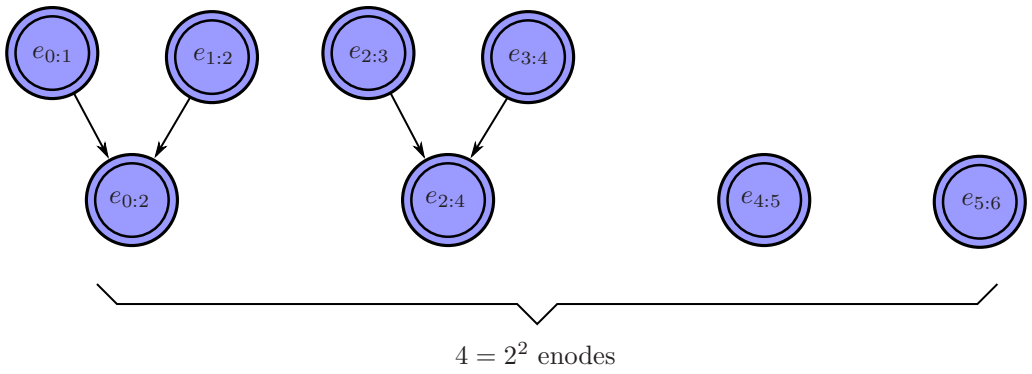


Figure 4-5: Illustration of Step 3 for $N = 6$. In this case, $n = \lceil \log_2 6 \rceil - 1 = 2$. We choose $N - 2^n = 2$ pairs of enodes. In this figure, the two chosen pairs are: $e_{0:1}$ and $e_{1:2}$; $e_{2:3}$ and $e_{3:4}$. Entanglement swapping between $e_{0:1}$ and $e_{1:2}$ results in the enode $e_{0:2}$; entanglement swapping between $e_{2:3}$ and $e_{3:4}$ results in the enode $e_{2:4}$. We now have a chain consisting of $2^2 = 4$ segments, separated by the repeaters 0, 2, 4, 5, and 6. These five repeaters are relabelled as $\overline{0}$, $\overline{1}$, $\overline{2}$, $\overline{3}$, and $\overline{4}$.

in $\mathscr{P}_{\mathrm{s}}$. These EQPs are generated directly from quantum channels instead of entanglement swapping. In the rest of the section, these EQPs are referred to as "crude entanglements." We now consider a homogeneous repeater chain with infinite quantum channels. For this chain, let $h(L)$ denote the minimum expected number of crude entanglements required to distribute one EQP shared between nodes that are connected by L quantum channels. The upper bound for the optimal value of $\mathscr{P}_{\mathrm{s}}$ relies on the lower bound for $h(L)$. The next proposition provides a lower bound for $h(L)$.

Proposition 4.3. For $N \in \mathbb{N}_+$, the minimum expected number of crude entanglements required to distribute one EQP shared between nodes that are connected by N quantum channels, denoted by $h(N)$, is lower bounded as

$$g(N) \leqslant h(N) \tag{4.17}$$

where $g(N)$ is defined in (4.16).

Proof. See Appendix C.5. □

Theorem 4.3. For homogeneous repeater chains with N quantum channels, if $N \in$

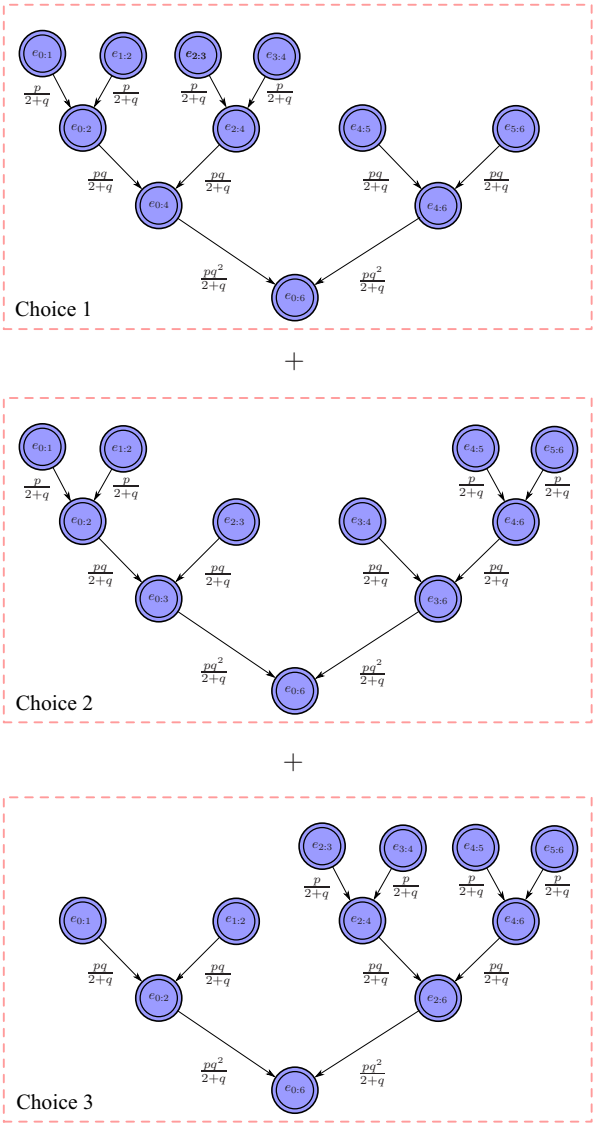


Figure 4-6: An illustration of the solution for $N = 6$.

$\mathbb{Z}_{\mathrm{e}}$, the maximum EDR is

$$\lambda^* = \frac{Npq^{n+1}}{2(N-2^n)+q(2^{n+1}-N)}$$

where $n = \lceil \log_2 N \rceil - 1$.

Proof. See Appendix C.7. □

4.3.3 Scenarios with an Odd N

We consider the scenario where N is odd, i.e., $N \in \mathbb{Z}_{\mathrm{o}}$. We first present a solution of $\mathscr{P}_{\mathrm{s}}$, denoted by $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i,j) \in \mathcal{E}\}$, and then prove that it is the optimal solution of $\mathscr{P}_{\mathrm{s}}$.

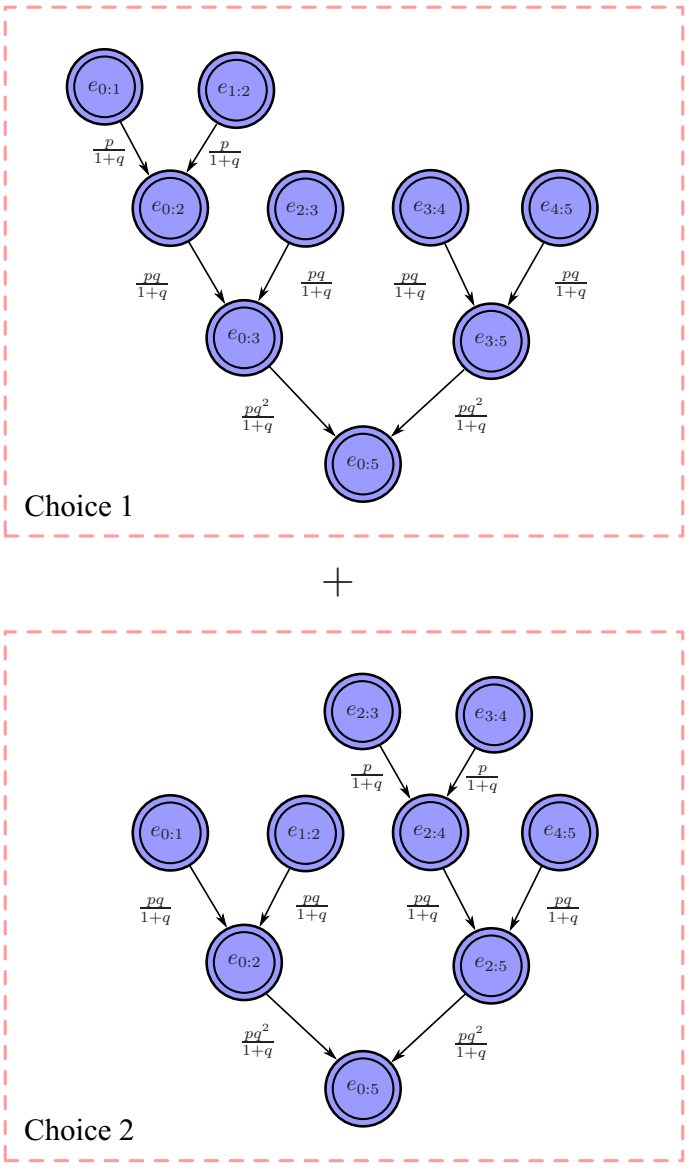


Figure 4-7: An illustration of the solution for $N = 5$. The quantities near the edges represent the eflow.

Let $n = \lceil \log_2 N \rceil - 1$. The solution $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i,j) \in \mathcal{E}\}$ can be obtained in six steps:

1. $\mathring{u}_{k:k+1} = 1$, $k \in \mathcal{K}_{0:N-2}$ and

$$\mathring{u}_{N-1:N} = q\frac{(N-1)}{2N - 2^{n+1} + q(2^{n+1} - N - 1)}$$

2. We consider the enodes: $e_{0:1}$, $e_{1:2}$, ..., $e_{N-2:N-1}$ and divide them into $(N-1)/2$ pairs of enodes: $e_{0:1}$ and $e_{1:2}$, $e_{2:3}$ and $e_{3:4}$, ..., $e_{N-3:N-2}$ and $e_{N-2:N-1}$;

3. There are $\binom{(N-1)/2}{N-2^n}$ possible ways to choose $(N-2^n)$ out of $(N-1)/2$ enode pairs. These choices are labeled as 1, 2, ..., $\binom{(N-1)/2}{N-2^n}$;

4. For the choice l, we have $(N-2^n)$ enode pairs chosen. Entanglement swapping is performed between two enodes in each chosen pair, resulting in an enode that is the common child of these two enodes; now we have a chain consisting of 2^n quantum channels and we can relabel the nodes in the chain as $\overline{0}$, $\overline{1}$, ..., $\overline{2^n}$;

5. For the newly labeled chain, we determine the following eflow:

$$\begin{aligned}
&\mathring{f}^{\overline{2^k a}:\overline{2^k a+2^{k-1}}}_{\overline{2^k a}:\overline{2^k a+2^k}}(l) = \mathring{f}^{\overline{2^k a+2^{k-1}}:\overline{2^k a+2^k}}_{\overline{2^k a}:\overline{2^k a+2^k}}(l) \\
&= \frac{1}{\binom{(N-1)/2}{N-2^n}} \frac{N-1}{2N - 2^{n+1} + q(2^{n+1} - N - 1)} pq^k, \\
&\qquad a = 0, 1, 2, \ldots, 2^{n-k} - 1, \quad k = 1, 2, \ldots, n+1
\end{aligned}$$

6. We repeat Steps 4 and 5 until all $\binom{(N-1)/2}{N-2^n}$ combinations are iterated. We add all the eflows to obtain the optimal solution.

In Step 3, we transfer the problem from the scenario with $N \in \mathbb{Z}_{\text{o}}$ to the scenario where N is a power of two. The method of determining $\{\mathring{f}^{i:k}_{i:j} : i, j, k \in \mathcal{N}\}$ is illustrated in Fig. 4-7. In this case, $n = \lceil \log_2 5 \rceil - 1 = 2$. We divide four enodes to two pairs: $e_{0:1}$ and $e_{1:2}$; as well as $e_{2:3}$ and $e_{3:4}$. There are two possible ways of choosing $N - 2^n = 1$ out of $(N-1)/2 = 2$ pairs. The choices are labeled as 1 and 2, and are shown in two red rectangles. For example in the first rectangle, the chosen pair is: $e_{0:1}$ and $e_{1:2}$. Entanglement swapping between $e_{0:1}$ and $e_{1:2}$ results in the enode $e_{0:2}$. We now have a chain consisting of $2^2 = 4$ segments, separated by the repeaters 0, 2, 3, 4, and

5. These five repeaters are relabelled as $\overline{0}$, $\overline{1}$, $\overline{2}$, $\overline{3}$, and $\overline{4}$ and we can perform the Steps 4 and 5. After we perform the Steps 4 and 5 in each rectangle, the eflow in two rectangles can then be added to obtain the optimal solution of $\mathscr{P}_{\rm s}$.

Remark 7. The feasibility of the $\{\mathring{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i, j) \in \mathcal{E}\}$ as a solution of $\mathscr{P}_{\rm s}$ can be verified by checking conditions (4.5)-(4.8). The value of the objective function corresponding to the solution is

$$\mathring{I}(s,t) = \frac{N-1}{2N - 2^{n+1} + q(2^{n+1} - N - 1)} pq^{n+1}.$$

We next determine an upper bound for the objective function of $\mathscr{P}_{\rm s}$ for $N \in \mathbb{Z}_{\rm o}$. If this upper bound coincides with $\mathring{I}(s,t)$, then the optimality of $\{\mathring{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i, j) \in \mathcal{E}\}$ will be proved.

Recall that the EQPs generated in Phase I are referred to as "crude entanglements." We now refer to the crude entanglements $\boldsymbol{\Xi}_{2k:2k+1}$, $k \in \mathbb{Z}$ as odd crude entanglements; analogously, we refer to the crude entanglements $\boldsymbol{\Xi}_{2k-1:2k}$, $k \in \mathbb{Z}$ as even crude entanglements. We now consider a homogeneous repeater chain with infinite quantum channels. For this chain, let $h_{\rm o}(L)$ and $h_{\rm e}(L)$ denote the minimum expected number of odd crude entanglements and even crude entanglements required to distribute one entangled pair shared between nodes that are connected by L quantum channels. The upper bound for the optimal value of $\mathscr{P}_{\rm s}$ relies on the lower bounds for $h_{\rm o}(L)$ and $h_{\rm e}(L)$. The next proposition provides lower bounds for $h_{\rm o}(L)$ and $h_{\rm e}(L)$.

Proposition 4.4. For $N \in \mathbb{N}_+$, the minimum expected number of odd crude entanglements and even crude entanglements required to distribute one entangled pair shared between nodes that are connected by N quantum channels, denoted by $h_{\rm o}(N)$ and $h_{\rm e}(N)$, respectively, satisfy the following conditions: $h_{\rm o}(1) = 1$, $h_{\rm e}(1) = 0$, if $N > 1$,

$$g_{\rm o}(N) \leqslant h_{\rm o}(N) \text{ and } g_{\rm e}(N) \leqslant h_{\rm e}(N)$$

where $g_{\mathrm{o}}(\cdot)$ and $g_{\mathrm{e}}(\cdot)$ are functions defined as

$$g_{\mathrm{o}}(L) = \begin{cases} \dfrac{2(L-2^l)+(2^{l+1}-L-1)q}{2q^{l+1}} + \dfrac{1}{q^l} & \text{if } L \in \mathbb{Z}_{\mathrm{o}} \\ \dfrac{2(L-2^l)+(2^{l+1}-L)q}{2q^{l+1}} & \text{if } L \in \mathbb{Z}_{\mathrm{e}} \end{cases}$$

$$g_{\mathrm{e}}(L) = \begin{cases} \dfrac{2(L-2^l)+(2^{l+1}-L-1)q}{2q^{l+1}} & \text{if } L \in \mathbb{Z}_{\mathrm{o}} \\ \dfrac{2(L-2^l)+(2^{l+1}-L)q}{2q^{l+1}} & \text{if } L \in \mathbb{Z}_{\mathrm{e}} \end{cases}$$

where $l = \lceil \log_2 L \rceil - 1$.

Proof. See Appendix C.8. □

Theorem 4.4. For homogeneous repeater chains with N quantum channels, if $N \in \mathbb{Z}_{\mathrm{o}}$, the maximum EDR is

$$\lambda^* = \frac{(N-1)pq^{n+1}}{2(N-2^n)+q(2^{n+1}-N-1)}$$

where $n = \lceil \log_2 N \rceil - 1$.

Proof. See Appendix C.10. □

4.4 Numerical Results

This section illustrates the performance of the proposed RED protocols through numerical examples.

4.4.1 General Networks

We evaluate the maximum EDR for the following general networks. Consider a region of 60×60 km^2. The nodes are deployed in this region according to a Poisson process. Let $\mathcal{N}$ denote an instantiation of node deployment. For $i, j \in \mathcal{N}$, $(i, j) \in \mathcal{E}$ if the distance between i and j, $D_{i:j}$, is less than 30 km; the parameter $p_{i:j}$ for an edge (i, j)

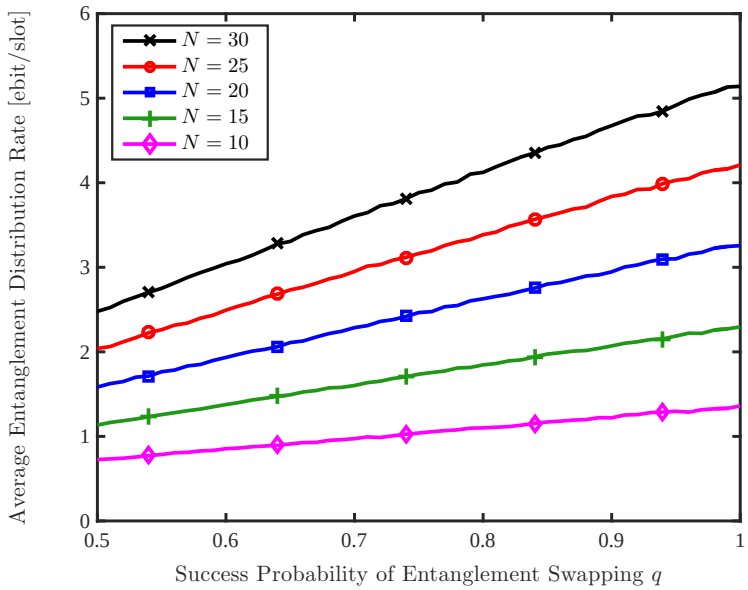


Figure 4-8: The average EDR in a random network.

is

$$p_{i:j} = 10^{-\gamma D_{i:j}/10}$$

where $\gamma = 0.2$ dB/km is the loss rate [107]. The pair of source node and sink node is randomly selected with equal probability among the nodes in the network.

Consider the performance metric as the average EDR, i.e., the empirical mean of the EDR achieved by solving $\mathscr{P}_{\mathrm{s}}$ averaging over instantiations of node deployments. Fig. 4-8 shows the average EDR as a function of q for different values of the average node number N.[5] Theorem 4.2 is used to generate results in Fig. 4-8. First, the average EDR increases with N. For example, when $q = 0.8$, the average EDR is 1.10 ebit/slot for $N = 10$, whereas it is 4.13 ebit/slot for $N = 30$. This corresponds to an increase of 2.75 times. This is because more nodes and more edges can provide more crude entanglements for distributing the target EQP $\boldsymbol{\Xi}_{s:t}$. Second, the average EDR increases with q. For example, when $N = 20$, the average EDR is 1.59 ebit/slot for $q = 0.5$, whereas it is 3.26 ebit/slot for $q = 1.0$. This corresponds to an increase

[5]Here, the amount of entanglement is quantified by the bit of entanglement (ebit), for example, one ebit corresponds to one entangled qubit pair.

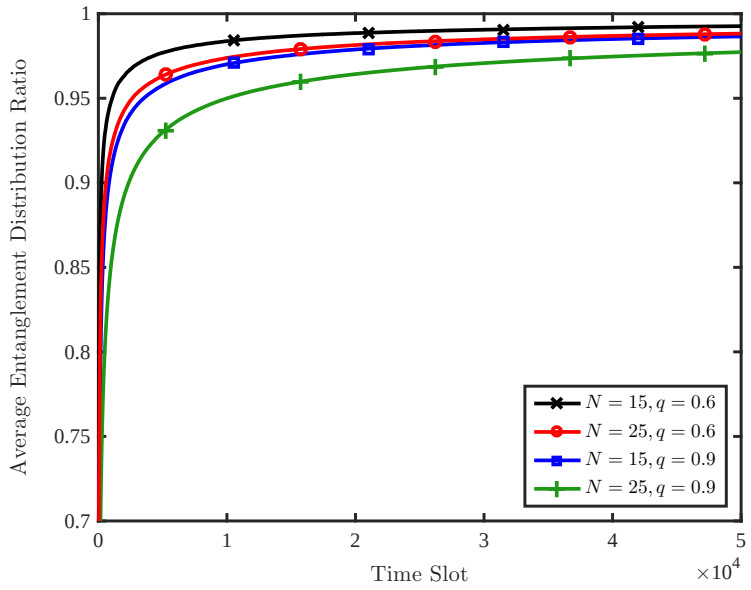


Figure 4-9: The average entanglement distribution ratio in a random network.

of 1.05 times. This observation agrees with the intuition, since unsuccessful entanglement swapping wastes the entanglements and larger q implies less unsuccessful entanglement swapping.

We showed in Section 4.2.3 that the stationary protocol approaches the maximum EDR for large time slot T. To characterize the behavior of the stationary protocol as a function of T, we consider the average entanglement distribution ratio at the time slot T, i.e., the empirical mean of the following random variable $\boldsymbol{\eta}_T$:

$$\boldsymbol{\eta}_T := \frac{\sum_{\tau=1}^{T} \mathbf{g}_{s:t}^{\mathring{\pi}}(\tau)}{T\,\boldsymbol{\lambda}^*}$$

where $\mathbf{g}_{s:t}^{\mathring{\pi}}(\tau)$ is the number of EQPs at the time slot τ distributed by the stationary protocol $\mathring{\pi}$, and $\boldsymbol{\lambda}^*$ is the optimal value of $\mathscr{P}_{\mathrm{s}}$. The randomness of $\boldsymbol{\eta}_T$ originates from node deployment, parameters $p_{i:j}$, $(i,j) \in \mathcal{E}$, and the probabilistic nature of the protocol. Fig. 4-9 shows the average entanglement distribution ratio as a function of T for different values of N and q. First, the average entanglement distribution ratio converges to one for different values of N and q. This verifies that the proposed stationary protocol achieves the maximal EDR. Second, the convergence speed de-

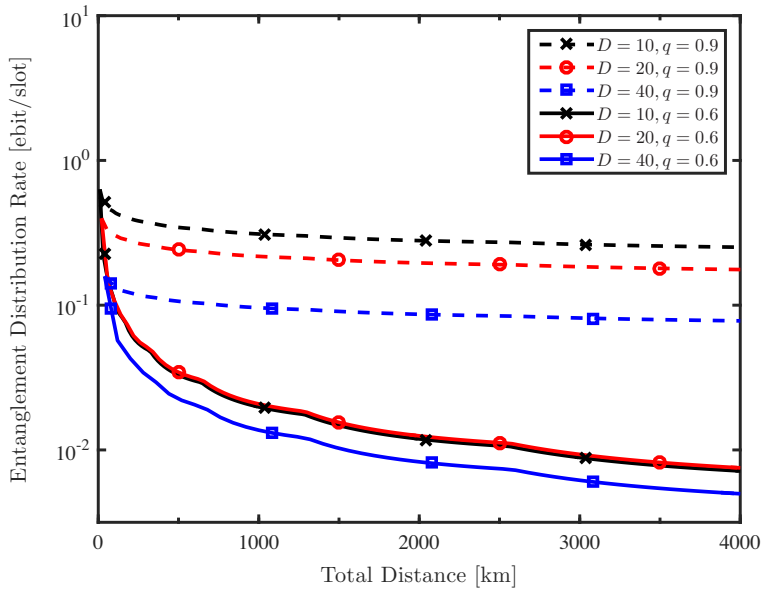


Figure 4-10: The EDR in a repeater chain with $\gamma = 0.2$ (dB/km).

creases with N. For example, when $q = 0.6$ and $T = 3 \cdot 10^4$, the average entanglement distribution ratio is 0.9848 for $N = 25$, whereas it is 0.9905 for $N = 15$. This is because more edges implies that there are more short paths from the source to the sink, and short paths require less entanglement swapping, so that the convergence speed can be increased.

