

The Quantum Internet — A New Frontier

www.quantum-internet.tech

Peter P. Rohde,^{1,2,*} Zixin Huang,³ He-Liang Huang,^{4,5,6,7} Zu-En Su,^{4,8,6} Simon Devitt,¹ Rohit Ramakrishnan,⁹ Atul Mantri,^{10,11} Nana Liu,¹² Scott Harrison,¹³ Darryl Vietch,¹⁴ Chandrashekhar Radhakrishnan,¹⁵ Jonathan P. Dowling,¹⁶ Tim Byrnes,^{17,18,19,15,20} and William J. Munro^{21,19}

¹Centre for Quantum Software & Information (QSI), Faculty of Engineering & Information Technology, University of Technology Sydney, Australia

²Hearne Institute for Theoretical Physics, Louisiana State University, Baton Rouge, United States

³Department of Physics & Astronomy, University of Sheffield, United Kingdom

⁴Hefei National Laboratory for Physical Sciences at Microscale and Department of Modern Physics, University of Science & Technology of China, Hefei, China

⁵CAS Centre for Excellence and Synergetic Innovation Centre in Quantum Information and Quantum Physics, University of Science & Technology of China, Hefei, China

⁶CAS-Alibaba Quantum Computing Laboratory, Shanghai, China

⁷Henan Key Laboratory of Quantum Information & Cryptography, Zhengzhou Information Science & Technology Institute, Zhengzhou, China

⁸CAS Centre for Excellence and Synergetic Innovation Centre in Quantum Information & Quantum Physics, University of Science & Technology of China, Hefei, China

⁹Indian Institute of Science, Bangalore, India

¹⁰Singapore University of Technology & Design, Singapore

¹¹Centre for Quantum Technologies, National University of Singapore, Singapore

¹²John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, Shanghai, China

¹³Leibniz Institute for Research & Information in Education, Frankfurt am Main, Germany

¹⁴Faculty of Engineering & Information Technology, University of Technology Sydney, Australia

¹⁵New York University Shanghai, China

¹⁶Hearne Institute for Theoretical Physics and Department of Physics & Astronomy, Louisiana State University, Baton Rouge, United States

¹⁷State Key Laboratory of Precision Spectroscopy, School of Physical & Material Sciences, East China Normal University, Shanghai, China

¹⁸NYU-ECNU Institute of Physics at NYU Shanghai, China

¹⁹National Institute of Informatics, Tokyo, Japan

²⁰Department of Physics, New York University, New York, United States

²¹NTT Basic Research Laboratories, Atsugi, Japan

(Dated: May 25, 2020)

The desire to share and unite remote digital assets motivated the development of the classical internet, the enabler of the entire 21st century economy and our modern way of life. As we enter the quantum era, it is to be expected 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. Socially, economically, politically and geostrategically, the upcoming era of quantum supremacy will be as significant for the 21st century as the transistor was for the 20th.

The inherently different scaling in the computational power of quantum computers fundamentally changes the dynamics of how they will operate in the future. Given their high expected initial cost, a client/server model for outsourcing computation will be essential for enabling the accessibility and proliferation of this technology, and ensuring its economic viability. We therefore anticipate the emergence of *cloud quantum computing*, a model for outsourcing quantum computations to the network. We argue that economic efficiency will mandate that all future quantum computers be united into a single global *virtual quantum computer*, offering exponentially more power to all network participants than if they were to keep their resources to themselves. This model for the allocation of computational resources is uniquely quantum, with no classical analogue, completely altering the economic landscape for the future of computation.

Given the sensitivity of much of the data to which future quantum computers are going to foreseeably be applied, protocols for encrypted quantum computation will be essential – the outsourcing of computations that neither an eavesdropper nor even the server performing the computation can spy upon. This will enable new models for the commercialisation and proliferation of quantum technologies, unlike any existing models for classical computing.

We argue that the quantum internet will not create a system of winners and losers, but will rather be of benefit to all of humanity, rich and poor alike, to an even greater extent than the classical internet. It is therefore essential that it be imminently established and pursued.

While this work is only an early step in a rapidly developing field, still in its infancy, the central concepts we present will be highly relevant to future developments, laying the groundwork for this blossoming field.

We present both original ideas, as well as an extensive review of relevant and related background material. The work is divided into technical sections (requiring only a basic knowledge of the notation of quantum mechanics), for those interested in mathematical details, as well as extensive, entirely non-technical sections for the less technically inclined.

We target this work very broadly at quantum and classical computer scientists, classical computer systems, software and network engineers, physicists, economists, artists, musicians, and those just generally curious about the future of quantum technologies and what they might bring to humanity.

— *Carpe futurum. Ad astra per alas fideles. Scientia potentia est. Benedictus benedicat.*



This work is licensed under a
Creative Commons Attribution 4.0 International License.

* To whom correspondence should be addressed:
dr.rohde@gmail.com; <http://www.peterrohde.org>



*To those who die doing what they love.
— Finis vitae sed non amoris.*

CONTENTS**I Introduction**

I. Mission statement	10
II. Foreword	10
III. Introduction	11
IV. The Quantum Internet Project	14

II Mathematical foundations

V. Introduction to quantum mechanics	15
A. Features of quantum mechanics	15
B. Quantum states	16
1. State vectors	16
2. Composite systems	17
3. Entangled states	17
4. Density operators	18
5. Reduced states	18
C. Evolution	19
1. Discrete-time evolution	19
2. Continuous-time evolution	19
3. Schrödinger vs Heisenberg pictures	19
D. Measurement	19
E. Example derivation	20

VI. Introduction to quantum optics	21
A. Discrete-variables	21
1. Photon-number states	21
2. Measurement	21
3. Creation & annihilation operators	21
4. Photon-number operators	22
B. Continuous-variables	22
1. Position & momentum operators	22
2. Position & momentum representations	22
3. Coherent states	22
4. Phase-space representations	23

VII. Introduction to quantum information theory	24
A. Probability, information and classical correlation measures	24
B. Quantum correlation measures	24

III Classical networks

VIII. Classical networking protocols	25
A. TCP/IP	26
1. Internet Protocol	26
2. User Datagram Protocol	26
3. Transmission Control Protocol	27
B. Ethernet	27
C. Gateway protocols & routing tables	28
D. Network hierarchies	28
IX. Mathematical representation of networks	29
A. Graph-theoretic representation	29
B. Cost vector analysis	29
C. Flow networks	30
D. Routing strategies	31
E. Strategy optimisation	31
1. Ad hoc operation vs. central authorities	32

X. Network topologies	33
A. Point-to-point	33
B. Linear	33
C. Complete	34
D. Lattice	34
E. Tree	35
1. Balanced tree	35
2. Random tree	36
3. Minimum spanning tree	36
F. Percolation	36
G. Random	37
H. Hybrid	38
I. Scale-free networks	39
J. The internet web-graph	40
K. Network robustness	40
XI. Network algorithms	41
A. Network exploration & pathfinding	42
B. Shortest-path	42
C. Constrained shortest-path	44
D. Single-source shortest-path	45
E. Minimum spanning tree	45
F. Minimum-cost flow	45
G. Maximum flow	46
H. Multi-commodity flow	46
I. Vehicle routing problem	46
J. Vehicle rescheduling problem	47
K. Improving network algorithms using quantum computers	47
IV. Quantum networks	47
XII. Quantum channels	48
A. Quantum processes	48
B. Quantum process matrices	49
C. Quantum processes in quantum networks	50
D. Characterising quantum states & channels	51
1. Quantum state tomography	51
2. Quantum process tomography	51
XIII. Optical encoding of quantum information	52
A. Single photons	52
B. Photon-number	53
C. Spatio-temporal	53
1. Transverse electro-magnetic modes	54
2. Time-bins	54
D. Thermal states	55
E. Phase-space	55
1. Coherent states	56
2. Cat states	57
3. Squeezed states	57
F. Non-optical encoding	58
XIV. Errors in quantum networks	58
A. Known unitaries	58
B. Unknown imperfect unitaries	58
C. Loss	58
1. Model	59
2. Linear optics networks	59
3. Single-photon encoding	59
4. Polarisation & dual-rail encoding	59
5. Photon-number encoding	59
6. Coherent state encoding	60
7. Cat state encoding	60
8. NOON states	60
9. Scaling	61

D. Dephasing	61	XX. Interconnecting & interfacing quantum networks	94
1. Qubits	61	A. Optical interfacing	95
2. T_2 -times	62	1. Two-level systems	95
3. Optical states	62	2. λ -configuration systems	95
E. Depolarisation	63	3. Atomic ensembles	96
F. Amplitude damping	63	4. Superconducting qubits	96
1. T_1 -times	63		
G. Mode-mismatch	63		
H. Dispersion	65	XXI. Optical routers	98
I. Spectral filtering	65	A. Mechanical switches	99
J. Phase-space	65	B. Interferometric switches	99
		1. Electro-optic modulators	99
XV. Quantum cost vector analysis	66	2. Acousto-optic modulators	101
A. Costs	66	3. Magneto-optic modulators	101
1. Efficiency	66	C. Two-channel two-port switches	101
2. Spectral filtering	66	D. Multiplexers & demultiplexers	102
3. Decoherence	66	E. Single-channel multi-port switches	103
4. Mode-mismatch	66	F. Multi-channel multi-port switches	103
5. Distance measures	67	G. Crossbar switches	104
6. Purity	67		
7. Entanglement	68	XXII. Optical stability in quantum networks	105
8. Phase-space	68	A. Photon wave-packets	105
9. Channel capacity	68	B. Mach-Zehnder interference	105
10. Latency	70	C. Hong-Ou-Mandel interference	106
11. Dollars	70	D. HOM vs MZ interference	107
B. Costs as distance metrics	70	E. Optical stabilisation	107
C. Non-trivial node operations	71		
D. Negative cost vectors	72		
XVI. Quantum Transmission Control Protocol (QTCP)	72	V. Protocols for the quantum internet	108
A. QTCP protocol stack	72		
1. Data (Message)	73	XXIII. State preparation	108
2. Packet	73	A. Coherent states	109
3. Strategy	75	B. Single-photons	109
4. Transport	75	C. NOON states	110
5. Reconstruction	75	D. Cluster states	111
6. Quality of service	76	E. Greenberger-Horne-Zeilinger states	111
7. Services & applications	76	F. W-states	111
B. Collision handling & classical errors	76	G. Bell states	112
C. Multi-packet operations	77	H. Cat states	112
1. Combining costs & attributes	78	I. Squeezed states	112
D. Extensibility of QTCP	79	J. Matter qubits	113
E. Scalability of QTCP	79		
XVII. Quality of service	79	XXIV. Measurement	113
A. Entanglement purification	79	A. Photo-detection	113
B. 3-qubit code	80	1. Mathematical representation	113
C. 9-qubit code	81	2. Avalanche photo-diodes	113
D. Stabiliser codes	83	3. Superconducting photo-detectors	114
XVIII. QEC with stabilizer codes	84	4. Quantum dot photo-detectors	114
1. Performing error correction	84	5. Photo-diodes	114
A. Surface codes	85	6. Experimental issues	114
B. Unitary error averaging	87	B. Multiplexed photo-detection	115
C. Qubit loss codes	88	C. Homodyne detection	116
D. Gate failure codes	88	D. Bell state & parity measurements	116
E. Decoherence-free subspaces	88	E. Matter qubits	117
F. Dynamical decoupling	89	F. Quantum non-demolition measurement	117
G. Continuous-variables	90	G. Weak measurement	117
XIX. Routing strategies	90	XXV. Evolution	118
A. Single user	90	A. Linear optics	118
B. Multiple users	91	B. Non-linear optics	119
C. Round robin	92	C. Non-optical systems	119
D. Data priority	92		
E. Randomisation	92	XXVI. Quantum memory	120
F. Cost priority	92	A. Network graph representation	120
G. All or nothing	93	B. Physical implementation	120
H. Optimal flow	94	C. Error correction	121

XXVII. High-level protocols	122	XXXVI. The end of classical cryptography?	156
A. Random number generation	122		
B. Entanglement purification	123		
C. Quantum state teleportation	124	XXXVII. Quantum cryptography	156
1. Open-destinations	126	A. Quantum key distribution	156
D. Quantum gate teleportation	127	1. BB84 protocol	157
E. Entanglement swapping	128	2. Privacy amplification	158
F. Quantum cryptography	130	3. E91 protocol	158
G. Superdense coding	130	4. Continuous-variable protocols	159
H. Quantum metrology	131	5. Security	159
I. Quantum state & process tomography	131	6. Public-key cryptography	159
J. Quantum clock synchronisation	131	B. Quantum Enigma machines	160
K. Quantum-enabled telescopy	133	C. Hybrid quantum/classical cryptography	161
		D. Quantum anonymous broadcasting	161
		E. Quantum voting	162
		F. Quantum secret sharing	163
VI Entanglement distribution	135		
XXVIII. Entanglement – The ultimate quantum resource	135	XXXVIII. Attacks on quantum cryptography	164
A. Bell states	135	A. Hacking discrete-variable protocols	164
B. GHZ states	135	1. Beamsplitter & photon-number-splitting attacks	164
C. Cluster states	136	2. Trojan horse and flash-back attacks	164
D. Why specialise in entanglement distribution?	136	3. Detector attacks	164
E. Why not distributed entangling measurements?	136	B. Attacks on continuous-variable protocols	165
XXIX. Quantum repeater networks	138	1. Attacks on the local oscillator	165
A. First-generation repeaters	138	2. Saturation attacks on the detector	165
1. Entanglement distribution	139	3. Trojan horse attacks	165
2. Entanglement purification	140	C. Quantum digital signatures	165
3. Entanglement swapping	142	1. The (classical) Lamport one-time signature	166
4. Performance	142	2. The Gottesman-Chuang QDS	166
B. Second-generation repeaters & error correction	143	3. A practical QDS protocol	166
C. Third-generation repeaters	144		
1. Loss-tolerant codes	144	XXXIX. Quantum crypto-assets	167
2. Operation	146	A. Secure quantum data	167
D. Resource scalings across repeater generations	147	B. Quantum atomic swaps	167
E. The transition to quantum networks	147	C. Quantum smart contracts	168
F. Repeater synchronisation	148		
XXX. The irrelevance of latency	149		
XXXI. The quantum SneakernetTM	150	VIII Quantum computing	168
	150	XL. Models for quantum computation	169
VII Quantum cryptography	150	A. Circuit model	169
XXXII. What is security?	150	B. Cluster states	170
XXXIII. Classical cryptography	151	C. Adiabatic quantum computation	172
A. Private-key cryptography	151	D. Restricted models for quantum computation	174
B. One-time pad cipher	151	E. Fault-tolerance	174
C. Public-key cryptography	152	F. The threshold theorem	175
D. Key exchange protocols	152		
E. Digital signatures	153	XLI. Quantum algorithms	175
F. Hashing	153	A. Deutsch-Jozsa	176
XXXIV. Attacks on classical cryptography	154	B. Quantum search	176
A. Classical attacks	154	C. Oracles	178
1. Brute-force	154	D. Quantum Fourier transform	178
2. Cryptanalysis	154	E. Phase-estimation	179
3. Integer factorisation	154	F. Quantum simulation	180
B. Quantum attacks	155	G. Integer factorisation	181
1. Brute-force	155	H. Quantum machine learning	181
2. Cryptanalysis	155	I. Topological data analysis	182
3. Integer factorisation	155	J. Sampling problems	183
XXXV. Bitcoin & the Blockchain	155	K. Shallow quantum circuits	184
		XLII. Physical architectures for quantum computing	186
		A. Universal linear optics	186
		B. Cluster state linear optics	188
		1. Fusion gates	188
		2. Fusion strategies	189
		C. Weak cross-Kerr non-linearities	193
		D. Passive linear optics	194
		1. BOSONSAMPLING	194

2. Coherent states	198	XLVIII. Time-sharing	231
3. Other linear optics sampling problems	199	XLIX. Economic model assumptions	232
4. Quantum walks	199	A. Efficient markets	232
E. Continuous-variables	202	B. Central mediating authority	232
1. Encoding quantum information using squeezed states	202	C. Network growth	233
2. Phase-shifters	202	D. Hardware cost	233
3. Bell pairs	203	L. Network power	233
4. Measurement	203	LI. Network value	233
5. Logical operations	203	LII. Rate of return	234
6. Cluster states	203	LIII. Market competitiveness	234
7. Fault-tolerance	204	LIV. Cost of computation	234
F. Hybrid light-matter architectures	204	A. Objective value	235
G. Atomic ensembles	205	B. Subjective value	235
H. Ion traps	205	LV. Arbitrage-free time-sharing model	236
I. Superconducting circuits	205	LVI. Problem size scaling functions	237
1. Charge qubits	207	LVII. Quantum computational leverage	237
2. Flux qubits	207	LVIII. Static computational return	239
3. Phase qubits	208	LIX. Forward contract pricing model	239
4. Quantum gates	208	LX. Political leverage	240
	209	LXI. QuantCoin TM – A quantum computation-backed cryptocurrency	240
XLIII. The Quantum Cloud TM	209	A. Spot market model	241
A. Outsourced quantum computation	209	B. Futures market model	242
B. Distributed quantum computation	210	LXII. Economic properties of the qubit marketplace	243
1. In-parallel & in-series computation	210	A. The concept of elasticity	243
2. Quantum enhancement	211	B. Elasticity of the qubit market	243
3. Quantum MapReduce	211	LXIII. Economic implications	243
4. Distributed quantum search algorithm	212	A. The price to pay for isolationism	244
5. Distributed unitary error averaging	213	B. Taxation	244
6. Delocalised computation	213	1. Computational perspective	244
C. Delegated quantum computation	214	2. Policy perspective	245
D. Modularised quantum computation	215	C. The quantum stock market	246
E. Outsourced quantum research	218	D. Geographic localisation	246
F. The globally unified quantum cloud	219	LXIV. Game theory of the qubit marketplace	247
XLIV. Encrypted cloud quantum computation	219	A. Key concepts	247
A. Classical computation	219	B. Strategies	248
1. Homomorphic encryption	220	C. Utility payoff behaviour	248
2. Blind computation	220	D. Cooperative payoff enhancement	250
B. Cluster states	220	1. ISOLATION strategies	250
1. Homomorphic encryption	220	2. LICENSE strategies	250
2. Blind quantum computation	221	3. UNIFY strategies	251
C. Circuit model	221	4. Strategic implications	251
1. Homomorphic encryption	221	5. Inefficient markets	252
2. Blind quantum computation	221	E. Mixed strategies	252
D. Passive optics	221	F. Taxation	253
1. Polarisation-key encoding	222	G. Resource asymmetry	253
2. Phase-key encoding	224	H. Multi-player games	254
3. Displacement-key encoding	226	I. Conclusions	254
E. One-time quantum programs	227	LXV. Summary of economic models	254
F. Authentication	227		
G. Digital signatures	227		
H. Computing on shared sections	227		
XLV. Verification of cloud quantum computing	227		
A. Randomised benchmarking	227		
B. Zero-knowledge proofs	227		
	229		
X. Economics & politics	229		
XLVI. Classical-equivalent computational power & computational scaling functions	229		
A. Virtual computational scaling functions	230	XI. History & state-of-the-art	257
B. Combined computational scaling functions	230		
XLVII. Per-qubit computational power	231	LXVI. Quantum teleportation & entanglement distribution	257

