THE QUANTUM INTERNET

THE SECOND QUANTUM REVOLUTION

PETER P. ROHDE

THE QUANTUM INTERNET

Following the emergence of quantum computing, the subsequent quantum revolution will be that of interconnecting individual quantum computers at the global level. In the same way that classical computers only realised their full potential with the emergence of the internet, a fully realised quantum internet is the next stage of evolution for quantum computation. This cutting-edge book examines in detail how the quantum internet would evolve in practice, focusing not only on the technology itself but also on the implications it will have economically and politically, with numerous non-technical sections throughout the text providing broader context to the discussion. The book begins with a description of classical networks before introducing the key concepts behind quantum networks, such as quantum internet protocols, quantum cryptography, and cloud quantum computing. This book is written in an engaging style and is accessible to graduate students in physics, engineering, computer science and mathematics.

PETER P. ROHDE is an ARC Future Fellow and Senior Lecturer in the Centre for Quantum Software & Information at the University of Technology Sydney. His theoretical proposals have inspired several world-leading experimental efforts in optical quantum information processing.

THE QUANTUM INTERNET

The Second Quantum Revolution

PETER P. ROHDE
University of Technology Sydney

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108491457
DOI: 10.1017/9781108868815

First published 2021

Printed in the United Kingdom by TJ Books Limited, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-49145-7 Hardback

In memory of Prof Jonathan P. Dowling

Contents

Preface

Quantum technologies are not just of interest to quantum physicists but will have transformative effects across countless areas – the next technological revolution. For this reason, this work is directed at a general audience of not only preexisting quantum computer scientists but also classical computer scientists, physicists, economists, artists, musicians, and computer, software and network engineers. More broadly, we hope that this work will be of interest to those who recognise the future significance of quantum technologies and the implications (or even just curiosities) that globally networking them might have – the creation of the global quantum internet [182, 99]. We expect that the answer to that question will look very different to what emerged from the classical internet.

A basic understanding of quantum mechanics [157], quantum optics [73], quantum computing and quantum information theory [127],[1] classical networking [177] and computer algorithms [48] are helpful, but not essential, to following our discussion. Some mathematical sections require a basic understanding of the mathematical notation of quantum mechanics, although the reader without this background ought to be able to nonetheless follow the broader arguments.

The entirely technically disinterested or mathematically incompetent reader may refer to just Parts I, IX and X – essentially brief, nontechnical, highly speculative essays about the motivation, applications and implications of the future quantum internet.

This work is partially a review of existing knowledge relevant to quantum networking and partially original ideas, to a large extent based on the adaptation of classical networking concepts and quantum information theory to the context of quantum networking. A reader with an existing background in these areas could skip the respective review sections.

[1] Throughout this book we use the Nielsen and Chuang convention for the pronunciation of 'zed' [127].

Our goal is to present a broadly accessible technical and nontechnical overview of how we foresee quantum technologies to operate in the era of quantum globalisation and the exciting possibilities and emergent phenomena that will evolve from it.

We do not shy away from making bold predictions about the future of the quantum internet, how it will manifest itself and what its implications will be for humanity and for science. Inevitably, some of our predictions will turn out to be accurate, whereas others will completely miss the mark entirely. We have no fear of controversy. How accurate our vision will be will have to be seen, but the most important goal in presenting grandiose predictions is to inspire new research directions, encourage future work and stimulate lively and rigorous scientific debate about future technology. If we succeed at achieving these things, yet every last one of our predictions turn out to be completely and utterly wrong, we will consider this work a resounding success. Our goal, first and foremost, is to inspire future science.

Acknowledgements

The desire to share and unite remote digital assets motivated the development of the classical internet, the enabler of the entire twenty-first century economy and our modern way of life. As we enter the quantum era, it is to be expected that there will be a similar demand for networking quantum assets, motivating a *global quantum internet* for bringing together the world's quantum resources, leveraging off their exponential trajectory in capability. We present models for quantum networking, how they might be applied in the future and the implications they will have.

Like the classical internet, it is to be expected that the implications of the quantum internet will be far more than technological, with far-reaching economic, political and geostrategic consequences, which to a large extent act as the driving force for how they will evolve. Although it is impossible to make concrete predictions for the future, we present our treatment of the topic holistically, discussing the interplay between the technology and its driving forces. This includes economic and strategic game-theoretic models that are unique to these quantum technologies, with no direct analogue in terms of conventional analyses. In short, the nonlinear scaling in the utility of quantum resources requires nonlinear economic and strategic models. The nonlinear nature in the utility of future quantum infrastructure implies 'quantum enhancement' not only from a physical perspective but in terms of their implications for humanity.