4.4.2 Homogeneous Repeater Chains

We evaluate the maximal EDR in homogeneous repeater chains. Let D (km) denote the distance of a quantum channel in the repeater chain. Then the total distance L between the node 0 and the node N is $L = D \cdot N$. Let the loss rate of the channel be γ (dB/km). The parameter $p_{i:j}$ for an edge (i, j) is determined by the distance D and the loss rate γ [107]. In particular,

$$p_{a:b} = 10^{-\gamma D/10}.$$

The success probability of entanglement swapping in the network is assumed to be the same for all nodes and denoted by q. In this subsection, Claim 1 is used to generate

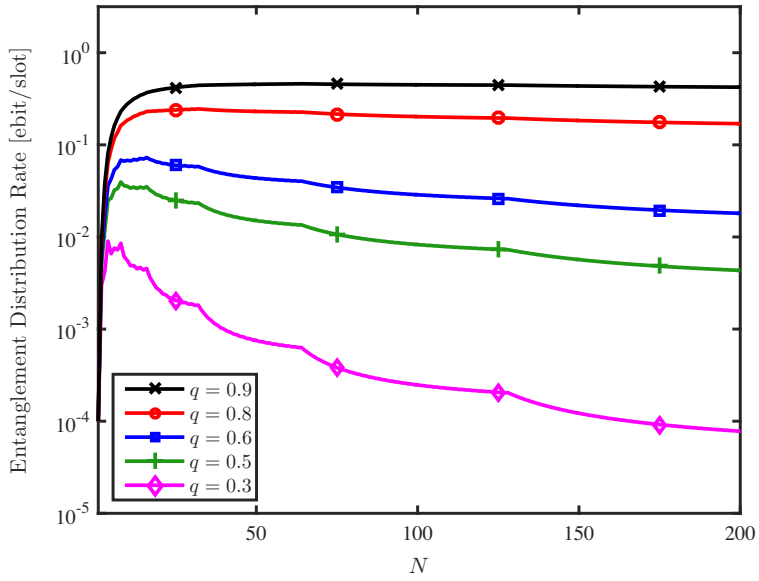


Figure 4-11: The EDR in a repeater chain with N segments. The total distance $L = 200$ km is fixed.

results in the figures.

Fig. 4-10 shows the EDR as a function of the total distance L for $\gamma = 0.2$ dB/km. First, the EDR increases with q. This observation is consistent with the results in general networks. For example, when $D = 20$ km and $L = 1500$ km, the EDR is 0.016 ebit/slot for $q = 0.6$, whereas it is 0.205 ebit/slot for $q = 0.9$. This corresponds to an increase of 12.212 times. Second, the EDR decreases with L and the rate decreases faster with smaller q. For example, for $D = 10$, the EDR decreases 1.18 times as the total distance L increases from 1000 km to 3000 km when $q = 0.9$, whereas it decreases 2.24 times when $q = 0.6$.

We next consider the scenario where the total distance L between the source and the sink is fixed. Fig. 4-11 shows the EDR as a function of N for $L = 200$ km and $\gamma = 0.2$ dB/km. A key observation is that the EDR first increases dramatically and then decreases slowly as a function of N. For example, when $q = 0.6$, the EDR is 10^{-4} ebit/slot, 0.0627 ebit/slot, and 0.0288 ebit/slot for $N = 1$, $N = 21$, and $N = 100$, respectively. This corresponds to an increase of 626 times from $N = 1$ to $N = 21$, and a decrease of 54.1% from $N = 21$ to $N = 100$. The EDR first increases with

N because the two neighboring quantum repeaters are far apart for a small N, and increasing N can significantly decrease D; the EDR then decreases with N because for a large N, increasing N does not significantly reduce D, and the imperfectness of quantum repeaters becomes the bottleneck of EDR. Therefore, more repeaters may not necessarily increase the EDR. The results in Fig. 4-11 provide guidance to the choice of repeater density in a chain and can offer insights into the design and implementation of general quantum networks.

Chapter 5

Concluding Remarks

This thesis studies the design and operation of quantum networks in the following three fields: state transmission, queueing delay, and remote entanglement distribution.

In Chapter 2, we develop protocols to broadcast quantum states from a transmitter to N different receivers in two scenarios, namely, non-oblivious with a known state and oblivious with multiple copies. To characterize the resources in this broadcast setting, we introduce two types of resources, broadcast cbits (bcbits) and broadcast qubits (bqubits). The protocols lead to a resource tradeoff inequality in the setting of quantum broadcasting: $\log N$ ebit $+ O(\log N)$ bcbits $\rightarrow 1$ bqubit and reduced the required cbits from $O(N)$ to $O(\log N)$ compared to sequentially sending quantum states. Moreover, we prove that $\Omega(\log N)$ bcbits are necessary for sending 1 bqubit to N receivers using shared entanglement. We also build circuits composed of single-quantum bit (qubit) gates and CNOT gates to show that the developed protocols can be implemented efficiently. As for future work, one may ask if the number of required bcbit can be further reduced, for example, from $2 \log N$ to $\log N$ for the non-oblivious scenario.

In Chapter 3, we introduce a formalism to analyze quantum queuing delay (QQD). The main methodologies used are dynamic programming and stochastic processes. We develop a cognitive-memory-based policy and show that this policy is optimal for the scenario with two receivers. For a general scenario, we derive an upper bound

for the expected average queue length achieved by the cognitive-memory-based policy. With this upper bound, we show that the expected average queuing delay can decrease exponentially with respect to the memory size. The performance of the proposed policies are evaluated in practical scenarios. Numerical results verify the optimality/near-optimality of the policies and show that the developed upper bound is tight. A key insight is that, unlike classical queuing delay, QQD can be significantly reduced by establishing entanglements before the quantum data arrive. The exponential decrease of QQD with respect to the memory size shows that a moderate memory size suffices to achieve a near-zero QQD.

In Chapter 4, we establish a framework of designing remote entanglement distribution (RED) protocols for quantum networks. We develop the optimal RED protocols for quantum networks based on the solutions of the linear programming problem. Moreover, we determine the maximum entanglement distribution rate (EDR) in a closed form for homogeneous repeater chains. The new vision developed in this thesis is the introduction of enodes and eflows. We transform the RED problem into linear programming and employ concepts and methods from the graph theory and classical flow networks. The performance of the proposed protocols is evaluated by simulation. The results for homogeneous repeater chains demonstrate that the issue of significant decay of communication capacity can be essentially solved by properly deploying quantum repeaters, even if the quantum repeaters are imperfect. Our results enable the distribution of entanglement over long distances and provide insights into the design and implementation of quantum networks. We hope that our results may incite some future work. For example, one may be interested in designing a protocol that converges to the maximum EDR faster than the protocol proposed in this thesis. It may also be worth investigating quantum networks in addition to the homogeneous repeater chains and trying to determine a closed-form maximum EDR. It is also interesting to see how to extend the results in this thesis to other models. For example, the parameters $p_{i:j}$ and q_k may vary over time; moreover, the entanglement generated or distributed may not be perfect. One may wonder how to determine the maximum EDR and how to obtain optimal protocol in these scenarios.

Appendix A

Proofs of the Results in Chapter 2

A.1 Proof of Theorem 2.1

Equation (2.3) is equivalent to the following claim: for every $|\varphi\rangle_{\mathrm{s}} \in \mathcal{H}$ and $|\psi\rangle \in \mathcal{H}_2$,

$$\sum_{k=1}^{K} \frac{N+1}{K} \operatorname{tr}\Big\{ \boldsymbol{\Xi}\left(|\varphi\rangle_{\mathrm{s}}\right) \boldsymbol{V}_k \, \boldsymbol{\Xi}\left(|\psi\rangle^{\otimes N}\right) \, \boldsymbol{V}_k^{\dagger} \Big\} \leqslant 1+\epsilon. \tag{A.1}$$

We next prove the claim above. For $\boldsymbol{V}_{\mathrm{s}} = \boldsymbol{V}^{\otimes N}$, where $\boldsymbol{V}$ is a unitary operator on $\mathcal{H}_2$, we have

$$\begin{aligned} &\operatorname{tr}\Big\{ \boldsymbol{V}_{\mathrm{s}} \boldsymbol{\Xi}\left(|\psi\rangle^{\otimes N}\right) \boldsymbol{V}_{\mathrm{s}}^{\dagger} \boldsymbol{\Xi}\left(|\varphi\rangle_{\mathrm{s}}\right) \Big\} \\ &= \operatorname{tr}\Big\{ \left(\boldsymbol{V}|\psi\rangle\right)^{\otimes N} \left(\langle\psi|\boldsymbol{V}^{\dagger}\right)^{\otimes N} |\varphi\rangle_{\mathrm{s}} \, {}_{\mathrm{s}}\langle\varphi| \Big\} \\ &= \Big| {}_{\mathrm{s}}\langle\varphi| \left(\boldsymbol{V}|\psi\rangle\right)^{\otimes N} \Big|^2 . \end{aligned} \tag{A.2}$$

Consider the random variable

$$\mathsf{x} = (N+1) \Big| {}_{\mathrm{s}}\langle\varphi| \left(\mathbf{V}|\psi\rangle\right)^{\otimes N} \Big|^2$$

where $\mathbf{V}$ is drawn from the Haar measure on the unitary group. We then bound the probability of the deviation for the sample mean of x as follows.

Lemma A.1. Let $\mathsf{x}_1, \mathsf{x}_2, \ldots, \mathsf{x}_K$ denote K i.i.d. samples of x, then for $\epsilon \in (0,1)$,

$$\mathbb{P}\Big\{ \frac{1}{K} \sum_{k=1}^{K} \mathsf{x}_k > 1 + \epsilon \Big\} \leqslant \exp\Big\{ - \frac{K\epsilon^2}{4(N+1)} \Big\}.$$

Proof. We first prove that $\mathbb{E}\{\mathsf{x}\} = 1$. Rewrite $\mathbf{V}\,|\psi\rangle = \mathsf{u}\,|0\rangle + \mathsf{v}\,|1\rangle$ and $|\varphi\rangle = \sum_{j=0}^{N} w_j\,|j\rangle_{\mathrm{s}}$. Then

$$\begin{aligned}
\mathbb{E}\{\mathsf{x}\} &= (N+1)\mathbb{E}\bigg\{ \bigg| \sum_{j=0}^{N} w_j \mathsf{u}^j \mathsf{v}^{N-j} \binom{N}{j}^{1/2} \bigg|^2 \bigg\} \\
&= (N+1)\mathbb{E}\bigg\{ \sum_{j=0}^{N} |w_j|^2 |\mathsf{u}|^{2j} |\mathsf{v}|^{2(N-j)} \binom{N}{j} \bigg\} \\
&= (N+1) \sum_{j=0}^{N} |w_j|^2 \frac{j!(N-j)!}{(N+1)!} \frac{N!}{j!(N-j)!} = 1
\end{aligned} \tag{A.3}$$

where the second equality is due to $\mathbb{E}\{\mathsf{u}^j (\mathsf{u}^*)^{j'} \mathsf{v}^k (\mathsf{v}^*)^{k'}\} = 0$ unless $j = j'$ and $k = k'$; and the third equality is because the joint probability density function (PDF) of $|\mathsf{u}|^2$ and $|\mathsf{v}|^2$ follows a Dirichlet distribution of order 2 with parameter $(1,1)$.

We then bound the moments of x as follows:

$$\begin{aligned}
\mathbb{E}\{\mathsf{x}^k\} &= (N+1)^k \mathbb{E}\Big\{ \big|\, {}_{\mathrm{s}}\langle\varphi|\, (\mathbf{V}\,|\psi\rangle)^{\otimes N} \big|^{2k} \Big\} \\
&\leqslant (N+1)^k \mathbb{E}\Big\{ \big|\, {}_{\mathrm{s}}\langle\varphi|\, (\mathbf{V}\,|\psi\rangle)^{\otimes N} \big|^{2} \Big\} \\
&= (N+1)^{k-1}
\end{aligned} \tag{A.4}$$

where the inequality is because $\big|\, {}_{\mathrm{s}}\langle\varphi|\, (\mathbf{V}\,|\psi\rangle)^{\otimes N} \big| \leqslant 1$, and the last equality is because of (A.3).

Let $\Lambda_{\mathsf{x}}(\cdot)$ denote that the Cramér function associated with x. We next show that for $\epsilon \in (0, \infty)$

$$\Lambda_{\mathsf{x}}(1+\epsilon) \geqslant \frac{\epsilon^2}{4(N+1)}. \tag{A.5}$$

By definition of the Cramér function, we have

$$\Lambda_{\mathsf{x}}(1+\epsilon) = \sup_{z\in\mathbb{R}} \big[z(1+\epsilon) - \ln \mathbb{E}\{e^{z\mathsf{x}}\}\big].$$

Let z take the value of $z^* = \epsilon/(2N+2)$. We have

$$\begin{aligned}
\Lambda_{\mathsf{x}}(1+\epsilon) &\geqslant z^*\cdot(1+\epsilon) - \ln \mathbb{E}\{e^{z^*\mathsf{x}}\} \\
&= z^*\cdot(1+\epsilon) - \ln\left(1+\sum_{k=1}^{\infty}\frac{(z^*)^k\mathbb{E}\{\mathsf{x}^k\}}{k!}\right) \\
&\geqslant z^*\cdot(1+\epsilon) - \sum_{k=1}^{\infty}\frac{(z^*)^k\mathbb{E}\{\mathsf{x}^k\}}{k!} \\
&\geqslant \frac{\epsilon(1+\epsilon)}{2N+2} - \sum_{k=1}^{\infty}\frac{\epsilon^k}{2^k k!(N+1)} \\
&= \frac{\epsilon^2}{2N+2} - \sum_{k=2}^{\infty}\frac{\epsilon^k}{2^k k!(N+1)} \\
&\geqslant \frac{\epsilon^2}{2N+2} - \sum_{k=2}^{\infty}\frac{\epsilon^2}{2\cdot 2^k(N+1)} \\
&= \frac{\epsilon^2}{4(N+1)}
\end{aligned}$$

where the first inequality is because the Cramér function is a superior for all possible $z\in\mathbb{R}$, the second inequality is because $\ln(1+\alpha)\leqslant\alpha$ for $\alpha>0$, the third inequality is because of (A.4), and the last inequality is because $k!\geqslant 2$ for $k\geqslant 2$. This proves (A.5).

According to Cramér's theorem [108], for i.i.d. real random variables x_1, x_2, …, x_K,

$$\mathbb{P}\bigg\{\frac{1}{K}\sum_{k=1}^{K}\mathsf{x}_k \geqslant 1+\epsilon\bigg\} \leqslant \exp\Big\{-K\inf_{a\geqslant 1+\epsilon}\Lambda_{\mathsf{x}}(a)\Big\} \leqslant \exp\Big\{-\frac{K\epsilon^2}{4(N+1)}\Big\}$$

where the second inequality is because of (A.5). □

Lemma A.1 implies that for any given $|\varphi\rangle_{\mathrm{s}} \in \mathcal{H}$ and $|\psi\rangle \in \mathcal{H}_2$,

$$\mathbb{P}\Big\{\frac{1}{K}\sum_{k=1}^{K}\mathrm{tr}\Big\{\mathbf{V}_{\mathrm{s},k}\boldsymbol{\Xi}\left(|\psi\rangle^{\otimes N}\right)\mathbf{V}_{\mathrm{s},k}^{\dagger}\boldsymbol{\Xi}\left(|\varphi\rangle_{\mathrm{s}}\right)\Big\} \geqslant \frac{2+\epsilon}{2(N+1)}\Big\} \leqslant \exp\Big\{-\frac{K\epsilon^2}{16(N+1)}\Big\} \tag{A.6}$$

where $\mathbf{V}_{\mathrm{s},k} = \mathbf{V}_k^{\otimes N}$ and the $\mathbf{V}_k$'s are i.i.d drawn from the Haar measure on the unitary group.

Lemma A.2. For every $\epsilon > 0$, there exists a subset of $\mathcal{H}_2$, denoted as $\mathcal{M}$, of cardinality

$$|\mathcal{M}| \leqslant \left(\frac{5N}{\epsilon}\right)^4$$

and for every state vector $|\phi\rangle \in \mathcal{H}_2$, there exists a state vector $|\widetilde{\phi}\rangle \in \mathcal{M}$ such that

$$\big\|\boldsymbol{\Xi}\left(|\phi\rangle^{\otimes N}\right) - \boldsymbol{\Xi}\left(|\widetilde{\phi}\rangle^{\otimes N}\right)\big\|_1 \leqslant 2\big\||\phi\rangle^{\otimes N} - |\widetilde{\phi}\rangle^{\otimes N}\big\|_2 \leqslant \epsilon. \tag{A.7}$$

Proof. By Lemma 4 in [31], there exists a set $\mathcal{M} \subseteq \mathcal{H}_2$ with

$$|\mathcal{M}| \leqslant \left(\frac{5N}{\epsilon}\right)^4$$

and for every state $|\phi\rangle \in \mathcal{H}_2$ there exists a state $|\widetilde{\phi}\rangle \in \mathcal{M}$ such that

$$\Big\||\phi\rangle - |\widetilde{\phi}\rangle\Big\|_2 \leqslant \frac{\epsilon}{2N}.$$

We next prove $\mathcal{M}$ constructed in this way satisfies the condition (A.7). In fact, we apply Lemma A.3 below repeatedly N times to obtain $\big\||\phi\rangle^{\otimes N} - |\widetilde{\phi}\rangle^{\otimes N}\big\|_2 \leqslant \epsilon/2$. □

Lemma A.3. Consider two arbitrary Hilbert spaces $\mathcal{H}$ and $\widetilde{\mathcal{H}}$. For any state vectors $|\psi_1\rangle, |\varphi_1\rangle \in \mathcal{H}$, $|\psi_2\rangle, |\varphi_2\rangle \in \widetilde{\mathcal{H}}$, $\| |\psi_1\rangle \otimes |\psi_2\rangle - |\varphi_1\rangle \otimes |\varphi_2\rangle \|_2 \leqslant \| |\psi_1\rangle - |\varphi_1\rangle \|_2 + \| |\psi_2\rangle - |\varphi_2\rangle \|_2$.

Proof.

$$\begin{aligned}
&\| \, |\psi_1\rangle \otimes |\psi_2\rangle - |\varphi_1\rangle \otimes |\varphi_2\rangle \, \|_2 \\
&= \| |\psi_1\rangle \otimes |\psi_2\rangle - |\varphi_1\rangle \otimes |\psi_2\rangle + |\varphi_1\rangle \otimes |\psi_2\rangle - |\varphi_1\rangle \otimes |\varphi_2\rangle \|_2 \\
&\leqslant \| (|\psi_1\rangle - |\varphi_1\rangle) \otimes |\psi_2\rangle \, \|_2 + \| \, |\varphi_1\rangle \otimes (|\psi_2\rangle - |\varphi_2\rangle) \|_2 \\
&= \| \, |\psi_1\rangle - |\varphi_1\rangle \, \|_2 + \| \, |\psi_2\rangle - |\varphi_2\rangle \, \|_2
\end{aligned}$$

where the inequality is due to triangular inequality, and the last equality is due to the fact that $|\psi_2\rangle$ and $|\varphi_1\rangle$ are unit vectors. □

Lemma A.2 shows that we can find a set $\mathcal{M}_1 \subseteq \mathcal{H}_2$ with $|\mathcal{M}_1| \leqslant \big(20N(N+1)/\epsilon\big)^4$ and for every pure state $|\psi\rangle \in \mathcal{H}_2$ there exists a state $|\widetilde{\psi}\rangle \in \mathcal{M}_1$ such that

$$\left\| \boldsymbol{\Xi} \left(|\psi\rangle^{\otimes N} \right) - \boldsymbol{\Xi} \left(|\widetilde{\psi}\rangle^{\otimes N} \right) \right\|_1 \leqslant 2 \left\| \, |\psi\rangle^{\otimes N} - \, |\widetilde{\psi}\rangle^{\otimes N} \, \right\|_2 \leqslant \frac{\epsilon}{4(N+1)}. \tag{A.8}$$

Moreover, by Lemma 4 in [31], we can find a set $\mathcal{M}_2 \subseteq \mathcal{H}$ with $|\mathcal{M}_2| \leqslant \big(20(N+1)/\epsilon\big)^{2(N+1)}$ and for every pure state $|\varphi\rangle_{\mathrm{s}} \in \mathcal{H}$, there exists a state $|\widetilde{\varphi}\rangle_{\mathrm{s}} \in \mathcal{M}_2$ such that

$$\left\| \boldsymbol{\Xi} \left(|\varphi\rangle_{\mathrm{s}} \right) - \boldsymbol{\Xi} \left(|\widetilde{\varphi}\rangle \right) \right\|_1 \leqslant 2 \left\| \, |\varphi\rangle - |\widetilde{\varphi}\rangle \, \right\|_2 \leqslant \frac{\epsilon}{4(N+1)}. \tag{A.9}$$

Using (A.6) and the union bound, we have

$$\begin{aligned}
&\mathbb{P}\left\{ \exists \, |\widetilde{\psi}\rangle \in \mathcal{M}_1, |\widetilde{\varphi}\rangle_{\mathrm{s}} \in \mathcal{M}_2, \frac{1}{K} \sum_{k=1}^{K} \mathrm{tr}\Big\{ \mathbf{V}_{\mathrm{s},k} \boldsymbol{\Xi} \left(|\widetilde{\psi}\rangle^{\otimes N} \right) \mathbf{V}_{\mathrm{s},k}^{\dagger} \boldsymbol{\Xi} \left(|\widetilde{\varphi}\rangle_{\mathrm{s}} \right) \Big\} \geqslant \frac{2+\epsilon}{2(N+1)} \right\} \\
&\qquad \leqslant |\mathcal{M}_1| \cdot |\mathcal{M}_2| \cdot \exp\left\{ - \frac{K\epsilon^2}{16(N+1)} \right\} < 1
\end{aligned}$$

where the inequality is because $|\mathcal{M}_1| \leqslant \big(20N(N+1)/\epsilon\big)^4$, $|\mathcal{M}_2| \leqslant \big(20(N+1)/\epsilon\big)^{2(N+1)}$, and K is as large as stated in the theorem. Therefore, there exist $\boldsymbol{V}_{\mathrm{s},k}$, $k \in \mathcal{K}_{1:K}$ such

that for all $|\widetilde{\psi}_0\rangle \in \mathcal{M}_1$, $|\widetilde{\varphi}_0\rangle_{\rm s} \in \mathcal{M}_2$,

$$\frac{1}{K}\sum_{k=1}^{K} \mathrm{tr}\Big\{ \boldsymbol{V}_{{\rm s},k} \boldsymbol{\Xi}\left(|\widetilde{\psi}_0\rangle^{\otimes N} \right) \boldsymbol{V}_{{\rm s},k}^{\dagger} \boldsymbol{\Xi}\left(|\widetilde{\varphi}_0\rangle_{\rm s} \right) \Big\} < \frac{2+\epsilon}{2(N+1)}. \tag{A.10}$$