LXVII. Entanglement swapping & quantum repeaters	257	1. Classical machine learning for quantum data	284
LXVIII. Quantum key distribution	257	2. Quantum learning protocols	285
LXIX. Entanglement purification	257	LXXX. Optimising the world	285
LXX. State preparation	258	A. Optimisation in everyday life	285
A. Coherent states	258	1. Traffic networks	286
B. Single-photons	258	2. Public transport scheduling	286
C. Entangled states based on non-linear optics	258	3. Economics	286
LXXI. Non-optical systems	258	4. Supply chain networks	286
A. Atomic ensembles	258	B. Classical optimisation techniques	287
B. Single atoms	258	C. Quantum enhancement	288
C. Quantum dots	258	1. Satisfiability problems	288
D. Nitrogen-vacancy centres	259	2. Graph colouring	288
E. Superconducting rings	259	3. Non-satisfiability-based optimisation algorithms	289
LXXII. Measurement	259	4. Curve fitting	289
A. Photo-detection	259	5. Semidefinite programming	290
B. Homodyning	259	D. An example of a poorly constructed joke	290
LXXIII. Evolution of optical states	259	E. Approximate optimisation	290
LXXIV. Quantum memory	259	F. How beneficial is quantum-enhanced optimisation?	291
LXXV. Quantum computation	259	G. Implications for the future	291
LXXVI. Timelines & references	260	XIII Roadmap – The way forward	292
XII To the future & beyond	270	LXXXI. The tech-web for the quantum future	292
LXXVII. Space-based quantum networks	270	XIV Essays – A new frontier	292
A. International efforts	271	LXXXII. The era of quantum supremacy	294
1. China: QUESS	271	LXXXIII. The global virtual quantum computer	294
2. Japan: SOCRATES	273	LXXXIV. The economics of the quantum internet	295
3. Singapore: Cubesat	273	LXXXV. The quantum future of cryptocurrencies	297
4. Europe: Retroreflector satellite	273	LXXXVI. Security implications of the global quantum internet	298
5. Canada: QEYSSat	273	LXXXVII. Geostrategic quantum politics	300
B. Future challenges	273	LXXXVIII. The quantum space race	300
1. Coverage	274	LXXXIX. The near future: Noisy intermediate-scale quantum technology (NISQ)	302
2. High-precision applications	275	A. Quantum optimisers	302
LXXVIII. Quantum genetic medicine	275	B. Quantum machine learning	303
A. The human genome project	276	C. Quantum semidefinite programming	303
1. Short-read sequencing	276	D. Quantum dynamics	303
2. <i>De novo</i> sequencing	276	XC. The future of quantum cryptography	303
3. Mapping	277	A. Performance	304
B. Genetic medicine	277	1. Key-rate	304
C. Quantum chemistry	278	2. Range	304
D. Quantum drug trial simulations	278	3. Cost & robustness	305
E. Cyborgs – The era of genetic editing	279	B. New protocols	305
F. In the cloud	279	C. Challenges in security	305
LXXIX. The Quantum Mind TM – Quantum machine learning	280	1. QKD with imperfect sources	305
A. Overview	280	2. Measurement-device-independent QKD	305
B. Classical machine learning in classical networks	280	XCI. The quantum ecosystem	306
1. Machine learning basics	280	XV The end	308
2. Security & fault management	281		
3. Traffic management & routing	282		
4. Distributed machine learning	282		
C. Machine learning on classical data with quantum resources	282		
1. Quantum machine learning overview	283		
2. Security & other applications	283		
3. Distributed quantum machine learning	284		
D. Machine learning with quantum data	284		

XCII. Conclusion – The vision of the quantum internet	308
XCIII. How to learn more & get your hands dirty	310
XCIV. Epilogue	311
Acknowledgments	311
Statement of contributions	311
Recommended wine pairings	312
About the authors	312



Part I

Introduction

“Any sufficiently advanced technology is indistinguishable from magic.” — Arthur C. Clarke.

“There may be babblers, wholly ignorant of mathematics, who dare to condemn my hypothesis, upon the authority of some part of the Bible twisted to suit their purpose. I value them not, and scorn their unfounded judgement.” — Nicolaus Copernicus.

“Imagination is more important than knowledge.” — Albert Einstein.

I. MISSION STATEMENT

THE biological evolution of the human race is reaching a plateau. The only mechanism for us to advance further is via technology. The 20th century was that of the transistor, gifting us unprecedented progress as a species, on an exponential trajectory. However, that trajectory

is rapidly approaching saturation. The 21st century will usher in entirely new technologies, far beyond just incremental developments in our current ones. Quantum technology will be one of them, transforming the world as profoundly as the transistor did. Our mission is to contribute to the realisation of this, propelling the human race into the future, benefiting all of humanity, blind to demographics and national borders. Together we will make the future.

— *Docendo discimus.*

II. FOREWORD

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 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 (??). We expect the answer to that question will look very different to what emerged from the classical internet.

A basic understanding of quantum mechanics (?), quantum optics (?), quantum computing and quantum information theory (?)¹, classical networking (?), and computer algorithms (?) 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. To bring readers from a mathematical but non-quantum background up to scratch, in Part. II we present introductory tutorials on quantum mechanics and quantum optics, covering the essential mathematics necessary for following this book.

The entirely technically disinterested or mathematically incompetent reader may refer to just Parts I, XIV & XV – essentially brief non-technical, 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

¹ Throughout this manuscript we use the Nielsen & Chuang convention for the pronunciation of ‘zed’ (?).

existing background in these areas could calmly skip the respective review sections.

Our goal is to present a broadly accessible technical and non-technical 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 don't 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, whilst 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.

The theme music for this work may be found at <http://soundcloud.com/peter-rohde/wir-sind-ein-volk> ©².

— *Liber magnus est.*

III. INTRODUCTION

“Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less.” — Marie Curie.

THE internet is one of the key technological achievements of the 20th century, an enabling factor in every aspect of our everyday use of modern technology. While digital computing was the definitive technology of the 20th century, quantum technologies will be for the 21st (??).

Perhaps the most exciting prospect in the quantum age is the development of quantum computers. Richard Feynman (?) 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 (?) demonstrated for the first time that algorithms can run on a quantum

² We use the copyright (©) and trademark (™) symbols liberally throughout this text when referring to things where we believe commercial opportunities may exist in the future. The authors do not own copyrights or trademarks to anything in this text.

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 one³, with many important applications having been found, including, amongst many others:

- Searching unstructured databases (Sec. XLI.B):
 - Grover's algorithm (?).
 - Quadratic speedup.
- Satisfiability & optimisation problems⁴ (Sec. LXXX):
 - Grover's algorithm.
 - Quadratic speedup.
 - Includes solving **NP**-complete problems, and brute-force cracking of private encryption keys.
 - Many optimisation problems are **NP**-complete or can be approximated in **NP**-complete.
- Period finding and integer factorisation (Sec. XLI.G):
 - Shor's algorithm (?).
 - Exponential speedup.
 - This compromises both Rivest, Shamir & Adleman (RSA) and elliptic-curve public-key cryptography (?), 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 which is not **NP**-complete (the ‘hardest’ of the **NP** problems).
- Simulation of quantum systems (Sec. XLI.F):
 - Lloyd's algorithm (?).
 - 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.

³ See the Quantum Algorithm Zoo for a comprehensive summary of the current state of knowledge on quantum algorithms (<http://math.nist.gov/quantum/zoo/>).

⁴ A satisfiability problem is one where 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.

- Simulation of quantum field theories:
 - Jordan-Lee-Preskill algorithm (??)
 - Exponential speedup.
 - A key area of fundamental physics research.
- Topological data analysis (Sec. XLI.I):
 - Lloyd’s algorithm (??).
 - 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 (??).
 - Exponential speedup.
 - Widespread applications in linear algebra and calculus.
- Quantum machine learning (Sec. LXXIX):
 - Lloyd’s algorithm (?).
 - This includes putting an end to humanity — *Ne obliviscaris*.

An elementary technical overview of some of these archetypal algorithms is presented in Sec. XLI.

It’s likely we haven’t 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*⁵ 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 (?)⁶. The demonstration of large-scale quantum computers, while unable to prove or disprove the ECT thesis⁷,

⁵ 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.

⁶ We have discovered a truly marvellous proof of this, which this footnote is too narrow to contain.

⁷ When one talks 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.

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** & **BPP**) and quantum computers (**BQP**) are distinct. Specifically, it is believed that **BPP** ⊂ **BQP**. If this conjecture is correct, it implies the existence of quantum algorithms super-polynomially faster than the best classical ones, and that the ECT thesis is not correct. More specifically, Fig. 1 illustrates the believed relationships between some of the most important complexity classes relevant to quantum computing.

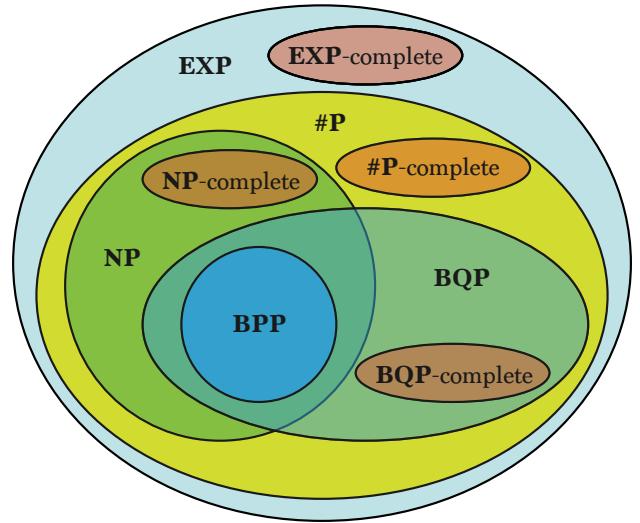


Figure 1: Believed relationships between the complexity classes most relevant to quantum computing. **BPP** is the class of polynomial-time probabilistic classical algorithms. **NP** is the class of problems verifiable in polynomial time using classical algorithms. **NP**-complete are the subset of **NP** 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 (**P** is the same as **BPP**, but deterministic, rather than probabilistic). **EXP** is the class of all algorithms that require exponential time. Note that it is actually unproven whether **P** = **BPP** or **P** ⊂ **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 that we simply haven’t tried hard enough to find the best deterministic algorithms. Furthermore, while it is known that **P** ⊆ **NP**, it is not known whether **BPP** ⊆ **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 that efficiently solvable on universal quantum computers. The computational superiority of quantum computers is based on the (strongly believed, yet unproven) assumption that **BPP** ⊂ **BQP**.

Aside from 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 our understanding of quantum mechanics is correct, quantum cryptographic protocols exist, which 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 notably, quantum cryptography and *cloud quantum computing* will be pivotal in the proliferation of quantum technology, which necessarily requires reliable quantum communications channels.

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 (?). 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 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 was 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 allow 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.

Here we present a treatment for such Quantum Transmission Control Protocols (QTCPs) as a theoretical foundation for a future quantum internet. We consider how such ad hoc networks may be described mathematically, how to quantify network performance, and present a QTCP stack for operating it. While the goals of QTCP are similar as for classical TCP, there are major conceptual differences between the classical and quantum internets, owing to the unique properties of quantum states with no classical analogue.

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 non-optical 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 non-optical systems. Indeed, it is almost certain that future large-scale quantum computers will not be all-optical, necessitating interfacing different physical architectures. We accommodate for this requirement in the design of the QTCP.

Shared quantum entanglement is a primitive resource with direct applications in countless protocols. This warrants special treatment of quantum networks, which do not implement a full QTCP 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 notably, 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 from consumers who can’t afford quantum computers, 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 super-classical scaling in its computational power, brings with it many important economic and strategic considerations that are extremely important to give attention

to in the post-classical 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, 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.

— *Vincit qui se vincit.*

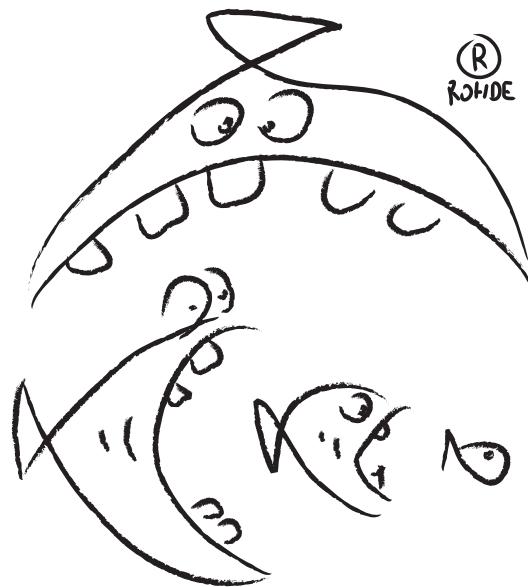
IV. THE QUANTUM INTERNET PROJECT

LUNCHING in parallel with this book is the Quantum Internet Project website, <http://www.quantum-internet.tech>. The purpose of this site is to facilitate ongoing discussion about the subject material covered in the book, and act as a news source, disseminating the latest developments, both in theory and experiment, related to the future quantum internet. The site comprises a blog, and an open online forum for community-driven discussion. The book authors will attempt to remain as actively involved in the conversation as possible.

The key objectives of the forum, which is divided into a set of topic-specific groups, are to facilitate the following:

- Updates on the latest developments in the field.
- Discussion of future research directions.
- Networking researchers and encouraging new research collaborations between them.
- Presentation of new ideas for discussion.
- Act as a central point of reference for all things quantum internet.
- Q&A about material covered in the book.
- Book errata.
- Suggestions for future editions of the book.

— *Citius altius forties.*



Part II

Mathematical foundations

“I come in peace. I didn’t bring artillery. But I’m pleading with you, with tears in my eyes: If you fuck with me, I’ll kill you all.” — General James “Mad Dog” Mattis.

TO make the upcoming discourse accessible to those from a non-quantum background, we will briefly level the playing-field upfront by presenting an overview of the key concepts and mathematical formulations for quantum mechanics and quantum optics, both relied upon heavily throughout this work. All of this mathematical formulation fits within the confines of familiar basic undergraduate linear algebra.

This part ought provide the mathematical background necessary to follow the upcoming discussions. Those already familiar with the laws and formulation of quantum mechanics and quantum optics can calmly skip this part, without risk of offending the authors.

However, rather than just present the mathematical formulations and expect you to reverently accept the maths as the truth, the light, and the way, we will try as best we can to inject some intuition and meaningful interpretation of the mathematics into our tutorial, so that hopefully the reader is able to walk away with not just the ability to follow basic quantum mechanical derivations, but also grasp the intuition of the underlying physical reality.

V. INTRODUCTION TO QUANTUM MECHANICS

“I think I can safely say that nobody understands quantum mechanics.” — Richard Feynman.

THE laws of quantum mechanics are inherently different to those of classical mechanics in virtually every respect, and bring with them their own unique mathematical notation for:

- The representation of quantum states.
- The evolution of quantum systems.
- The measurement of quantum states.

We review all of these formalisms in the upcoming sections, and present the necessary mathematical framework to enable following the derivations presented here.

A. Features of quantum mechanics

Classical and quantum physics are fundamentally different in a number of extremely important ways. These differences are what allow quantum mechanics to enable the entire new field of quantum technologies, most notably quantum computing, which offers substantial computational gains over traditional classical computing. The same applies to the novel and groundbreaking promises of the other key quantum technologies, such as quantum cryptography, quantum teleportation, and quantum-enhanced metrology. All of these gains are achieved by exploiting these unique new characteristics offered by quantum systems, that are physically impossible in a classical context – all technologies must obey the laws of physics, and therefore new laws of physics necessarily imply new technological possibilities.

The key conceptual differences between classical and quantum mechanics are:

- Superposition: the ability of quantum systems to be in multiple states simultaneously (simultaneously, not probabilistically!). This arises since quantum states can always be interpreted as waves, which are inherently delocalised and non-particle-like.
- Quantum amplitudes: quantum states are not governed by classical probabilities, but by *quantum amplitudes*, which are complex numbers rather than real numbers.
- Entanglement: a type of non-local quantum correlation, stronger than any possible classical correlation.
- Unitary evolution: the free evolution of quantum systems is restricted to being described by unitary operators. This implies that all quantum evolution is perfectly time-reversible, since all unitary operators are invertible.

- Measurement collapse: the measurement of a quantum system changes its state in general, *collapsing* it to the state corresponding to the measurement outcome⁸.

- Non-determinism: quantum systems exhibit genuine non-determinism, associated with the randomness of the measurement collapse process⁹.

- Exponential complexity: the number of unique parameters required to fully describe a quantum system grows exponentially with its ‘size’ (e.g the number of constituent subsystems, such as quantum particles or qubits). This is in contrast to classical systems whose state complexity grows only linearly with system size, since it’s always possible to describe constituent subsystems independently (there is no notion of entanglement). Therefore the number of parameters required to characterise classical systems accumulates additively, whereas for quantum systems it grows multiplicatively.

- Wave-particle duality: all quantum objects can simultaneously be considered to be particles and waves, where the wave-like and particle-like characteristics express themselves differently depending on the application.

- Uncertainty principle: unlike classical systems, where we can essentially know everything about them to arbitrary accuracy, quantum systems necessarily exhibit uncertainties. Heisenberg’s uncertainty principle places strict physical bounds on the amount of information that can be known about quantum systems, exhibiting tradeoffs between how

⁸ An open and intuitively confusing conundrum in quantum theory is to resolve the question of what exactly it is that constitutes ‘measurement’ – the *measurement paradox*. Given that all quantum evolution is supposed to be governed by unitary dynamics, whereas measurements are not, at what point in physical reality does an operation transition from one to the other? This is an extremely important open question, particularly from the perspective of quantum foundations and the philosophy of physics. This has led to the emergence of countless theories for the interpretation of quantum mechanics, which is an essentially untestable field of research, more akin to fundamentalist religious movements than physical science. However, the resolution is largely unimportant in a practical technological context, where it suffices to define ‘measurement’ as being ‘when quantum information becomes classical information’, e.g once we are able to read it off a measurement device’s LCD screen!

⁹ When we say ‘random’ we *really* mean random, not ‘quasi-random’ but actually deterministic’. This true randomness is another of the unique and surprising features of quantum mechanics. Conversely to classical physics, quantum mechanics is inherently and provably non-deterministic! Specifically, it has been proven theoretically (?) that so-called ‘hidden variable theories’ are incompatible with quantum physics, which rules out any form of quasi-randomness and mandates true randomness.

much we can know about different parameters simultaneously. The best-known example of this is the position-momentum uncertainty principle, which tells us that the more information we know about an object's position, the less we can simultaneously know about its momentum, and vice versa. It is not possible even in principle to have perfect information about both (this is a physical limitation, not an engineering one!).

All of these phenomena will become more clear and intuitive as we introduce the basic relevant mathematical constructions in the upcoming sections.

B. Quantum states

A quantum state is a mathematical description of a quantum object, fully characterising its quantum mechanical properties, much like classical attributes like momentum and position characterise the state of classical objects. This captures both the wave- and particle-like properties of quantum objects, given the quantum principle of *wave-particle duality*. The key mathematical difference is that quantum objects exhibit unique, non-classical properties like superposition and entanglement. This requires introducing new constructs for their mathematical representation.