This work is based on the combined efforts of a highly interdisciplinary team. Zixin Huang contributed to multiple sections of the book, in particular to those on cryptography and quantum algorithms. He-Liang Huang and Zu-En Su contributed to the sections on experimental quantum optics, the Chinese quantum satellite developments and the early structure of the book. Simon Devitt contributed the sections on the quantum SneakerNet, error correction and fault tolerance. Rohit Ramakrishnan and Chandrashekar Radhakrishnan contributed to the sections on optical interfacing and switching, quantum memories and experimental quantum optics. Si-Hui Tan and Atul Mantri contributed to the sections on secure cloud

quantum computing. Nana Liu contributed on quantum machine learning and quantum algorithms. Scott Harrison contributed to the sections on economics and game theory. Tim Byrnes contributed the sections on clock synchronisation and telescopy. William J. Munro contributed the sections on quantum repeater networks and provided editorial assistance. Jonathan P. Dowling acted as co-editor, although his recent passing implies that a number of differences in editorial opinion now swing in Peter Rohde's favour, who acted as lead author and editor.

1

Introduction

The internet is one of the key technological achievements of the twentieth century, an enabling factor in every aspect of our everyday use of modern technology. Whereas digital computing was the definitive technology of the twentieth century, quantum technologies will be for the 21st [127, 23].

Perhaps the most exciting prospect in the quantum age is the development of quantum computers. Richard Feynman [65] was the first to ask the question *'If quantum systems are so exponentially complex that we are unable to simulate them on our classical computers, can those same quantum systems be exploited in a controlled way to exponentially outperform our classical computers?'* Subsequently, the Deutsch-Jozsa algorithm [52] demonstrated for the first time that algorithms can run on a quantum computer, exponentially outperforming any classical algorithm. Since then, an enormous amount of research has been dedicated to finding new quantum algorithms, and the search has indeed been a very fruitful one,[1] with many important applications having been found, including, amongst many others:

- Searching unstructured databases:
  - Grover's algorithm [83].
  - Quadratic speedup.
- Satisfiability and optimisation problems:[2]
  - Grover's algorithm.
  - Quadratic speedup.
  - Includes solving **NP**-complete problems and brute-force cracking of private encryption keys.

[1] See the Quantum Algorithm Zoo for a comprehensive summary of the current state of knowledge on quantum algorithms.

[2] A satisfiability problem is one in which we search a function's input space for a solution(s) satisfying a given output constraint. The hardest such problems, like the archetypal 3-SAT problem, are **NP**-complete.

  - Many optimisation problems are **NP**-complete or can be approximated in **NP**-complete.
- Period finding and integer factorisation:
  - Shor's algorithm [165].
  - Exponential speedup.
  - This compromises both Rivest, Shamir and Adleman (RSA) and elliptic-curve public-key cryptography [141], the most widely used cryptographic protocols on the internet today.
  - This problem is believed to be **NP**-intermediate – an **NP** problem that lies outside **P** (and is therefore classically hard) but that is not **NP**-complete (the 'hardest' of the **NP** problems).
- Simulation of quantum systems:
  - Lloyd's algorithm [107].
  - Exponential speedup.
  - This includes simulation of molecular and atomic interactions in the study of quantum chemistry or nuclear physics; interactions between drug molecules and organic molecules for drug design; genetic interactions for the study of genetics and genetic medicine; nanoscale semiconductor physics for integrated circuit design; and much more.
- Simulation of quantum field theories:
  - Jordan-Lee-Preskill algorithm [94, 34].
  - Exponential speedup.
  - A key area of fundamental physics research.
- Topological data analysis:
  - Lloyd's algorithm [108].
  - Exponential speedup.
  - Broad applications including social media network analysis; consumer behaviour; behavioural dynamics; neuroscience; and higher-dimensional signal and image processing.
- Solving linear systems of equations:
  - Algorithms by [84, 26].
  - Exponential speedup.
  - Widespread applications in linear algebra and calculus.
- Quantum machine learning:
  - Lloyd's algorithm [109].
  - This includes putting an end to humanity.