For any $|\psi\rangle \in \mathcal{H}_2$ and $|\varphi\rangle_{\rm s} \in \mathcal{H}$, we can find $|\widetilde{\psi}\rangle \in \mathcal{M}_1$ and $|\widetilde{\varphi}\rangle_{\rm s} \in \mathcal{M}_2$ such that (A.8) and (A.9) hold. Then

$$\begin{aligned}
&\frac{1}{K}\sum_{k=1}^{K} \mathrm{tr}\Big\{ \boldsymbol{V}_{{\rm s},k} \boldsymbol{\Xi}\left(|\psi\rangle^{\otimes N} \right) \boldsymbol{V}_{{\rm s},k}^{\dagger} \boldsymbol{\Xi}\left(|\varphi\rangle_{\rm s} \right) \Big\} \\
&\leqslant \frac{1}{K}\sum_{k=1}^{K} \left| \langle\psi|^{\otimes N} \boldsymbol{V}_{{\rm s},k}^{\dagger} \big[\boldsymbol{\Xi}\left(|\varphi\rangle_{\rm s} \right) - \boldsymbol{\Xi}\left(|\widetilde{\varphi}\rangle_{\rm s} \right) \big] \boldsymbol{V}_{{\rm s},k} |\psi\rangle^{\otimes N} \right| \\
&\quad + \frac{1}{K}\sum_{k=1}^{K} \left| {}_{\rm s}\langle\widetilde{\varphi}| \boldsymbol{V}_{{\rm s},k} \big[\boldsymbol{\Xi}\left(|\psi\rangle^{\otimes N} \right) - \boldsymbol{\Xi}\left(|\widetilde{\psi}\rangle^{\otimes N} \right) \big] \boldsymbol{V}_{{\rm s},k}^{\dagger} |\widetilde{\varphi}\rangle_{\rm s} \right| \\
&\quad + \frac{1}{K}\sum_{k=1}^{K} \mathrm{tr}\Big\{ \boldsymbol{V}_{{\rm s},k} \boldsymbol{\Xi}\left(|\widetilde{\psi}\rangle^{\otimes N} \right) \boldsymbol{V}_{{\rm s},k}^{\dagger} \boldsymbol{\Xi}\left(|\widetilde{\varphi}\rangle_{\rm s} \right) \Big\} \\
&\leqslant \frac{1}{K}\sum_{k=1}^{K} \Big\| \boldsymbol{\Xi}\left(|\varphi\rangle_{\rm s} \right) - \boldsymbol{\Xi}\left(|\widetilde{\varphi}\rangle_{\rm s} \right) \Big\|_1 + \frac{1}{K}\sum_{k=1}^{K} \Big\| \boldsymbol{\Xi}\left(|\psi\rangle^{\otimes N} \right) - \boldsymbol{\Xi}\left(|\widetilde{\psi}\rangle^{\otimes N} \right) \Big\|_1 + \frac{2+\epsilon}{2(N+1)} \\
&\leqslant \frac{\epsilon}{4(N+1)} + \frac{\epsilon}{4(N+1)} + \frac{2+\epsilon}{2(N+1)} = \frac{1+\epsilon}{N+1}
\end{aligned}$$

where the first inequality is due to repeated use of the triangular inequality, the second inequality is because $\|\cdot\|_1$ denotes the sum of the absolute values of eigenvalues and (A.10); and the last inequality is because of (A.8) and (A.9). This proves the claim (A.1).

A.2 Proof of Theorem 2.2

Equation (2.6) is equivalent to the following claim: for any $|\varphi\rangle \in \mathcal{H} \otimes \mathcal{H}$, there exist I unitary operators $\boldsymbol{U}_i$'s such that

$$\frac{1}{I}\sum_{i=1}^{I} \mathrm{tr}(\boldsymbol{\Xi}\left(|\varphi\rangle \right) |\xi_i\rangle \langle\xi_i|) \leqslant \frac{1+\epsilon}{(N+1)^2}. \tag{A.11}$$

We next prove the claim above.

Suppose $|\xi_i\rangle$ are drawn i.i.d. according to the random vector $|\boldsymbol{\xi}\rangle$, where

$$|\boldsymbol{\xi}\rangle = \frac{1}{\sqrt{N+1}}(\mathbf{U})^{\otimes N} \otimes \mathbb{1}^{\otimes N} \sum_{j=0}^{N} |j\rangle_{\mathrm{s}} |j\rangle_{\mathrm{s}}$$

in which $\mathbf{U}$ is distributed according to the Haar measure. Let $\mathsf{y} = (N+1)^2 \operatorname{tr}\{\boldsymbol{\Xi}(|\varphi\rangle)|\boldsymbol{\xi}\rangle\langle\boldsymbol{\xi}|\}$. In order to study the statistical property of y, we first present the following lemma, where the proof is given in Appendix A.3.

Lemma A.4. Suppose a unitary operator $\mathbf{U} \in \mathbb{C}^{2\times 2}$ is distributed according to the Haar measure, the following equality holds:

$$(N+1)\mathbb{E}\big\{\mathbf{U}^{\otimes N}|x_1\rangle\langle x_2|\mathbf{U}^{\dagger\otimes N}\big\} = \langle x_2|x_1\rangle \mathbb{1}_{\mathrm{s}}$$

where $|x_1\rangle$ and $|x_2\rangle$ are pure states in $\mathcal{H}_{\mathrm{s}}$.

With Lemma A.4, we can determine the mean of y as follows.

$$\begin{aligned}\mathbb{E}\{\mathsf{y}\} &= (N+1)\mathbb{E}\Big\{\operatorname{tr}\Big\{\boldsymbol{\Xi}(|\varphi\rangle)\mathbf{U}^{\otimes N}\otimes\mathbb{1}^{\otimes N}\sum_{j=0}^{N}\sum_{j'=0}^{N}|j\rangle_{\mathrm{s}}\,{}_{\mathrm{s}}\langle j'|\,|j\rangle_{\mathrm{s}}\,{}_{\mathrm{s}}\langle j'|\,(\mathbf{U})^{\dagger\otimes N}\otimes\mathbb{1}^{\otimes N}\Big\}\Big\} \\ &= \mathbb{E}\Big\{\operatorname{tr}\Big\{\boldsymbol{\Xi}(|\varphi\rangle)\sum_{j=0}^{N}\sum_{j'=0}^{N}\delta_{jj'}|j\rangle_{\mathrm{s}}\,{}_{\mathrm{s}}\langle j'|\Big\}\Big\} = 1. \end{aligned} \tag{A.12}$$

Furthermore, we can bound the probability of the deviation for the sample mean of y as follows.

Lemma A.5. Let $\mathsf{y}_1, \mathsf{y}_2, \ldots, \mathsf{y}_J$ denote J i.i.d. samples of y, then for $\epsilon \in (0,1)$,

$$\mathbb{P}\Big\{\frac{1}{J}\sum_{j=1}^{J}\mathsf{y}_j \geqslant 1+\epsilon\Big\} \leqslant \exp\Big\{-\frac{J\epsilon^2}{4(N+1)^2}\Big\}.$$

Proof. We first bound the moments of y as follows:

$$\mathbb{E}\{\mathsf{y}^k\} = (N+1)^{2k}\mathbb{E}\big\{|\langle\varphi|\boldsymbol{\xi}\rangle|^{2k}\big\} \leqslant (N+1)^{2k}\mathbb{E}\big\{|\langle\varphi|\boldsymbol{\xi}\rangle|^2\big\} = (N+1)^{2k-2}$$

where the first inequality is because $|\langle\varphi|\xi\rangle| \leqslant 1$, and the last equality is because of (A.12). The rest of the proof is analogous to that in Lemma A.1 by replacing x and $N+1$ with y and $(N+1)^2$, respectively. □

Lemma A.5 implies that for a pure state $|\varphi\rangle \in \mathcal{H}\otimes\mathcal{H}$,

$$\mathbb{P}\Big\{\Big|\frac{1}{I}\sum_{i=1}^{I}\mathrm{tr}\{\boldsymbol{\Xi}\left(|\varphi\rangle\right)|\xi_i\rangle\langle\xi_i|\} \geqslant \frac{2+\epsilon}{2(N+1)^2}\Big\} \leqslant \exp\Big\{-\frac{I\epsilon^2}{16(N+1)^2}\Big\}. \tag{A.13}$$

By Lemma 4 in [31], there exists a set $\mathcal{M}\subseteq\mathcal{H}\otimes\mathcal{H}$ with $|\mathcal{M}| \leqslant \big(10(N+1)^2/\epsilon\big)^{2(N+1)^2}$ and for every state $|\varphi\rangle\in\mathcal{H}\otimes\mathcal{H}$, there exists a state vector $|\widetilde{\varphi}\rangle\in\mathcal{M}$ such that

$$\big\|\boldsymbol{\Xi}\left(|\varphi\rangle\right)-\boldsymbol{\Xi}\left(|\widetilde{\varphi}\rangle\right)\big\|_1 \leqslant \frac{\epsilon}{2(N+1)^2}. \tag{A.14}$$

Using (A.13) and the union bound, we have

$$\begin{aligned}\mathbb{P}\Big\{\exists\,|\varphi\rangle\in\mathcal{M}, \frac{1}{I}\sum_{i=1}^{I}\mathrm{tr}\{\boldsymbol{\Xi}\left(|\varphi\rangle\right)|\xi_k\rangle\langle\xi_k|\} \geqslant \frac{2+\epsilon}{2(N+1)^2}\Big\} &\leqslant |\mathcal{M}|\cdot\exp\Big\{-\frac{I\epsilon^2}{16(N+1)^2}\Big\}\\ &< 1\end{aligned}$$

where the inequality is because $|\mathcal{M}| \leqslant \big(10(N+1)^2/\epsilon\big)^{2(N+1)^2}$ and I is as large as stated in the theorem. Therefore, there exist $|\boldsymbol{\xi}_i\rangle$, $i\in\mathcal{K}_{1:I}$ such that for all $|\widetilde{\varphi}_0\rangle\in\mathcal{M}$,

$$\frac{1}{I}\sum_{i=1}^{I}\mathrm{tr}\{\boldsymbol{\Xi}\left(|\widetilde{\varphi}\rangle_0\right)|\boldsymbol{\xi}_i\rangle\langle\boldsymbol{\xi}_i|\} < \frac{2+\epsilon}{2(N+1)^2}. \tag{A.15}$$

For any $|\varphi\rangle\in\mathcal{H}\otimes\mathcal{H}$, we can find $|\widetilde{\varphi}\rangle\in\mathcal{M}$ such that (A.14) holds. Then

$$\begin{aligned}&\frac{1}{I}\sum_{i=1}^{I}\mathrm{tr}\{\boldsymbol{\Xi}\left(|\varphi\rangle\right)|\boldsymbol{\xi}_i\rangle\langle\boldsymbol{\xi}_i|\}\\ &\leqslant \frac{1}{I}\sum_{i=1}^{I}\Big|\langle\boldsymbol{\xi}_i|\left[\boldsymbol{\Xi}\left(|\varphi\rangle\right)-\boldsymbol{\Xi}\left(|\widetilde{\varphi}\rangle\right)\right]|\boldsymbol{\xi}_i\rangle\Big| + \frac{1}{I}\sum_{i=1}^{I}\mathrm{tr}\{\boldsymbol{\Xi}\left(|\widetilde{\varphi}\rangle\right)|\boldsymbol{\xi}_i\rangle\langle\boldsymbol{\xi}_i|\}\\ &\leqslant \frac{1}{I}\sum_{i=1}^{I}\big\|\boldsymbol{\Xi}\left(|\varphi\rangle\right)-\boldsymbol{\Xi}\left(|\widetilde{\varphi}\rangle\right)\big\|_1 + \frac{2+\epsilon}{2(N+1)^2} \leqslant \frac{\epsilon}{2(N+1)^2}+\frac{2+\epsilon}{2(N+1)^2} = \frac{1+\epsilon}{(N+1)^2}\end{aligned}$$

where the first inequality is due to repeated use of the triangular inequality, the second inequality is because $\|\cdot\|_1$ denotes the sum of the absolute values of eigenvalues and (A.15); and the last inequality is because of (A.14). This proves the claim (A.11).

A.3 Proof of Lemma A.4

The proof can be divided into four steps. In particular, we will prove the following three claims sequentially.

Claim 1: $(N+1)\mathbb{E}\big\{\mathsf{U}^{\otimes N}\left|0\right\rangle^{\otimes N}\left\langle 0\right|^{\otimes N}(\mathsf{U})^{\dagger\otimes N}\big\} = \mathbb{1}_{\mathrm{s}}$.

To prove Claim 1, we rewrite $\mathsf{U}\left|0\right\rangle = \mathsf{u}\left|0\right\rangle + \mathsf{v}\left|1\right\rangle$. Then

$$
\begin{aligned}
&(N+1)\mathbb{E}\big\{\mathsf{U}^{\otimes N}\left|0\right\rangle^{\otimes N}\left\langle 0\right|^{\otimes N}(\mathsf{U})^{\dagger\otimes N}\big\} \\
&= (N+1)\mathbb{E}\big\{(\mathsf{u}\left|0\right\rangle + \mathsf{v}\left|1\right\rangle)^{\otimes N}(\mathsf{u}^*\left\langle 0\right| + \mathsf{v}^*\left\langle 1\right|)^{\otimes N}\big\} \\
&= (N+1)\mathbb{E}\Bigg\{\sum_{k=0}^{N}\mathsf{u}^k\mathsf{v}^{N-k}\sqrt{\frac{N!}{k!(N-k)!}}\left|k\right\rangle_{\mathrm{s}}\sum_{k=0}^{N}(\mathsf{u}^*)^k(\mathsf{v}^*)^{N-k}\sqrt{\frac{N!}{k!(N-k)!}}\left\langle k\right|_{\mathrm{s}}\Bigg\} \\
&= (N+1)\sum_{k=0}^{N}\sum_{k'=0}^{N}\mathbb{E}\{\mathsf{u}^k(\mathsf{u}^*)^{k'}\mathsf{v}^{N-k}(\mathsf{v}^*)^{(N-k')}\}\sqrt{\frac{N!}{k!(N-k)!}}\sqrt{\frac{N!}{k'!(N-k')!}}\left|k\right\rangle_{\mathrm{s}}\,{}_{\mathrm{s}}\left\langle k'\right| \\
&\overset{(a)}{=} (N+1)\sum_{k=1}^{N}\mathbb{E}\big\{|\mathsf{u}|^{2k}|\mathsf{v}|^{2(N-k)}\big\}\frac{N!}{k!(N-k)!}\left|k\right\rangle_{\mathrm{s}}\,{}_{\mathrm{s}}\left\langle k\right| \\
&\overset{(b)}{=} (N+1)\sum_{k=1}^{N}\frac{k!(N-k)!}{(N+1)!}\frac{N!}{k!(N-k)!}\left|k\right\rangle_{\mathrm{s}}\,{}_{\mathrm{s}}\left\langle k\right| \\
&= \mathbb{1}_{\mathrm{s}}
\end{aligned}
$$

where (a) is because $\mathbb{E}\{\mathsf{u}^k(\mathsf{u}^*)^{k'}\mathsf{v}^{N-k}(\mathsf{v}^*)^{(N-k')}\} = 0$ for $k \neq k'$; and (b) is because $|\mathsf{u}|^{2k}$ and $|\mathsf{v}|^{2(N-k)}$ follow a Dirichlet distribution of order 2 with parameters $k+1$ and $N-k+1$.

Claim 2: For arbitrary a, b,

$$\mathbb{E}\big\{\mathsf{U}^{\otimes N}(a\left|0\right\rangle + b\left|1\right\rangle)^{\otimes N}\left\langle 0\right|^{\otimes N}(\mathsf{U})^{\dagger\otimes N}\big\} = \frac{a^N}{N+1}\mathbb{1}_{\mathrm{s}}.$$

To prove Claim 2, we denote $\mathsf{U}\left|0\right\rangle = \left|\mathsf{v}\right\rangle$, $\mathsf{U}\left|1\right\rangle = \left|\mathsf{\mu}\right\rangle$. Moreover, let $\left|\mathsf{v},\mathsf{\mu}\right\rangle_j$ denote

the completely symmetric state of N qubits with j of them in state $|\nu\rangle$ and $N-j$ of them in state $|\mu\rangle$. Then

$$\begin{aligned}
&\mathbb{E}\big\{\mathbf{U}^{\otimes N}(a\,|0\rangle+b\,|1\rangle)^{\otimes N}\,\langle 0|^{\otimes N}\,(\mathbf{U})^{\dagger\otimes N}\big\}\\
&=\mathbb{E}\big\{(a\,|\nu\rangle+b\,|\mu\rangle)^{\otimes N}\,\langle\nu|^{\otimes N}\,\big\}\\
&=\sum_{j=0}^{N}\sqrt{\frac{N!}{j!(N-j)!}}a^j b^{N-j}\mathbb{E}\big\{\,|\nu,\mu\rangle_j\,\langle\nu|^{\otimes N}\,\big\}.
\end{aligned}$$

Note that because of the right invariance of the Haar measure, the joint distribution of $|\mu\rangle$ and $|\nu\rangle$ remains the same if we replace $|\mu\rangle$ with $\exp\{\mathrm{i}\pi/(N-j)\}\,|\mu\rangle$ provided that $j\neq N$. Consequently, for $j\neq N$,

$$\mathbb{E}\big\{\,|\nu,\mu\rangle_j\,\langle\nu|^{\otimes N}\,\big\}=\mathbb{E}\big\{\,\exp\{\mathrm{i}\pi\}\,|\nu,\mu\rangle_j\,\langle\nu|^{\otimes N}\,\big\}=-\mathbb{E}\big\{\,|\nu,\mu\rangle_j\,\langle\nu|^{\otimes N}\,\big\}=\mathbf{0}.$$

Then

$$\mathbb{E}\big\{\mathbf{U}^{\otimes N}(a\,|0\rangle+b\,|1\rangle)^{\otimes N}\,\langle 0|^{\otimes N}\,(\mathbf{U})^{\dagger\otimes N}\big\}=a^N\mathbb{E}\big\{\mathbf{U}^{\otimes N}\,|0\rangle^{\otimes N}\,\langle 0|^{\otimes N}\,(\mathbf{U})^{\dagger\otimes N}\big\}=\frac{a^N}{N+1}\mathbb{1}_{\mathrm{s}}$$

where the last inequality is because of Claim 1.

Claim 3: for any $|u\rangle\,,|v\rangle\in\mathcal{H}_2$,

$$\mathbb{E}\big\{\mathbf{U}^{\otimes N}\,|u\rangle^{\otimes N}\,\langle v|^{\otimes N}\,(\mathbf{U}^{\mathrm{sub}})^{\dagger\otimes N}\big\}=\frac{\langle v|u\rangle^N}{N+1}\mathbb{1}_{\mathrm{s}}.$$

Consider a unitary operator $\boldsymbol{U}(v)$ such that $\boldsymbol{U}(v)\,|v\rangle=|0\rangle$. Then

$$\begin{aligned}
&\mathbb{E}\big\{\mathbf{U}^{\otimes N}\,|u\rangle^{\otimes N}\,\langle v|^{\otimes N}\,(\mathbf{U})^{\dagger\otimes N}\big\}\\
&=\mathbb{E}\big\{(\mathbf{U}\boldsymbol{U}(v))^{\otimes N}\,|u\rangle^{\otimes N}\,\langle v|^{\otimes N}\,(\mathbf{U}\boldsymbol{U}(v))^{\dagger\otimes N}\big\}\\
&=\mathbb{E}\big\{\mathbf{U}^{\otimes N}\big(\boldsymbol{U}(v)\,|u\rangle\big)^{\otimes N}\,\langle 0|^{\otimes N}\,(\mathbf{U})^{\dagger\otimes N}\big\}=\frac{\langle 0|\,\boldsymbol{U}(v)\,|u\rangle^N}{N+1}\mathbb{1}_{\mathrm{s}}=\frac{\langle v|u\rangle^N}{N+1}\mathbb{1}_{\mathrm{s}},
\end{aligned}$$

where the first equality is due to the right invariance of the Haar measure and the third equality is because of Claim 2.

With these three claims, we can prove the original lemma. Since $\{(a\,|0\rangle+b\,|1\rangle)^{\otimes N} : a,b \in \mathbb{C}, |a|^2+|b|^2=1\}$ span the space $\mathcal{H}$, we can find a basis of $\mathcal{H}$ as follows: $(a_k\,|0\rangle+b_k\,|1\rangle)^{\otimes N}$, $k \in \mathcal{K}_{1:N+1}$. We can then write $|x_1\rangle$ and $|x_2\rangle$ as

$$|x_1\rangle = \sum_{k=1}^{N+1} \lambda_k (a_k\,|0\rangle + b_k\,|1\rangle)^{\otimes N}$$
$$|x_2\rangle = \sum_{k=1}^{N+1} \lambda'_k (a_k\,|0\rangle + b_k\,|1\rangle)^{\otimes N}.$$

Then

$$\begin{aligned}
&(N+1)\mathbb{E}\big\{\mathbf{U}^{\otimes N}\,|x_1\rangle_{\mathrm{s}}\,{}_{\mathrm{s}}\langle x_2|\,(\mathbf{U})^{\dagger\otimes N}\big\} \\
&= (N+1)\mathbb{E}\big\{\mathbf{U}^{\otimes N}\sum_{k=1}^{N+1}\lambda_k(a_k\,|0\rangle+b_k\,|1\rangle)^{\otimes N}\sum_{l=1}^{N+1}(\lambda'_l)^*(a_l^*\,\langle 0|+b_l^*\,\langle 1|)^{\otimes N}(\mathbf{U})^{\dagger\otimes N}\big\} \\
&= (N+1)\sum_{k=1}^{N+1}\sum_{l=1}^{N+1}\lambda_k(\lambda'_l)^*\mathbb{E}\big\{\mathbf{U}^{\otimes N}(a_k\,|0\rangle+b_k\,|1\rangle)^{\otimes N}(a_l^*\,\langle 0|+b_l^*\,\langle 1|)^{\otimes N}(\mathbf{U})^{\dagger\otimes N}\big\} \\
&= (N+1)\sum_{k=1}^{N+1}\sum_{l=1}^{N+1}\lambda_k(\lambda'_l)^*(a_ka_l^*+b_kb_l^*)^N\mathbb{1}_{\mathrm{s}} \\
&= (N+1)\,\langle x_2|x_1\rangle\,\mathbb{1}_{\mathrm{s}}
\end{aligned}$$

where the third equality is because of Claim 3.