1. State vectors

The most fundamental mathematical object in representing quantum systems is the *state vector*, a vector representation for the quantum wave-function.

A state vector resides in an N -dimensional complex Hilbert space, $\mathcal{H} \in \mathbb{C}^N$, where N denotes the number of basis states supporting the system. This support could be, for example, the available electron energy levels in an atom, or the two orthogonal polarisations available to a photon in an optical mode.

These vector representations for quantum states can quickly become unwieldy, since their dimensionality grows exponentially with the number of constituent subsystems. For example, a single qubit resides in a 2-dimensional Hilbert space, $\mathcal{H} \in \mathbb{C}^2$, whereas an n -qubit system resides in a 2^n -dimensional one, $\mathcal{H} \in \mathbb{C}^{2^n}$. This is in contrast to n classical bits, which require just n parameters to collectively describe. Note that the difference in scaling between the vector representations of quantum systems and classical ones is exponential in nature. It is from this simple observation that the first intuitive inkling arises that perhaps this inherent astronomical complexity can be exploited for our benefit! — *Ad astra!*

Mathematically, we can employ the following alternate state vector representations, where the latter is referred

to as *Dirac notation* (with $|n\rangle$ denoting the basis states),

$$|\psi\rangle = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \sum_{n=1}^N \alpha_n |n\rangle. \quad (5.1)$$

Here $\alpha_n \in \mathbb{C}$ are referred to as the *amplitudes* of the state, which for normalisation satisfy,

$$\sum_n |\alpha_n|^2 = 1. \quad (5.2)$$

Contrast this with the mathematical definition for the normalisation of classical probability distributions,

$$\sum_n p_n = 1. \quad (5.3)$$

The mathematical objects denoted as $|\cdot\rangle$ are commonly referred to as *kets*.

In this book we will largely be dealing with qubits, the simplest unit of quantum information, which are supported by exactly two basis states, denoted $|0\rangle$ and $|1\rangle$. A qubit can therefore be expressed as,

$$|\psi\rangle = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \alpha|0\rangle + \beta|1\rangle. \quad (5.4)$$

These are referred to as the *logical basis states*, which are treated abstractly, detached from a particular physical realisation (e.g by associating them with electron energy levels, photon polarisation, or indeed *any* two-level quantum system), much in the same way that classical bits are treated abstractly as a unit of information, detached from physical reality.

Also mathematically useful is the *dual* of a state vector, which is simply the complex conjugate transpose (\dagger) of the vector, capturing the same information about the state in conjugate form,

$$\langle\psi| = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}^\dagger = (\alpha^*, \beta^*) = \alpha^*|0\rangle + \beta^*|1\rangle. \quad (5.5)$$

The objects $\langle\cdot|$ are referred to as *bras*.

a. *State overlap* For two generic state vectors, supported by the same basis,

$$|\psi\rangle = \sum_n \alpha_n |n\rangle, \quad |\phi\rangle = \sum_n \beta_n |n\rangle, \quad (5.6)$$

the *overlap* between them is defined as,

$$\langle \psi | \phi \rangle = \sum_n \alpha_n \beta_n^*, \quad (5.7)$$

which is mathematically equivalent to their vector dot product. The magnitude of the overlap of two quantum states lies in the range,

$$0 \leq |\langle \psi | \phi \rangle| \leq 1, \quad (5.8)$$

where $|\langle \psi | \phi \rangle| = 0$ indicates that the two states were orthogonal, and $|\langle \psi | \phi \rangle| = 1$ implies they are identical (up to an irrelevant global phase¹⁰). The object $\langle \cdot | \cdot \rangle$ is referred to as a *braket*¹¹. The normalisation condition from Eq. (5.2) can therefore equivalently be written,

$$\langle \psi | \psi \rangle = 1. \quad (5.9)$$

Owing to the cyclic property of the trace operator, it follows that the trace,

$$\text{tr}(|\psi\rangle\langle\phi|) = \langle\psi|\phi\rangle, \quad (5.10)$$

is a state overlap. This quantity has the useful intuitive interpretation as the *distinguishability* between quantum states, which plays an important role in measurement theory and the theory of quantum information. It is also widely used to characterise the quality of quantum experiments, as a way of comparing quantum states prepared in the lab with the ideal states obtained from the theory.

Because basis states are orthonormal, this implies that for basis states $|m\rangle$ and $|n\rangle$ we have the trivial delta function overlap identity,

$$\langle m | n \rangle = \delta_{m,n}. \quad (5.11)$$

2. Composite systems

The joint state vector of a multi-partite system is represented using a tensor (or Kronecker) product Hilbert space, which has dimensionality equal to the product of their individual dimensions¹², hence yielding the dreaded (cherished!) exponential growth in the complexity of quantum systems,

$$\mathcal{H}_{A,B} = \mathcal{H}_A \otimes \mathcal{H}_B, \quad (5.12)$$

where,

$$\begin{aligned} \mathcal{H}_A &\in \mathbb{C}^M \text{ (system } A), \\ \mathcal{H}_B &\in \mathbb{C}^N \text{ (system } B), \\ \mathcal{H}_{A,B} &\in \mathbb{C}^{MN} \text{ (composite system).} \end{aligned} \quad (5.13)$$

The state vector of two *independent* subsystems is simply given by their tensor product¹³,

$$\begin{aligned} |\psi\rangle_{A,B} &= |\psi\rangle_A \otimes |\phi\rangle_B \\ &= \sum_{m,n} \alpha_m \beta_n |m\rangle_A \otimes |n\rangle_B. \end{aligned} \quad (5.14)$$

This is a so-called *separable state*, since the state vector can be factorised into independent terms characterising each of the subsystems – each subsystem has its own complete and independent physical description in isolation.

3. Entangled states

More generally, we may also have states of the form,

$$|\psi\rangle_{A,B} = \sum_{m,n} \lambda_{m,n} |m\rangle_A \otimes |n\rangle_B, \quad (5.15)$$

where in general $\lambda_{m,n}$ may not be separable as,

$$\lambda_{m,n} = \alpha_m \beta_n. \quad (5.16)$$

In this case no complete physical description exists for either subsystem in isolation – they may only be described collectively. Now the state is said to be *entangled*, a uniquely quantum class of physical states with no classical analogue. The simplest and best-known example is the *Bell state* on two qubits,

$$\begin{aligned} |\Phi\rangle_{A,B}^+ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} (|0\rangle_A |0\rangle_B + |1\rangle_A |1\rangle_B). \end{aligned} \quad (5.17)$$

It's obvious upon inspection that the Bell state cannot be written in separable form,

$$|\Phi\rangle_{A,B}^+ \neq |\psi\rangle_A \otimes |\phi\rangle_B, \quad (5.18)$$

and is hence entangled.

Entanglement is one of the key unique features of quantum physics, which underpins the operation of most quantum information processing protocols, including quantum

¹⁰ In quantum mechanics, all quantum states are physically invariant under the introduction of a *global* phase-factor, i.e $|\psi\rangle \rightarrow e^{i\theta} |\psi\rangle$. Thus, global phases may always be safely ignored without changing the underlying physical reality. This most certainly does not apply to *local* phases, which are a part of physical reality, and critically affect interference dynamics.

¹¹ We never quite appreciated the corny sense of humour of the author of the terms *bra*, *ket*, and *braket*. But somehow it stuck.

¹² This is in contrast to the dimensionality of composite classical systems, which grow additively.

¹³ Often the \otimes symbol is omitted for brevity and left implicit.

computing in particular, where extremely complex many-qubit entanglement is present and exploited for computational advantage. This immediately hints that in future quantum information systems, where information will be encoded into entangled states in general, information will be highly non-local, in contrast to most standard methods for representing classical information.

4. Density operators

The state vector formalism presented above applies to *pure states*, i.e ones which exhibit perfect quantum coherence (i.e superposition), and no classical randomness. But realistic physical systems combine both classical probabilities *and* quantum superposition amplitudes, so-called *mixed states*. To capture both of these features simultaneously we employ *density operators*. For an N -dimensional Hilbert space, the density operator, $\hat{\rho}$, is an $N \times N$ complex Hermitian matrix of the form,

$$\hat{\rho} = \hat{\rho}^\dagger, \quad (5.19)$$

which satisfies,

$$\text{tr}(\hat{\rho}) = \sum_i \hat{\rho}_{i,i} = 1, \quad (5.20)$$

for normalisation.

Thus, in the special case of a qubit, density operators are 2×2 complex matrices,

$$\hat{\rho} = \begin{pmatrix} a & b \\ b^* & 1-a \end{pmatrix}, \quad (5.21)$$

where diagonal elements are necessarily real, and off-diagonal ones may be complex in general.

The intuitive physical interpretation of density matrices is as follows:

- Rows and columns enumerate the basis states of the Hilbert space.
- Diagonal elements represent the classical probabilities of the respective basis states.
- Off-diagonal elements represent *coherences* between pairs of basis states, i.e whether they exist as a classical probability distribution (when the coherence is zero) or as a quantum superposition (when non-zero). The $\hat{\rho}_{m,n}$ matrix element (for $m \neq n$) represents the degree of coherence between basis states $|m\rangle$ and $|n\rangle$.

Therefore, referring back to the qubit example in Eq. (5.21), a and $1 - a$ can be regarded as the *classical* probability amplitudes associated with being in the $|0\rangle$ or $|1\rangle$ basis state, while b characterises to what extent it is a classical mixture of these two states (for $b = 0$) or a coherent superposition of them (for $|b| = \sqrt{a(1-a)}$).

The density operator captures *all* information (both classical and quantum) that can be known about a physical system under the standard laws of quantum physics, and is therefore the most general representation for quantum states, and a very powerful representation to employ.

For a pure state (a qubit in this example), the density operator takes the form of a vector outer product,

$$\begin{aligned} \hat{\rho} &= |\psi\rangle\langle\psi| \\ &= \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}. \end{aligned}$$

A classical mixture of quantum states $\hat{\rho}_i$ with probability distribution p_i takes the form,

$$\hat{\rho} = \sum_i p_i \hat{\rho}_i, \quad (5.22)$$

where the probabilities are normalised such that,

$$\sum_i p_i = 1. \quad (5.23)$$

Eq. (5.22) has the elegant interpretation that the state $\hat{\rho}$ is in quantum state $\hat{\rho}_i$ with classical probability p_i .

a. *Purity* An important measure on density operators is their *purity*, which quantifies to what extent the system is coherent (i.e *not* classically probabilistic),

$$\mathcal{P} = \text{tr}(\hat{\rho}^2), \quad (5.24)$$

where,

$$\frac{1}{N} \leq \mathcal{P} \leq 1, \quad (5.25)$$

and $\mathcal{P} = 1$ only for pure states.

5. Reduced states

Suppose there exists a bipartite system, to which we only have access to one subsystem. Since in general quantum states are not always separable, we now necessarily have an incomplete physical description of what we have access to. This physical description is obtained by taking the *partial trace* of the joint system, tracing out the subsystem to which we don't have access,

$$\hat{\rho}_A = \text{tr}_B(\hat{\rho}_{A,B}). \quad (5.26)$$

$\hat{\rho}_A$ is referred to as the *reduced state* of subsystem A , and captures all that can be known about A in isolation of B .

In general, if A and B are entangled, the reduced state of either will provide an incomplete physical description, since it ignores the entanglement between them, capturing only local properties of the state. For this reason,

attempting to reconstruct a joint density operator from its reduced density operators, in general does not faithfully reproduce the original joint density operator, unless the two subsystems were perfectly separable to begin with. That is, for,

$$\begin{aligned}\hat{\rho}_A &= \text{tr}_B(\hat{\rho}_{A,B}), \\ \hat{\rho}_B &= \text{tr}_A(\hat{\rho}_{A,B}),\end{aligned}\quad (5.27)$$

in general,

$$\hat{\rho}_{A,B} \neq \hat{\rho}_A \otimes \hat{\rho}_B. \quad (5.28)$$

The partial trace is a linear operator defined to satisfy the following properties,

$$\begin{aligned}\text{tr}_B(\hat{\rho}_A \otimes \hat{\rho}_B) &= \text{tr}(\hat{\rho}_B) \cdot \hat{\rho}_A, \\ \text{tr}_B(|i\rangle_A \langle j|_A \otimes |k\rangle_B \langle l|_B) &= \langle k|l\rangle \cdot |i\rangle_A \langle j|_A, \\ \text{tr}_B(\hat{\rho} + \hat{\sigma}) &= \text{tr}_B(\hat{\rho}) + \text{tr}_B(\hat{\sigma}).\end{aligned}\quad (5.29)$$

A useful property of the reduced state of a larger pure state is that its purity characterises how entangled that subsystem is with the peripheral system. For example, a joint system which is separable and pure exhibits reduced states which are also pure. But the reduced state of an entangled state, such as that shown in Eq. (5.17), is necessarily mixed. Thus, the purity \mathcal{P} of the reduced states of a joint system may be used as a metric for quantifying the degree of entanglement in the system. This provides an extremely important link between entanglement and decoherence – the loss of coherence in an observed quantum system can always be attributed to it becoming entangled to the external environment, to which we don't have access.

C. Evolution

Thus far we have only provided a description for the state of quantum systems. What about their evolution? We can't execute any algorithm without evolution! Evolution in quantum mechanics has two forms, discrete-time and continuous-time.

1. Discrete-time evolution

In quantum mechanics the discrete-time evolution of an N -dimensional state is given by arbitrary $N \times N$ -dimensional unitary transformations, $U(N)$, satisfying,

$$\hat{U}^\dagger \hat{U} = \hat{\mathbb{1}}. \quad (5.30)$$

Intuitively, this corresponds to the class of higher-dimensional complex rotations that preserve the orthonormality of quantum states – i.e just a change of basis in the higher-dimensional Hilbert space. In principle, any

unitary operator is physically allowed, although most are challenging to artificially engineer, motivating quantum engineers to typically decompose quantum protocols into *universal gate sets* – a restricted but accessible set of operations that are practical to implement, yet sufficient to constructively build up whatever complex transformations we need.

The evolution of quantum state vectors under unitary transformation is simply given by matrix multiplication,

$$|\psi'\rangle = \hat{U}|\psi\rangle, \quad (5.31)$$

or for density operators by conjugation,

$$\hat{\rho}' = \hat{U} \hat{\rho} \hat{U}^\dagger. \quad (5.32)$$

2. Continuous-time evolution

Sometimes an alternate continuous-time description is employed. Here the system is governed by a *Hamiltonian* operator, \hat{H} , the direct quantum analogue of a classical Hamiltonian, which simply relates to unitary evolution via exponentiation,

$$\hat{U} = e^{-i\hat{H}t}, \quad (5.33)$$

over evolution time t .

3. Schrödinger vs Heisenberg pictures

The above formalism describes quantum evolution by acting the unitary operator on the quantum state directly,

$$|\psi'\rangle = \hat{U}|\psi\rangle. \quad (5.34)$$

This is known as the *Schrödinger picture*. An alternate, but entirely mathematically equivalent approach is the *Heisenberg picture*, whereby we evolve the operators acting upon a state, rather than the state itself. Now the operator evolution is described by conjugating the operators with the unitary operator describing the evolution. That is, representing the quantum state in terms of some operator \hat{o} , the operator evolution proceeds as,

$$\hat{o}' = \hat{U} \hat{o} \hat{U}^\dagger. \quad (5.35)$$

For example, representing photonic states using creation operators \hat{a}^\dagger (introduced in Sec. VI.A.3), the evolved creation operator is simply given by $\hat{U} \hat{a}^\dagger \hat{U}^\dagger$. In quantum optics in particular, where creation operator representations are highly convenient, this is a very useful formulation for describing quantum evolution.

D. Measurement

The final ingredient in our toolbox of allowed quantum operations is measurement. After all, what use is a quantum protocol if we can't read out the answer?

So-called *projective measurements* in quantum mechanics are described by sets of *measurement projectors*,

$$\hat{M}_i = |m_i\rangle\langle m_i|, \quad (5.36)$$

where,

$$\sum_i \hat{M}_i = \hat{\mathbb{1}}, \quad (5.37)$$

for normalisation. Each projector in the set corresponds to a possible allowed measurement outcome. After measuring a given outcome the state is disturbed and collapses (i.e. is *projected*) into being in the state corresponding to the measurement outcome.

If the set of projectors corresponds to the eigenbasis of some unitary operator \hat{U} , we can say that the measurement is performed in the \hat{U} -basis.

Importantly – and this is a key feature of quantum mechanics – we cannot control what the measurement outcome is. It occurs randomly according to some classical probability distribution, given by the overlaps between the measurement projectors and the measured quantum state.

The set of possible projected states following measurement is given by,

$$\begin{aligned} |\psi_i\rangle &= \frac{\hat{M}_i|\psi\rangle}{\sqrt{p_i}}, \\ \hat{\rho}_i &= \frac{\hat{M}_i\hat{\rho}\hat{M}_i^\dagger}{p_i}, \end{aligned} \quad (5.38)$$

where those outcomes occur with probabilities,

$$\begin{aligned} p_i &= \langle\psi|\hat{M}_i^\dagger\hat{M}_i|\psi\rangle, \\ p_i &= \text{tr}(\hat{M}_i\hat{\rho}\hat{M}_i^\dagger). \end{aligned} \quad (5.39)$$

Taking the state vector representation from Eq. (5.1), performing a measurement in the basis in which the vector is expressed (i.e. measurement projectors $|n\rangle\langle n|$), the measurement probabilities are simply given by the absolute squares of the respective amplitudes,

$$p_i = |\alpha_i|^2. \quad (5.40)$$

More general classes of measurements are allowed, given by measurement operators \hat{M}_i comprised of linear combinations of projectors. The same rules above for the action of these measurement operators apply, Eqs. (5.38, 5.39).

The described formalism in terms of measurement projectors gives us ‘single-shot’ information about a measurement – what is the probability of a given single measurement outcome, and what is the state of the system following this measurement? Alternately, we might be interested not in single-shot events, but statistical measurement information, such as the average measurement outcome in the limit of a large number of repeated independent trials.

We define a measurement *observable* as an operator \hat{O} acting on a state-space, with eigenvalues O_i over some range of i . The eigenvalues O_i denote the allowed values of the measurement outcome for the associated eigenstate, which corresponds to the projected post-measurement state. The statistical expectation value of these measurement outcomes for a given state $|\psi\rangle$ may now be obtained as the expectation value of the observable,

$$\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle, \quad (5.41)$$

and their variance by,

$$\langle\hat{O}^2\rangle - \langle\hat{O}\rangle^2. \quad (5.42)$$

Both of these are in the limit of an infinite number of independent trials of the measurement performed on identical copies of the state.

E. Example derivation

To illustrate the basic techniques presented above, we now present a derivation of the operation of a very simple quantum protocol comprising state preparation, evolution, and measurement.

Let us begin by preparing a single-qubit state into the logical basis state,

$$|\phi\rangle = |0\rangle, \quad (5.43)$$

where we have employed Dirac notation for simplicity.

Next we will apply the so-called Hadamard gate, given by the 2×2 unitary matrix,

$$\hat{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (5.44)$$

Applying this gate to our input state yields,

$$\begin{aligned} |\psi\rangle &= \hat{H}|0\rangle \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \end{aligned} \quad (5.45)$$

which is an equal superposition of the two logical basis states.