Some of these are discussed in more detail in Chapter 28.

It is likely we have not yet begun to fully recognise the capabilities of quantum computers and the full plethora of applications they may have in the future. We stand at the beginning of the emergence of an entirely new type of technology.

In addition to many practical applications, the onset of quantum computing carries with it deep philosophical implications, specifically, the extended Church-Turing (ECT) thesis hypothesises that any physically realisable system can be *efficiently*[3] simulated by a universal Turing machine (i.e., classical computer). The believed exponential complexity of quantum systems inclines quantum computer scientists to believe that the ECT thesis is therefore false [50].[4] The demonstration of large-scale quantum computers, though unable to prove or disprove the ECT thesis,[5] could at least provide some convincing evidence against the ECT conjecture.

From a computational complexity theorist's perspective, it is strongly believed that the complexity classes of problems efficiently solvable on classical computers (**P** and **BPP**) and quantum computers (**BQP**) are distinct. Specifically, it is believed that **BPP** $\subset$ **BQP**. If this conjecture is correct, it implies the existence of quantum algorithms superpolynomially faster than the best classical ones and that the ECT thesis is not correct. More specifically, Figure 1.1 illustrates the believed relationships between some of the most important complexity classes relevant to quantum computing.

In addition to quantum computing, quantum cryptography holds the promise of uncrackable cryptographic protocols, guaranteed not by the assumed complexity of solving certain mathematical problems like integer factorisation or brute-force searching but by the laws of quantum mechanics. That is, provided that our understanding of quantum mechanics is correct, quantum cryptographic protocols exist that cannot be cracked, irrespective of the computational resources of an adversary.

Already we are beginning to see elementary realisations of essential quantum technologies such as quantum computing, cryptography and metrology. As these technologies become increasingly viable and more ubiquitous, the demand for networking them and sharing quantum resources between them will become a pressing issue. Most notable, quantum cryptography and *cloud quantum computing* will be pivotal in the proliferation of quantum technology, which necessarily requires reliable quantum communications channels.

[3] The term 'efficient' is one coined by the computer scientist to mean that a problem can be solved in time at most polynomial in the size of the problem.

[4] We have discovered a truly marvellous proof of this, which this footnote is too narrow to contain.

[5] When we talk about 'scalability' or the 'ECT thesis' we are talking about asymptotic relationships. Clearly no finite-sized experiment can prove asymptotic scaling with certainty. But with a sufficiently large quantum computer at our disposal, demonstrating exponentially more computational power than its classical sibling, we might be reasonably satisfied in convincing ourselves about the nature of the scaling of different computational models.