A.4 Proof of Theorem 2.4

We will prove Theorem 2.4 in the following steps.

Step 1: Consider a unitary transformation $\boldsymbol{F}_N^i$ that operates on an input state with N qubits by performing the following operation only on the first, the ith, and

the $(i+1)$th qubits:

$$|010\rangle \to \sqrt{\frac{\binom{N-1}{i-1}}{\binom{N}{i}}}\,|110\rangle + \sqrt{\frac{\binom{N-1}{i}}{\binom{N}{i}}}\,|001\rangle\,,$$
$$|110\rangle \to |010\rangle\,, \qquad |011\rangle \to |011\rangle\,,$$
$$|000\rangle \to |000\rangle\,, \qquad |100\rangle \to |100\rangle\,,$$
$$|101\rangle \to |101\rangle\,, \qquad |111\rangle \to |111\rangle\,.$$

Since the operation involves only three qubits and leaves other qubits unchanged, $\boldsymbol{F}_N^i$ can be implemented using C_1 gates [109], where C_1 is a constant that does not depend on N.

Step 2: Consider a unitary transformation $\boldsymbol{X}_N^i$ that operates on an input state with N qubits by performing the following operation only on the first and the ith qubit:

$$|00\rangle \to |00\rangle\,, \qquad |01\rangle \to |01\rangle$$
$$|10\rangle \to \sqrt{\frac{1}{N}}\,|10\rangle + \sqrt{\frac{N-1}{N}}\,|11\rangle$$
$$|11\rangle \to \sqrt{\frac{N-1}{N}}\,|10\rangle - \sqrt{\frac{1}{N}}\,|11\rangle\,.$$

Since the operation involves only two qubits and leaves other qubits unchanged, $\boldsymbol{X}_N^i$ can be implemented using C_2 gates [109], where C_2 is a constant that does not depend on N.

Step 3: Consider a unitary transformation $\boldsymbol{G}_N$ that operates on an input state

with N qubits as follows:

$$
\begin{aligned}
&|N,0\rangle \to |N,0\rangle \\
&|N,1\rangle \to \sqrt{\frac{1}{N}}\,|N,1\rangle + \sqrt{\frac{N-1}{N}}\,|N,2\rangle \\
&|N,n\rangle \to \sqrt{\frac{\binom{N-1}{n-1}}{\binom{N}{k}}}\,|1\rangle\,|N-1,n-1\rangle + \sqrt{\frac{\binom{N-1}{n}}{\binom{N}{n}}}\,|N,n+1\rangle\,, \quad n \in \mathcal{K}_{1:N} \setminus \{1,N\} \\
&|N,N\rangle \to |1\rangle\,|N-1,N-1\rangle
\end{aligned}
$$

where $|i,j\rangle$ $(i,j \in \mathbb{N}^*)$ denotes a pure state consisting of i qubits, and its jth qubit is $|1\rangle$ and all the remaining qubits are $|0\rangle$; and $|i,0\rangle_{\mathrm{L}}$ denotes a pure state consisting of i $|0\rangle$'s; and $\mathcal{K}_{1:N}$.

Let $\mathrm{CNOT}(i,j)$ denotes the CNOT gate that operates on the ith and the jth qubits with the ith qubit serving as the control bit. Since

$$\boldsymbol{G}_N = \boldsymbol{F}_N^2 \boldsymbol{F}_N^3 \ldots \boldsymbol{F}_N^{N-1} \mathrm{CNOT}(N,1) \boldsymbol{X}_N^2$$

$\boldsymbol{G}_N$ can be implemented with $(N-2)C_1 + 1 + C_2$ basic gates. Consequently, the transformation $\boldsymbol{G}_N$ can be implemented using NC_3 gates, where $C_3 := \max\{1, C_1, C_2\}$ is a constant that does not depend on N.

Step 4: Consider a unitary transformation $\boldsymbol{U}_N$ that operates on an input state with N qubits as follows:

$$|N,n\rangle \to |n\rangle_{\mathrm{s}}\,, \qquad i \in \{0\} \cup \mathcal{K}_{1:N}.$$

One can verify that

$$(\mathbb{1} \otimes \boldsymbol{U}_{N-1}) \boldsymbol{G}_N = \boldsymbol{U}_N. \tag{A.16}$$

Next we will show by induction that the transformation $\boldsymbol{U}_N$ can be implemented with $N(N+1)C_3/2$ basic gates. For the base case: building $\boldsymbol{U}_1$ does not require any

gate. For the induction step: $\boldsymbol{U}_{N-1}$ can be implemented with $(N-1)NC_3/2$ basic gates by the induction hypothesis; moreover, building $\boldsymbol{G}_N$ requires NC_3 basic gates as shown in Step 3; and thus, $\boldsymbol{U}_N$ in (A.16) can be implemented with $N(N+1)C_3/2$ basic gates.

Step 5: Consider a unitary transformation $\boldsymbol{T}_N$ that operates on an input state with N qubits as follows:

$$|N, n\rangle \rightarrow |n\rangle_{\mathrm{s}}\,, \qquad n \in \{0, 1\}.$$

One can verify that

$$(\mathbb{1} \otimes \boldsymbol{T}_{N-1})\boldsymbol{A}_N = \boldsymbol{T}_N, \tag{A.17}$$

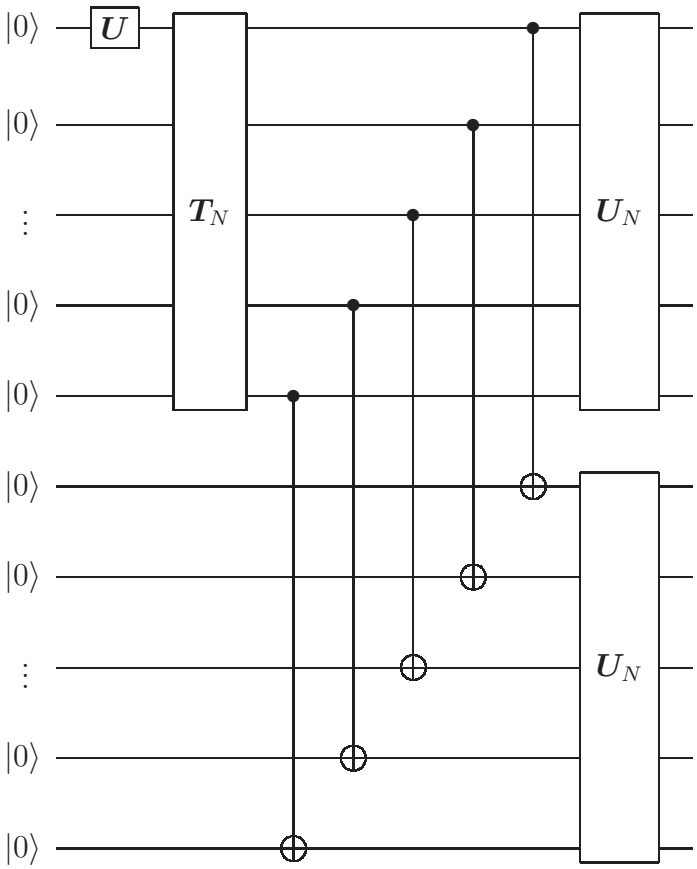
where $\boldsymbol{A}_N$ is a unitary transformation that operates on an input state with N qubits by performing the following operation only on the first and second qubits:

$$\begin{aligned} |00\rangle &\rightarrow |00\rangle \\ |10\rangle &\rightarrow \frac{1}{\sqrt{N}}\,|10\rangle + \sqrt{\frac{N-1}{N}}\,|01\rangle\,. \end{aligned}$$

Since the operation involves only two qubits and leaves other qubits unchanged, $\boldsymbol{A}_N$ can be implemented using C_2 gates [109].

Next we will show by induction that the transformation $\boldsymbol{T}_N$ can be implemented with NC_2 basic gates. For the base case: building $\boldsymbol{T}_1$ doe not require any gate. For the induction step: $\boldsymbol{T}_{N-1}$ can be implemented with $(N-1)C_2$ basic gates by the induction hypothesis; moreover, building $\boldsymbol{A}_N$ requires C_2 qubits as shown in the previous paragraph; and thus, $\boldsymbol{T}_N$ in (A.17) can be implemented with NC_2 basic gates.

Step 6: The state $\frac{1}{\sqrt{N+1}}\sum_{j=0}^{N}|j\rangle_{\mathrm{s}}\,|j\rangle_{\mathrm{s}}$ can be prepared from the state $|00\ldots0\rangle$ using the following circuit with $\left[N(N+1)C_3 + NC_2 + N + 1\right]$ basic gates.



where $\boldsymbol{U}$ is given by

$$|0\rangle \rightarrow \frac{1}{\sqrt{N+1}}|0\rangle + \sqrt{\frac{N}{N+1}}|1\rangle .$$

Note that if the output of the first N qubits above are applied with $\boldsymbol{U}_N^\dagger$ (or equivalently, removing $\boldsymbol{U}_N$ that operates on the first N qubits in the circuit above), the following state can be prepared from the state $|00\ldots0\rangle$ with $\left[N(N+1)C_3/2+NC_2+1\right]$ basic gates:

$$\frac{1}{\sqrt{N+1}}\sum_{n=0}^{N}|N,n\rangle\,|n\rangle_{\rm s} . \tag{A.18}$$

Step 7: Consider a unitary transformation $\boldsymbol{P}$ that operates on an input state with $N+\lceil\log(N+1)\rceil$ qubits:

$$|N,n\rangle\,|\lceil\log(N+1)\rceil,0\rangle \rightarrow |N,0\rangle\,|n\rangle_{\rm b}\,, n\in\{0\}\cup\mathcal{K}_{1:N}$$

Recall that $|k\rangle_{\rm b}$ is the binary representation of k with $\lceil \log(N+1) \rceil$ qubits (e.g., for $N=6$, $k=5$, $|k\rangle_{\rm b} = |101\rangle$).

The transformation $\boldsymbol{P}$ can be implemented with $N^3 C_4$ basic gates. To see this, we consider the unitary transformation $\boldsymbol{P}_i$ that operates on an input state with $N + \lceil \log(N+1) \rceil$ qubits by performing the following operation only on the nth qubit and the last $\lceil \log(N+1) \rceil$ qubits:

$$
\begin{aligned}
|\lceil \log(N+1) \rceil, 1\rangle &\rightarrow |0\rangle\, |n\rangle_{\rm b} \\
|0\rangle\, |k\rangle_{\rm b} &\rightarrow |0\rangle\, |k\rangle_{\rm b}\,, \quad 0 \leqslant k \leqslant N, k \neq n\,.
\end{aligned}
$$

Since the operation $\boldsymbol{P}_i$ involves only $\lceil \log(N+1) \rceil + 1$ qubits, it can be implemented with $N^2 C_4$ basic gates by [109]. Note that $\boldsymbol{P}$ can be concatenated as $\boldsymbol{P} = \boldsymbol{P}_1 \boldsymbol{P}_2 \ldots \boldsymbol{P}_N$. Thus, $\boldsymbol{P}$ can be implemented with $N^3 C_4$ basic gates.

Step 8: With the results above, we can efficiently prepare the state $\frac{1}{\sqrt{N+1}} \sum_{j=0}^{N} |j\rangle_{\rm b}\, |j\rangle_{\rm s}$. In fact, this can be accomplished by attaching a $\lceil \log(N+1) \rceil$-qubit ancillary state $|00\ldots0\rangle$ to (A.18), applying $\boldsymbol{P}$ to the first N qubits of (A.18) as well as the ancillary state, and keeping only the ancillary state and the last N qubits of (A.18). In total, $\left[N(N+1)C_3/2 + NC_2 + N + 1 + N^3 C_4\right]$ basic gates are needed. This gives the proof of the part (a) of Theorem 2.4.

Step 9: Note also that $\mathring{\boldsymbol{H}}_{\rm failure}$ is at most rank $(N+1)$; if $\mathring{\lambda}_k > 0$, $\mathring{\boldsymbol{H}}_{{\rm b},k}$ is rank 1 and there are at most $(N+1)(N+2)/2$ such matrices according to the sparsity analysis in [84]. Such positive-operator valued measures (POVMs) can be implemented by a unitary gate on $\lceil \log{(N+1)(N+4)/2} \rceil$ qubits, together with standard basis measurement [110]. This unitary gate can be implemented with $O(4^{\lceil \log{(N+1)(N+4)/2} \rceil}) = O(N^4)$ gates [109]. Thus, the POVMs $\mathring{\boldsymbol{H}}_{{\rm b},k}$ and $\mathring{\boldsymbol{H}}_{\rm b,failure}$ can be implemented with poly(N) basic gates and standard basis measurements. Similarly, $\mathring{\boldsymbol{M}}_{{\rm b},k}$ and $\mathring{\boldsymbol{M}}_{\rm b,failure}$ can be implemented with poly(N) basic gates and standard basis measurements. This gives the proof of the part (b) of Theorem 2.4.

Step 10: We first apply $\boldsymbol{U}_N^{-1}$ to $|\psi\rangle^{\otimes N}$, then attach a $\lceil \log(N+1) \rceil$-qubit ancillary state $|00\ldots0\rangle$ to the output, resulting in a $(N + \lceil \log(N+1) \rceil)$-qubit state. We then

apply $\boldsymbol{P}$ in Step 7 to this state and keep only the last $\lceil \log(N+1) \rceil$ qubits to obtain the desired state $|\psi\rangle_{\rm b}$. Steps 4 and 7 show that $\boldsymbol{U}_N^{-1}$ and $\boldsymbol{P}$ can be implemented with poly(N) basic gates. Therefore, $|\psi\rangle^{\otimes N}$ can be transformed to $|\psi\rangle_{\rm b}$ with poly(N) basic gates. This gives the proof of the part (c) of Theorem 2.4.

Appendix B

Proofs of the Results in Chapter 3

B.1 Proof of Theorem 3.1

The proof for Theorem 3.1 is organized in five steps.

Step 1: Connection to the Blackwell optimal policy

We first introduce the $\beta-$discounted dynamic programming problem in the context of quantum queuing system: Given an initial state $\boldsymbol{x}^{(1)}$ and $\beta \in (0,1)$, the aim is to find a policy $\pi = \{u^{(0)}, u^{(1)}, \dots\}$, that minimizes the cost function

$$J_\beta(\boldsymbol{x}^{(1)}) = \limsup_{N\to\infty} \mathbb{E}\left\{ \sum_{n=0}^{N-1} \sum_{i=1}^{N_\mathrm{r}} \beta^n \max\{0, s_i^{(n)}\} \middle| \boldsymbol{x}^{(1)} \right\}$$

subject to the system equation constraint (3.1).

The definition of Blackwell optimal policy is as follows:

Definition 3 ([94], Chapter 4)**.** A stationary policy μ is said to be Blackwell optimal if it is simultaneously optimal for all the $\beta-$discounted problems with β in an interval $(\beta^*, 1)$, where β^* is some scalar with $0 < \beta^* < 1$.

Lemma B.1 ([94], Proposition 4.1.3)**.** There exists a Blackwell optimal policy.

Lemma B.2 ([94], Proposition 4.1.7)**.** A Blackwell policy is optimal over all policies in the average cost problem.

Based on Lemma B.1, we can consider a Blackwell optimal policy $u^{*(n)}$. Noting that the Blackwell optimal policy is stationary, i.e., independent on the time index n, we can drop the superscript (n) and write the optimal control as u^*. By Definition 3 and Lemma B.1, there exists $\beta^* \in (0,1)$ such that u^* is optimal for $\beta-$discounted problems with $\beta \in (\beta^*, 1)$. We next show that either u^* is threshold-based control as stated in Theorem 3.1, or we can create threshold-based control that is also a Blackwell optimal policy.

Step 2: Introduction of some notations

Since $N_{\mathrm{r}} = 2$, $\boldsymbol{x}^{(1)} \in \mathbb{R}^4$. With a little abuse of notation, we write J_β as a function of four variables, i.e.,

$$J_\beta(\boldsymbol{x}) = J_\beta(s_1, s_2, c_1, c_2) \tag{B.1}$$

where $\boldsymbol{x} = [\, s_1, s_2, c_1, c_2 \,]^{\mathrm{T}}$. We also introduce a function that represents the future cost

$$\widetilde{J}_\beta(s_1, s_2) := \mathbb{E}\{J_\beta(s_1, s_2, \mathsf{c}_1, \mathsf{c}_2)\} = \sum_{c_1, c_2 \in \{-1,0,1\}} \mathbb{P}\{\mathsf{c}_1^{(1)} = c_1, \mathsf{c}_2^{(1)} = c_2\} J_\beta(s_1, s_2, c_1, c_2).$$

One can verify that

$$J_\beta(s_1, s_2, c_1, c_2) = \max\{0, s_1\} + \max\{0, s_2\} + \beta \widetilde{J}_\beta(u_\beta^*(s_1, s_2, c_1, c_2))$$

where u_β^* denotes the optimal control policy for the β-discounted problem.[1]

Step 3: Connection to the "convexity" of $\widetilde{J}_\beta$

Note that for the $\beta-$discounted problem, for a state $[\, s_1, s_2, c_1, c_2 \,]^{\mathrm{T}}$, if

$$\sum_{i=1}^{N_{\mathrm{r}}} \max\{0, -s_i - c_i\} > M$$

$$s_i + c_i < 0 \quad i = 1, 2$$

[1]Note that the optimal control policy for the β-discounted problem is stationary and does not rely on the time index n. Similarly to (B.1), we write u_β^* as a function of four variables with an output of two variables.

the optimal control is

$$\begin{aligned}[\tilde{s}_1, \tilde{s}_2]^{\mathrm{T}} &= \underset{[\tilde{s}_1,\tilde{s}_2]^{\mathrm{T}} \in \mathcal{F}(s_1,s_2,c_1,c_2)}{\arg\min} \beta\, \mathbb{E}\{J_\beta(\tilde{s}_1, \tilde{s}_2, \mathsf{c}_1, \mathsf{c}_2)\} \\ &= \underset{[\tilde{s}_1,\tilde{s}_2]^{\mathrm{T}} \in \mathcal{F}(s_1,s_2,c_1,c_2)}{\arg\min} \widetilde{J}_\beta(\tilde{s}_1, \tilde{s}_2)\end{aligned}$$

where $\mathcal{F}(s_1, s_2, c_1, c_2)$ is the feasible set, given by

$$\mathcal{F}(s_1, s_2, c_1, c_2) := \{[x, y]^{\mathrm{T}} \in \mathbb{Z}^2 : x \geqslant s_1 + c_1, y \geqslant s_2 + c_2, x + y \geqslant -M\}.$$

Note that the feasible set has a physical meaning: a quantum node can discard established entangled qubits, but cannot store entangled qubits beyond the size of the entanglement memory.

In the remaining parts of the proof, we will show that $\widetilde{J}_\beta$ has the following property

$$2\widetilde{J}_\beta(s_1, s_2) \leqslant \widetilde{J}_\beta(s_1 + 1, s_2 - 1) + \widetilde{J}_\beta(s_1 - 1, s_2 + 1) \tag{B.2}$$

provided that $s_1 \leqslant -1$, $s_2 \leqslant -1$, and $s_1 + s_2 = -M$. The property (B.2) implies that $\widetilde{J}_\beta(s_1, -M - s_1)$ is discretely convex as a function of s_1 for $s_1 \in \mathcal{K}_{-M:0}$. One can easily verify that if (B.2) holds, then either u^*_β is threshold-based control or there exists threshold-based control that achieves the same performance as u^*_β (the latter occurs only when $\widetilde{J}_\beta(s_1, -M - s_1)$ first decreases, stays constant, and then increases as a function of s_1 when $s_1 \in \mathcal{K}_{-M:0}$).

We will prove a stronger claim than (B.2) as follows:

Lemma B.3. There exists $\beta^* \in (0, 1)$ such that for $\beta \in (\beta^*, 1)$,

$$\widetilde{J}_\beta(s_1, s_2 - 1) - \widetilde{J}_\beta(s_1 - 1, s_2) \leqslant \widetilde{J}_\beta(s_1, s_2 - 2) - \widetilde{J}_\beta(s_1 - 1, s_2 - 1) \tag{B.3}$$

$$\widetilde{J}_\beta(s_1 - 1, s_2) - \widetilde{J}_\beta(s_1, s_2 - 1) \leqslant \widetilde{J}_\beta(s_1 - 2, s_2) - \widetilde{J}_\beta(s_1 - 1, s_2 - 1) \tag{B.4}$$

provided that $s_1 \leqslant 1$, $s_2 \leqslant 1$, and s_1 and s_2 are valid numbers.[2]

[2]In this thesis, "valid numbers" are defined as numbers that lead to valid state inputs for the function $\widetilde{J}_\beta$.

If (B.3) and (B.4) hold, then adding these two equations gives

$$2\widetilde{J}_\beta(s_1-1,s_2-1) \leqslant \widetilde{J}_\beta(s_1,s_2-2)+\widetilde{J}_\beta(s_1-2,s_2)$$

for $s_1 \leqslant 1$ and $s_2 \leqslant 1$, which is stronger than (B.2).

Step 4: Value iteration of J_β and $\widetilde{J}_\beta$

We consider the value iteration of J_β and $\widetilde{J}_\beta$. In particular, consider

$$J_\beta^{(0)}(s_1,s_2,c_1,c_2) = \max\{0,s_1\}+\max\{0,s_2\}$$

and

$$\begin{aligned}
&J_\beta^{(k+1)}(s_1,s_2,c_1,c_2)\\
&= \max\{0,s_1\}+\max\{0,s_2\}+\beta \min_{u(s_1,s_2,c_1,c_2)\in\mathcal{F}(s_1,s_2,c_1,c_2)} \mathbb{E}\{J_\beta^{(k)}(u(s_1,s_2,c_1,c_2),\mathsf{c}_1,\mathsf{c}_2)\}\\
&= \max\{0,s_1\}+\max\{0,s_2\}+\beta\mathbb{E}\{J_\beta^{(k)}(u_k(s_1,s_2,c_1,c_2),\mathsf{c}_1,\mathsf{c}_2)\}
\end{aligned}$$

where u_k denotes the policy that achieves the minimum in the first equality. Correspondingly, consider

$$\widetilde{J}_\beta^{(0)}(s_1,s_2) = \mathbb{E}\{J_\beta^{(0)}(s_1,s_2,\mathsf{c}_1,\mathsf{c}_2)\} = \max\{0,s_1\}+\max\{0,s_2\} \tag{B.5}$$

and

$$\widetilde{J}_\beta^{(k)}(s_1,s_2) = \mathbb{E}\{J_\beta^{(k)}(s_1,s_2,\mathsf{c}_1,\mathsf{c}_2)\}.$$

One can verify that the function $\widetilde{J}_\beta^{(k)}$ has the following iteration:

$$\widetilde{J}_\beta^{(k+1)}(s_1,s_2) = \max\{0,s_1\}+\max\{0,s_2\}+\beta\mathbb{E}\Big\{\widetilde{J}_\beta^{(k)}\big(u_k(s_1,s_2,\mathsf{c}_1,\mathsf{c}_2)\big)\Big\}.$$

Value iteration shows that $\lim_{k\to\infty} J_\beta^{(k)} = J_\beta$ [94]; consequently, $\lim_{k\to\infty}\widetilde{J}_\beta^{(k)} = \widetilde{J}_\beta$.