Applying our measurement identities, we obtain measurement probabilities,

$$\begin{aligned} p_0 &= |\langle 0|\psi\rangle|^2 = \frac{1}{2}, \\ p_1 &= |\langle 1|\psi\rangle|^2 = \frac{1}{2}, \end{aligned} \quad (5.46)$$

and respective post-measurement projected states,

$$\begin{aligned} |\psi_0\rangle &= \sqrt{2}|0\rangle\langle 0|\psi\rangle = |0\rangle, \\ |\psi_1\rangle &= \sqrt{2}|1\rangle\langle 1|\psi\rangle = |1\rangle. \end{aligned} \quad (5.47)$$

Thus, we see that with 50:50 probability the measurement randomly collapses the measured state onto the $|0\rangle$ or $|1\rangle$ states.

VI. INTRODUCTION TO QUANTUM OPTICS

"All the fifty years of conscious brooding have brought me no closer to answer the question, 'What are light quanta?'. Of course today every rascal thinks he knows the answer, but he is deluding himself." — Albert Einstein.

MUCH as the classical internet is heavily dependent upon optics to mediate communication, any future quantum internet will almost inevitably rely heavily on quantum optics to mediate quantum communication. It is therefore worth introducing the language and basic set of tools employed by quantum opticians.

A. Discrete-variables

Like most things in the quantum world, light is discretised into fundamental, indivisible particles called *photons* – light quanta. This yields the most common representation for quantum states of light, referred to as the *discrete-variables* (DV) picture.

1. Photon-number states

The most basic optical quantum state comprising photons is the *photon-number state* or *Fock state*. These states form a discrete basis, labelled by the integer number of photons in the state,

$$|n\rangle, n \in \mathbb{Z}^+. \quad (6.1)$$

The special case of $|0\rangle$ is referred to as the *vacuum state*, since it contains no photons.

The Fock states can be thought of as the energy levels in a quantum harmonic oscillator, and the energy of the n -photon Fock state is therefore given by,

$$E_n = \left(n + \frac{1}{2}\right)\hbar\omega, \quad (6.2)$$

where ω is the optical frequency (in radians per second).

Any optical state in a single mode can be represented in the photon-number basis, which generalises to the multi-mode case in the obvious way, using a basis of composite photon-number states,

$$|\vec{n}\rangle = |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_m\rangle, \quad (6.3)$$

for m optical modes.

2. Measurement

A measurement in the photon-number basis is represented using photon-number projectors,

$$\hat{\Pi}_n = |n\rangle\langle n|, \quad (6.4)$$

which obey the usual completeness relation for measurement operators,

$$\sum_{n=0}^{\infty} \hat{\Pi}_n = \hat{\mathbb{1}}. \quad (6.5)$$

3. Creation & annihilation operators

In many cases it's convenient to represent states using photonic *creation* (\hat{a}^\dagger) and *annihilation* (\hat{a}) operators. These (non-commuting) operators have the effect of acting upon a photon-number state and incrementing or decrementing its photon-number. Note that these operators are not unitary, and therefore do not represent legitimate quantum evolutions on their own. However, they may be employed to construct unitary operators representing physically legitimate evolution processes.

These operators satisfy the following basic algebraic properties:

$$\begin{aligned} \hat{a}^\dagger|n\rangle &= \sqrt{n+1}|n+1\rangle, \\ \hat{a}|n\rangle &= \sqrt{n}|n-1\rangle, \\ \hat{a}|0\rangle &= 0. \end{aligned} \quad (6.6)$$

From this, it follows that an arbitrary Fock state can be expressed in terms of creation operators as,

$$|n\rangle = \frac{1}{\sqrt{n!}}(\hat{a}^\dagger)^n|0\rangle. \quad (6.7)$$

The creation and annihilation operators obey the commutation relation,

$$[\hat{a}, \hat{a}^\dagger] = 1, \quad (6.8)$$

where,

$$[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}, \quad (6.9)$$

is the *commutator* of two operators (operators that commute obviously necessarily exhibit $[\hat{A}, \hat{B}] = 0$). For multiple modes (i and j) the different creation operators commute, as do the respective annihilation operators,

$$\begin{aligned} [\hat{a}_i^\dagger, \hat{a}_j^\dagger] &= 0, \\ [\hat{a}_i, \hat{a}_j] &= 0. \end{aligned} \quad (6.10)$$

The physical interpretation of this property is that because photons are bosons, they are exchange-symmetric, and thus the physical state is invariant under swapping them.

4. Photon-number operators

A particularly useful and ubiquitous operator is the photon-number operator, defined as,

$$\hat{n} = \hat{a}^\dagger \hat{a}, \quad (6.11)$$

which satisfies the eigenvalue relation with photon-number states,

$$\hat{n}|n\rangle = n|n\rangle. \quad (6.12)$$

Treating it as a measurement observable, the expectation value of the photon-number operator, $\langle \hat{n} \rangle$, gives us the average measured photon-number when a state is measured in the photon-number basis, yielding the statistical average of the photon-number measurement outcome.

B. Continuous-variables

A completely alternate, but entirely equivalent formalism for the representation of quantum optical states is in the *continuous-variables* (CV) picture. Here we no longer think in terms of a discretised photon-number basis, but in terms of a continuous basis in the complex plane, referred to as *phase-space*, completely analogous to the familiar classical phase-space. This alternate representation is introduced purely as a matter of convenience, since some states (typically referred to as CV states), are particularly well-suited to elegant representation in this picture, as are some types of evolution.

1. Position & momentum operators

In the CV picture, instead of representing states using photonic creation operators, we express them in terms of the *position* (\hat{x}) and *momentum* (\hat{p}) operators, which are related to the creation and annihilation operators via,

$$\begin{aligned} \hat{x} &= \sqrt{\frac{\hbar}{2\omega}}(\hat{a} + \hat{a}^\dagger), \\ \hat{p} &= -i\sqrt{\frac{\hbar\omega}{2}}(\hat{a} - \hat{a}^\dagger), \end{aligned} \quad (6.13)$$

where ω is the optical frequency. These operators obey the commutation relation,

$$[\hat{x}, \hat{p}] = i\hbar, \quad (6.14)$$

and satisfy the eigenvalue relations with the position and momentum eigenstates,

$$\begin{aligned} \hat{x}|x\rangle &= x|x\rangle, \\ \hat{p}|p\rangle &= p|p\rangle. \end{aligned} \quad (6.15)$$

These eigenstates satisfy the completeness relations,

$$\begin{aligned} \int_{-\infty}^{\infty} |x\rangle\langle x| dx &= \hat{1}, \\ \int_{-\infty}^{\infty} |p\rangle\langle p| dp &= \hat{1}. \end{aligned} \quad (6.16)$$

The position and momentum operators represent the *quadratures* of a mode, and correspond to the real and imaginary components of a harmonic oscillator's amplitude.

2. Position & momentum representations

The *position representation* of an optical state corresponds to expressing a state vector $|\psi\rangle$ in the position basis,

$$|\psi\rangle = \int \psi(x)|x\rangle dx. \quad (6.17)$$

Here the wave-function in the position representation is defined as,

$$\psi(x) = \langle x|\psi\rangle. \quad (6.18)$$

The state can similarly be expressed via a *momentum representation*,

$$|\psi\rangle = \int \tilde{\psi}(p)|p\rangle dp, \quad (6.19)$$

s with wave-function,

$$\tilde{\psi}(p) = \langle p|\psi\rangle, \quad (6.20)$$

The quadrature eigenstates are mutually related to one another by a Fourier transform,

$$\begin{aligned} |x\rangle &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-2ixp} |p\rangle dp, \\ |p\rangle &= \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{2ixp} |x\rangle dx. \end{aligned} \quad (6.21)$$

3. Coherent states

A particularly ubiquitous state that emerges in CV representations is the *coherent state*, which closely approximates classical laser light,

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle, \quad (6.22)$$

where $\alpha \in \mathbb{C}$ is the complex coherent amplitude. Alternately, coherent states can be defined as the eigenstates of the annihilation operator,

$$\hat{a}|\alpha\rangle = \alpha|\alpha\rangle. \quad (6.23)$$

Writing the coherent amplitude as,

$$\alpha = re^{i\theta}, \quad (6.24)$$

r can be interpreted as the strength of the classical field, and θ as its local phase.

Note that the coherent state has indeterminate photon-number (and hence energy) – it is in fact a coherent superposition of *all* photon-numbers (except when $\alpha = 0$). However, the *mean* photon-number is a useful quantity, given by,

$$\bar{n} = \langle \hat{n} \rangle = |\alpha|^2. \quad (6.25)$$

4. Phase-space representations

We can graphically represent phase-space on the complex plane, where the two axes denote expectation values of the \hat{x} and \hat{p} operators. Unlike classical phase-space, in quantum phase-space states are ‘blobs’ rather than points. The variance of these blobs represents the measurement uncertainty in a given direction, which is strictly non-zero in the quantum world, owing to the Heisenberg uncertainty principle.

The phase-space representation for a coherent state is a circular blob with mean α in the complex plane, shown in Fig. 2. Its variance is indicative of the quantum uncertainty of the field in the two quadrature directions. A coherent state is a *minimum uncertainty state*, where the product of the variances in the two orthogonal directions is minimised. For other states it will be greater in general, but strictly not less.

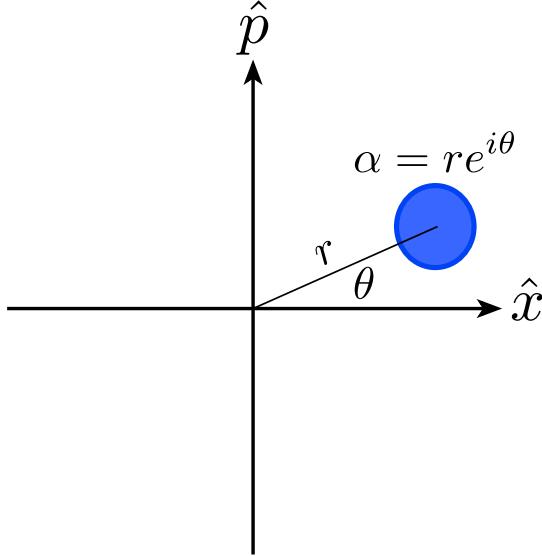


Figure 2: Phase-space representation for a coherent state of complex amplitude $\alpha = re^{i\theta}$, where r can be regarded as the field strength, and θ its phase.

Beyond coherent states, represented by circular blobs, all manner of other geometries are available, allowing

states such as *squeezed states* (another useful type of CV state), which are rather complex to work with in the DV picture, to be elegantly graphically represented.

In phase-space, the evolution of a state under the accumulation of optical phase simply manifests itself as a rotation of the complex plane around the origin over time. Thus, the coherent state blob simply revolves around the origin. On the other hand, photon-number states manifest themselves as concentric rings in phase-space. Since this geometry is circular-symmetric, its phase-space representation is invariant under phase evolution.

VII. INTRODUCTION TO QUANTUM INFORMATION THEORY

“When you find yourself in a room surrounded by your enemies you tell yourself, ‘I am not locked in here with you, you are locked in here with me’. This is the kind of mindset you should have if you want to succeed in life. Get rid of that victim mentality.” — Bruce Lee.

A. Probability, information and classical correlation measures

“Information theory begins with the observation that there is a fundamental link between probabilities and information”(?).

Suppose that there are two parties trying to communicate, conventionally named Alice and Bob. Alice sends information using an alphabet $\{x_j\}$, and each letter occurs with probability p_j . If some letter occur less likely than others, then Bob will be more surprised receiving those, and one might naturally expect that the information content of those events are higher.

This suggests a way to measure the *information content* of an outcome, or how surprised Bob is:

$$i(x_j) \equiv \log \left(\frac{1}{p(x_j)} \right) = -\log(p(x_j)) \quad (7.1)$$

We can model Alice’s source as a random variable X , whose outcomes are $\{x_j\}$ each occurring with probability p_j . Then the expected information content of the source, or *Shannon entropy* associated with the X is defined as

$$H(X) = - \sum_i p_i \log_2(p_i). \quad (7.2)$$

Entropy plays a central role in information theory. The intuitive interpretation is that it quantifies an experimenter’s uncertainty about X before measuring it, and his expected information gain is $H(X)$ bits upon learning the outcome. The information $H(X)$ is zero if and only if one of the probabilities $p(x)$ is unity, with the others being zero. In this case the value of X is already known and so there is no information to be gained from observing it.

If there are two random variables X and Y , the joint entropy of the two is simply given by,

$$H(X, Y) = - \sum_{x,y} p_{x,y} \log_2(p_{x,y}). \quad (7.3)$$

Note that the joint entropy $H(X, Y) = H(X) + H(Y)$ if and only if X and Y are independent, i.e. the occurrence of one does not change the probability of the other.

If the two distributions are not independent, i.e. knowing something about X reveals information on Y , the two

Figure 3

variables are said to be *correlated*. Suppose Alice possesses the random variable X and Bob has the random variable Y , the *mutual information* specifies the number of bits in common between the two distributions. Equivalently, this represents the maximum number of bits that one party can learn about the other just by inspecting their own information.

For two classical distributions, the classical mutual information is given by,

$$I(X; Y) = H(X) + H(Y) - H(X, Y). \quad (7.4)$$

It is a measure of the correlation between the events X and Y . The quantity in Eq. (7.4) is important because it upper bounds the amount of information Alice and Bob can reliably communicate. For a discrete random variable with d possible outcomes, the maximum mutual information between X and Y is $I(X, Y) = \log_2 d$ bits.

Another quantity of interest is the *conditional entropy*, which describes the entropy of y should X is known. The entropy of Y conditioned on X is given by

$$H(X|Y) = H(X, Y) - H(Y) \quad (7.5)$$

$$= - \sum_{x,y} p(x, y) \log_2 \frac{p(x, y)}{p(x)}. \quad (7.6)$$

Classically the conditional entropy is always positive, but as we will see later, this is not the case if quantum states are involved.

The relationship between the quantities mentioned above is graphically summarised in Figure 3.

B. Quantum correlation measures

The von Neuman entropy (?) for quantum density operators, $S(\hat{\rho})$, is defined analogously, replacing probabilities with density operator eigenvalues,

$$\begin{aligned} S(\hat{\rho}) &= - \sum_x \lambda_x \log_2(\lambda_x) \\ &= -\text{tr}(\hat{\rho} \log \hat{\rho}), \end{aligned} \quad (7.7)$$

where $\{\lambda\}$ is the eigenvalue spectrum of $\hat{\rho}$. This modification is logically justified, as the eigenvalues can be interpreted directly as a purely classical probability distribution of orthogonal states when the density operator is transformed into a basis with no coherences between basis states (i.e. a diagonal basis or spectral decomposition). In that case, the Shannon and von Neuman entropies essentially have identical physical interpretations.

There are a few properties of the Von Neumann entropy which are often useful ():

1. For pure states $\hat{\rho} = |\psi\rangle\langle\psi|$, $S(\hat{\rho}) = 0$.

2. The Von Neumann entropy is unchanged by a unitary change of basis,

$$S(U\hat{\rho}U^\dagger) = S(\hat{\rho}) \quad (7.8)$$

3. For a state living on a Hilbert space of dimension d , $S(\hat{\rho}) \leq \log_2 d$, the equality is achieved when the quantum state is maximally mixed.

4. For a given bipartite system AB

$$S(\hat{\rho}_{AB}) \leq S(\hat{\rho}_A) + S(\hat{\rho}_B). \quad (7.9)$$

Equality holds if $\rho_{AB} = \rho_A \otimes \rho_B$.

Analogously, the quantum mutual information for bipartite state $\hat{\rho}_{A,B}$ is defined as,

$$I(A; B)_{\hat{\rho}} = S(\hat{\rho}_A) + S(\hat{\rho}_B) - S(\hat{\rho}_{A,B}), \quad (7.10)$$

using the von Neuman entropy.

The mutual information between two quantum states is invariant under local unitary transformations,

$$I(A; B)_{\hat{\rho}} = I(\hat{U}_A \hat{\rho}_A \hat{U}_A^\dagger; \hat{U}_B \hat{\rho}_B \hat{U}_B^\dagger), \quad (7.11)$$

since the eigenvalue spectrum of a density operator is invariant under unitary transformations. Therefore, the mutual information represents the maximum amount of information Bob can learn about Alice's state under *any* local operations.

Now, in sharp contrast with the classical case, the conditional entropy for quantum states

$$H(A|B)_{\hat{\rho}} = S(\hat{\rho}_{AB}) - S(\hat{\rho}_B) \quad (7.12)$$

can be negative! If ρ_{AB} is a maximally entangled state, then $S(\hat{\rho}_{AB}) = 0$, and $H(A|B) = -S(\hat{\rho}_{AB}) < 0$. Intuitively, this can be interpreted as the fact that bipartite maximally entangled states can be more strongly correlated than classically possible, and that knowing the state of A reveals an "unnatural" amount of information about the state of B .

A quantum process cannot increase the mutual information between two parties. This yields the *data processing inequality* that, for a sequence of channels $X \rightarrow Y \rightarrow Z$,

$$\begin{aligned} I(X : Z) &\leq I(X : Y), \\ I(X : Z) &\leq I(Y : Z), \end{aligned} \quad (7.13)$$

with equality if and only if the channel not specified in the identity on the right hand side ($Y \rightarrow Z$ or $X \rightarrow Y$ respectively) is unitary, i.e one of the links in the chain perfectly preserves information content. The progression is shown in Fig. 4.

The mutual information is defined as being between a particular known pair of states. Of course, in a quantum channel we have the flexibility to manipulate the input state and the measurement at the output. From this, we

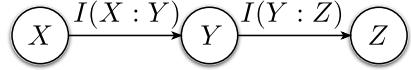


Figure 4: A sequence of events $X \rightarrow Y \rightarrow Z$. The data processing inequality states that the mutual information from beginning to end is upper-bounded by the mutual information between neighbouring stages, as per Eq. (7.13).

can then define the *classical information of the channel*, $I_c(\hat{\rho}_{A,B})$ – the maximum of the mutual information, optimised over *all* possible measurement settings. Suppose $\hat{\rho}_{A,B}$ is shared between Alice and Bob, from which they would like to extract maximal correlations (i.e joint information). They measure using local measurement bases $\{\Lambda_A^x\}$ and $\{\Lambda_B^x\}$, yielding random variables A and B . The quantum mutual information of the channel \mathcal{E} is defined as,

$$I_c(\mathcal{E}) = \max_{\rho, \{\Lambda_A^x\}, \{\Lambda_B^x\}} I(A; B)_\rho. \quad (7.14)$$

Eq. (7.14) quantifies the maximum amount of correlation that can be established between input and output states that can be achieved across the channel. This dictates the physical upper bound on the achievable bitrate (classical communication) or bandwidth of the channel.

All of the correlations discussed above concern classical information. We now move on to discuss quantum correlations. Whilst classical correlation measures have a clear operational meaning, quantum channel capacities are less so – these quantities can be interpreted as how well the channel preserves the quantum state.

The conditional entropy for quantum states can be negative is significant, in fact, it has its own name and notation, the *coherent information*

$$I(A\rangle B) = H(B)_{\hat{\rho}} - H(A, B)_{\hat{\rho}}, \quad (7.15)$$

For a maximally entangled state, the coherent information is equal to one. It measures to what extent we know less about a subsystem compared to the whole.