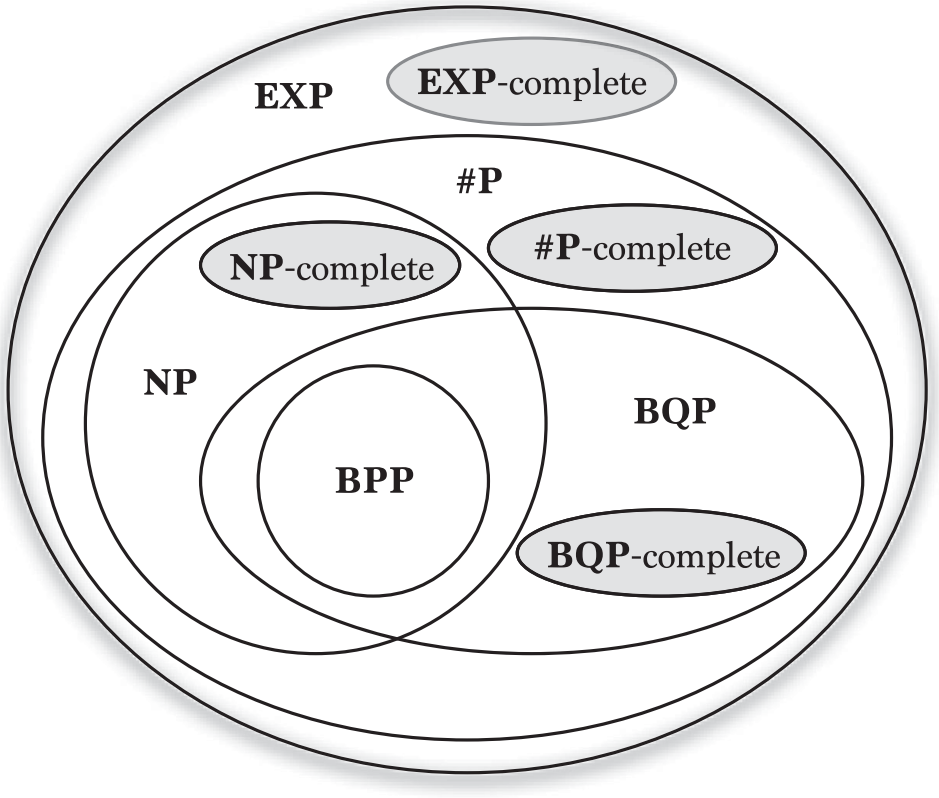

Figure 1.1 Believed relationships between the complexity classes most relevant to quantum computing. **BPP** is the class of polynomial-time probabilistic classical algorithms. **NP** is the class of problems verifiable in polynomial time using classical algorithms. **NP**-complete are the subset of **NP** problems polynomial-time reducible to any other problem in **NP**, similarly for other 'complete' problems. **BQP** is the class of probabilistic algorithms solvable in polynomial time on universal quantum computers. **#P** is the set of counting problems that count satisfying solutions to **P** problems (**P** is the same as **BPP** but deterministic rather than probabilisitic). **EXP** is the class of all algorithms that require exponential time. Note that it is actually unproven whether $\mathbf{P} = \mathbf{BPP}$ or $\mathbf{P} \subset \mathbf{BPP}$. There are examples where the best known **BPP** algorithms outperform the best known **P** algorithms, which could arise because the two classes are inequivalent or because we simply have not tried hard enough to find the best deterministic algorithms. Furthermore, though it is known that $\mathbf{P} \subseteq \mathbf{NP}$, it is not known whether $\mathbf{BPP} \subseteq \mathbf{NP}$. For the sake of illustration in our Venn diagram we have taken the view that it is. **BPP** is regarded as the class of problems efficiently solvable on universal Turing machines (i.e., classical computers), whereas **BQP** is the class efficiently solvable on universal quantum computers. The computational superiority of quantum computers is based on the (strongly believed, yet unproven) assumption that $\mathbf{BPP} \subset \mathbf{BQP}$.

The first demonstrations of digital computer networks were nothing more than simple two-party, point-to-point (P2P) communication. However, the internet we have today extends far beyond this, allowing essentially arbitrary worldwide networking across completely ad hoc networks comprising many different mediums, with any number of parties, in an entirely plug-and-play and decentralised fashion. Similarly, elementary demonstrations of quantum communication have been performed across a small number of parties, and much work has been done on analysing quantum channel capacities in this context. But, as with digital computing, demand for a future *quantum internet* is foreseeable, enabling the arbitrary communication of quantum resources, between any number of parties, over ad hoc networks.

The digital internet may be considered a technology stack, such as TCP/IP (Transmission Control Protocol/Internet Protocol), comprising different levels of abstraction of digital information [177]. At the lowest level we have raw digital data we wish to communicate across a physical medium. Above this, we decompose the data into packets. The packets are transmitted over a network, and TCP is responsible for routing the packets to their destination and guaranteeing data integrity and Quality of Service (QoS). Finally, the packets received by the recipient are combined and the raw data are reconstructed.

The TCP layer remains largely transparent to the end-user, enabling virtual software interfaces to remote digital assets that behave as though they were local. This allows high-level services such as the File Transfer Protocol (FTP), the worldwide web, video and audio streaming and outsourced computation on supercomputers, as though everything were taking place locally, with the end-user oblivious to the underlying networking protocols, which have been abstracted away. To the user, YouTube videos or Spotify tracks behave as though they were held as local copies. And FTP or DropBox allows storage on a distant data centre to be mounted as though it were a local volume. We foresee a demand for these same criteria in the quantum era.

In the context of a quantum internet, packets of data will instead be quantum states, and the transmission control protocol is responsible for guiding them to their destination and ensuring quality control.

Our treatment of quantum networks will be optics heavy, based on the reasonable assumption that communications channels will almost certainly be optical, albeit with many possible choices of optical states and mediums. However, this does not preclude nonoptical systems from representing quantum information that is not in transit, and we consider such ‘hybrid’ architectures in detail, as well as the interfacing between optical and nonoptical systems. Indeed, it is almost certain that future large-scale quantum computers will not be all-optical, necessitating interfacing different physical architectures.