Therefore, for proving Lemma B.3, it suffices to prove that for all k,

$$F^{(k)}(s_1, s_2) \leqslant 0, \quad s_1 \leqslant Q_{\mathrm{t}}, s_2 \leqslant 0$$
$$G^{(k)}(s_1, s_2) \leqslant 0, \quad s_1 \leqslant 0, s_2 \leqslant Q_{\mathrm{t}}$$

where

$$\begin{aligned} F^{(k)}(s_1, s_2) := & \left[\widetilde{J}_{\beta}^{(k)}(s_1, s_2 - 1) - \widetilde{J}_{\beta}^{(k)}(s_1 - 1, s_2)\right] \\ & - \left[\widetilde{J}_{\beta}^{(k)}(s_1, s_2 - 2) - \widetilde{J}_{\beta}^{(k)}(s_1 - 1, s_2 - 1)\right] \\ G^{(k)}(s_1, s_2) := & \left[\widetilde{J}_{\beta}^{(k)}(s_1 - 1, s_2) - \widetilde{J}_{\beta}^{(k)}(s_1, s_2 - 1)\right] \\ & - \left[\widetilde{J}_{\beta}^{(k)}(s_1 - 2, s_2) - \widetilde{J}_{\beta}^{(k)}(s_1 - 1, s_2 - 1)\right]. \end{aligned}$$

The proof will be shown in the next step.

Step 5: Induction method to prove Lemma B.3

We use the induction method to prove the following claims simultaneously:

Lemma B.4. For all k, the following inequalities hold:

$$F^{(k)}(s_1, s_2) \leqslant 0, \quad s_1 \leqslant Q_{\mathrm{t}}, s_2 \leqslant 0 \tag{B.6}$$
$$H^{(k)}(s_1, s_2) \leqslant 0, \quad s_1 \leqslant Q_{\mathrm{t}}, s_2 \leqslant 0 \tag{B.7}$$
$$G^{(k)}(s_1, s_2) \leqslant 0, \quad s_1 \leqslant 0, s_2 \leqslant Q_{\mathrm{t}} \tag{B.8}$$
$$I^{(k)}(s_1, s_2) \leqslant 0, \quad s_1 \leqslant 0, s_2 \leqslant Q_{\mathrm{t}} \tag{B.9}$$
$$F^{(k)}(s_1, s_2) \leqslant \eta_{s_2}^{(k)}, \quad s_1 \leqslant Q_{\mathrm{t}}, s_2 \geqslant 1 \tag{B.10}$$
$$H^{(k)}(s_1, s_2) \leqslant \eta_{s_2}^{(k)}, \quad s_1 \leqslant Q_{\mathrm{t}}, s_2 \geqslant 1 \tag{B.11}$$
$$G^{(k)}(s_1, s_2) \leqslant \zeta_{s_1}^{(k)}, \quad s_1 \geqslant 1, s_2 \leqslant Q_{\mathrm{t}} \tag{B.12}$$
$$I^{(k)}(s_1, s_2) \leqslant \zeta_{s_1}^{(k)}, \quad s_1 \geqslant 1, s_2 \leqslant Q_{\mathrm{t}} \tag{B.13}$$
$$\widetilde{J}_{\beta}^{(k)}(s_1, s_2) - \widetilde{J}_{\beta}^{(k)}(s_1, s_2 - 1) \leqslant \nu_{s_2}^{(k)}, \ s_1 \leqslant Q_{\mathrm{t}}, s_2 \geqslant 1 \tag{B.14}$$
$$\widetilde{J}_{\beta}^{(k)}(s_1, s_2) - \widetilde{J}_{\beta}^{(k)}(s_1 - 1, s_2) \leqslant \upsilon_{s_1}^{(k)}, \ s_1 \geqslant 1, s_2 \leqslant Q_{\mathrm{t}} \tag{B.15}$$

where

$$
\begin{aligned}
H^{(k)}(s_1,s_2) &:= \big[\widetilde{J}_\beta^{(k)}(s_1,s_2-1)-\widetilde{J}_\beta^{(k)}(s_1,s_2)\big]\\
&\quad-\big[\widetilde{J}_\beta^{(k)}(s_1,s_2-2)-\widetilde{J}_\beta^{(k)}(s_1,s_2-1)\big]\\
I^{(k)}(s_1,s_2) &:= \big[\widetilde{J}_\beta^{(k)}(s_1-1,s_2)-\widetilde{J}_\beta^{(k)}(s_1,s_2)\big]\\
&\quad-\big[\widetilde{J}_\beta^{(k)}(s_1-2,s_2)-\widetilde{J}_\beta^{(k)}(s_1-1,s_2)\big].
\end{aligned}
$$

Moreover, $\eta_i^{(k)}$ and $\nu_i^{(k)}$ are auxiliary sequences given by

$$
\begin{aligned}
&\eta_i^{(0)} = \eta_\beta^*(i), \quad i\in\mathcal{K}_{1:Q_{\mathrm{t}}}\\
&\eta_1^{(k)} = -1+\beta\big[\,\tilde{p}_2\eta_2^{(k-1)}+(1-\tilde{p}_2-\tilde{q}_2)\eta_1^{(k-1)}\,\big]\\
&\eta_i^{(k)} = \beta\big[\,\tilde{p}_2\eta_{i+1}^{(k-1)}+(1-\tilde{p}_2-\tilde{q}_2)\eta_i^{(k-1)}+\tilde{q}_2\eta_{i-1}^{(k-1)}\,\big], \qquad i\in\mathcal{K}_{2:Q_{\mathrm{t}}-1}\\
&\eta_{Q_{\mathrm{t}}}^{(k)} = \beta\big[\tilde{p}_2\nu_{Q_{\mathrm{t}}}^{(k-1)}+(1-\tilde{p}_2-\tilde{q}_2)\eta_{Q_{\mathrm{t}}}^{(k-1)}+\tilde{q}_2\eta_{Q_{\mathrm{t}}-1}^{(k-1)}\,\big]
\end{aligned}
$$

and

$$
\begin{aligned}
&\nu_i^{(0)} = \nu_\beta^*(i), \qquad i\in\mathcal{K}_{1:Q_{\mathrm{t}}}\\
&\nu_1^{(k)} = 1+\beta\big[\,\tilde{p}_2\nu_2^{(k-1)}+(1-\tilde{p}_2)\nu_1^{(k-1)}\,\big]\\
&\nu_i^{(k)} = 1+\beta\big[\,\tilde{p}_2\nu_{i+1}^{(k-1)}+(1-\tilde{p}_2-\tilde{q}_2)\nu_i^{(k-1)}+\tilde{q}_2\nu_{i-1}^{(k-1)}\,\big], \qquad i\in\mathcal{K}_{2:Q_{\mathrm{t}}}
\end{aligned}
$$

where $\tilde{p}_2=\mathbb{P}\{\mathsf{c}_2^{(1)}=1\}$, $\tilde{q}_2=\mathbb{P}\{\mathsf{c}_2^{(1)}=-1\}$; $\eta_\beta^*(i)$ and $\nu_\beta^*(i)$ are steady states of the system above, i.e., $\eta_\beta^*(i)=\lim_{k\to\infty}\eta_i^{(k)}$ and $\nu_\beta^*(i)=\lim_{k\to\infty}\nu_i^{(k)}$;[3] and $\zeta_i^{(k)}$ and $\upsilon_i^{(k)}$ are obtained by replacing η, ν, $\tilde{p}_2$, $\tilde{q}_2$ with ζ, υ, $\tilde{p}_1$, $\tilde{q}_1$, respectively.

Proof. See Appendix B.2. □

The completion of proving Lemma B.4 finishes the proof of Theorem 3.1.

[3]The expressions of $\eta_\beta^*(i)$ and $\nu_\beta^*(i)$ will be explicitly given in Appendix B.2.

B.2 A Proof of Lemma B.4

We first prove a lemma that will be used in the later proof.

Lemma B.5. There exists β^* such that for any $\beta \in (\beta^*, 1)$

$$\begin{aligned}
&\eta^*_\beta(1) \in (-1, 0)\\
&\eta^*_\beta(i) > 0, i \neq 0\\
&\nu^*_\beta(i) > 0.
\end{aligned}$$

Proof. Denote $Q_{\mathrm{t}} \times Q_{\mathrm{t}}$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ as follows:

$$\boldsymbol{A} = \begin{bmatrix}
1-\tilde{p}_2 & \tilde{p}_2 & & & & \\
\tilde{q}_2 & 1-\tilde{p}_2-\tilde{q}_2 & \tilde{p}_2 & & & \\
& \tilde{q}_2 & 1-\tilde{p}_2-\tilde{q}_2 & \tilde{p}_2 & & \\
& & & \ddots & & \\
& & & \tilde{q}_2 & 1-\tilde{p}_2-\tilde{q}_2
\end{bmatrix}$$

and

$$\boldsymbol{B} = \boldsymbol{A} - \tilde{q}_2 \boldsymbol{E}_{1,1}.$$

Consider the vectors $\boldsymbol{\nu}^*_\beta$ and $\boldsymbol{\eta}^*_\beta$ that represent steady states, i.e., $[\boldsymbol{\nu}^*_\beta]_i = \nu^*_\beta(i)$ and $[\boldsymbol{\eta}^*_\beta]_i = \eta^*_\beta(i)$.

The steady states of $\nu^{(i)}_k$ and $\nu^{(i)}_k$ are given by

$$\begin{aligned}
\boldsymbol{\nu}^*_\beta &= (\boldsymbol{I} - \beta \boldsymbol{A})^{-1}\mathbf{1}\\
\boldsymbol{\eta}^*_\beta &= (\boldsymbol{I} - \beta \boldsymbol{B})^{-1}\big(-\boldsymbol{e}_1 + \beta\tilde{p}_2 \boldsymbol{E}_{Q_{\mathrm{t}},Q_{\mathrm{t}}}\boldsymbol{\nu}^*_\beta\big).
\end{aligned}$$

One can easily verify that $\boldsymbol{I} - \beta\boldsymbol{A}$ and $\boldsymbol{I} - \beta\boldsymbol{B}$ are invertible for $\beta \in (0, 1]$.

Note that for $\beta = 1$,

$$[\boldsymbol{\nu}^*_\beta]_i = \sum_{m=i}^{Q_\mathrm{t}} \sum_{n=0}^{m-1} \frac{\tilde{p}_2^n \tilde{q}_2^{m-1-n}}{\tilde{p}_2^m} > 0. \tag{B.16}$$

Moreover, for $\beta = 1$,

$$[\boldsymbol{\eta}^*_\beta]_i = \frac{1}{\sum_{m=1}^{Q_\mathrm{t}} \tilde{p}_2^m \tilde{q}_2^{Q_\mathrm{t}-m}} \Big(- \sum_{m=i-1}^{Q_\mathrm{t}-1} \tilde{p}_2^{Q_\mathrm{t}-1-m} \tilde{q}_2^m + \frac{\sum_{m=0}^{Q_\mathrm{t}-1} \tilde{p}_2^m \tilde{q}_2^{Q_\mathrm{t}-1-m}}{\tilde{p}_2^{Q_\mathrm{t}-1}} \sum_{m=0}^{i-1} \tilde{p}_2^{Q_\mathrm{t}-1-m} \tilde{q}_2^m \Big).$$

It is straightforward to verify that

$$[\boldsymbol{\eta}^*_\beta]_1 = 0 \text{ and } [\boldsymbol{\eta}^*_\beta]_i > 0, \quad i > 1. \tag{B.17}$$

Moreover,

$$\left.\frac{\partial \boldsymbol{\eta}^*}{\partial \beta}\right|_{\beta=1} = (\boldsymbol{I} - \boldsymbol{B})^{-1} \boldsymbol{B} \boldsymbol{\eta}^*_\beta + p_2 (\boldsymbol{I} - \boldsymbol{B})^{-1} \boldsymbol{E}_{Q_\mathrm{t}, Q_\mathrm{t}} \boldsymbol{\nu}^*_\beta \Big|_{\beta=1}.$$

Noticing that

$$(\boldsymbol{I} - \boldsymbol{B})^{-1} = \sum_{j=0}^{\infty} \boldsymbol{B}^j$$

one can see that each element $(\boldsymbol{I} - \boldsymbol{B})^{-1}$ is positive. As a consequence, one can easily derive that

$$\left.\frac{\partial [\boldsymbol{\eta}^*]_i}{\partial \beta}\right|_{\beta=1} > 0. \tag{B.18}$$

Combining (B.16), (B.17), and (B.18), together with the continuity of $\boldsymbol{\eta}^*_\beta$ and $\boldsymbol{\nu}^*_\beta$ as functions of β, we arrive at the desired results. □

Lemma B.6. The functions $\widetilde{J}^{(k)}_\beta(s_1, \cdot)$ and $\widetilde{J}^{(k)}_\beta(\cdot, s_2)$ are increasing functions for all valid s_1 and s_2.

The proof of such monotonicity is omitted.

We now prove Lemma B.4 by induction. The base case can be easily verified with the initial values (B.5). We next consider the induction step. Suppose (B.7) to (B.15) hold for the case k, we next consider the case $k+1$. Note that adding (B.6) and (B.8) as well as adding (B.10) and (B.12) give

$$2\widetilde{J}_{\beta}^{(k)}(s_1-1,s_2-1) \leqslant \widetilde{J}_{\beta}^{(k)}(s_1,s_2-2)+\widetilde{J}_{\beta}^{(k)}(s_1-2,s_2) \tag{B.19}$$

for $s_1 \leqslant 1$ and $s_2 \leqslant 1$.

For brevity, we only show the proof of (B.7) and (B.14) for the case $k+1$. The other cases can be proved similarly.

For (B.7), we first show that for $c_1, c_2 \in \{-1,0,1\}$ and $s_1 \leqslant Q_{\mathrm{t}}, s_2 \leqslant 0$,

$$2\widetilde{J}_{\beta}^{(k)}\big(u_k(s_1,s_2-1,c_1,c_2)\big) \leqslant \widetilde{J}_{\beta}^{(k)}\big(u_k(s_1,s_2-2,c_1,c_2)\big)+\widetilde{J}_{\beta}^{(k)}\big(u_k(s_1,s_2,c_1,c_2)\big). \tag{B.20}$$

Note that $\max\{0,-s_1\}+\max\{0,-(s_2-2)\} \leqslant M$, we have $\max\{0,-(s_1+c_1)\}+\max\{0,-(s_2-1+c_2)\} \leqslant M+1$.

If $\max\{0,-(s_1+c_1)\}+\max\{0,-(s_2-1+c_2)\} \leqslant M$, then

$$\begin{aligned}
&\widetilde{J}_{\beta}^{(k)}\big(u_k(s_1,s_2-1,c_1,c_2)\big)-\widetilde{J}_{\beta}^{(k)}\big(u_k(s_1,s_2,c_1,c_2)\big)\\
&=\widetilde{J}_{\beta}^{(k)}(Q_{\mathrm{m}},s_2-1+c_2)-\widetilde{J}_{\beta}^{(k)}(Q_{\mathrm{m}},s_2+c_2)\\
&\leqslant \min\Big\{\widetilde{J}_{\beta}^{(k)}(Q_{\mathrm{m}},\max\{s_2-2+c_2,-M\}),\\
&\quad \widetilde{J}_{\beta}^{(k)}(s_1+c_1+1,s_2+c_2-2),\widetilde{J}_{\beta}^{(k)}(s_1+c_1,s_2+c_2-1)\Big\}\\
&\qquad -\widetilde{J}_{\beta}^{(k)}(Q_{\mathrm{m}},s_2-1+c_2)\\
&=\widetilde{J}_{\beta}^{(k)}\big(u_k(s_1,s_2-2,c_1,c_2)\big)-\widetilde{J}_{\beta}^{(k)}\big(u_k(s_1,s_2-1,c_1,c_2)\big)
\end{aligned} \tag{B.21}$$

where $Q_{\mathrm{m}} := \min\{Q_{\mathrm{t}}, s_1+c_1\}$. The equalities are due to the definition of u_k and the assumption that $\max\{0,-(s_1+c_1)\}+\max\{0,-(s_2-1+c_2)\} \leqslant M$. The inequality can be verified using (B.6) and (B.7) for the case k and Lemma B.6. For example, if

$\widetilde{J}_\beta^{(k)}(s_1+c_1+1, s_2+c_2-2)$ achieves the minimum in (B.21), then $s_1+c_1+1 \leqslant 0$ and

$$\begin{aligned}
&\widetilde{J}_\beta^{(k)}(Q_{\mathrm{m}}, s_2-1+c_2) - \widetilde{J}_\beta^{(k)}(Q_{\mathrm{m}}, s_2+c_2) \\
&= \widetilde{J}_\beta^{(k)}(s_1+c_1, s_2-1+c_2) - \widetilde{J}_\beta^{(k)}(s_1+c_1, s_2+c_2) \\
&\leqslant \widetilde{J}_\beta^{(k)}(s_1+c_1+1, s_2-1+c_2) - \widetilde{J}_\beta^{(k)}(s_1+c_1, s_2+c_2) \\
&\leqslant \widetilde{J}_\beta^{(k)}(s_1+c_1+1, s_2+c_2-2) - \widetilde{J}_\beta^{(k)}(s_1+c_1, s_2-1+c_2) \\
&= \widetilde{J}_\beta^{(k)}(s_1+c_1+1, s_2+c_2-2) - \widetilde{J}_\beta^{(k)}(Q_{\mathrm{m}}, s_2-1+c_2)
\end{aligned}$$

where the first inequality is because of Lemma B.6 and the second inequality is because of (B.6). If the minimum in (B.21) is achieved by other terms, the discussions are similar.

If $\max\{0, -(s_1+c_1)\} + \max\{0, -(s_2-1+c_2)\} = M+1$, then $c_1 = -1$ and $s_1 \leqslant 0$. One can verify that $-s_1 > M$. As a consequence, then $c_2 = -1$, $s_2+c_2-1 < 0$, and

$$\begin{aligned}
&\widetilde{J}_\beta^{(k)}(u_k(s_1, s_2-1, c_1, c_2)) - \widetilde{J}_\beta^{(k)}(u_k(s_1, s_2, c_1, c_2)) \\
&\leqslant \min\left\{\widetilde{J}_\beta^{(k)}(s_1-1, s_2-1), \widetilde{J}_\beta^{(k)}(s_1, s_2-2), \widetilde{J}_\beta^{(k)}(s_1+1, s_2-3)\right\} \\
&\qquad - \min\left\{\widetilde{J}_\beta^{(k)}(s_1, s_2-2), \widetilde{J}_\beta^{(k)}(s_1-1, s_2-1)\right\} \\
&= \widetilde{J}_\beta^{(k)}(u_k(s_1, s_2-2, c_1, c_2) - \widetilde{J}_\beta^{(k)}(u_k(s_1, s_2-1, c_1, c_2))
\end{aligned}$$

where the inequality can be verified by discussing all the cases and using (B.19), (B.20), and (B.7) for the case k, as well as Lemma B.6. This proves (B.20).

We now prove (B.7).

$$
\begin{aligned}
&H^{(k+1)}(s_1, s_2) \\
&= \Big[\widetilde{J}_\beta^{(k+1)}(s_1, s_2 - 1) - \widetilde{J}_\beta^{(k+1)}(s_1, s_2)\Big] - \Big[\widetilde{J}_\beta^{(k+1)}(s_1, s_2 - 2) - \widetilde{J}_\beta^{(k+1)}(s_1, s_2 - 1)\Big] \\
&= 2\max\{0, s_2 - 1\} - \max\{0, s_2\} - \max\{0, s_2 - 2\} + 2\mathbb{E}\Big\{\widetilde{J}_\beta^{(k)}\big(u_k(s_1, s_2 - 1, \mathsf{c}_1, \mathsf{c}_2)\big)\Big\} \\
&\quad - \mathbb{E}\Big\{\widetilde{J}_\beta^{(k)}\big(u_k(s_1, s_2 - 2, \mathsf{c}_1, \mathsf{c}_2)\big)\Big\} - \mathbb{E}\Big\{\widetilde{J}_\beta^{(k)}\big(u_k(s_1, s_2, \mathsf{c}_1, \mathsf{c}_2)\big)\Big\} \\
&\leqslant \beta \sum_{c_1, c_2 \in \{-1, 0, +1\}} \mathbb{P}\{\mathsf{c}_1^{(1)} = c_1, \mathsf{c}_2^{(1)} = c_2\} \times \Big[J_\beta^{(k)}\big(u_k(s_1, s_2 - 1, c_1, c_2)\big) \\
&\quad - J_\beta^{(k)}\big(u_k(s_1, s_2, c_1, c_2)\big) - J_\beta^{(k)}\big(u_k(s_1, s_2 - 2, c_1, c_2)\big) + J_\beta^{(k)}\big(u_k(s_1, s_2 - 1, c_1, c_2)\big)\Big] \\
&\leqslant 0
\end{aligned}
$$

where the inequality is due to $2\max\{0, s_2 - 1\} - \max\{0, s_2\} - \max\{0, s_2 - 2\} \leqslant 0$ and (B.20). This proves (B.7) for the case $k + 1$.

Proof of (B.14) for the case $k + 1$:

We first show that

$$
\widetilde{J}_\beta^{(k)}\big(u_k(s_1, 1, c_1, -1)\big) - \widetilde{J}_\beta^{(k)}(u_k(s_1, 0, c_1, -1)) \leqslant \nu_1^{(k)} \tag{B.22}
$$

If $s_1 + c_1 \leqslant -M$, then

$$
\begin{aligned}
&\widetilde{J}_\beta^{(k)}\big(u_k(s_1, 1, c_1, -1)\big) - \widetilde{J}_\beta^{(k)}\big(u_k(s_1, 0, c_1, -1)\big) \\
&= \widetilde{J}_\beta^{(k)}(-M, 0) - \min\Big\{\widetilde{J}_\beta^{(k)}(-M, 0), \widetilde{J}_\beta^{(k)}(-M + 1, -1)\Big\} \\
&\leqslant \widetilde{J}_\beta^{(k)}(-M, 1) - \widetilde{J}_\beta^{(k)}(-M, 0) \\
&\leqslant \nu_1^{(k)}
\end{aligned}
$$

where the first inequality is due to $F^{(k)}(-M + 1, 1) \leqslant 0$ from (B.10) and Lemma B.6, and the second inequality is due to (B.14).

If $s_1 + c_1 > -M$, then

$$\begin{aligned}
&\widetilde{J}_\beta^{(k)}\big(u_k(s_1, 1, c_1, -1)\big) - \widetilde{J}_\beta^{(k)}\big(u_k(s_1, 0, c_1, -1)\big) \\
&= \widetilde{J}_\beta^{(k)}(Q_{\mathrm{m}}, 0) - \widetilde{J}_\beta^{(k)}(Q_{\mathrm{m}}, -1) \\
&\leqslant \widetilde{J}_\beta^{(k)}(Q_{\mathrm{m}}, 1) - \widetilde{J}_\beta^{(k)}(Q_{\mathrm{m}}, 0) \\
&\leqslant \nu_1^{(k)}
\end{aligned}$$

where the first inequality is due to (B.11) and the second inequality is due to (B.14). Hence, we have proved (B.22).