— *Alis volat prorilis.*



Part III

Classical networks

“Mobile phones should be left to the kids. They’re the only ones who can operate them.” — Anthony Hincks.

To set the context for our upcoming treatment of quantum networks, we begin by discussing *classical* networks, and some of the key protocols behind their operation.

VIII. CLASSICAL NETWORKING PROTOCOLS

HERE have been numerous approaches employed in the past for sharing communications links between multiple users. This includes:

- Channel-switching: an entire communications channel is designated for exclusive use by a given user.
- Packet-switching: data is divided into packets, which are routed independently by the network, being reconstructed by the recipient once all packets have been received.
- Time- or frequency-multiplexing: each user is designated a particular frequency spectrum or series of time-slots exclusively for their use.
- Code Division Multiple Access (CDMA): all users can broadcast over a channel simultaneously, and the construction of the coding technique enables demultiplexing of the distinct signals, despite their interference with one another.
- Ethernet: all users are free to broadcast over a shared channel at will, and *collision detection* identifies when packets interfere, after which they are discarded and rebroadcast following a random waiting period, repeating until success.

Nowadays packet-switched networks have become the norm in most digital networks, as they facilitate far greater efficiency in the use of network bandwidth, and are more easily scaled to greater numbers of users in a dynamic and ad hoc manner. It is foreseeable the same trend will continue with quantum technologies, especially given their initial high cost, where maximising network utility is paramount.

In this work we will focus on packet-switched networks when we later introduce our quantum networking protocols. However, with sufficient flexibility in the design of our upcoming quantum protocols, packet-switched networks can easily be made to effectively implement channel-switched, or time-/frequency-multiplexed communication.

A. TCP/IP

The present-day internet is built on top of a protocol stack comprising primarily the Internet Protocol (IP), User Datagram Protocol (UDP), and Transmission Control Protocol (TCP). Most commonly, these are simply referred to as TCP/IP. These define a stack of different layers of abstraction for communicating data packets between nodes in a network, determining their routing, and enforcing any quality of service requirements.

1. Internet Protocol

IP is the standard protocol employed in the internet for P2P communication of data packets. It is a low-level protocol that encapsulates digital data into packets containing a header field, which specifies routing information, most notably the IP addresses of the source and destination. IP does not enforce any kind of quality control, which is instead delegated to higher-level protocols like TCP (Sec. [VIII.A.3](#)), a higher-level of abstraction built on top of IP (Sec. [VIII.A.3](#)).

Multiple packets with the same source and destination needn’t follow the same route – the routing is determined dynamically in realtime by routers, based on network characteristics such as load or latency. Thus, packets belonging to the same underlying data may arrive out of order, or some may go missing altogether. IP does not address these issues, instead engaging in only ‘best-effort’ delivery.

In IP there is no central authority with knowledge of the state of the entire network, which tells routers in the network how to best route packets. Thus, IP must be complemented with up-to-date routing tables, held by routers/nodes in the network, which make routing decisions on a per-packet basis. This is achieved using gateway protocols, discussed next.

2. User Datagram Protocol

The UDP is a simple protocol built on top of IP, based on a ‘send-and-forget’ principle for sending data packets. That is, there is no quality of service guarantee, and no notifications are provided to the sender as to whether packets successfully reached their destination. However, a checksum (hash) forms a part of the packet headers to enable error detection by the recipient. The lack of quality control bypasses the associated latency, making it particularly useful in time-critical applications, where the late arrival of a packet is useless and therefore needn’t be retransmitted.

UDP is connectionless, meaning that no designated connection is established between hosts. Instead data is simply transmitted and then forgotten about. The receiver

may not even be operational on the network, in which case the packets are lost without notice.

Key examples for the use of UDP are realtime audio and video transmission. If a packet associated with a frame in a video link is delayed and arrives several frames late, it is useless, since it is associated strictly with a previous frame in the video that has already passed. Quality control, in the form of contacting the sender to request a retransmission, would therefore achieve nothing. This applies similarly to live audio streaming, such as voice over IP (VoIP), where the late arrival of a packet cannot possibly be correctly inserted into the audio playback and might as well be discarded.

Therefore, UDP prioritises latency over reliability, and is best suited to time-critical applications where quality of service is not relevant.

3. Transmission Control Protocol

TCP differs from UDP in that it intrinsically supports quality control. The protocol is able to determine whether a packet successfully reached its destination, and if not, retransmit it as often as necessary to guarantee packet delivery. A checksum is also included in packet headers to enable error detection. This quality control has made TCP the dominant protocol employed in the present-day internet, where, in most scenarios, we wish to guarantee that data has been correctly delivered – if an email is missing random segments of its text, users will become irate very quickly!

TCP is connection-oriented, meaning that a handshaking protocol establishes a dedicated bidirectional channel between two hosts. It also enforces packet reordering, to counter out-of-order packet arrival.

However, the enforced quality control and handshaking protocols incur a network performance overhead that UDP does not, since handshaking protocols consume bandwidth. Thus, TCP should not be used instead of UDP if there are no quality of service requirements.

B. Ethernet

Ethernet is a networking protocol based on ‘broadcasting’ on a shared network. This model is particularly suited to local area networks (LANs), where all users share a single communications channel rather than dedicated P2P links, as shown in Fig. 5.

In the Ethernet protocol, every user is free to broadcast data onto the shared channel as they please – all users transmit to, and receive from a single shared channel. However, clearly sometimes packet ‘collisions’ will occur, resulting in packet corruption. To overcome this, Ethernet packets contain a checksum that can be used to verify upon arrival whether a packet has been corrupted by a

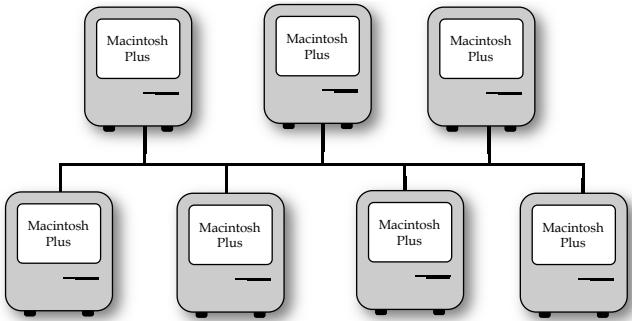


Figure 5: The topology of an Ethernet network, whereby all users share a common channel, which they can broadcast to at their leisure. If data packets collide, it is detected via the packets’ checksums, and the corrupted packets may be re-broadcast after a random ‘backoff’ waiting period, repeating this process until packet delivery is successful. Obviously the chance of collisions occurring increases with the number of connected users, thus network performance is inversely related to the number of nodes.

collision. If a collision is detected, the respective users are able to re-broadcast, following a randomly chosen waiting period (known as ‘backoff’). Collisions therefore reduce network performance, and it follows that network bandwidth decreases with the number of users competing for bandwidth¹⁴.

From this protocol, any given packet will eventually be successfully transmitted uncorrupted, collision-free, albeit with uncertain timing that grows with the number of competing users. For this reason, the Ethernet protocol is not ideal for time-critical applications requiring hard guarantees on network latency.

The beauty of this approach is that only a single channel is required for connecting all users. No dedicated P2P connections are required. As the number of users increases, the complexity of the network topology does not – requiring only the addition of a node to the existing backbone. For small LANs this is clearly reasonably functional. However, as the size of networks increases, the

¹⁴ Think of that awkward dinner table conversation, where two people start talking simultaneously (Peter & Jon). It’s immediately obvious to them both that they are interfering with one another, and if they were to just talk over one another (packet collision), no one would understand either of them. So, they both awkwardly pause, before starting to speak again. In a *really* awkward conversation, they will both start again simultaneously, after which there will be an even longer awkward pause before recommencing. Eventually, this self-regulating system will resolve itself probabilistically, with a sole victor controlling the airwaves, commanding the attention of the listeners. Provided that all dinner guests adhere to social etiquette and backoff appropriately, with repeated conversations, all guests will statistically receive an equitable share of attention, albeit with some wastage of conversation time owing to the periods of silence. Clearly, the proportion of the time wasted due to collisions will scale up with the number of guests, limiting the protocol to not-too-large tables (or very quiet guests).

rate at which packet collisions occur increases, resulting in a reduction in network bandwidth. Thus, the Ethernet protocol is ideally suited to small LANs, but is clearly not viable at a global level, where network competition is astronomical and the overhead from backoff would reduce network performance to a standstill, wasting most of the bandwidth.

Another elegant feature of the Ethernet protocol is that bandwidth allocation is self-regulating, with bandwidth fairly and equitably allocated between users, not prioritising any user over another. This applies even in completely ad hoc networks, with users joining and leaving the network willy nilly. Provided all users are correctly and honestly implementing the BROADCAST AND BACKOFF protocol, network bandwidth is allocated evenly amongst users, and no mediating, overriding central authority is needed to oversee network resource allocation. This allows Ethernet networks to be truly ‘plug-and-play’.

C. Gateway protocols & routing tables

In the absence of a central mediating authority, routing decisions must be made by individual nodes in the network, upon receipt of packets. For routers to make sensible routing decisions, they must have some idea of the overall structure and state of the network. This is achieved using gateway protocols, which communicate information about the state of the network on a nearest-neighbour basis. There are various gateway protocols in use, with the Exterior Gateway Protocol (EGP) and Border Gateway Protocol (BGP) being very common.

We let every node in the network have a routing table, initially empty, that will ultimately be populated with information on how to best route incoming packets further along the route to their destination.

To mitigate the need for a central authority, nodes engage in only nearest neighbour communication, sharing their routing tables with one another, to query about the distance metrics (Sec. IX.B) associated with routes to different destinations. This communication is taking place regularly, and as nodes’ routing tables become populated, updating in real-time, they will (hopefully) reach a steady-state. From these tables, single-source shortest path algorithms (Sec. XI.D) can be applied by nodes to construct a complete picture of costs to every point in the network. Such a nearest neighbour algorithm is effectively a distributed breadth-first-search algorithm (Sec. XI.A).

D. Network hierarchies

The disadvantage of Ethernet’s BROADCAST AND BACKOFF principle is that packets are often wasted – whenever a collision occurs. Because there is no mediating central authority, packet collisions are a certainty in

a heavily-utilised shared network, each time resulting in packet loss, and an associated reduction in usable network bandwidth.

To the other extreme, we could have dedicated P2P channels between every pair of users. Then there would be guaranteed no packet collisions, and therefore maximum bandwidth efficiency, but the network would be extremely costly, and plug-and-play extremely challenging.

To address this dilemma, the topology and subdivision of networks need to be carefully designed. If we consider a large organisation, for example, potentially networking thousands of desktop PCs, the bandwidth wastage associated with packet collisions could grind the entire network to a halt, were all thousands of PCs to be communicating large amounts of data simultaneously. However, if a hierarchy of subnetworks could be implemented, rather than a single monolithic network, efficiency could be improved drastically.

Suppose our hypothetical organisation had several different departments, and users had a tendency to communicate primarily with other users in the same department. By defining distinct departmental subnets, which individually implement Ethernet, but interconnect with one another using an alternate routing framework, we can easily see that many unnecessary packet collisions may be entirely avoided. That is, why broadcast data to users who we know don’t want it? A simple example of this is shown in Fig. 6.

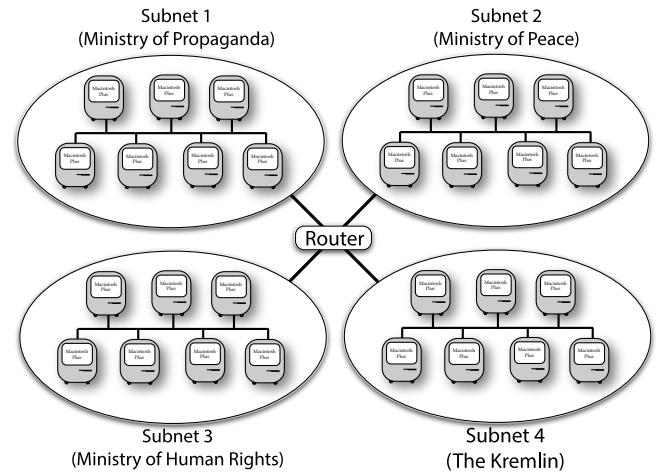


Figure 6: Simple example of a network with two levels in its hierarchy. At the lowest level are 4 different subnets belonging to different departments within an organisation, each of which implements Ethernet networking. Above this, the subnets connect together in a star network via a central router. By subdividing the network hierarchy as such, if most traffic emanating within a subnet stays within that same subnet, collisions between packets on different subnets are avoided, thereby improving the efficiency of the subnets’ Ethernet implementations.

Extending upon this simple intuitive example, enormous amounts of research and development have been

invested into the design of network hierarchies, and how to optimise their efficiency. A pressing consideration in the design of network protocol stacks is therefore to accommodate for multiple routing protocols, and enabling their inter-compatibility.

— *Te futuo et caballum tuum.*

IX. MATHEMATICAL REPRESENTATION OF NETWORKS

We now turn our attention to defining a mathematical construction for the representation of (quantum and/or classical) networks, that 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.

A. Graph-theoretic representation

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

$$G = (V, E), \quad (9.1)$$

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 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* and *attributes*, quantifying the relative performance of different links.

A key feature of the global internet is redundancy. In a packet-switched environment, 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, since 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, as they do not accommodate redundancy.

B. Cost vector analysis

The edge weights in G represent the *costs* (\vec{c}) and *attributes* (\vec{a}) associated with using that link. In general these needn't be single numbers, but would rather be sets or data-structures, representing different types of costs and attributes of links, of which there may be many. These could include, for example, latency, bandwidth, dollar cost, and quality measures.

The distinction between costs and attributes, is that costs may be expressed in terms of units which may be interpreted as distances metrics in a Euclidean sense, obeying the following requirements:

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

- *Identity operations: If a channel performs nothing, its associated cost is zero, $c(\mathbb{1}) = 0$.*
- *Triangle inequality: $c(v_1 \rightarrow v_2 \rightarrow v_3) \leq c(v_1 \rightarrow v_2) + c(v_2 \rightarrow v_3)$, across all paths $v_1 \rightarrow v_2 \rightarrow v_3$. 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 is not somehow better off than if it hadn't traversed that channel at all.*

Attributes, on the other hand do not have a distance interpretation, and may have arbitrary structure. A detailed discussion on the relationship between costs and attributes is presented in Sec. XVI.D.

The reason we demand costs have a distance interpretation is so that graph-theoretic pathfinding algorithms (Sec. XI.B) are applicable, allowing us to build upon the vast pre-existing understanding of graph theory. Ideally we would like equality in costs' triangle inequality, which yields an exact cost. But often this isn't 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 and attributes that realistic quantum networks will be subject to is presented in Sec. XV.

A *route* between two nodes, Alice (A) and Bob (B), of the network, G , is an acyclic subgraph connecting those nodes, $R_{A \rightarrow B} \subseteq G$. In general ad hoc networks there will

typically be multiple paths between two nodes $A \rightarrow B$. For a particular cost metric, the cost of an entire route is simply the sum of the costs of each of the constituent links,

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

$$c(R_{A \rightarrow B}) = \sum_{i=1}^{|R_{A \rightarrow B}| - 1} c(v_i \rightarrow v_{i+1}), \quad (9.2)$$

where v_i is the i th node in the route $R_{A \rightarrow B}$.

Fig. 7 illustrates a simple example network with all of its available routes, $R_{A \rightarrow B} \subseteq G$. Fig. 8 illustrates the optimal path for $A \rightarrow B$ based on edge weights.

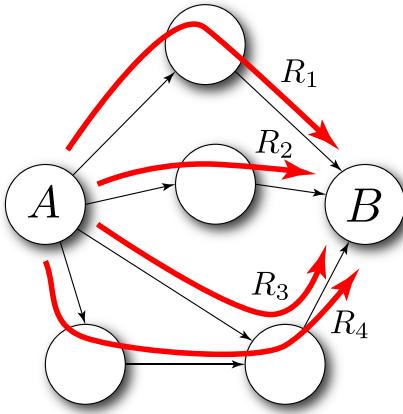


Figure 7: Example of a simple network with multiple routes $A \rightarrow B$. Note that R_3 and R_4 are competing with one another for use of the last link, which the routing strategy, \mathcal{S} , will need to resolve if multiple simultaneous transmissions are taking place.

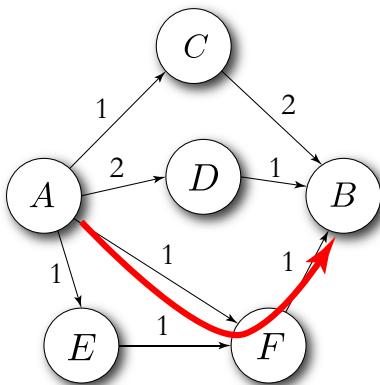


Figure 8: The same network graph from Fig. 7, with links weighted by some arbitrary cost metric. Applying a shortest-path algorithm yields the optimal route between Alice and Bob to be $A \rightarrow F \rightarrow B$, which incurs a net cost of $c = 2$, as opposed to all other routes, which incur a net cost of $c = 3$.

In a given network, it is unlikely that only a single cost metric or attribute will be of interest when determining optimal routings. There may be a tradeoff 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, while a carrier pigeon is slow but cheap (and prohibited by PETA). What is the best tradeoff between the two?

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

Def 3 (Net routing cost) *The net cost of a route $A \rightarrow B$ is given by,*

$$c_{\text{net}}(R) = f_{\text{cost}}(\vec{c}(R), \vec{a}(R)), \quad (9.3)$$

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 and attributes.

Note that the net routing cost needn't be a metric, as 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, since it already captures all these accumulations.

Eq. (9.3) 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,

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

$$c_{\text{total}}(\vec{R}_{A \rightarrow B}) = \sum_{r \in \vec{R}_{A \rightarrow B}} c_{\text{net}}(r), \quad (9.4)$$

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

C. Flow networks

On a shared network with many users utilising the network simultaneously, it may be the case that the preferred goal for the network is to maximise *flow* – the total amount of information that can be transmitted per unit time, i.e the net utilisation of the network's resources, summed over all users. In this case we can build on the existing theory of *flow networks* (?), which characterise the load of links within the network.

A flow network is easily obtained from the network graph by associating a ‘capacity’ attribute with each link and defining the graph weighted by the capacities as the flow network, preserving the underlying structure of the network graph.

When a route within the graph is utilised, we decrement the capacities of each link in that route, generating the so-called *residual network* (?), which will now take the place of the original network in subsequent calculations. This process effectively tallies the links' utilisation, and when the tally hits zero, the link can no longer be used for any new routes. This forms a basic building block for more complex flow network algorithms.

There are many variations on flow networks. The simplest case is of a single user transmitting multiple packets simultaneously to a recipient. Depending on link capacities, different packets may need to follow different routes through the network, if network performance is to be maximised. Alternately, it may not be possible to send the desired number of packets simultaneously if the network capacity saturates.

The more complex (and realistic) scenario is of multiple users each transmitting from distinct starting nodes to distinct recipient nodes across a shared network. This is known as a *multi-commodity flow network* (?), and is likely to be the dominant class of networks in real-world networking applications.