Shared quantum entanglement is a primitive resource with direct applications in countless protocols. This warrants special treatment of quantum networks that do not implement a full network stack but instead specialise in just this one task – entanglement distribution. We will see that such a specialised network will already be immensely useful for a broad range of applications, and its simplicity brings with it many inherent advantages.

The quantum internet will enable advances in the large-scale deployment of quantum technologies. Most notable, in the context of quantum computing it will allow initially very expensive technology to be economically viable and broadly accessible via the outsourcing of computations for both consumers who cannot afford quantum computers and to well-resourced hosts who can – *cloud quantum computing*.

With the addition of recent advances in homomorphic encryption and blind quantum computing, such cloud quantum computing can be performed securely, guaranteeing privacy of both data and algorithms, secure even against the host performing the computation. This opens up entirely new economic models and applications for the licensing of compute time on future quantum computers in the cloud.

The unique behaviour of quantum computing, in terms of the superclassical scaling in its computational power, brings with it many important economic and strategic considerations that are extremely important to give attention to in the postclassical world.

But quantum technologies extend far beyond computation. Many other exciting applications for controlled quantum systems exist, with new ones frequently emerging. Thus, the quantum internet will find utility beyond cloud quantum computing, enabling the global exchange of quantum resources and assets. This could include the networking of elementary quantum resources such as state preparation, entanglement sharing and teleportation and quantum measurements or scale all the way up to massively distributed quantum computation or a global quantum cryptography network.

It is hard to foresee the future trajectory of quantum technology, much as no one foresaw the advances digital technology has made over the last half century. But it is certain that as the internet transformed digital technology, the quantum internet will define the future of quantum technologies.

Part I

Classical Networks

2

Mathematical Representation of Networks

We begin by turning our attention to defining a mathematical construction for the representation of (quantum and/or classical) networks, which we will subsequently rely on heavily in our framework for quantum networks. This encompasses representing networks as graphs, representing the cost of communications within the network and how to optimise network routing to minimise costs. These notions will be essential in our treatment of quantum networks.

2.1 Graph-Theoretic Representation

We consider a classical network to be a weighted, directed graph,

$$G = (V, E), \tag{2.1}$$

where vertices represent *nodes* ($v \in V$) in the network and the weighted edges represent communication *links* ($e \in E$) between neighbouring nodes.

A node could be, for example, data storage, a classical computer implementing a computation, a router that switches the connections between incoming and outgoing links or an end-user – anything that communicates with the network, sender or receiver. A link, on the other hand, is any arbitrary means of communication between nodes, such as optical fibre, satellite, radio, electrical, smoke signals, tin cans connected by a taut piece of string or a well-trained carrier pigeon. In the protocols to be described here, it is completely irrelevant what the specific mediums for communication are. Rather what matters are *costs* quantifying the relative performance of different links.

A key feature of the global internet is redundancy. In a packet-switched environment, when sending identical packets twice might each follow entirely different routes to their common destination. Node-to-node redundancy is easily accommodated for in the graph-theoretic model by allowing multiple distinct edges between nodes. It is extremely important to accommodate multiple edges in network graphs,

because redundant routes provide a direct means by which to load-balance a route. So, for example, a hub in Australia might connect to a sister hub in New Zealand using both a fibre-optic undersea cable and simultaneously via a satellite uplink. If the faster of the two connections is running out of capacity, a proportion of the packets can simply be switched to the other link, thereby balancing the load. For this reason we abstain from using an adjacency matrix representation for network graphs, because they do not accommodate redundancy.

2.2 Cost Vector Analysis

The edge weights in G represent the *costs* ($\vec{c}$) associated with using that link.

Definition 1 (Network cost metrics) *Cost metrics satisfy the following properties:*

- *Identity operations: If a channel performs nothing, its associated cost is zero,* $c(\hat{I}) = 0$.
- *Triangle inequality:*
  $c(v_1 \to v_2 \to v_3) \leq c(v_1 \to v_2) + c(v_2 \to v_3)$,
  across all paths $v_1 \to v_2 \to v_3$. *In the case of strict equality under addition we refer to the cost as a strictly additive cost.*
- *Positivity:* $c \geq 0$. *This ensures that shortest-path algorithms will function correctly. It is also congruent with the intuitive expectation that data traversing a communications channel are not somehow better off than if they had not traversed that channel at all.*

The reason we demand that costs have a distance interpretation is so that graph-theoretic pathfinding algorithms are applicable, allowing us to build upon the vast preexisting understanding of graph theory. Ideally we would like equality in costs' triangle inequality, which yields an exact cost. But often this is not possible and we are satisfied with the inequality, which simply dictates an upper bound on cost.