Next, for $s_2 = 1$, we prove (B.14) for the case $k+1$:

$$\begin{aligned}
&\widetilde{J}_\beta^{(k+1)}(s_1, 1) - \widetilde{J}_\beta^{(k+1)}(s_1, 0) \\
&= 1 + \beta \sum_{c_1, c_2 \in \{-1,0,+1\}} \mathbb{P}\{\mathsf{c}_1^{(1)} = c_1, \mathsf{c}_2^{(1)} = c_2\} \times \Big[\widetilde{J}_\beta^{(k)}\big(u_k(s_1, 1, c_1, c_2)\big) \\
&\qquad - \widetilde{J}_\beta^{(k)}\big(u_k(s_1, 0, c_1, c_2)\big)\Big] \\
&= 1 + \beta \sum_{c_1 \in \{-1,0,+1\}, c_2 = 1} \tilde{p}_2 \Big[\widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, 2\big) - \widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, 1\big)\Big] \\
&\qquad + \beta \sum_{c_1 \in \{-1,0,+1\}, c_2 = 0} (1 - \tilde{p}_2 - \tilde{q}_2) \Big[\widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, 1\big) - \widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, 0\big)\Big] \\
&\qquad + \beta \sum_{c_1 \in \{-1,0,+1\}, c_2 = -1} \tilde{q}_2 \Big[\widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, 0\big) - \widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, -1\big)\Big] \\
&\leqslant 1 + \beta \tilde{p}_2 \nu_2^{(k)} + \beta(1 - \tilde{p}_2 - \tilde{q}_2)\nu_1^{(k)} + \beta \tilde{q}_2 \nu_1^{(k)} \\
&= 1 + \beta \tilde{p}_2 \nu_2^{(k)} + \beta(1 - \tilde{p}_2)\nu_1^{(k)} \\
&= \nu_1^{(k+1)}
\end{aligned}$$

where the inequality is due to (B.14) for the case k and (B.22).

Next, for $s_2 > 1$, we prove (B.14) for the case $k+1$:

$$
\begin{aligned}
&\widetilde{J}_\beta^{(k+1)}(s_1, s_2) - \widetilde{J}_\beta^{(k+1)}(s_1, s_2 - 1) \\
&= 1 + \beta \sum_{c_1, c_2 \in \{-1,0,1\}} \mathbb{P}\{\mathsf{c}_1^{(1)} = c_1, \mathsf{c}_2^{(1)} = c_2\} \times \Big[\widetilde{J}_\beta^{(k)}\big(u_k(s_1, s_2, c_1, c_2)\big) \\
&\qquad - \widetilde{J}_\beta^{(k)}\big(u_k(s_1, s_2 - 1, c_1, c_2)\big) \Big] \\
&= 1 + \beta \sum_{c_1 \in \{-1,0,+1\}, c_2 = 1} \tilde{p}_2 \Big[\widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, s_2 + 1\big) - \widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, s_2\big) \Big] \\
&\qquad + \beta \sum_{c_1 \in \{-1,0,+1\}, c_2 = 0} (1 - \tilde{p}_2 - \tilde{q}_2) \Big[\widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, s_2\big) - \widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, s_2 - 1\big) \Big] \\
&\qquad + \beta \sum_{c_1 \in \{-1,0,+1\}, c_2 = -1} \tilde{q}_2 \Big[\widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, s_2 - 1\big) - \widetilde{J}_\beta^{(k)}\big(Q_{\mathrm{m}}, s_2 - 2\big) \Big] \\
&\leqslant 1 + \beta \tilde{p}_2 \nu_{s_2+1}^{(k)} + \beta(1 - \tilde{p}_2 - \tilde{q}_2) \nu_{s_2}^{(k)} + \beta q_2 \nu_{s_2-1}^{(k)} \\
&= \nu_{s_2}^{(k+1)}
\end{aligned}
$$

We then finish the proof of (B.14) for the case $k+1$.

Appendix C

Proofs of the Results in Chapter 4

C.1 Proof of Theorem 4.1

Consider the RED protocol π^* that achieves the optimal entanglement rate. Let $\mathsf{F}_{i:j}^{i:k}(\tau)$ and $\mathsf{F}_{i:j}^{k:j}(\tau)$ denote the number of entangled qubit pairs (EQPs) $\boldsymbol{\Xi}_{i:k}$ and $\boldsymbol{\Xi}_{k:j}$ used for distributing $\boldsymbol{\Xi}_{i:j}$ after τ time slots, respectively; let $\mathsf{h}_{ij}^{ijk}(\tau)$ denote the number of EQPs $\boldsymbol{\Xi}_{i:j}$ distributed by entanglement swapping that consumes $\boldsymbol{\Xi}_{i:k}$ and $\boldsymbol{\Xi}_{k:j}$ after τ time slots. If $(i,j) \in \mathcal{E}$, let $\mathsf{P}_{i:j}(\tau)$ denote the number of EQPs $\boldsymbol{\Xi}_{i:j}$ generated in Phase I after τ time slots. At the time slot T, since the number of EQPs $\boldsymbol{\Xi}_{i:j}$ is nonnegative, we have

$$1_{\mathcal{E}}(i,j)\mathsf{p}_{i:j}(T) + \sum_{k\in\mathcal{N}\setminus\{i,j\}} \mathsf{h}_{ij}^{ijk}(T) \geqslant \sum_{k\in\mathcal{N}\setminus\{i,j\}} (\mathsf{F}_{i:k}^{i:j}(T) + \mathsf{F}_{k:j}^{i:j}(T)).$$

Taking the expectation on both sides and dividing them by T, we have

$$1_{\mathcal{E}}(i,j)p_{i:j} + \frac{1}{T}\sum_{k\in\mathcal{N}\setminus\{i,j\}} \mathbb{E}\big\{\mathsf{h}_{ij}^{ijk}(T)\big\} \geqslant \frac{1}{T}\sum_{k\in\mathcal{N}\setminus\{i,j\}} \big(\mathbb{E}\big\{\mathsf{F}_{i:k}^{i:j}(T)\big\} + \mathbb{E}\big\{\mathsf{F}_{k:j}^{i:j}(T)\big\}\big).$$

Note that the success probability for distributing $\boldsymbol{\Xi}_{i:j}$ with $\boldsymbol{\Xi}_{i:k}$ and $\boldsymbol{\Xi}_{k:j}$ is q_k, and $\mathsf{F}_{i:j}^{i:k}(T) = \mathsf{F}_{i:j}^{k:j}(T)$. Hence,

$$\mathbb{E}\big\{\mathsf{h}_{ij}^{ijk}(T)\big\} = q_k \frac{\mathbb{E}\big\{\mathsf{F}_{i:j}^{i:k}(T)\big\} + \mathbb{E}\big\{\mathsf{F}_{i:j}^{k:j}(T)\big\}}{2}$$

and we have

$$\begin{aligned} 1_{\mathcal{E}}(i,j)p_{i:j} + \sum_{k\in\mathcal{N}\setminus\{i,j\}} q_k \frac{\mathbb{E}\big\{\mathsf{F}_{i:j}^{i:k}(T)\big\} + \mathbb{E}\big\{\mathsf{F}_{i:j}^{k:j}(T)\big\}}{2T} \\ \geqslant \frac{1}{T} \sum_{k\in\mathcal{N}\setminus\{i,j\}} \big(\mathbb{E}\big\{\mathsf{F}_{i:k}^{i:j}(T)\big\} + \mathbb{E}\big\{\mathsf{F}_{k:j}^{i:j}(T)\big\}\big). \end{aligned} \tag{C.1}$$

Let

$$f_{i:j}^{i:k} = \frac{1}{T}\mathbb{E}\{\mathsf{F}_{i:j}^{i:k}(T)\}.$$

Evidently, $f_{i:j}^{i:k}$ satisfies the definition of the eflow. Moreover, $\{f_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ defined above satisfies the constraint (4.2) using (C.1). One can also easily verify that $\{f_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ satisfies the constraints (4.3)-(4.4). Therefore, $\{f_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ is a feasible solution of $\mathscr{P}$.

Note that

$$\sum_{\tau=1}^{T} \mathsf{g}_{s:t}^{\pi^*}(\tau) = T \cdot 1_{\mathcal{E}}(s,t)\mathsf{p}_{s:t}(T) + \sum_{k\in\mathcal{N}\setminus\{s,t\}} \mathsf{h}_{st}^{stk}(T)$$

and hence

$$\begin{aligned} \frac{1}{T}\sum_{\tau=1}^{T} \mathbb{E}\{\mathsf{g}_{s:t}^{\pi^*}(\tau)\} &= 1_{\mathcal{E}}(s,t)\mathbb{E}\{\mathsf{p}_{s:t}(T)\} + \frac{1}{T}\sum_{k\in\mathcal{N}\setminus\{s,t\}} \mathbb{E}\{\mathsf{h}_{st}^{stk}(T)\} \\ &= 1_{\mathcal{E}}(s,t)p_{s:t} + \frac{1}{T}\sum_{k\in\mathcal{N}\setminus\{s,t\}} q_k \frac{\mathbb{E}\big\{\mathsf{F}_{s:t}^{s:k}(T)\big\} + \mathbb{E}\big\{\mathsf{F}_{s:t}^{k:t}(T)\big\}}{2} \\ &= 1_{\mathcal{E}}(s,t)p_{s:t} + \sum_{k\in\mathcal{N}\setminus\{s,t\}} q_k \frac{f_{s:t}^{s:k} + f_{s:t}^{k:t}}{2} \end{aligned}$$

which is exactly the objective function of $\mathscr{P}$. Let λ^* denote the optimal value for $\mathscr{P}$. Since λ^* is the upper bound for the objective function that corresponds to a feasible solution, we have

$$\frac{1}{T}\sum_{\tau=1}^{T}\mathbb{E}\{\mathbf{g}_{s:t}^{\pi^*}(\tau)\} \leqslant \lambda^*$$

for all T and π^*. Taking the lim inf over T on the left side, we arrive at the desired result.

C.2 Proof of Proposition 4.1

Consider the optimal solution $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ such that the graph $\mathcal{G}$ corresponding to $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ has the minimum number of edges. If the graph $\mathcal{G}$ has no directed cycles, the proof is finished. Otherwise, we will find another optimal solution of $\mathscr{P}$ and the corresponding graph of the new solution has fewer edges compared to $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$. This will lead to a contradiction and finish the proof.

Consider one of the directed cycles in $\mathcal{G}$ as in Fig. C-1, consisting of nodes $e_{x_{2l-1}:x_{2l}}$, $l \in \mathcal{K}_{1:K}$, where K is the number of nodes in the cycle. For notational convenience, let $x_1 = x_{2K+1}$ and $x_2 = x_{2K+2}$. Note that the structure of entanglement swapping requires that $\text{Card}(\{x_{2l-1}, x_{2l}\} \cap \{x_{2l+1}, x_{2l+2}\}) = 1$ and $\text{Card}(\{y_{2l-1}, y_{2l}\} \cap \{x_{2l+1}, x_{2l+2}\}) = 1$, where y_{2l-1} and y_{2l} are shown in Fig. C-1. From the definition of the edge, we have that $f_{x_{2l+1}:x_{2l+2}}^{x_{2l-1}:x_{2l}} > 0$, $l \in \mathcal{K}_{1:K}$.

Consider a number δ:

$$\delta = \min\left\{f_{x_{2l+1}:x_{2l+2}}^{x_{2l-1}:x_{2l}}, l \in \mathcal{K}_{1:K}\right\}. \tag{C.2}$$

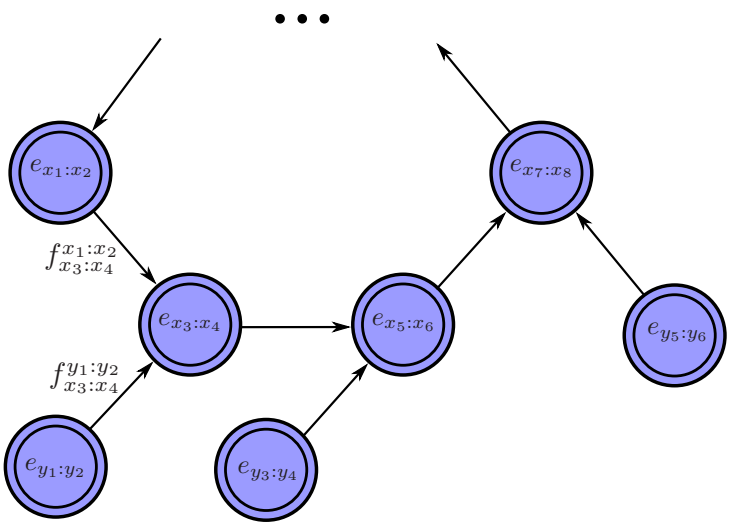


Figure C-1: One of the directed cycles in $\mathcal{G}$.

Evidently $\delta > 0$. We construct a solution of $\mathscr{P}$ as follows:

$$\begin{aligned}
\tilde{f}^{x_{2l-1}:x_{2l}}_{x_{2l+1}:x_{2l+2}} &= f^{x_{2l-1}:x_{2l}}_{x_{2l+1}:x_{2l+2}} - \delta,\ l \in \mathcal{K}_{1:K} \\
\tilde{f}^{y_{2l-1}:y_{2l}}_{x_{2l+1}:x_{2l+2}} &= f^{y_{2l-1}:y_{2l}}_{x_{2l+1}:x_{2l+2}} - \delta,\ l \in \mathcal{K}_{1:K} \\
\tilde{f}^{i:j}_{i:k} &= f^{i:j}_{i:k},\ \text{for other eflows.}
\end{aligned}$$

We next show $\{\tilde{f}^{i:k}_{i:j} : i,j,k \in \mathcal{N}\}$ is an optimal solution of $\mathscr{P}$. Regarding the constraint (4.2), if $\{i,j\} = \{x_{2l-1}, x_{2l}\}$,

$$\begin{aligned}
& p_{i:j} 1_{\mathcal{E}}(i,j) + \sum_{k \in \mathcal{N}\setminus\{i,j\}} q_k \frac{\tilde{f}^{i:k}_{i:j} + \tilde{f}^{k:j}_{i:j}}{2} \\
&= 1_{\mathcal{E}}(i,j) p_{i:j} - \delta \cdot q_{\{x_{2l-3},x_{2l-2}\}\cap\{x_{2l-1},x_{2l}\}} \\
&\quad + \sum_{k \in \mathcal{N}\setminus\{i,j\}} q_k \frac{f^{i:k}_{i:j} + f^{k:j}_{i:j}}{2} \\
&\geqslant -\delta \cdot q_{\{x_{2l-3},x_{2l-2}\}\cap\{x_{2l-1},x_{2l}\}} + \sum_{k \in \mathcal{N}\setminus\{i,j\}} (f^{i:j}_{i:k} + f^{i:j}_{k:j}) \\
&\geqslant -\delta + \sum_{k \in \mathcal{N}\setminus\{i,j\}} (f^{i:j}_{i:k} + f^{i:j}_{k:j}) \\
&= \sum_{k \in \mathcal{N}\setminus\{i,j\}} (\tilde{f}^{i:j}_{i:k} + \tilde{f}^{i:j}_{k:j})
\end{aligned}$$

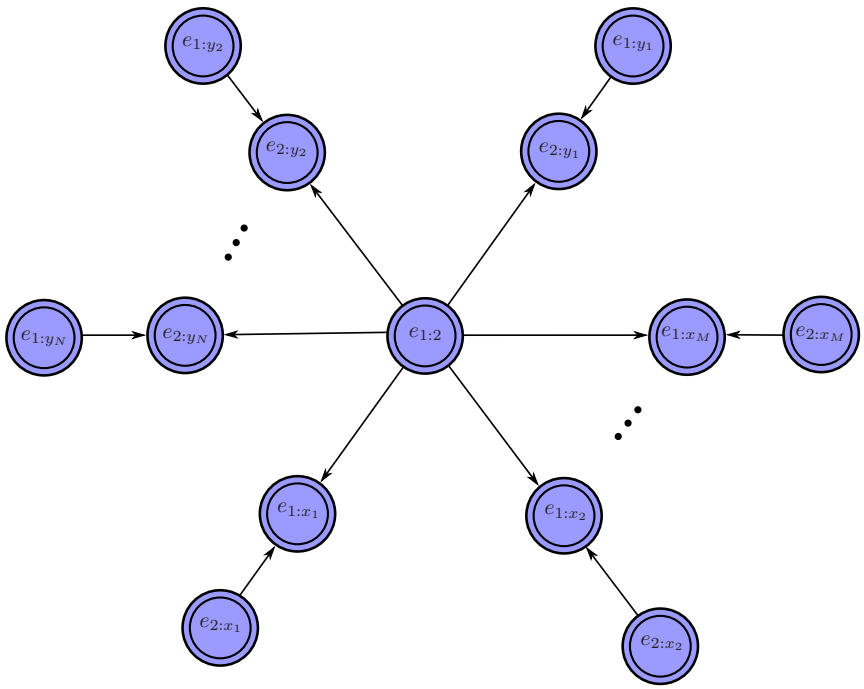


Figure C-2: Children of the non-isolated enode $e_{1:2}$.

where the first inequality is because $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ satisfies the constraint (4.2) and the second inequality is because $q_k \leqslant 1$ for all $k \in \mathcal{N}$. If $\{i, j\} = \{y_{2l-1}, y_{2l}\}$,

$$\begin{aligned}
&p_{i:j} 1_{\mathcal{E}}(i,j) + \sum_{k \in \mathcal{N} \setminus \{i,j\}} q_k \frac{\tilde{f}_{i:j}^{i:k} + \tilde{f}_{i:j}^{k:j}}{2} \\
&\quad = 1_{\mathcal{E}}(i,j) p_{i:j} + \sum_{k \in \mathcal{N} \setminus \{i,j\}} q_k \frac{f_{i:j}^{i:k} + f_{i:j}^{k:j}}{2} \\
&\quad \geqslant -\delta + \sum_{k \in \mathcal{N} \setminus \{i,j\}} (f_{i:k}^{i:j} + f_{k:j}^{i:j}) \\
&\quad = \sum_{k \in \mathcal{N} \setminus \{i,j\}} (\tilde{f}_{i:k}^{i:j} + \tilde{f}_{k:j}^{i:j})
\end{aligned}$$

where the inequality is because $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ satisfies the constraint (4.2) and the fact $\delta > 0$. If $\{i, j\} \neq \{x_{2l-1}, x_{2l}\}$ and $\{i, j\} \neq \{y_{2l-1}, y_{2l}\}$, then the constraint (4.2) trivially holds. Regarding the constraint (4.3), since $f_{x_{2l+1}x_{2l+2}}^{x_{2l-1}x_{2l}} = f_{x_{2l+1}x_{2l+2}}^{y_{2l-1}y_{2l}} > 0$ and δ is selected according to (C.2), the constraint (4.3) holds.

The constraint (4.4) also trivially holds since $f_{i:j}^{s:t} = 0$ and the enode $e_{s:t}$ does not

belong to the set $\{e_{x_{2l-1}:x_{2l}}, e_{y_{2l-1}:y_{2l}}, l \in \mathcal{K}_{1:K}\}$. Using the same argument, one can show that the value of the objective function remains unchanged if $f_{i:j}^{i:k}$ is replaced with $\tilde{f}_{i:j}^{i:k}$. This proves that $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ is an optimal solution of $\mathscr{P}$.

One can easily find that replacing $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ with $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ does not add additional edges in the corresponding graphs. Moreover, due to (C.2), at least one of the elements in $\left\{\tilde{f}_{x_{2l+1}:x_{2l+2}}^{x_{2l-1}:x_{2l}}, l \in \mathcal{K}_{1:K}\right\}$ is zero. Consequently, the graph corresponding to the optimal solution $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ has fewer edges than that corresponding to $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$. This gives the desired contradiction that $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ has the minimum number of edges and finishes the proof.

C.3 Proof of Proposition 4.2

Consider the optimal solution $\{\mathring{f}_{i:j}^{i:k}\}_{i,j,k\in\mathcal{N}}$ that has the minimum number of edges. Proposition 4.1 shows that the associated graph is acyclic. We will prove that this optimal solution is efficient by contradiction.

Suppose there exists a non-isolated enode $e_{1:2}$ in $\mathcal{G}$ that is not an ancestor of $e_{s:t}$. Consider the children of $e_{1:2}$, denoted as $e_{1:x_1}, e_{1:x_2}, \ldots, e_{1:x_M}$, $e_{2:y_1}, e_{2:y_2}, \ldots, e_{2:y_N}$ shown in Fig. C-2. We construct a solution $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ as follows:

$$\tilde{f}_{i:j}^{i:k} = \begin{cases} 0, & \text{if } e_{i:k} \text{ is a descendant of } e_{1:2}; \\ 0 & \text{if } e_{i:j} \text{ is a descendant of } e_{1:2}; \\ 0, & \text{if } e_{i:j} = e_{1:2}; \\ f_{i:j}^{i:k}, & \text{otherwise.} \end{cases}$$

Evidently, the graph $\widetilde{\mathcal{G}}$ corresponding to $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ can be obtained from $\mathcal{G}$ by removing some edges so that the enode $e_{1:2}$ does not have descendants in $\widetilde{\mathcal{G}}$. We next show that $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ is an optimal solution of $\mathscr{P}$. One can find that the constraints (4.2)-(4.4) trivially hold. Regarding the objective function, consider $f_{s:t}^{s:k}$ for some $k \in \mathcal{N}$. If $f_{s:t}^{s:k} > 0$, then $e_{s:t}$ is a descendant of $e_{s:k}$. Since $e_{s:t}$ is not a descendant of $e_{1:2}$, $e_{s:k}$ is not a descendant of $e_{1:2}$. Consequently, $\tilde{f}_{s:t}^{s:k} = f_{s:t}^{s:k}$.

If $f_{s:t}^{s:k} = 0$, then $\tilde{f}_{s:t}^{s:k} = f_{s:t}^{s:k}$ trivially. This shows that $\tilde{f}_{s:t}^{s:k} = f_{s:t}^{s:k}$ for all $k \in \mathcal{N}$. Similarly, $\tilde{f}_{s:t}^{k:t} = f_{s:t}^{k:t}$ for all $k \in \mathcal{N}$. As a result,

$$\sum_{k \in \mathcal{N} \setminus \{s,t\}} q_k \frac{\tilde{f}_{s:t}^{s:k} + \tilde{f}_{s:t}^{k:t}}{2} = \sum_{k \in \mathcal{N} \setminus \{s,t\}} q_k \frac{f_{s:t}^{s:k} + f_{s:t}^{k:t}}{2}$$

showing that $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ is the optimal solution of $\mathscr{P}$. However, the graph $\widetilde{\mathcal{G}}$ corresponding to $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ has fewer edges than $\mathcal{G}$, leading to the desired contradiction.

C.4 Proof of Theorem 4.2

The proof of Proposition 4.1 and Proposition 4.2 shows that we can construct an optimal solution of $\mathscr{P}$, denoted as $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$, such that the graph $\mathcal{G}$ corresponding to $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ is acyclic and efficient. We will show that given $\{f_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$, we can provide an optimal solution of $\mathscr{P}_{\rm s}$, denoted as $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ and $\{\tilde{u}_{i:j} : (i, j) \in \mathcal{E}\}$.