D. 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 vectors of costs and attributes within the network associated with the utilisation of that route,

Def 5 (Routing strategies) A routing strategy is defined by,

$$\begin{aligned} \mathcal{S}(i, j, \vec{c}, \vec{a}) &\rightarrow \{k, \vec{c}', \vec{a}'\}, \\ i, j &\in V, \\ k &\in \{R_{v_i \rightarrow v_j}\}, \end{aligned} \quad (9.5)$$

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} and \vec{a} are vectors of associated costs and attributes.

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. Alternately, 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, which may also be time-dependent,

Def 6 (Time-dependent routing strategies)

A time-dependent strategy, \mathcal{S}_t , updates the network graph, G_t , at each time-step t ,

$$G_{t+1} = \mathcal{S}_t(G_t). \quad (9.6)$$

\mathcal{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.

A detailed exposition of routing strategies is provided in Sec. XIX.

E. Strategy optimisation

Clearly the goal when choosing routing strategies is to minimise the total cost, Eq. (9.3). That is, solving the optimisation problem,

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

$$\begin{aligned} c_{\min} &= \min_{\mathcal{S}}(c_{\text{total}}), \\ \mathcal{S}_{\text{opt}} &= \operatorname{argmin}_{\mathcal{S}}(c_{\text{total}}). \end{aligned} \quad (9.7)$$

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 (discussed in Sec. XI.B) 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**.

1. 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 (Sec. [XI.A](#)) or gateway protocols (Sec. [VIII.C](#)).

On the other hand, for **CENTRAL** algorithms, either network exploration could be employed, or alternately 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, since 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.

2. 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. [\(9.4\)](#), 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, since it requires knowledge of the entire state of the network, which is collectively optimised.

Since **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 $\text{LOCAL} \subset \text{GLOBAL}$. 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), discussed in Sec. [XIB](#).

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 LANs, 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, since 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, is highly questionable.

We will not aim to comprehensively characterise the computational complexity of **GLOBAL** strategies. However, in Sec. [XIX](#) we will present some elementary analyses of several toy models for realistic strategies. Some such strategies are efficient although not optimal, but 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 won’t even begin to delve into here.

F. Message- vs. packet-level routing

In Eq. [\(9.6\)](#) we defined the action of a strategy, \mathcal{S} , on a network, G . However, we were intentionally ambiguous in our introduction of the time-dependence, given by t . This is to allow us to consider changes at one of two different time-scales: the packet level, or the message level. The *message* is the entire data stream transmitted from Alice to Bob, whereas the *packet* is a small block of data taken from the message, where each packet may be

independently routed.

When defining the action of strategies, we could do so at either of these time-scales. We could choose routes in their entirety, from start to finish, at the beginning of the message transmission, under the assumption that the costs in the network will be constant over that duration and no one will misbehave. We refer to such strategies as *message-level strategies*. Alternately, and perhaps more realistically in many scenarios, the costs and attributes of a network could be highly dynamic and readily change within the transmission time-window. In that case, we will employ *packet-level strategies*, which reevaluate the strategy independently for each packet and for each of their hops between nodes.

In our future discussions on routing strategies, context will make it clear when we are referring to packet- or message-level strategies.

— *Nil sine magno labore.*

X. NETWORK TOPOLOGIES

As 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,

$$G = (V, E), \quad (10.1)$$

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 QTCP 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 in our QTCP protocol (Sec. XI) 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, since they strongly impact important features such as construction cost of the network infrastructure, routing cost vector analysis (Sec. XV), likelihood of successful routing, and transmission time.

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

A. Point-to-point

The most trivial network topology, which also acts as the elementary primitive from which our other topologies will be constructed is a simple dedicated point-to-point (P2P) connection between two parties, where the sender and recipient of a packet reside on neighbouring nodes.

Such P2P connections may be reserved exclusively for the two connected neighbouring nodes. In this instance, the packets' ROUTING QUEUES trivially specify just the recipient. Alternately, the P2P link may be an intermediate step between more distant sender/recipient pairs.

In the case whereby the P2P connection is reserved exclusively for a particular sender/recipient pair, the link has the property that there is no competition between multiple users sharing the channel, and the QTCP stack needn't concern itself with dynamic routing strategies¹⁵. This significantly simplifies network scheduling algorithms (Sec. XIX), and a FIRST-COME FIRST-SERVED (i.e chronologically ordered FIFO queue) strategy may be employed. Furthermore, packet collisions cannot occur, thereby improving network efficiency.

In the case whereby the P2P connection is not reserved for exclusive use between a single sender/recipient pair, but shared between different competing routes in the network, the importance of network routing strategies manifests itself. Now competition for access to the channel will reduce network efficiency, scaling inversely against the number of network participants, and the priorities and costs of packets must be tallied for the purpose of implementing routing strategies.

B. Linear

A linear graph topology, shown in Fig. 9, has very simple properties. The number of edges simply scales as,

$$|E| = |V| - 1, \quad (10.2)$$

and the graph diameter is simply the number of vertices,

$$d = |V|. \quad (10.3)$$

There are limited routing considerations for such a topology since there is always exactly one route between two

¹⁵ Assuming the P2P channel has sufficient capacity to meet demand and exhibits better cost metrics than other potential redundant, indirect routes.

Topology	Vertices ($ V $)	Edges ($ E $)	Diameter (d)
Point-to-point	2	1	1
Linear	$ V $	$O(V)$	$ V $
Complete	$ V $	$O(V ^2)$	1
Lattice	mn	$O(mn)$	$O(m + n)$
Tree	$ V $	$O(V)$	$O(\log V)$
Percolation	$p_{\text{vertex}} \cdot V $	$p_{\text{edge}} \cdot E $	variable
Random	$p_{\text{vertex}} \cdot V $	$p_{\text{edge}} \cdot E $	variable
Scale-free	$ V $	$O(V ^2)$	$O(\log \log V)$

Table I: Summary of the mathematical characteristics of different network topologies.

points, although buffering issues may still arise under congestion.



Figure 9: A simple linear graph topology with $|V| = 10$ vertices.

Because there is no path redundancy, linear graphs are vulnerable to node failures, since the deletion of a single node makes disconnects the network.

C. Complete

The complete graph, denoted $K_{|V|}$, is a $|V|$ -vertex graph where every vertex has an undirected link to every other. From a networking point of view, this can be regarded as the extremity of exclusive-use P2P networking, whereby every node has a direct link with every other. Thus, any sender can directly communicate with any receiver, via a dedicated direct channel, with no need to utilise any indirect routes. This topology has the favourable property that although any node can communicate with any other, by exclusively utilising direct P2P links we achieve several benefits:

- Packet collisions can be mitigated entirely, thereby maximising network efficiency.
- Competition for the use of links can be eliminated, minimising congestion and the need for buffering (i.e quantum memory).
- Network costs can typically be minimised, as every route only traverses a single link, and there will be no accumulation of costs.
- The network has maximal route redundancy, making it the most tolerant against link failures¹⁶.

- A trivial FIRST-COME FIRST-SERVED routing strategy can be employed, eliminating the need for any dynamic or computationally complex strategies.
- If the network allows indirect routes to be established, the maximal redundancy of the topology also maximises the ability for routing strategies to engage in load-balancing across routes.
- In the special case of a symmetric complete graph, whereby all edge weights are approximately equal, the shortest path between any two nodes is trivially the P2P link between them, and no complex scheduling algorithms are required.

However, these highly desirable benefits come at the expense of requiring the most elaborate and expensive network, with maximal interconnectedness.

This type of topology could arise in, for example, international-scale networks, where links of very high bandwidth (and value) between nations or continents need to be maximally utilised, which would be undermined by sparse, shared network topologies. Additionally, in this instance route redundancy will be highly valued, as the isolation of one continent from another would be catastrophic to the functioning of the global network.

Fig. 10 illustrates the K_{15} graph. The number of edges scales as,

$$|E| = O(n^2). \quad (10.4)$$

Clearly route-finding is trivial, since there is always a direct P2P link from sender to receiver, with no possibility of collisions with other packets, requiring $O(1)$ search time (assuming all users are communicating only via their direct links with one another, which may not strictly be the case when costs are factored into strategies).

D. Lattice

A lattice graph is simply an $n \times m$ lattice of vertices (of any geometry, e.g squares), connecting each vertex to its immediate geometric neighbours. The number of edges scales obviously as,

$$|E| = O(mn). \quad (10.5)$$

¹⁶ To disconnect a given node v from the network, all $|v|$ links emanating from it must be broken, otherwise redundant routes to the remainder of the network will exist.

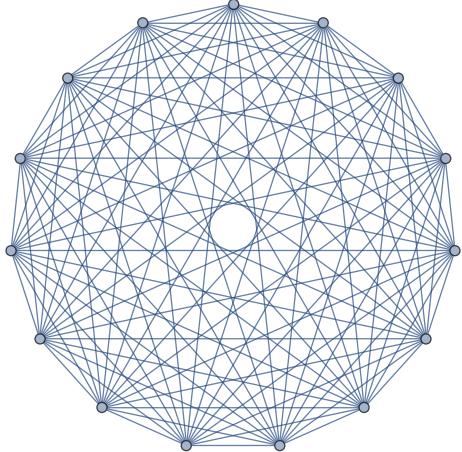


Figure 10: The 15-vertex complete graph, K_{15} . Every vertex has an edge to every other, with a total of 105 edges.

This type of graph is useful when link costs are measured in terms of Euclidean distances, and nodes have nearest neighbour links, as per Fig. 11.

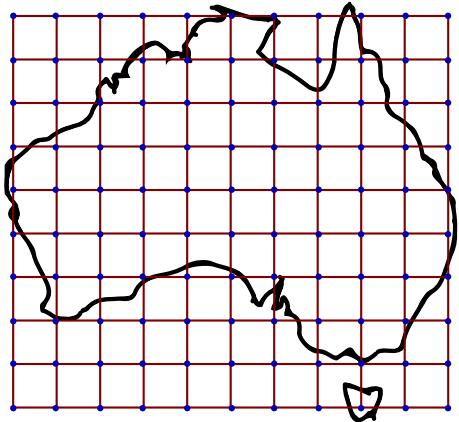


Figure 11: A 10×11 square lattice graph, and how it might represent a network topology with geographically associated costs. Notice that Hobart has no internet connection (why even include Tasmania at all?).

A slightly distorted lattice graph, in which vertices have been dragged around geometrically to match, for example, cities within a country, closely resembles the topology of the network. Similarly, if the nodes represent houses in the street layout of a highly regular city like Manhattan, a lattice may be a good approximation.

In the case of a balanced lattice, in which all edges are of equal weight, the cost of a route is the sum of the number of steps in the vertical and horizontal directions, also known as the Manhattan or L_1 distance,

$$L_1 = |x_{\text{start}} - x_{\text{finish}}| + |y_{\text{start}} - y_{\text{finish}}|. \quad (10.6)$$

In this case, route finding is simplified, since *all* routes, which strictly traverse in one direction vertically and one direction horizontally, are optimal and of equal distance.

Thus, the diameter (maximum number of hops between any two points) on the network is,

$$d = O(m + n). \quad (10.7)$$

E. Tree

A tree is a graph containing no cycles, only *branches*. There are many uses for tree graphs, but one property is of particular convenience in many applications: because the graph is acyclic, there is always exactly one path from any vertex to any other. This mitigates the need for shortest-path algorithms designed for general graphs, and simplifies route-finding algorithms (to be discussed in Sec. XI.A). However, this brings with it the drawback that the topology is most vulnerable to link failures, since the removal of any link from the tree will separate it into a multipartite graph, making communication between the disjoint subgraphs (which are also trees) impossible, as there are no redundant routes. In a sense, tree graphs can be considered the polar opposites of complete graphs.

Trees are specified entirely by *branching parameters* (b_i) – the number of child nodes emanating from a given node, i . In general, branching parameters may be distinct for each node, although often trees with symmetries in their branching structures are considered, such as the balanced trees discussed in Sec. X.E.1. A node terminates a branch if its branching parameter is zero (i.e. it has no children).

The *depth* (d) of a tree is the maximum number of steps from the root node to a terminating node with no children. The depth scales between $d = O(|V|)$, for the trivial linear tree ($b_i = 1$), and $d = O(\log |V|)$ for non-trivial branching parameters ($b_i \neq 1$).

The worst-case number of edges that must be traversed to reach any vertex from any other is,

$$O(\log |V|), \quad (10.8)$$

known as the *diameter* of the graph, which implies that accumulated cost metrics scale similarly. Trees are the most frugal graphs in their number of edges, which are fixed at,

$$|E| = |V| - 1, \quad (10.9)$$

irrespective of the branching parameters, since because the graph is strictly acyclic, every addition of an edge requires the addition of exactly a single vertex. This makes tree graphs the cheapest to construct in terms of physical resource usage.

1. Balanced tree

A balanced tree is a tree with a regular, self-similar structure, in which every node at a given depth is the parent of the same number of sub-nodes, all separated by

the same edge weights. That is, the network has a hierarchical structure, subdividing into identically structured subnetworks. Such a network is characterised by just two parameters – the branching parameter, b , and the depth, d . Some examples of balanced trees with different b and d are shown in Fig. 12.

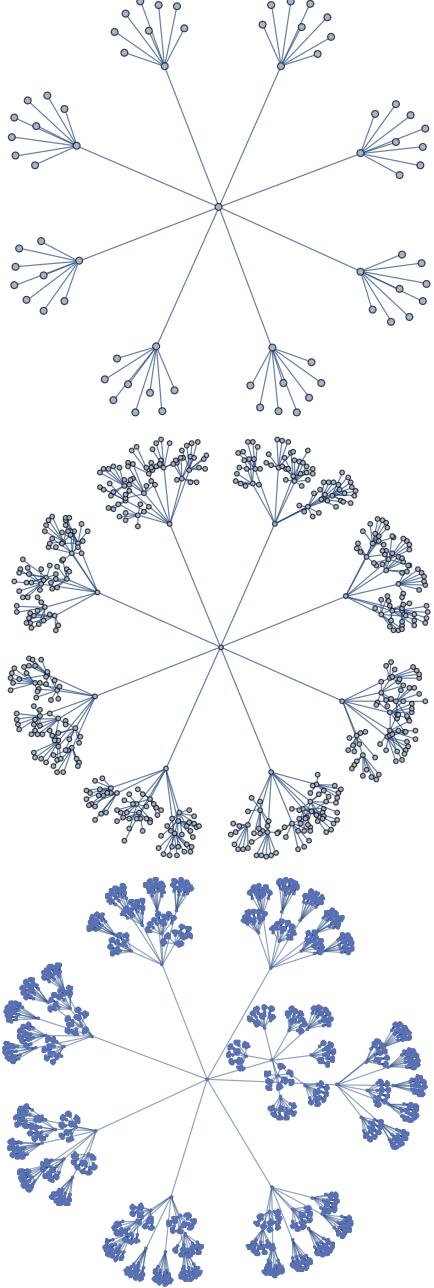


Figure 12: Balanced tree graphs with branching factor $b = 8$, and depths $d = 3, 4, 5$. Despite having no redundant paths, the hierarchical structure of balanced trees somewhat resembles that of real-world networks, which are typically decomposed into a pyramid scheme of progressively smaller subnetworks.

This type of structure is (approximately) natural in many realistic scenarios. Consider for example a network

containing a hierarchy of clusters of nodes representing a LAN, followed by a neighbouring internet router, followed by a city-wide router, followed by a country-wide router. In such a case, this type of general structure is typical (although more realistically one might expect the branching parameter to vary with depth).

A special case is when $d = 1$, which we refer to as a *star* graph. This might arise naturally when a series of subnets are connected together via a central router (e.g. Fig. 6), with no further hierarchy in the network.

2. Random tree

While balanced trees accurately capture the hierarchical nature of realistic networks, they are somewhat contrived in their perfect symmetry. The subnetworks in a given network are not likely to actually all be identical. Random trees are perhaps more realistic, in that their tree structure captures the hierarchical nature of real-world networks, and also their highly ad hoc nature.

To construct a random tree we simply randomly choose a branching parameter, according to some arbitrary distribution, for every node. When a node has $b_i = 0$, it terminates the lineage. Some examples of random trees are shown in Fig. 13.

3. Minimum spanning tree

A *spanning tree* S , of a graph G , is a tree subgraph $S \subset G$, containing every vertex of G . The *weight* of a spanning tree is the sum of all its constituent edge weights. Thus, the *minimum spanning tree* (MST) is a spanning tree that minimises net weight. An example is shown in Fig. 14. See Sec. XI.E for a discussion on MST algorithms.

The calculation of MSTs is most likely to come into consideration when actually performing the initial construction of networks, where we wish to connect all nodes in the network, but using the most frugal possible physical resources. MSTs serve this purpose, and since they are trees, inherit all the same properties of tree networks.

In general, the MST of a graph is not unique, and there may be an arbitrarily large number of completely differently structured MSTs all with the same minimum weight.

F. Percolation

A variation on any graph is to instead have a randomised implementation of it, whereby each of the possible edges or vertices occur with some probability, p_{edge} or p_{vertex} , otherwise deleted. These are referred to as *edge percolation* and *site percolation* graphs respectively.

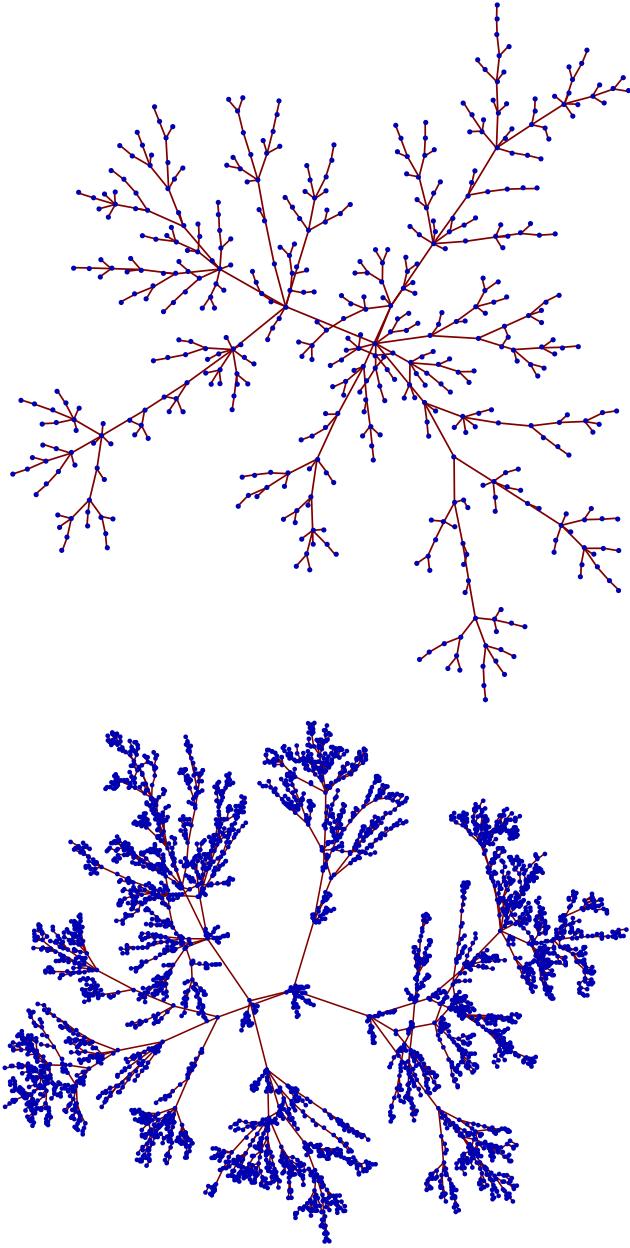


Figure 13: Random trees with different randomised branching parameters (higher b at the bottom). When a node has zero branches, it terminates the branch. This type of graph topology qualitatively captures the hierarchical, yet ad hoc qualities of many real-world networks, and may act as a useful test model for simulations.