A detailed discussion of some of the major costs that realistic quantum networks will be subject to is presented in Chapter 8.

A *route* between two nodes, Alice (A) and Bob (B), of the network, G, is an acyclic subgraph connecting those nodes, $R_{A\to B} \subseteq G$. In general ad hoc networks there will typically be multiple paths between two nodes $A \to B$. For a particular cost metric, the cost of an entire route is simply the sum of the costs of each of the constituent links,

Definition 2 (Route costs) *The net cost of a route $A \to B$, using cost metric $c(A \to B)$, traversing nodes v_i is*

$$c(R_{A\to B}) = \sum_{i=1}^{|R_{A\to B}|-1} c(v_i \to v_{i+1}), \tag{2.2}$$

where v_i is the ith node in the route $R_{A\to B}$.

Figure 2.1 illustrates a simple example network with all of its available routes, $R_{A\to B} \subseteq G$. Figure 2.2 illustrates the optimal path for $A \to B$ based on edge weights.

In a given network, it is unlikely that only a single cost metric will be of interest when determining optimal routings. There may be a trade-off between different measures. For example, for time-critical applications the cost of a route might be considered a combination of both dollar cost and latency – a satellite has very low latency but is extremely expensive, whereas a carrier pigeon is slow but cheap (and prohibited by PETA). What is the best trade-off between the two?

To accommodate this, we allow the *net cost* of a route to be defined as an arbitrary function of other primitive costs of the route,

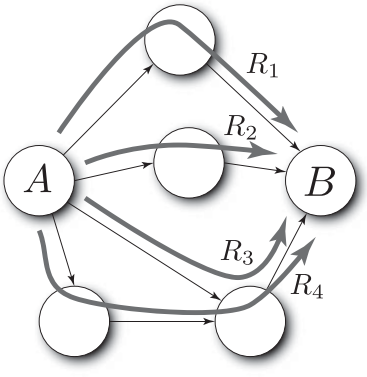

Figure 2.1 Example of a simple network with multiple routes $A \to B$. Note that R_3 and R_4 are competing with one another for use of the last link, which the routing strategy, $\mathcal{S}$, will need to resolve if multiple simultaneous transmissions are taking place.

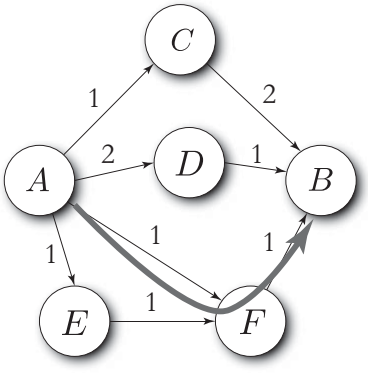

Figure 2.2 The same network graph from Figure 2.1, with links weighted by some arbitrary cost metric. Applying a shortest-path algorithm yields the optimal route between Alice and Bob to be $A \to F \to B$, which incurs a net cost of $c = 2$, as opposed to all other routes, which incur a net cost of $c = 3$.

Definition 3 (Net routing cost) *The net cost of a route $A \to B$ is given by*

$$c_{\text{net}}(R) = f_{\text{cost}}(\vec{c}(R)), \tag{2.3}$$

where c_{net} is a single numeric value representing the net cost as calculated from an arbitrary cost function, f_{cost}, of the vector of associated costs.

Note that the net routing cost need not be a metric, because the cost function could be arbitrary. The net cost can be thought of as a ranking for routes but not necessarily as a metric that accumulates across routes, because it already captures all of these accumulations.

Equation (2.3) gives us the net cost of a given route. For multiple users we would like to simultaneously optimise the cost across all users of the network. Thus, we define the routing cost for the entire network to be the following.

Definition 4 (Network routing cost) *The net routing cost of all costs over all active routes $\vec{R}$ is*

$$c_{\text{total}}(\vec{R}_{\vec{A}\to\vec{B}}) = \sum_{r\in\vec{R}_{\vec{A}\to\vec{B}}} c_{\text{net}}(r), \tag{2.4}$$

where $\vec{R}_{\vec{A}\to\vec{B}}$ is a set of active routes connecting each pair $A_i \to B_i \ \forall\, i$.