We will start from the solution $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ and $\{\tilde{u}_{i:j} : (i, j) \in \mathcal{E}\}$, where $\tilde{f}_{i:j}^{i:k} = f_{i:k}^{i:j}$, $\forall i, j, k$ and $\tilde{u}_{i:j} = 1$, $\forall i, j$. Evidently, the solution $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ satisfies the constraints (4.2)-(4.4), but $\{\tilde{f}_{i:j}^{i:k} : i, j, k \in \mathcal{N}\}$ and $\{\tilde{u}_{i:j} : (i, j) \in \mathcal{E}\}$ may not satisfy the constraints in $\mathscr{P}_{\rm s}$. We next update each of the enodes. Here, for an enode $e_{i:j}$, updating $e_{i:j}$ means updating the incoming eflow of $e_{i:j}$ $\tilde{f}_{i:j}^{i:k}$, $k \in \mathcal{N} \setminus \{i, j\}$ and determining $\tilde{u}_{i:j}$. After updating an enode, we will make sure that three requirements are satisfied: 1) conditions (4.5)-(4.8) hold for all the updated enodes; 2) conditions (4.2)-(4.4) hold for all the nodes; and 3) the value of the objective function remains unchanged. In this way, after updating all the enodes, we will obtain a solution of $\mathscr{P}_{\rm s}$ with the same optimal value of $\mathscr{P}$.

We first update isolated enodes. If $e_{i:j}$ is an isolated enode in $\mathcal{G}$, then we set $\tilde{f}_{i:k}^{i:j} = \tilde{f}_{k:j}^{i:j} = \tilde{f}_{i:j}^{i:k} = \tilde{f}_{i:j}^{k:j} = 0$ and $\tilde{u}_{i:j} = 0$. Note that such an update satisfies the three requirements in the previous paragraph.

We then update the ancestors of $e_{s:t}$. Since $\mathcal{G}$ is acyclic, we can determine a

topological ordering of $\mathcal{G}$ starting from $e_{s:t}$ and update the nodes with the order. For an enode $e_{i:j}$, since the condition (4.2) holds as required, we have

$$\begin{aligned} p_{i:j}1_{\mathcal{E}}(i,j) + \sum_{k\in\mathcal{N}\setminus\{i,j\}} q_k \frac{\tilde{f}_{i:j}^{i:k} + \tilde{f}_{i:j}^{k:j}}{2} \\ \geqslant \sum_{k\in\mathcal{N}\setminus\{i,j\}} (\tilde{f}_{i:k}^{i:j} + \tilde{f}_{k:j}^{i:j}),\ i,j\in\mathcal{N}. \end{aligned}$$

We then determine $\tilde{u}_{i:j}$ and update the incoming flow $\tilde{f}_{i:j}^{i:k}$ and $\tilde{f}_{i:j}^{k:j}$ as follows:

$$\begin{aligned} \varrho_{i:j} &\to \frac{\sum_{k\in\mathcal{N}\setminus\{i,j\}}(\tilde{f}_{i:k}^{i:j} + \tilde{f}_{k:j}^{i:j})}{1_{\mathcal{E}}(i,j)p_{i:j} + \sum_{k\in\mathcal{N}\setminus\{i,j\}} q_k(\tilde{f}_{i:j}^{i:k} + \tilde{f}_{i:j}^{k:j})/2} \\ \tilde{u}_{i:j} &\to \varrho_{i:j},\quad \tilde{f}_{i:j}^{i:k} \to \varrho_{i:j}\tilde{f}_{i:j}^{i:k},\quad \tilde{f}_{i:j}^{k:j} \to \varrho_{i:j}\tilde{f}_{i:j}^{k:j}. \end{aligned}$$

One can easily verify that such an update satisfies the three requirements. This finishes the proof.

C.5 Proof of Proposition 4.3

We show that

$$h(N) \geqslant g(N) \tag{C.3}$$

for all $N \in \mathbb{N}_+$ by induction. The base case with $N = 1$ can be easily verified since $h(1) \geqslant 1$. For the induction step, suppose (C.3) holds for $N = 1, 2, \ldots, N_1$. We next show that (C.3) holds for $N = N_1 + 1$.

Evidently, the EQP $\boldsymbol{\Xi}_{0:N_1+1}$ is distributed based on entanglement swapping between EQPs $\boldsymbol{\Xi}_{0:a}$ and $\boldsymbol{\Xi}_{a:N_1+1}$ for some $a \in \mathcal{K}_{1:N_1}$. Let x_a denote the fraction of the EQPs $\boldsymbol{\Xi}_{0:N_1+1}$ that is distributed based on entanglement swapping between $\boldsymbol{\Xi}_{0:a}$ and $\boldsymbol{\Xi}_{a:N_1+1}$. For a fixed a, we consider the expected numbers of crude entanglements required to distribute one EQP $\boldsymbol{\Xi}_{0:a}$ and one EQP $\boldsymbol{\Xi}_{a:N_1+1}$. These two numbers can

be lower-bounded by $h(a)$ and $h(N_1+1-a)$. Then

$$\begin{aligned} h(N_1+1) &\geqslant \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}} x_a\Big[h(a)+h(N_1+1-a)\Big] \\ &\geqslant \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}} x_a\Big[g(a)+g(N_1+1-a)\Big] \\ &\geqslant \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}} x_a\cdot q\cdot g(N_1+1) \\ &= \sum_{a\in\mathcal{K}_{1:N_1}} x_a g(N_1+1) \\ &= g(N_1+1) \end{aligned}$$

where the first inequality has been explained, the second inequality is because of the induction hypothesis, the third inequality is because of Lemma C.1 below, and the second equality is because $\sum_{a\in\mathcal{K}_{1:N_1}} x_a = 1$. This completes the proof for (C.3) for all $N\in\mathbb{N}_+$.

Lemma C.1. For K, $M\in\mathbb{N}_+$,

$$\frac{g(K)+g(M)}{q} \geqslant g(K+M). \tag{C.4}$$

Proof. See Appendix C.6. □

C.6 Proof of Lemma C.1

Without loss of generality, we assume $K\leqslant M$. We prove (C.4) in three cases: $K=M$; $K=M-1$; and $K\leqslant M-2$.

Case 1: $K=M$.

Let $k = \lceil \log_2 K \rceil - 1$. Then

$$\begin{aligned}
\frac{g(K) + g(M)}{q} &= \frac{2}{q} \cdot g(K) \\
&= \frac{2}{q} \cdot \frac{2(K - 2^k) + (2^{k+1} - K)q}{q^{k+1}} \\
&= \frac{2(2K - 2^{k+1}) + (2^{k+2} - 2K)q}{q^{k+2}} \\
&= g(2K) \\
&= g(K + M)
\end{aligned}$$

where the fourth equality is because $k = \lceil \log_2 2K \rceil - 2$. Therefore, the inequality (C.4) holds.

Case 2: $K = M - 1$.

Let $k = \lceil \log_2 K \rceil - 1$. We discuss two subcases: $K = 2^{k+1}$ and $K < 2^{k+1}$. If $K = 2^{k+1}$, then

$$\begin{aligned}
\frac{g(K) + g(M)}{q} &= \frac{2(K - 2^k) + (2^{k+1} - K)q}{q^{k+2}} \\
&\quad + \frac{2(K + 1 - 2^{k+1}) + (2^{k+2} - K - 1)q}{q^{k+3}} \\
&= \frac{2^{k+1}}{q^{k+2}} + \frac{2 + (2^{k+1} - 1)q}{q^{k+3}} \\
&= \frac{2 + (2^{k+2} - 1)q}{q^{k+3}} \\
&= \frac{2(2K + 1 - 2^{k+2}) + (2^{k+3} - 2K - 1)q}{q^{k+3}} \\
&= g(2K + 1) \\
&= g(K + M)
\end{aligned}$$

where the first equality is because $k + 1 = \lceil \log_2 M \rceil - 1$ and the fifth equality is

because $k+2 = \lceil \log_2 (K+M) \rceil - 1$. If $K < 2^{k+1}$, then

$$\begin{aligned}
\frac{g(K)+g(M)}{q} &= \frac{2(K-2^k)+(2^{k+1}-K)q}{q^{k+2}} \\
&\quad + \frac{2(K+1-2^k)+(2^{k+1}-K-1)q}{q^{k+2}} \\
&= \frac{2(2K+1-2^{k+1})+(2^{k+2}-2K-1)q}{q^{k+2}} \\
&= g(2K+1) \\
&= g(K+M)
\end{aligned}$$

where the third equality is because $k+1 = \lceil \log_2 (K+M) \rceil - 1$. Therefore, the inequality (C.4) holds.

Case 3: $K \leqslant M-2$.

Note that

$$\begin{aligned}
g(K+1) - g(K) &= \frac{2-q}{q^{\lceil \log_2 (K+1) \rceil}} \\
&\leqslant \frac{2-q}{q^{\lceil \log_2 M \rceil}} = g(M) - g(M-1) \qquad \text{(C.5)}
\end{aligned}$$

where the equalities can be verified by some calculation, and the inequality is because $K+1 \leqslant M$ and $2-q > 0$. The inequality (C.5) implies

$$g(K) + g(M) \geqslant g(K+1) + g(M-1).$$

Similarly, we have

$$g(K+1) + g(M-1) \geqslant g(K+2) + g(M-2)$$

provided that $K+1 \leqslant (M-1)-2$. Such a process can be repeated and we will have

$$g(K)+g(M) \geqslant \begin{cases} 2g\left(\dfrac{K+M}{2}\right) & \text{if } K+M \in \mathbb{Z}_{\mathrm{e}} \\ g\left(\dfrac{K+M-1}{2}\right)+g\left(\dfrac{K+M+1}{2}\right) & \text{if } K+M \in \mathbb{Z}_{\mathrm{o}}. \end{cases}$$

The results from Case 1 and Case 2 show that

$$q \cdot g(K+M) = \begin{cases} 2g\left(\dfrac{K+M}{2}\right) & \text{if } K+M \in \mathbb{Z}_{\mathrm{e}} \\ g\left(\dfrac{K+M-1}{2}\right)+g\left(\dfrac{K+M+1}{2}\right) & \text{if } K+M \in \mathbb{Z}_{\mathrm{o}}. \end{cases}$$

Combining the results above, we have that the inequality (C.4) holds in Case 3.

C.7 Proof of Theorem 4.3

In the homogeneous repeater chain, the expected number of available crude entanglements generated across N quantum channels is at most pN per time slot. As a consequence, the expected number of available crude entanglements generated across N quantum channels is at most pNT after T time slots. Moreover, by definition of $h(\cdot)$, it requires $h(N)$ crude entanglements on average to obtain one EQP between nodes s and t. Therefore, the expected number of EQPs between s and t is at most $pNT/h(N)$ after T time slots. This implies for any protocol π,

$$\begin{aligned} \lambda^{\pi} &= \liminf_{T\to\infty} \frac{1}{T} \sum_{\tau=1}^{T} \mathbb{E}\{\mathrm{g}_{s:t}^{\pi}(\tau)\} \\ &\leqslant \liminf_{T\to\infty} \frac{1}{T} \cdot \frac{pNT}{h(N)} = \frac{Np}{h(N)}. \end{aligned}$$

Together with Proposition 4.3, we have the upper bound of the EDR for any protocol:

$$\lambda^{\pi} \leqslant \frac{Np}{g(N)} = \frac{Npq^{n+1}}{2(N-2^n)+q(2^{n+1}-N)}.$$

Note that $\mathring{I}(s,t)$ in Remark 6 coincides with the upper bound above. Therefore, $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i,j) \in \mathcal{E}\}$ is the optimal solution of $\mathscr{P}_{\mathrm{s}}$, and this upper bound is indeed the maximum EDR because of Theorem 4.2.

C.8 Proof of Proposition 4.4

We next show that

$$h_{\mathrm{o}}(N) \geqslant g_{\mathrm{o}}(N) \tag{C.6}$$

$$h_{\mathrm{e}}(N) \geqslant g_{\mathrm{e}}(N) \tag{C.7}$$

for all $N \in \mathbb{N}_+$ by induction. The base case with $N = 1$ can be easily verified since $h_{\mathrm{o}}(1) \geqslant 1$, $h_{\mathrm{e}}(1) \geqslant 0$. For the induction step, suppose (C.6) and (C.7) hold for $N = 1, 2, \ldots, N_1$. We next show that (C.6) and (C.7) hold for $N = N_1 + 1$.

Evidently, the EQP $\boldsymbol{\Xi}_{0:N_1+1}$ is distributed based on entanglement swapping between EQPs $\boldsymbol{\Xi}_{0:a}$ and $\boldsymbol{\Xi}_{a:N_1+1}$ for some $a \in \mathcal{K}_{1:N_1}$. Let x_a denote the fraction of EQPs $\boldsymbol{\Xi}_{0:N_1+1}$ that is distributed based on entanglement swapping between $\boldsymbol{\Xi}_{0:a}$ and $\boldsymbol{\Xi}_{a:N_1+1}$. For a fixed a, we consider the expected numbers of odd crude entanglements required to distribute one entangled pair $\boldsymbol{\Xi}_{0:a}$ and one EQP $\boldsymbol{\Xi}_{a:N_1+1}$. If $a \in \mathbb{Z}_{\mathrm{o}}$, these two numbers can be lower-bounded by $h_{\mathrm{o}}(a)$ and $h_{\mathrm{e}}(N_1+1-a)$, respectively; if $a \in \mathbb{Z}_{\mathrm{e}}$, these two numbers can be lower-bounded by $h_{\mathrm{o}}(a)$ and $h_{\mathrm{o}}(N_1+1-a)$, respectively.

Then

$$
\begin{aligned}
h_{\mathrm{o}}(N_1+1) &\geqslant \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{o}}} x_a\Big[h_{\mathrm{o}}(a)+h_{\mathrm{e}}(N_1+1-a)\Big] \\
&\quad+\frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{e}}} x_a\Big[h_{\mathrm{o}}(a)+h_{\mathrm{o}}(N_1+1-a)\Big] \\
&\geqslant \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{o}}} x_a\Big[g_{\mathrm{o}}(a)+g_{\mathrm{e}}(N_1+1-a)\Big] \\
&\quad+\frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{e}}} x_a\Big[g_{\mathrm{o}}(a)+g_{\mathrm{o}}(N_1+1-a)\Big] \\
&\geqslant \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{o}}} x_a\cdot q\cdot g_{\mathrm{o}}(N_1+1) \\
&\quad+\frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{e}}} x_a\cdot q\cdot g_{\mathrm{o}}(N_1+1) \\
&= \sum_{a\in\mathcal{K}_{1:N_1}} x_a g_{\mathrm{o}}(N_1+1) \\
&= g_{\mathrm{o}}(N_1+1) \qquad\qquad (\mathrm{C}.8)
\end{aligned}
$$

where the first inequality has been explained, the second inequality is because of the induction hypothesis, the third inequality is because of (C.10) and (C.12) in Lemma

C.2 below, and the last equality is because $\sum_{a\in\mathcal{K}_{1:N_1}} x_a = 1$. Similarly,

$$\begin{aligned}
h_{\mathrm{e}}(N_1+1) &= \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{o}}} x_a\Big[h_{\mathrm{e}}(a)+h_{\mathrm{o}}(N_1+1-a)\Big] \\
&\quad + \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{e}}} x_a\Big[h_{\mathrm{e}}(a)+h_{\mathrm{e}}(N_1+1-a)\Big] \\
&\geqslant \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{o}}} x_a\Big[g_{\mathrm{e}}(a)+g_{\mathrm{o}}(N_1+1-a)\Big] \\
&\quad + \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{e}}} x_a\Big[g_{\mathrm{e}}(a)+g_{\mathrm{e}}(N_1+1-a)\Big] \\
&\geqslant \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{o}}} x_a\cdot q\cdot g_{\mathrm{e}}(N_1+1) \\
&\quad + \frac{1}{q}\sum_{a\in\mathcal{K}_{1:N_1}\cap\mathbb{Z}_{\mathrm{e}}} x_a\cdot q\cdot g_{\mathrm{e}}(N_1+1) \\
&= \sum_{a\in\mathcal{K}_{1:N_1}} x_a g_{\mathrm{e}}(N_1+1) \\
&= g_{\mathrm{e}}(N_1+1). \qquad \text{(C.9)}
\end{aligned}$$

Equations (C.8) and (C.9) complete the proof for (C.6) and (C.7) for all $N\in\mathbb{N}_+$.

Lemma C.2. For $K, M\in\mathbb{N}_+$. If $K\in\mathbb{Z}_{\mathrm{e}}$, then

$$\frac{g_{\mathrm{o}}(K)+g_{\mathrm{o}}(M)}{q} \geqslant g_{\mathrm{o}}(K+M) \qquad \text{(C.10)}$$

$$\frac{g_{\mathrm{e}}(K)+g_{\mathrm{e}}(M)}{q} \geqslant g_{\mathrm{e}}(K+M). \qquad \text{(C.11)}$$

If $K\in\mathbb{Z}_{\mathrm{o}}$, then

$$\frac{g_{\mathrm{o}}(K)+g_{\mathrm{e}}(M)}{q} \geqslant g_{\mathrm{o}}(K+M) \qquad \text{(C.12)}$$

$$\frac{g_{\mathrm{e}}(K)+g_{\mathrm{o}}(M)}{q} \geqslant g_{\mathrm{e}}(K+M). \qquad \text{(C.13)}$$

Sketch of the Proof. See Appendix C.9. □

C.9 Sketch of the Proof of Lemma C.2

We only prove the first inequality (C.10) and the proof for the remaining inequalities is similar.

Recall that $K \in \mathbb{Z}_{\mathrm{e}}$. We prove (C.10) in three cases: $M \in \mathbb{Z}_{\mathrm{e}}$; $M \in \mathbb{Z}_{\mathrm{o}}$ and $K \geqslant M-1$; and $M \in \mathbb{Z}_{\mathrm{o}}$ and $K < M-1$.

Case 1: $M \in \mathbb{Z}_{\mathrm{e}}$.

Since $K \in \mathbb{Z}_{\mathrm{e}}$ and $M \in \mathbb{Z}_{\mathrm{e}}$, $K+M \in \mathbb{Z}_{\mathrm{e}}$. Then

$$\begin{aligned}\frac{g_{\mathrm{o}}(K)+g_{\mathrm{o}}(M)}{q} &= \frac{g(K)+g(M)}{2q} \\ &\geqslant \frac{g(K+M)}{2} = g_{\mathrm{o}}(K+M)\end{aligned}$$

where the equalities are due to the definition of $g(\cdot)$ in (4.16), and the inequality is due to Lemma C.1.

Case 2: $M \in \mathbb{Z}_{\mathrm{o}}$ and $K \geqslant M-1$.

We have that

$$\begin{aligned}&\frac{g_{\mathrm{o}}(K)+g_{\mathrm{o}}(M)}{q} \\ &\quad= \frac{g_{\mathrm{o}}(K)+g_{\mathrm{o}}(M+1)}{q} + \frac{q-1}{q^{m+2}} \\ &\quad\geqslant g_{\mathrm{o}}(K+M+1) + \frac{q-1}{q^{m+2}} \\ &\quad= g_{\mathrm{o}}(K+M) + \frac{1-q}{q^{\lceil \log_2 (K+M) \rceil}} + \frac{q-1}{q^{m+2}} \\ &\quad\geqslant g_{\mathrm{o}}(K+M)\end{aligned}$$

where $m = \lceil \log_2 M \rceil - 1$. The equalities are due to the expression of $g_{\mathrm{o}}(\cdot)$ and the fact that $M+1 \in \mathbb{Z}_{\mathrm{e}}$ and $M+K \in \mathbb{Z}_{\mathrm{o}}$; the first inequality is due to the proof in Case 1; and the last inequality is because

$$\lceil \log_2(K+M) \rceil \geqslant \lceil \log_2(2M-1) \rceil = \lceil \log_2(2M) \rceil = m+2.$$

Case 3: $M \in \mathbb{Z}_{\mathrm{o}}$ and $K < M - 1$.

Since $K \in \mathbb{Z}_{\mathrm{e}}$ and $M \in \mathbb{Z}_{\mathrm{o}}$, we have $K \leqslant M - 3$.

$$\begin{aligned} & g_{\mathrm{o}}(K+2) - g_{\mathrm{o}}(K) \\ & \qquad = \frac{2-q}{2q^{\lceil \log_2(K+2) \rceil}} + \frac{2-q}{2q^{\lceil \log_2(K+1) \rceil}} \\ & \qquad \leqslant \frac{2-q}{q^{\lceil \log_2(K+2) \rceil}} \\ & \qquad \leqslant \frac{2-q}{q^{\lceil \log_2(M-1) \rceil}} \\ & \qquad = \frac{2-q}{q^{\lceil \log_2(M-2) \rceil}} \\ & \qquad \leqslant \frac{1}{q^{\lceil \log_2 M \rceil}} + \frac{1-q}{q^{\lceil \log_2(M-2) \rceil}} \\ & \qquad = g_{\mathrm{o}}(M) - g_{\mathrm{o}}(M-2) \end{aligned} \tag{C.14}$$

where the equalities can be verified by some calculation, and the second inequality is because $K \leqslant M - 3$ and $2 - q > 0$. The inequality (C.14) implies

$$g_{\mathrm{o}}(M) + g_{\mathrm{o}}(K) \geqslant g_{\mathrm{o}}(M-2) + g_{\mathrm{o}}(K+2).$$

Similarly, we have

$$g_{\mathrm{o}}(M-2) + g_{\mathrm{o}}(K+2) \geqslant g_{\mathrm{o}}(M-4) + g_{\mathrm{o}}(K+4)$$

provided that $K + 2 \leqslant (M-2) - 3$. Such a process can be repeated and we will have

$$g_{\mathrm{o}}(M) + g_{\mathrm{o}}(K) \geqslant g_{\mathrm{o}}\left(\frac{K+M-1}{2}\right) + g_{\mathrm{o}}\left(\frac{K+M+1}{2}\right).$$

Since $M \in \mathbb{Z}_{\mathrm{o}}$ and $K \in \mathbb{Z}_{\mathrm{e}}$, one of $(K + M \pm 1)/2$ is odd and the other is even; their difference is no more than 1. The result from Case 2 shows that

$$g_{\mathrm{o}}\left(\frac{K+M-1}{2}\right) + g_{\mathrm{o}}\left(\frac{K+M+1}{2}\right) \geqslant q \cdot g_{\mathrm{o}}(K+M).$$

Combining the results above, we have that the inequality (C.10) holds in Case 3.