For any given graph, its associated percolation graph has average vertex and edge counts,

$$\begin{aligned} |E|_{\text{av}} &= p_{\text{edge}} \cdot |E|, \\ |V|_{\text{av}} &= p_{\text{vertex}} \cdot |V|. \end{aligned} \quad (10.10)$$

Adjusting $p_{\text{edge/vertex}}$ allows us to tune between the desired graph G (when $p_{\text{edge/vertex}} = 1$) and the completely disconnected graph (when $p_{\text{edge/vertex}} = 0$).

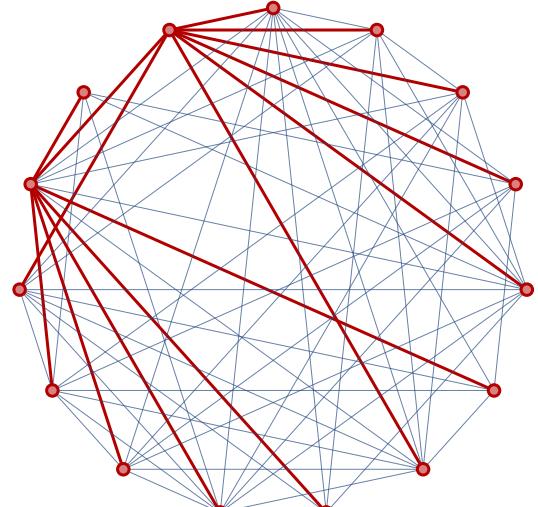


Figure 14: A random graph (blue) with an MST highlighted (red).

This model is very useful in real-world applications, allowing unreliable channels/nodes to be incorporated into our network model. The analysis of such percolation networks is invaluable for understanding the robustness of such networks to channel and node failures.

Note that percolation graphs might be disjoint with sufficient defects, in which case the respective network becomes unreliable. Specifically, with sufficiently low $p_{\text{edge/vertex}}$, ‘islands’ may form in the network topology – small segregated networks, which are unable to interface with the remainder of the network.

For asymptotically large percolation graphs, *percolation theory* (?) provides thresholds for $p_{\text{edge/vertex}}$ such that routes across the network exist in asymptotic limits (?).

Fig. 15 illustrates several square lattice graphs with different percolation probabilities, and how the larger network segregates into smaller disconnected islands as failure rates increase.

G. Random

We refer to a random graph as being one in which edges between each pair of vertices occur with some probability p_{edge} . No vertices are removed from the network, although some may have order $|v| = 0$, i.e $p_{\text{vertex}} = 1$. This can be thought of as the edge percolation graph of the complete graph $K_{|V|}$.

The average number of edges in such a network scales as,

$$|E|_{\text{av}} = p_{\text{edge}} \cdot O(|V|^2). \quad (10.11)$$

Some examples are shown in Fig. 16.

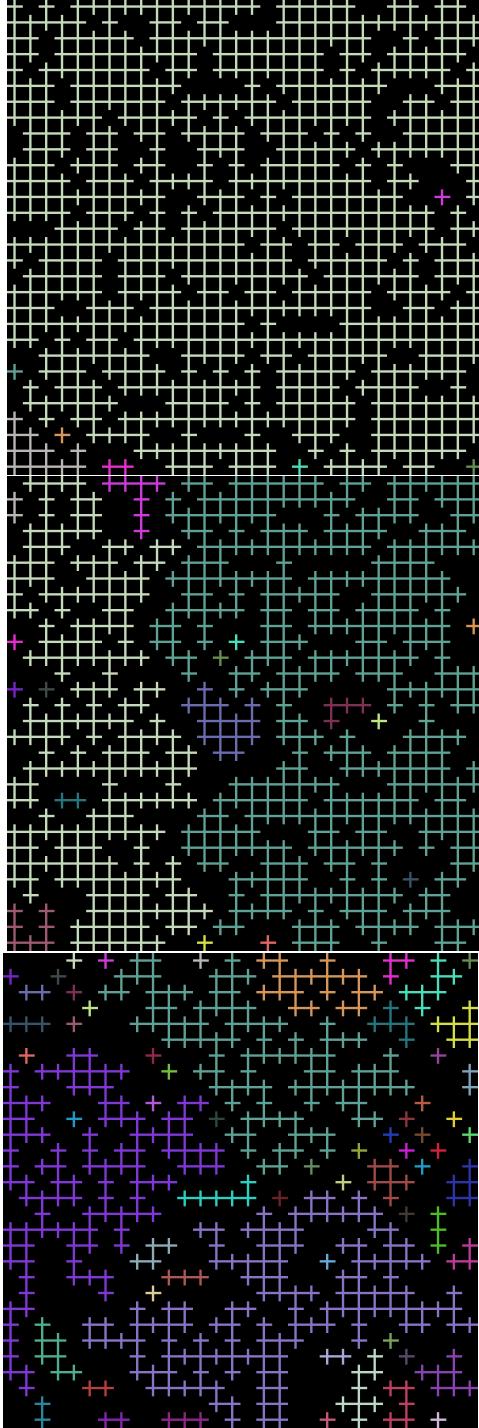


Figure 15: A square lattice graph subject to different percolation rates (node defects). As the failure rate increases (top to bottom), the larger network segregates into a multipartite graph of smaller disjoint islands (denoted by colour).

H. Hybrid

Real networks are highly unlikely to fit the exact form factor of any of the classes of graphs presented above. Rather, a truly global internet is inevitably going to

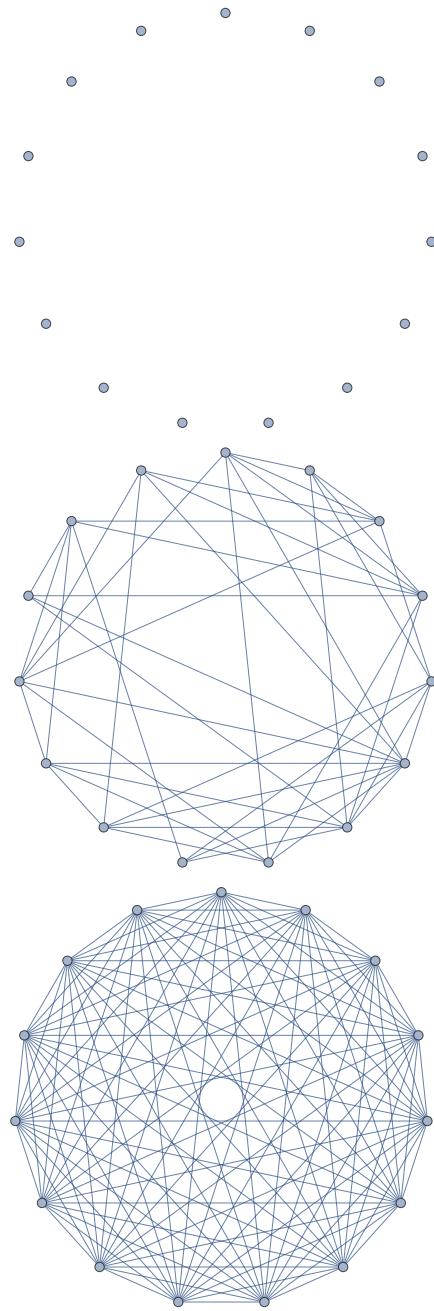


Figure 16: The 15-vertex random graph. This is the same as K_{15} in Fig. 10, but where edges are present with probabilities $p_{\text{edge}} = 0, 0.5, 1$ (top to bottom).

comprise many subnetworks, each structured completely independently of one another, with little consistency or large-scale planning between them. Who thinks about the broader structure of the global internet when setting up their office network?

For example, at the global scale, it is entirely plausible that the internet might take on a random tree-like structure. But when we get down to a lower level, the tree structure vanishes and is replaced by all manner of different network topologies, run and maintained by different

organisations in their own distinct ways.

Furthermore, the real-world internet is not simply a hierarchy of different types of well-known graph structures. Rather, it takes the form of ‘glued’ graphs, whereby networks running over different mediums, or via different operators, each exhibit their own independent graph topologies, meeting at interconnect points that join the different networks. Typically this yields redundancy in the routes between different nodes, ushering in the need for combinatorial optimisation techniques when allocating network resources.

This hybrid network topology is the norm today in our classical internet, and it is entirely foreseeable that a similar trend will emerge in the future quantum internet as quantum technologies become more mainstream, their networking less well structured, and competing, redundant links are in place.

I. Scale-free networks

Scale-free networks are not defined as obeying a specific topological structure, but rather as following a particular statistical distribution in the connectedness of their nodes. Specifically, the probability distribution function for the order of vertices (degree distribution) roughly follows a Pareto distribution or power law,

$$P(k) \sim k^{-\gamma}, \quad (10.12)$$

where $P(k)$ is the probability of a vertex having order k (up to normalisation), and $\gamma > 1$. Most commonly $2 \leq \gamma \leq 3$. Fig. 17 illustrates this scaling behaviour for different power coefficients.

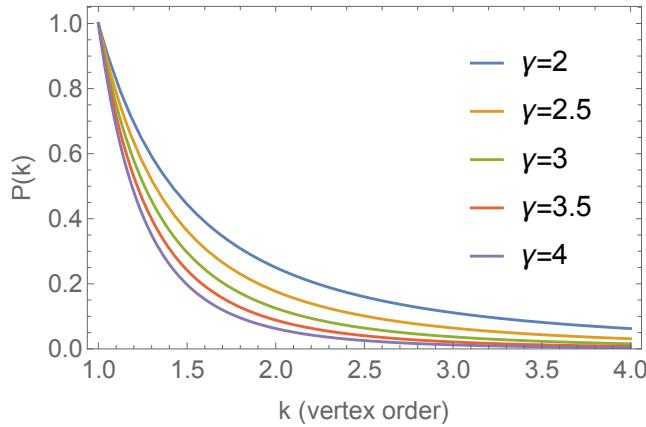


Figure 17: Examples of the power law, characteristic of the vertex order distribution in scale-free networks.

This distribution is observed empirically in many real-world networks and sociological structures, and as such is more an observation about the typical behaviour of naturally occurring and evolving human-made networks than an explicit definition for their construction. However,

owing to this particular statistical behaviour, and the underlying causations for their Pareto distribution, much has been researched and is known about the properties of scale-free networks.

Scale-free networks arise naturally in systems exhibiting *preferential attachment*, i.e when a new node is added to the system it preferentially attaches to nodes that are already more highly connected. This yields a so-called *fitness model*.

According to the Bianconi-Barabási fitness model (??), let $\eta_i > 0$ be the *fitness factor* of node i , which follow a distribution $\rho(\eta)$, a characteristic of the network. Then the fitness parameters are defined to be normalised such that,

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \quad (10.13)$$

Upon adding a new node of degree m to the network, the temporal dynamics will satisfy,

$$\frac{\partial P(k_i)}{\partial t} = m\Pi_i. \quad (10.14)$$

The probability distribution can then be shown to have solution,

$$P(k) \approx \int \rho(\eta) \frac{C}{\eta} \left(\frac{m}{k} \right)^{\frac{C}{\eta+1}} d\eta, \quad (10.15)$$

where,

$$C = \int \frac{\rho(\eta) \cdot \eta}{1 - \beta(\eta)} d\eta, \\ \beta(\eta) = \frac{\eta}{C}, \quad (10.16)$$

which is a linear combination of power law relationships, as required for the definition for scale-free.

Intuitively, why would we expect computer networks (classical or quantum) to be scale-free? To answer this, we simply must establish whether the preferential attachment property will hold. In computer networks there are several reasons why we might expect this to be the case:

- **Distance:** connecting to more highly-connected nodes reduces (on average) the number of channels data packets must traverse to reach their destination, making them ‘cheaper’ in terms of their cost vector analysis (Sec. XV).
- **Availability:** larger nodes are more likely to have unused network sockets available for use by new nodes, and they are likely to be more readily accessible. For example, one is more likely to be able to successfully sign up for an internet connection with a major national ISP than a small, local upstart player.

- Economies of scale: the dollar cost per connection of a larger node is likely to be less than for a smaller one, owing to economies of scale. For example, the cost per FLOP of a large-scale supercomputer is far less than for a desktop PC, and Google's data-centres experience lower cost-per-bandwidth on their internet connections than home-users connecting via their ISPs.

One notable characteristic of scale-free networks is their hierarchical structure, with a small number of very highly-connected ‘hubs’ at the top of the food chain, which quickly connect onto smaller hubs, and so on down the food chain with decreasing connectivity, as shown in Fig. 18.

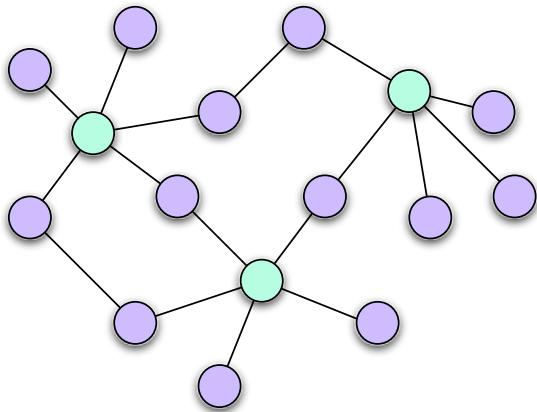


Figure 18: Example of a scale-free network. The graph exhibits a highly hierarchical structure, with a small number of ‘hub’ nodes that are highly connected (green, with vertex order 5), that connect downward to a larger number of less connected nodes (purple, with vertex order 1-2).

A feature of scale-free networks of especial interest in the context of computer networks is their robustness against node failure. Suppose we constructed a percolated (Sec. X.F) instance of a typical scale-free network. Such a network is highly robust against *random* node/edge deletions compared to many other graph constructions, in the sense that a relatively large number of failures must occur to disconnect the graph. This makes the scale-free network characteristic a particularly attractive one from the perspective of the failure-tolerance of a network. It should be noted that, on the other hand, a scale-free network is highly vulnerable to *targeted* node/edge deletions, that specifically target the highly-connected nodes. A targeted attack against major hubs could disconnect the network with relatively few successful attacks. This brings with it important geostrategic considerations when constructing network infrastructure.

Scale-free networks typically exhibit extremely small diameter (average distance between nodes), scaling as (?),

$$d = O(\log \log |V|). \quad (10.17)$$

That is, they exhibit (exponentially) smaller diameter than tree graphs (which already exhibit only logarithmic depth). Thus, expanding the network (in a manner consistent with the model) by adding a moderate number of new nodes effectively leaves graph diameter unchanged – the graph diameter is virtually a constant under modest evolution.

J. The internet web-graph

Of course, all the topological structures described until now are in-principle constructs. Of most relevance is the *internet web-graph*, the graph of the actual internet (or some other real-world network).

Fig. 19 illustrates some example web-graphs, constructed from subsets of data from the actual internet. The combination of random, densely and sparsely connected, and tree structures, and its clear hierarchy are all evident. This encourages our intuition of the different types of structures present in realistic networks.

The internet web-graph has been observed to be a scale-free network (Sec. X.I), observing its power law distribution in node connectivity, as per Eq. (10.12).

K. Network robustness

A key feature of any network topology is its robustness against node or channel failures. This is important from the perspective of naturally occurring hardware faults, and also from a geostrategic perspective, where adversaries may be launching attacks against the network. In general, there are two main contributing factors to network robustness:

- Redundancy: the number of redundant paths between two points in a graph stipulates how many backups there are to finding a route to a destination in the advent of one route failing.
- Diameter: the chance of a data packet encountering a faulty node/channel increases with the number of hops required to reach its destination. Graphs with smaller diameter are hence less vulnerable.

The extreme case of network robustness is the complete graph, K_n , which has P2P links between every pair of nodes. Therefore, if a single channel fails, there are *always* alternate paths taking us between nodes. On the opposing extreme are tree graphs, which contain no redundancy whatsoever, and just a single failure will disconnect the network, making certain routes impossible. Scale-free networks sit in the intermediate zone, but are relatively robust against the failure of random nodes/links, but are vulnerable to conspiratorial failures, which target the elite,

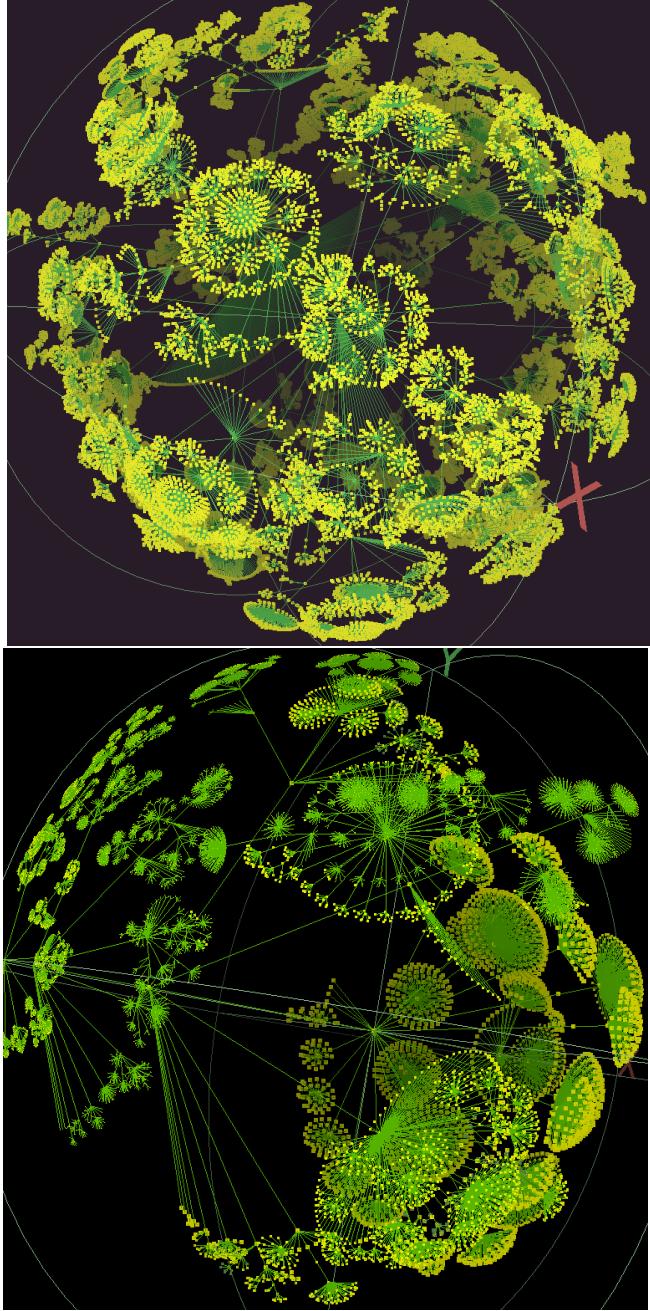


Figure 19: Examples of real-world web-graphs of the internet, capturing their high-level random tree-like structure. Graphics attributed to the Center for Applied Internet Data Analysis (CAIDA), <http://www.caida.org>.

highly connected hub-nodes¹⁷.