2.3 Routing Strategies

A *strategy*, $\mathcal{S}$, is simply an algorithm that chooses a route based on the starting and finishing nodes of a communication and also updates the cost vectors within the network associated with the utilisation of that route.

Definition 5 (Routing strategies) *A routing strategy is defined by*

$$\begin{aligned} \mathcal{S}(i,j,\vec{c}) &\rightarrow \{k,\vec{c}\,'\}, \\ i,j &\in V, \\ k &\in \{R_{v_i \rightarrow v_j}\}, \end{aligned} \tag{2.5}$$

where $\mathcal{S}$ denotes the strategy, k is a route, i and j are the source and destination nodes of the route and $\vec{c}$ is a vector of associated costs.

The goal of the strategy $\mathcal{S}$ is to minimise a chosen cost measure.

No particular route through a network is going to have infinite capacity and therefore we cannot typically always reemploy the same most cost-effective route for all data. Particularly in multi-user networks, as routes are employed for communicating quantum states, their cost metrics may change according to load or other external influences. Alternatively, some routes may come into and out of operation. For example, a satellite requiring line-of-sight communication may oscillate in and out of sight, thereby periodically enabling and disabling respective network routes. For this reason, it is important that strategies accommodate dynamic changes in the network. This is easily accounted for by letting the edge weights in our network graph be a function of time, G_t, which are updated via the application of a strategy that may also be time dependent.

Definition 6 (Time-dependent routing strategies) *A time-dependent strategy, $\mathcal{S}_t$, updates the network graph, G_t, at each time step t,*

$$G_{t+1} = \mathcal{S}_t(G_t). \tag{2.6}$$

*S_t could be any **BPP** algorithm, deterministic or probabilistic.*

For example, the network might have bandwidth restrictions on some links, in which case if more than a certain amount of data is transmitted through a link it is no longer available for use until previous transmissions have completed. Or, based on market dynamics, the dollar cost of utilising a link may change with its demand.

This type of cost minimisation approach to routing is analogous to *distance-vector routing protocols* in classical networking theory.

2.4 Strategy Optimisation

Clearly the goal when choosing routing strategies is to minimise the total cost, Eq. (2.3). That is, solving the following optimisation problem.

Definition 7 (Strategy optimisation) *The optimisation of strategies with a network comprising net costs* c_{total} *is given by*

$$c_{\min} = \min_{\mathcal{S}}(c_{\text{total}}),$$
$$\mathcal{S}_{\text{opt}} = \operatorname*{argmin}_{\mathcal{S}}(c_{\text{total}}). \tag{2.7}$$

Choosing optimal strategies is a challenging problem, potentially requiring complex, computationally inefficient optimisation techniques. Strategy optimisation is an example of resource allocation whose optimal solutions are often notoriously difficult to solve exactly, residing in complexity classes like **NP**-complete (or worse!). In general, the number of possible routes through a graph will grow exponentially with the number of vertices. Thus, explicitly enumerating each possible route is generally prohibitive for large networks, unless some known structure provides 'shortcuts' to optimisation. Having said this, Dijkstra's shortest-path algorithm is the perfect counterexample, demonstrating that although an exponential number of routes may exist between two points, an optimal one can be found in **P**.

Ad hoc Operation vs Central Authorities

When considering strategy optimisation, the first question to ask is 'Who performs the optimisation, and who has access to what information?'

In terms of who performs the optimisation, the two main options are that either each node is responsible for optimising the routes of packets passing through it (INDIVIDUAL algorithms) or there is a reliable and trusted central mediating authority who oversees network operation and performs all strategy decision making (CENTRAL algorithms).

In the case of INDIVIDUAL algorithms, the required knowledge of the state of the network could be obtained using network exploration algorithms or gateway protocols.

On the other hand, for CENTRAL algorithms, either network exploration could be employed or, alternatively, the network policy could require nodes to notify

the central authority upon joining or leaving the network. The former introduces an overhead in classical networking resource usage, because network exploration must be performed routinely to keep the ledger of nodes up to date. The latter, on the other hand, avoids this but introduces a point of failure, in that all network participants must be reliable in notifying the central authority as required by the network policy. Failure to do so could result in invalid or suboptimal strategies.