C.10 Proof of Theorem 4.4

In the homogeneous repeater chain, the expected number of the available even crude entanglements generated across N quantum channels is at most $p(N-1)/2$ per time slot. As a consequence, the expected number of available even crude entanglements generated across N quantum channels is at most $p(N-1)T/2$ after T time slots. Moreover, by definition of $h_{\mathrm{e}}(\cdot)$, it requires $h_{\mathrm{e}}(\cdot)$ even crude entanglements on average to obtain one entangled pair shared between nodes s and t. Therefore, the expected number of entangled qubit pairs shared between s and t is at most $(N-1)T/(2h_{\mathrm{e}}(N))$ after T time slots. This implies for any protocol π,

$$\begin{aligned}\lambda^{\pi} &= \liminf_{T\to\infty} \frac{1}{T} \sum_{\tau=1}^{T} \mathbb{E}\{\mathtt{g}_{s:t}^{\pi}(\tau)\} \\ &\leqslant \liminf_{T\to\infty} \frac{1}{T} \cdot \frac{T(N-1)p}{2h_{\mathrm{e}}(N)} = \frac{(N-1)p}{2h_{\mathrm{e}}(N)}.\end{aligned}$$

Together with Proposition 4.4, we have an upper bound of the EDR for any protocol:

$$\lambda^{\pi} \leqslant \frac{(N-1)p}{2g_{\mathrm{e}}(N)} = \frac{(N-1)pq^{n+1}}{2(N-2^n)+(2^{n+1}-N-1)q}.$$

Note that $\mathring{I}(s,t)$ in Remark 7 coincides with the upper bound above. Therefore, $\{\mathring{f}_{i:j}^{i:k} : i,j,k \in \mathcal{N}\}$ and $\{\mathring{u}_{i:j} : (i,j) \in \mathcal{E}\}$ is the optimal solution of $\mathscr{P}_{\mathrm{s}}$, and this upper bound is indeed the maximum EDR because of Theorem 4.2.

Bibliography

[1] J. P. Dowling and G. J. Milburn, "Quantum technology: the second quantum revolution," *Phil. Trans. R. Soc. Lond. A*, vol. 361, no. 1809, pp. 1655–1674, Jun. 2003.

[2] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, 2018.

[3] "National strategic overview for quantum information science," *National Science and Technology Council*, Sep. 2018.

[4] A. S. Fletcher, P. W. Shor, and M. Z. Win, "Channel-adapted quantum error correction for the amplitude damping channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5705–5718, Dec. 2008.

[5] S. Guerrini, M. Chiani, and A. Conti, "Secure key throughput of intermittent trusted-relay QKD protocols," in *Proc. IEEE Global Telecomm. Conf.*, Abu Dhabi, United Arab Emirates, United Arab Emirates, Dec. 2018, pp. 1–6.

[6] N. Hosseinidehaj, Z. Babar, R. Malaney, S. X. Ng, and L. Hanzo, "Satellite-based continuous-variable quantum communications: State-of-the-art and a predictive outlook," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 1, pp. 881–919, 2019.

[7] G. M. D'Ariano, P. L. Presti, and M. G. A. Paris, "Using entanglement improves the precision of quantum measurements," *Phys. Rev. Lett.*, vol. 87, no. 27, p. 270404, Dec. 2001.

[8] Z. Huang, C. Macchiavello, and L. Maccone, "Usefulness of entanglement-assisted quantum metrology," *Phys. Rev. A*, vol. 94, no. 1, p. 012101, Jul. 2016.

[9] R. Demkowicz-Dobrzański and L. Maccone, "Using entanglement against noise in quantum metrology," *Phys. Rev. Lett.*, vol. 113, no. 25, p. 250801, Dec. 2014.

[10] P. W. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proc. of 35th Annual Symposium on Foundations of Computer Science, IEEEPress*, Los Alamitos, CA, 1994.

[11] A. Y. Kitaev, "Quantum computations: algorithms and error correction," *Russ. Math. Surv.*, vol. 52, no. 6, pp. 1191–1249, 1997.

[12] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Physical review letters*, vol. 103, no. 15, p. 150502, 2009.

[13] D. Bruß and C. Macchiavello, "Multipartite entanglement in quantum algorithms," *Phys. Rev. A*, vol. 83, no. 5, p. 052313, May 2011.

[14] T. Peng, A. W. Harrow, M. Ozols, and X. Wu, "Simulating large quantum circuits on a small quantum computer." *arXiv*, 2019. [Online]. Available: https://arxiv.org/abs/1904.00102

[15] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum-enhanced positioning and clock synchronization," *Nature*, vol. 412, pp. 417–419, 2001.

[16] Y. Aharonov, J. Oppenheim, S. Popescu, B. Reznik, and W. G. Unruh, "Measurement of time of arrival in quantum mechanics," *Phys. Rev. A*, vol. 57, no. 6, pp. 4130–4139, Jun. 1998.

[17] P. Komar, E. M. Kessler, M. Bishof, L. Jiang, A. S. Sorensen, J. Ye, and M. D. Lukin, "A quantum network of clocks," *Nat. Phys.*, vol. 10, pp. 582–587, Jun. 2014.

[18] H. J. Kimble, "The quantum Internet," *Nature*, vol. 453, no. 7198, p. 1023, 2008.

[19] L. M. Duan, M. D. Lukin, J. I. Cirac, and P. Zoller, "Long-distance quantum communication with atomic ensembles and linear optics," *Nature*, vol. 414, pp. 413–418, Nov. 2001.

[20] S. Lloyd, J. H. Shapiro, F. N. C. Wong, P. Kumar, S. M. Shahriar, and H. P. Yuen, "Infrastructure for the quantum Internet," in *ACM SIGCOMM Computer Commun. Review*, vol. 34, no. 5, New York, NY, USA, Oct. 2004, pp. 9–20.

[21] A. Reiserer and G. Rempe, "Cavity-based quantum networks with single atoms and optical photons," *Rev. Mod. Phys.*, vol. 87, no. 4, pp. 1380–1418, Jun. 2015.

[22] A. S. Cacciapuoti and M. Caleffi, "Toward the quantum Internet: A directional-dependent noise model for quantum signal processing," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Brighton, United Kingdom, May 2019, pp. 7978–7982.

[23] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. K. Wootters, "Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels," *Phys. Rev. Lett.*, vol. 70, no. 13, pp. 1895–1899, Mar. 1993.

[24] M. Żukowski, A. Zeilinger, M. A. Horne, and A. K. Ekert, "'event-ready-detectors' Bell experiment via entanglement swapping," *Phys. Rev. Lett.*, vol. 71, no. 26, pp. 4287–4290, Dec. 1993.

[25] H.-K. Lo, "Classical-communication cost in distributed quantum-information processing: A generalization of quantum-communication complexity," *Phys. Rev. A*, vol. 62, no. 1, p. 012313, Jul. 2000.

[26] A. K. Pati, "Minimum classical bit for remote preparation and measurement of a qubit," *Phys. Rev. A*, vol. 63, no. 1, p. 014302, Dec. 2000.

[27] B. Zeng and P. Zhangi, "Remote-state preparation in higher dimension and the parallelizable manifold S^{n-1}," *Phys. Rev. A*, vol. 65, no. 2, p. 022316, Feb. 2002.

[28] I. Devetak and T. Berger, "Low-entanglement remote state preparation," *Phys. Rev. Lett.*, vol. 87, no. 19, p. 197901, Nov. 2001.

[29] D. W. Berry and B. C. Sanders, "Optimal remote state preparation," *Phys. Rev. Lett.*, vol. 90, no. 5, p. 057901, Feb. 2003.

[30] C. H. Bennett, D. P. DiVincenzo, P. W. Shor, J. A. Smolin, B. M. Terhal, and W. K. Wootters, "Remote state preparation," *Phys. Rev. Lett.*, vol. 87, no. 7, p. 077902, Jul. 2001.

[31] C. H. Bennett, P. Hayden, D. W. Leung, P. W. Shor, and A. Winter, "Remote preparation of quantum states," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 56–74, Jan. 2005.

[32] W. Dür and J. I. Cirac, "Multiparty teleportation," *Journal of Modern Optics*, vol. 47, no. 2/3, pp. 247–255, 2000.

[33] N. Gisin and S. Massar, "Optimal quantum cloning machines," *Phys. Rev. Lett.*, vol. 79, no. 11, p. 2153, 1997.

[34] D. Bruβ, D. P. DiVincenzo, A. Ekert, C. A. Fuchs, C. Macchiavello, and J. A. Smolin, "Optimal universal and state-dependent quantum cloning," *Phys. Rev. A*, vol. 57, no. 4, p. 2368, Apr. 1998.

[35] M. Mio, J. Daniel, M. Plenio, and V. Vedral, "Quantum telecloning and multiparticle entanglement," *Phys. Rev. A*, vol. 59, no. 1, p. 156, Jan. 1999.

[36] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 2–14, 2002.

[37] V. Giovannetti, S. Lloyd, and L. Maccone, "Quantum metrology," *Phys. Rev. Lett.*, vol. 96, no. 1, p. 010401, 2006.

[38] I. Chuang and D. Gottesman, "Quantum digital signatures," U.S. Patent 7,246,240, Jul. 17, 2007.

[39] Y. Yang, G. Chiribella, and D. Ebler, "Efficient quantum compression for ensembles of identically prepared mixed states," *Phys. Rev. Lett.*, vol. 116, no. 8, p. 080501, 2016.

[40] Y. Yang, G. Chiribella, and M. Hayashi, "Optimal compression for identically prepared qubit states," *Phys. Rev. Lett.*, vol. 117, no. 9, p. 090502, 2016.

[41] M. Schlosshauer, "Decoherence, the measurement problem, and interpretations of quantum mechanics," *Rev. Mod. Phys.*, vol. 76, no. 4, pp. 1267–1305, Oct. 2004.

[42] K. Jagannathan, A. Chatterjee, and P. Mandayam, "Qubits through queues: The capacity of channels with waiting time dependent errors," in *National Conf. Commun.*, Bangalore, India, Feb. 2019, pp. 1–6.

[43] P. Mandayam, K. Jagannathan, and A. Chatterjee, "The classical capacity of a quantum erasure queue-channel," in *Proc. IEEE Workshop on Signal Process. Advances in Wireless Commun.*, Cannes, France, Jul. 2019, pp. 1–6.

[44] F.-G. Deng, G. L. Long, and X.-S. Liu, "Two-step quantum direct communication protocol using the Einstein-Podolsky-Rosen pair block," *Phys. Rev. A*, vol. 68, no. 4, p. 042317, Apr. 2003.

[45] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," *Rev. Modern Phys.*, vol. 81, no. 2, pp. 865–942, Jun. 2009.

[46] M. Epping, H. Kampermann, C. Macchiavello, and D. Bruß, "Multi-partite entanglement can speed up quantum key distribution in networks," *New Journal of Physics*, vol. 19, no. 9, p. 093012, Sep. 2017.

[47] Y. Xie, J. Yuan, and Q. T. Sun, "Design of quantum LDPC codes from quadratic residue sets," *IEEE Trans. Commun.*, vol. 66, no. 9, pp. 3721–3735, Sep. 2018.

[48] R. B. Cooper, *Introduction to Queueing Theory*, 2nd ed. New York: Elsevier North Holland, 1981.

[49] D. P. Bertsekas and R. G. Gallager, *Data Networks*, 2nd ed. Upper Saddle River, NJ: Prentice Hall, 1992.

[50] L. Kleinrock, *Queueing Systems*, 1st ed. New York: Wiley-Interscience, 1975, vol. 1.

[51] A. Ephremides, P. Varaiya, and J. Walrand, "A simple dynamic routing problem," *IEEE Trans. Autom. Control*, vol. AC-25, no. 4, pp. 690–693, Aug. 1980.

[52] L. Tassiulas and A. Ephremides, "Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks," *IEEE Trans. Autom. Control*, vol. 37, no. 12, pp. 1936–1948, Dec. 1992.

[53] J. S. Kaufman, "Blocking in a shared resource environment," *IEEE Trans. Commun.*, vol. 29, no. 10, pp. 1474–1481, Oct. 1981.

[54] L. Gyongyosi and S. Imre, "Decentralized base-graph routing for the quantum Internet," *Phys. Rev. A*, vol. 98, no. 2, p. 022310, Aug. 2018.

[55] L. Jiang, J. M. Taylor, N. Khaneja, and M. D. Lukin, "Optimal approach to quantum communication using dynamic programming," *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 44, pp. 17 291–17 296, Oct. 2007.

[56] A. Dahlberg, M. Skrzypczyk, T. Coopmans, L. Wubben, F. Rozpedek, M. Pompili, A. Stolk, P. Pawelczak, R. Knegjens, J. de Oliveira Filho, R. Hanson, and S. Wehner, "A link layer protocol for quantum networks," *arXiv*, 2019. [Online]. Available: https://arxiv.org/pdf/1903.09778.pdf

[57] S. Khatri, C. T. Matyas, A. U. Siddiqui, and J. P. Dowling, "Practical figures of merit and thresholds for entanglement distribution in quantum networks," *Phys. Rev. Res.*, vol. 1, no. 2, p. 023032, Sep. 2019.

[58] W. Dür, H. J. Briegel, J. I. Cirac, and P. Zoller, "Quantum repeaters based on entanglement purification," *Phys. Rev. A*, vol. 59, no. 1, pp. 169–181, Jan. 1999.

[59] L. Jiang, J. M. Taylor, K. Nemoto, W. J. Munro, R. V. Meter, and M. D. Lukin, "Quantum repeater with encoding," *Phys. Rev. A*, vol. 79, no. 3, p. 032325, Mar. 2009.

[60] L. Vaidman, "Teleportation of quantum states," *Phys. Rev. A*, vol. 49, no. 2, p. 1473, Feb. 1994.

[61] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, "Experimental quantum teleportation," *Nature*, vol. 390, pp. 575–579, Dec. 1997.

[62] W. Dai, T. Peng, and M. Z. Win, "Remote state preparation for multiple parties," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, Brighton, United Kingdom, May 2019, pp. 7983–7987.

[63] S. Pirandola, "Capacities of repeater-assisted quantum communications," *arXiv*, 2017. [Online]. Available: https://arxiv.org/pdf/1601.00966v4.pdf

[64] ——, "End-to-end capacities of a quantum communication network," *Nature Communications Physics*, vol. 2, no. 51, May 2019.

[65] S. Bäuml, K. Azuma, G. Kato, and D. Elkouss, "Linear programs for entanglement and key distribution in the quantum Internet," *arXiv*, 2019. [Online]. Available: https://arxiv.org/pdf/1809.03120.pdf

[66] E. Schoute, L. Mančinska, T. Islam, I. Kerenidis, and S. Wehner, "Shortcuts to quantum network routing," *arXiv*, 2016. [Online]. Available: https://arxiv.org/pdf/1610.05238.pdf

[67] M. Caleffi, "Optimal routing for quantum networks," *IEEE Access*, vol. 5, pp. 22 299–22 312, 2017.

[68] M. Pant, H. Krovi, D. Towsley, L. Tassiulas, L. Jiang, P. Basu, D. Englund, and S. Guha, "Routing entanglement in the quantum Internet," *npj, Quantum Information*, vol. 5, no. 1, pp. 1–9, Mar. 2019.

[69] H. J. Briegel, W. Dür, J. I. Cirac, and P. Zoller, "Quantum repeaters: The role of imperfect local operations in quantum communication," *Phys. Rev. Lett.*, vol. 81, no. 26, pp. 5932–5935, Dec. 1998.

[70] O. A. Collins, S. D. Jenkins, A. Kuzmich, and T. A. B. Kennedy, "Multiplexed memory-insensitive quantum repeaters," *Phys. Rev. Lett.*, vol. 98, p. 060502, Feb 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.98.060502

[71] N. Sangouard, C. Simon, J. Minář, H. Zbinden, H. De Riedmatten, and N. Gisin, "Long-distance entanglement distribution with single-photon sources," *Phys. Rev. A*, vol. 76, no. 5, p. 050301, 2007.

[72] Z.-B. Chen, B. Zhao, Y.-A. Chen, J. Schmiedmayer, and J.-W. Pan, "Fault-tolerant quantum repeater with atomic ensembles and linear optics," *Phys. Rev. A*, vol. 76, no. 2, p. 022329, 2007.

[73] N. Sangouard, C. Simon, H. de Riedmatten, and N. Gisin, "Quantum repeaters based on atomic ensembles and linear optics," *Rev. Mod. Phys.*, vol. 83, pp. 33–80, Mar 2011. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.83.33

[74] C. Jones, D. Kim, M. T. Rakher, P. G. Kwiat, and T. D. Ladd, "Design and analysis of communication protocols for quantum repeater networks," *New Journal of Physics*, vol. 18, no. 8, p. 083015, 2016.

[75] N. Sinclair, E. Saglamyurek, H. Mallahzadeh, J. A. Slater, M. George, R. Ricken, M. P. Hedges, D. Oblak, C. Simon, and W. Sohler, "Spectral multiplexing for scalable quantum photonics using an atomic frequency comb quantum memory and feed-forward control," *Phys. Rev. Lett.*, vol. 113, no. 5, p. 053603, 2014.

[76] S. Guha, H. Krovi, C. A. Fuchs, Z. Dutton, J. A. Slater, C. Simon, and W. Tittel, "Rate-loss analysis of an efficient quantum repeater architecture," *Phys. Rev. A*, vol. 92, no. 2, p. 022357, 2015.

[77] X. Liu, Z.-Q. Zhou, Y.-L. Hua, C.-F. Li, and G.-C. Guo, "Semihierarchical quantum repeaters based on moderate lifetime quantum memories," *Phys. Rev. A*, vol. 95, no. 1, p. 012319, 2017.

[78] E. Shchukin, F. Schmidt, and P. van Loock, "On the waiting time in quantum repeaters with probabilistic entanglement swapping," *arXiv*, 2018. [Online]. Available: https://arxiv.org/pdf/1710.06214.pdf

[79] N. K. Bernardes, L. Praxmeyer, and P. van Loock, "Rate analysis for a hybrid quantum repeater," *Phys. Rev. A*, vol. 83, no. 1, p. 012323, 2011.

[80] IBM, "The Quantum Experience," https://quantumexperience.ng.bluemix.net/qx/experience, 2018.

[81] B. Schumacher and M. D. Westmoreland, "Sending classical information via noisy quantum channels," *Phys. Rev. A*, vol. 56, no. 1, pp. 131–138, Jul. 1997.

[82] A. S. Fletcher, P. W. Shor, and M. Z. Win, "Optimum quantum error recovery using semidefinite programming," *Phys. Rev. A*, vol. 75, no. 1, p. 012338, Jan. 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.75.012338

[83] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Review*, vol. 38, no. 1, pp. 49–95, Mar. 1996.

[84] W. Dai, Y. Shen, and M. Z. Win, "A computational geometry framework for efficient network localization," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1317–1339, Feb. 2018.

[85] W. Dür, H.-J. Briegel, J. Cirac, and P. Zoller, "Quantum repeaters based on entanglement purification," *Phys. Rev. A*, vol. 59, no. 1, p. 169, 1999.

[86] N. M. Linke, D. Maslov, M. Roetteler, S. Debnath, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe, "Experimental comparison of two quantum computing architectures," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3305–3310, 2017.

[87] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell *et al.*, "Superconducting quantum circuits at the surface code threshold for fault tolerance," *Nature*, vol. 508, no. 7497, p. 500, 2014.

[88] N. Kalb, A. A. Reiserer, P. C. Humphreys, J. J. W. Bakermans, S. J. Kamerling, N. H. Nickerson, S. C. Benjamin, D. J. Twitchen, M. Markham, and R. Hanson, "Entanglement distillation between solid-state quantum network nodes," *Science*, vol. 356, no. 6341, pp. 928–932, 2017.

[89] R. Reichle, D. Leibfried, E. Knill, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, "Experimental purification of two-atom entanglement," *Nature*, vol. 443, pp. 838–841, Oct. 2006.

[90] L. Ruan, W. Dai, and M. Z. Win, "Adaptive recurrence quantum entanglement distillation for two-Kraus-operator channels," *Phys. Rev. A*, vol. 97, no. 5, p. 052332, May 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.97.052332

[91] S. L. Mouradian, T. Schröder, C. B. Poitras, L. Li, J. Goldstein, E. H. Chen, M. Walsh, J. Cardenas, M. L. Markham, D. J. Twitchen, M. Lipson, and D. Englund, "Scalable integration of long-lived quantum memories into a photonic circuit," *Phys. Rev. X*, vol. 5, no. 3, p. 031009, Jul. 2015.

[92] J. D. C. Little, "A proof for the queuing formula: $L = \lambda W$," *Oper. Res.*, vol. 9, no. 3, pp. 296–435, May 1961.

[93] W. Pfaff, B. J. Hensen, H. Bernien, S. B. van Dam, M. S. Blok, T. H. Taminiau, M. J. Tiggelman, R. N. Schouten, M. Markham, D. J. Twitchen, and R. Hanson, "Unconditional quantum teleportation between distant solid-state quantum bits," *Science*, vol. 345, no. 6196, pp. 532–535, Aug. 2014.

[94] D. P. Bertsekas, *Dynamic programming and optimal control, volume II.* Belmont, MA: Athena Scientific, 2011.

[95] A. S. Fletcher, P. W. Shor, and M. Z. Win, "Structured near-optimal channel-adapted quantum error correction," *Phys. Rev. A*, vol. 77, no. 1, p. 012320, Jan. 2008. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.77.012320

[96] S. Croke, S. M. Barnett, and G. Weir, "Optimal sequential measurements for bipartite state discrimination," *Phys. Rev. A*, vol. 95, no. 5, p. 052308, May 2017.

[97] W. Dai, T. Peng, and M. Z. Win, "Optimal remote entanglement distribution," *IEEE J. Sel. Areas Commun.*, 2020, to appear.

[98] D. P. Bertsekas, *Dynamic Programming and Optimal Control, volume I*, 4th ed. Belmont, MA: Athena Scientific, 2017.

[99] B. T. Kirby, S. Santra, V. S. Malinovsky, and M. Brodsky, "Entanglement swapping of two arbitrarily degraded entangled states," *Phys. Rev. A*, vol. 94, no. 1, p. 012336, Jul. 2016.

[100] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*, 1st ed. Belmont, MA 02178-9998: Athena Scientific, 1997.

[101] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge, UK: Cambridge University Press, 2004.

[102] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 3rd ed. Springer, 2008.

[103] R. J. Vanderbei, *Linear Programming: Foundations and Extensions*, 2nd ed. Norwell, MA 02061: Kluwer Academic Publishers, 2001.

[104] W. Dai, T. Peng, and M. Z. Win, "Quantum queuing delay," *IEEE J. Sel. Areas Commun.*, 2020, to appear.

[105] D. S. Gunderson, *Handbook of Mathematical Induction: Theory and Applications*, 1st ed. Boca Raton, FL: Chapman and Hall/CRC, 2010.

[106] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 1st ed. New York: Wiley-Interscience, 1984.

[107] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, "Fundamental limits of repeaterless quantum communications," *Nature Communications*, vol. 8, no. 15043, 2017.

[108] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, NY: Springer-Verlag., 1998.

[109] J. J. Vartiainen, M. Möttönen, and M. M. Salomaa, "Efficient decomposition of quantum gates," *Phys. Rev. Lett.*, vol. 92, no. 17, p. 177902, 2004.

[110] P.-X. Chen, J. A. Bergou, S.-Y. Zhu, and G.-C. Guo, "Ancilla dimensions needed to carry out positive-operator-valued measurement," *Phys. Rev. A*, vol. 76, no. 6, p. 060303, Dec. 2007. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.76.060303