Local vs Global Optimisation

There are two general approaches one might consider when choosing strategies: *local optimisation* (LOCAL) and *global optimisation* (GLOBAL). LOCAL simply takes each state to be communicated, one by one, and allows it to individually choose an optimal routing strategy based on the state of the network at that moment. GLOBAL is far more sophisticated and simultaneously optimises the sum of the routing costs, Eq. (2.4), of all currently in-demand routes.

To implement LOCAL optimisation, either INDIVIDUAL or CENTRAL algorithms may be employed. On the other hand, GLOBAL optimisation necessarily requires a CENTRAL algorithm, because it requires knowledge of the entire state of the network, which is collectively optimised.

Because GLOBAL represents the class of all algorithms that take all network costs by all packets into consideration, it must clearly perform at least as well as LOCAL, which only takes into consideration the costs of a given packet. But we expect GLOBAL to perform better than LOCAL in general, owing to the additional information it takes into consideration. We express this as LOCAL$\subset$GLOBAL. However, GLOBAL requires solving a complex, simultaneous optimisation problem, which is likely to be computationally hard, whereas LOCAL can be efficiently solved using multiple independent applications of, for example, an efficient shortest-path algorithm (so-called GREEDY algorithms).

A further stumbling block for GLOBAL is that it requires some central authority, responsible for the global decision making, to have complete, real-time knowledge of the state of the entire network. This may be plausible for small local area networks but would clearly be completely implausible for the internet as a whole. So it is to be expected that different layers and subnets in the network hierarchy will employ entirely different strategy optimisation protocols. This is certainly reminiscent of the structure of the present-day internet.

Roughly speaking, we might intuitively guess that at lower levels in the network hierarchy, responsible for smaller subnets, there will be a tendency towards the adoption of GLOBAL strategies, as full knowledge of the state of the subnet is readily obtained and maintained. However, as we move to the highest levels of the network hierarchy (e.g., routing of data across international or intercontinental boundaries),

we might expect more laissez-faire (i.e., GREEDY) strategies to be adopted, because the prospects of enforcing a central authority with full knowledge of the state of the internet, who is also trusted by all nations to fairly and impartially allocate network resources and mediate traffic, are highly questionable.

We will not aim to comprehensively characterise the computational complexity of GLOBAL strategies. However, in Chapter 9 we will present some elementary analyses of several toy models for realistic strategies. Some such strategies are efficient and, although not optimal, nonetheless satisfy certain criteria we might expect.

Future developments in the optimisation techniques required for GLOBAL strategies may improve network performance, leaving our techniques qualitatively unchanged.

When employing LOCAL, on the other hand, things are often far simpler. If we are optimising over a cost metric satisfying the distance interpretation, we may simply employ a shortest-path algorithm to find optimal routes through the network.

If one were to become even more sophisticated, one might even envisage treating network resource allocation in a game theoretic context, which we will not even begin to delve into here.

3

Network Topologies

Because quantum (or classical) networks inherently reside on graphs, it is important to introduce some of the key graph structures of relevance to networking and some of their properties of relevance to quantum networking protocols.

Let the graph G representing the network be

$$G = (V, E), \tag{3.1}$$

with vertices V and edges E. In principle a network could be characterised by any connected graph whatsoever. However, there are certain structures and patterns that emerge very frequently and deserve special attention.

It is paramount that quantum networking protocols have the capacity to deal with the diverse network topologies that are likely to present themselves in the future real-world quantum internet. Some of the graph-theoretic algorithms that we rely on are computationally efficient for *arbitrary* graph topologies, even more so for certain classes of graphs exhibiting particular structure, such as tree graphs or complete graphs. Others, however, are computationally inefficient in general but may have efficient approximation algorithms for some or all classes of topologies.

We will now review some of the graph structures most likely to arise in quantum networks, learning from the structures that have become ubiquitous in classical networking. Understanding the basic mathematical properties of these different network topologies is extremely important to take into consideration when designing future quantum networks, because they strongly impact important features such as construction cost of the network infrastructure, routing cost vector analysis, likelihood of successful routing and transmission time.

A summary of the basic mathematical characteristics of the topologies presented is shown in Table 3.1, specifically showing the number of edges and vertices and *diameter* of the topologies (i.e., the distance between extremal points in the network).