Fig. 20 illustrates some examples of the robustness of these two extreme cases to link and node failure.

— *Vitam regit fortuna, non sapientia.*

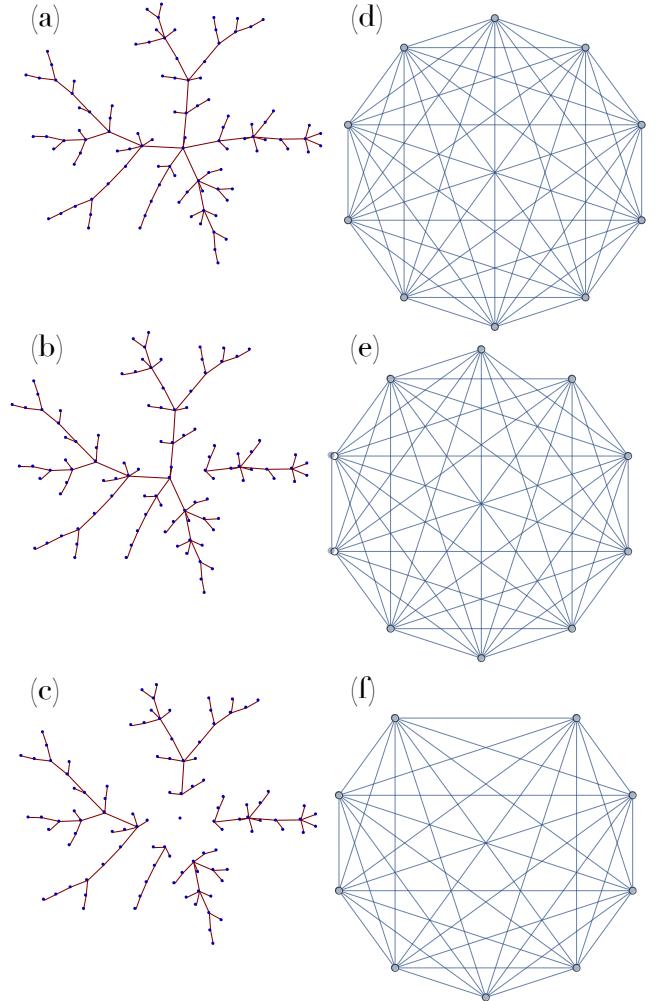


Figure 20: Robustness of network topologies to node and link deletion. Examples of a tree graph (a) and a complete graph, K_n , with $n = 10$ (d). (b,e) The same graphs subject to a single link failure. The failure disconnects the tree graph into a bipartite graph (b), whereas the complete graph's connectivity is unhindered as alternate routes exist between all nodes (e). A single node failure disconnects the tree graph into a $|v|$ -partite graph (c), where $|v|$ is the order of the vertex at which failure occurs. The complete graph, on the other hand, is simply reduced to a K_{n-1} graph, with no loss of connectivity (f). Thus, tree graphs are the most vulnerable network topologies to node/link failures, whereas complete graphs are the most robust.

XI. NETWORK ALGORITHMS

HAVING introduced some of the more relevant graph structures, we now introduce some of the key graph-theoretic algorithms of direct relevance to networking theory (?). In graph theory, many fundamental problems are believed to be computationally hard to solve, often NP-complete. However, there are several important graph algorithms that are (very) classically efficient to solve, and which are of great utility to us as network architects.

We will focus heavily on combinatorial optimisation

¹⁷ The 1%.

techniques, where the goal is to allocate network resources so as to optimise some cost metric. This includes both single- and multi-user algorithms, the latter being the far more relevant ones in the context of shared networks like the internet.

In Tab. II we summarise the upcoming discussion on important network algorithms, and their associated complexities.

A. Network exploration & pathfinding

Here the goal is to systematically explore every vertex in an unknown graph exactly once, so as to reconstruct the entire network graph, or to find a target node with unknown location (which can obviously be achieved if the former can be). The two main approaches are *breadth-first-search* (BFS) and *depth-first-search* (DFS) algorithms. In both cases we begin at a starting (root) node, from which we wish to explore the entire graph by only following edges to nearest neighbours one at a time.

In BFS we proceed from the root node to visit every one of its neighbours. Having done so, and created a list of those neighbours, we proceed onto the neighbours of the neighbours, and so on, until every vertex in the graph has been visited, or the target node found.

In DFS, on the other hand, we begin by following a single arbitrary path until we reach a dead-end, at which point we backtrack until we reach a branch leading to a vertex we hadn't previously visited.

Examples of these two algorithms are shown in Fig. 21.

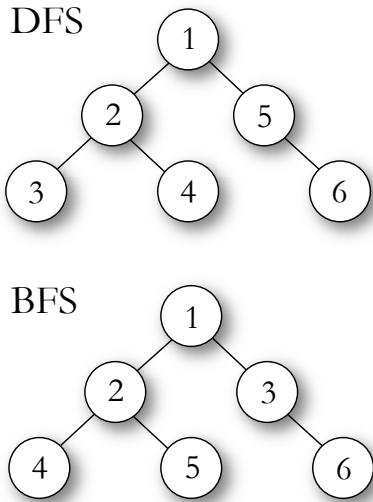


Figure 21: Comparison of the order in which vertices are explored, using the breadth-first-search (BFS) and depth-first-search (DFS) algorithms, where vertex 1 is the root vertex.

Both BFS and DFS guarantee visiting every vertex in a connected graph, and do so using only nearest neighbour transitions. Such algorithms are therefore very useful for network discovery.

The BFS algorithm is particularly applicable to pathfinding in ad hoc networks. Consider the situation where there is no central authority with full knowledge of the network, overseeing network operation. Rather, everyone needs to figure things out for themselves by only interrogating their neighbours, to whom they have direct connections. This directly leads to a BFS algorithm, where a node speaks to each of its neighbours in turn, who subsequently do the same thing, yielding a recursive algorithm. This can be naturally parallelised, as each node can be interrogating its neighbours independently, thereby implementing a distributed BFS algorithm. Note that, when searching for a target node, while the BFS algorithm obviously finds the target using the smallest number of hops (i.e a lowest-order route), it needn't necessarily find the route with the lowest cost (which is distinct from the number of hops in general). Shortest-path algorithms require *a priori* knowledge of the full network graph, discussed in Sec. XI.B.

Both BFS and DFS exhibit runtime ,

$$O(|V| + |E|), \quad (11.1)$$

where $|V|$ and $|E|$ are the number of vertices and edges respectively. Thus, these graph exploration algorithms reside in the complexity class **P**, and are classically efficient.

B. Shortest-path

In graph theory, the shortest-path problem is that of finding a subgraph of a given graph G , connecting two vertices, $A \rightarrow B \subset G$, such that the sum of its edge weights is minimised. In the context of our application to route-finding, this amounts to finding a route that minimises cost.

The first proven shortest-path algorithm was invented by and named after Dijkstra (?), which requires runtime,

$$O(|V|^2), \quad (11.2)$$

which has since been improved to,

$$O(|E| + |V| \log |V|), \quad (11.3)$$

by (?) using min-priority queues implemented via Fibonacci heaps. Thus, the shortest-path algorithm resides in **P** – one of the relatively few, and highly valuable optimisation problems that is classically efficient. Subsequently, a number of improvements and variations on Dijkstra's algorithm have been proposed, most notably the A^* algorithm (?), which has found widespread modern use, using a heuristic approach to improve performance over Dijkstra. An alternate implementation of Dijkstra using

Algorithm	Description	Complexity class	Scaling
Breadth-first-search	Explore all vertices in a graph	P	$O(V + E)$
Depth-first-search	(same as above)	P	$O(V + E)$
Shortest-path (Dijkstra)	Find the shortest route between two nodes in a directed graph	P	$O(E + V \log V)$
Shortest-path (A^*)	(same as above)	P	(varying)
Single-source shortest path	Find the shortest paths from a given node to <i>all</i> other nodes	P	$O(V \cdot E)$
Minimum spanning tree	Find a spanning tree of a graph that minimises the total of the edge weights	P	$O(E \log V)$
Minimum-cost flow (Orlin)	Minimise total costs in a network with a specified amount of flow	P	$O(V \log V (E + V \log V))$
Maximum flow (Ford-Fulkerson)	Maximise flow in a network, regardless of costs	P	$O(E \cdot c_{\max})$
Multi-commodity flow	Same as maximum flow, but generalised to arbitrary numbers of users	NP -complete (exactly), P (approximation using heuristics)	$O(\exp(V))$
Vehicle routing problem	Generalises the shortest-path algorithm to multiple users, with distinct sources and destinations	NP -complete	$O(\exp(V))$
Vehicle rescheduling problem	Same as above but with dynamically changing costs	NP -complete	$O(\exp(V))$

Table II: Summary of some important network algorithms and their complexities. The **NP**-complete algorithms are not believed to have efficient classical algorithms, and their exact scaling is not well understood.

Formally, let \vec{R} be the set of all routes $A \rightarrow B$. Then,

$$c_{\text{opt}} = \min_{r \in \vec{R}} \left(\sum_{i \in r} c_{\text{net}}(i) \right), \quad (11.4)$$

where $i \in r$ denotes the i th edge in the route r . Intuitively, the (in general) exponential number of possible paths through a graph might lead one to believe the above optimisation problem is a computationally inefficient one (such as **NP**-complete, or worse). However, perhaps surprisingly, Dijkstra's algorithm cleverly manages to reduce this to a polynomial-time problem. A sketch of the algorithm is provided in Alg. I, which needn't be understood by the reader desperate to read further.

Fig. 8 illustrates a directed, edge-weighted graph. A shortest-path algorithm applied between vertices A and B would return $R_{\text{shortest}} = A \rightarrow F \rightarrow B$ as the minimum-cost route.

When introducing network graphs earlier, we insisted upon all costs being associated with edges rather than vertices, and presented a trivial means by which to convert vertex costs to edge costs in Fig. 28. This adundance arose because the presently described shortest-path algorithms operate purely in terms of edge weights, not vertex weights. But the mapping we presented from the latter to the former obviates this issue.

This is the motivating factor behind representing network graphs purely in terms of edge weights (Sec. XII.C), thereby enabling compatibility with shortest-path algorithms.

For the purposes of the QTCP protocol, we are interested in the case of directed graphs (recall that in terms of cost metrics, undirected graphs can be converted to

directed graphs by replacing undirected edges with a pair of identical edges in opposite directions).

Shortest-path techniques find widespread application in many areas. Computer networks are an obvious candidate, since networks are inherently graph-theoretic by nature.

To implement the shortest-path algorithms discussed above, the party performing the calculation requires knowledge of the full network graph. In an ad hoc network, where users might be added to or removed from the network arbitrarily, this isn't necessarily the case.

One solution is for a central authority to be responsible for maintaining a ledger of all network participants and their connectivity, which users are required to notify upon joining or leaving the network. The central authority may then apply shortest-path calculations, which may be queried by users. However, a disruption in connection to the central authority, or failure of nodes to notify the central authority upon joining or leaving the network, introduces a point of failure into the operation of the protocol.

Another approach, which does not require a reliable central authority, is for users to implement network exploration algorithms each time they wish to perform a shortest-path calculation. This facilitates truly ad hoc networking, but incurs the cost overhead associated with nodes frequently implementing network exploration. However, network exploration is a purely classical algorithm, which may run entirely over the classical network, and therefore incurs no cost in quantum resources.

With this approach, a new node can join the network, without having to know anything about the topology of the network. Similarly, upon leaving the network, it needn't notify anyone, since a future interrogation by a

```

function DijkstraShortestPath(G, A, B):
    1. currentNode = A
    2. tentativeDistances[A] = 0
    3. tentativeDistances[others] =  $\infty$ 
    4. nodesVisited[A] = True
    5. nodesVisited[others] = False
    6. loopStart:
        7. neighbours = currentNode.neighbourhood
        8. nodesVisited[neighbours] = True
        9. for(n in neighbours) {
            10. newTentativeDist =
                min(tentativeDistances[currentNode]
                    + edgeWeight[currentNode, n],
                    tentativeDistances[n])
            11. nodesVisited[currentNode] = True
        }
        12. }
        13. if(nodesVisited[B] = True) {
            14. return(tentativeDistances[B])
            15.  $\square$ 
        }
        16. }
        17. currentNode =
            tentativeDistances[unvisitedNodes].
            nodeWithSmallest()
        18. goto(loopStart)
        19.  $\square$ 

```

Algorithm I: Dijkstra's original shortest-path algorithm for finding the lowest weight path through a graph, G , between two vertices, A (source) and B (destination). The algorithm has $O(|E|+|V|\log |V|)$ runtime (in **P**). An example application of this algorithm is shown in Fig. 22.

neighbour will be detected as a non-existent node. The BFS is therefore highly suited to ad hoc operation. In fact, present-day internet gateway protocols (Sec. VIII.C) essentially implement a distributed version of BFS.

C. Constrained shortest-path

In some scenarios we may wish to find a shortest-path through a graph, subject to some constraints. In general, the addition of constraints can make such algorithms far more computationally complex, undermining the efficiency of Dijkstra's algorithm. However, in some circumstances such constraints can easily be incorporated, without undermining the performance of the algorithm.

In particular, if there are constraints on the relationships between nearest-neighbours in the graph, this can be incorporated by pre-processing the graph via deletions of edges that violate the constraints. Then the usual shortest-path algorithm may be applied to the reduced graph. The rather trivial algorithm is shown in Alg. II,

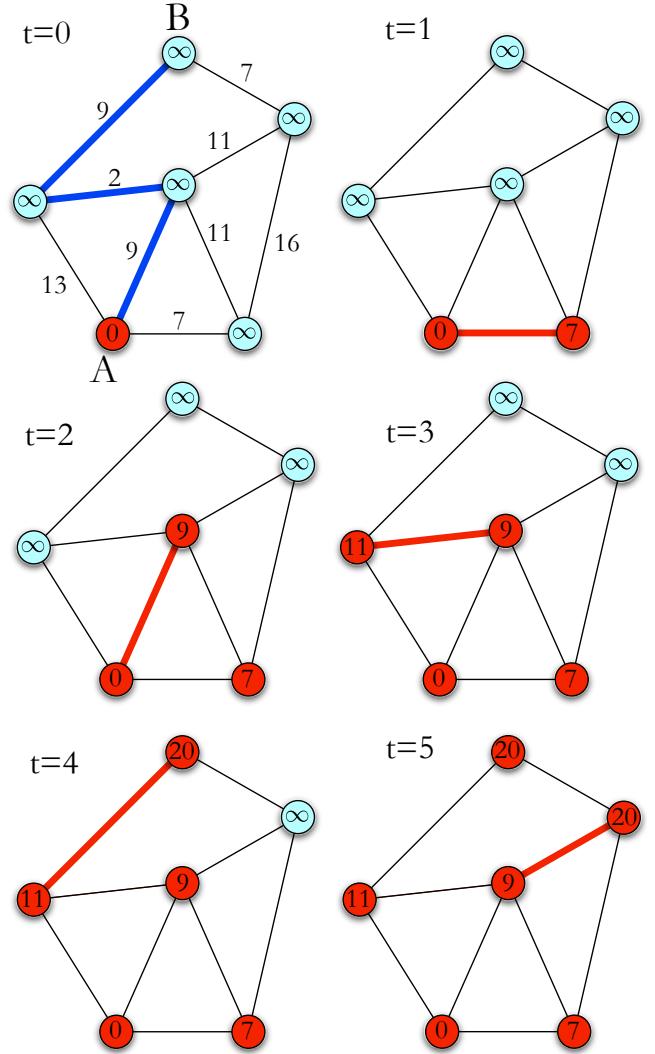


Figure 22: Example of Dijkstra's shortest-path algorithm for finding the shortest route from A to B (blue route), yielding a shortest route of distance 20 using 5 algorithmic steps. Initially ($t = 0$) all nodes are designated a tentative distance of ∞ and marked as unvisited (blue vertices), except the starting node A , which is designated distance 0 and visited (red vertices). At each step we choose the lowest cost path to a previously unexplored node (red edges). This updates the tentative distance of the newly-visited neighbour, which is now marked as visited. This is iterated until all nodes have been visited. Each node is visited exactly once.

with a simple example shown in Fig. 23.

A notable example of where this technique finds applicability is in entanglement swapping networks, discussed later in Secs. XXVII.E & XXIX. Consider the network shown in Fig. 109. Here the entanglement swapping network comprises nodes alternating between two different functionalities: Bell pair preparation, and entanglement swapping. Representing these as colours, this effectively introduces the edge constraint that edges between nearest neighbours of the same colour ought to be removed. Additionally, the end-points of the network must neighbour

```

function ConstrainedShortestPath( $G, A, B, C$ ):
    1. For the set of nearest-neighbour
       constraints  $C$ , generate the set of edges
        $E(C)$  that violate the constraints.
    2.  $G' = G - E(C)$ 
    3. route = DijkstraShortestPath( $G', A, B$ )
    4. return(route)
    5.  $\square$ 

```

Algorithm II: Efficient algorithm for finding a constrained shortest-path, where the constraints are in terms of nearest-neighbour relationships.

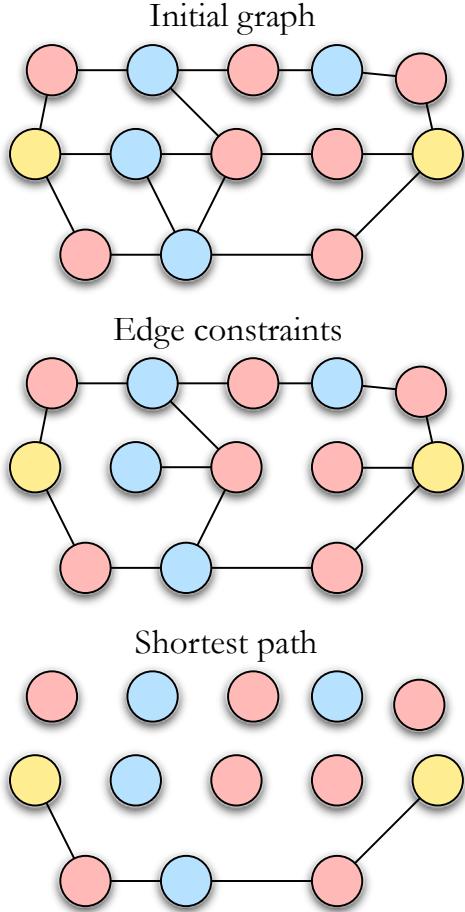


Figure 23: Example of a constrained shortest-path algorithm. Beginning with the initial graph we eliminate edges that do not satisfy imposed nearest-neighbour constraints. This is performed as a pre-processing stage. Then a usual shortest-path algorithm is applied to the reduced graph, yielding the optimal route subject to the required constraints.

Bell pair sources, not entanglement swappers. This enforces the additional constraint of removing edges between end-points with the wrong coloured neighbours. Having applies these edge deletions enforcing these constraints, the shortest-path algorithm will now find the constrained optimal route. Thus, this example of an entanglement swapping network is isomorphic to the example presented

in Fig. 23.

D. Single-source shortest-path

The shortest-path algorithm by Dijkstra presented above finds the shortest route between two specified nodes in a network. However, when employing INDIVIDUAL routing strategies, where there is no central mediation of the network, each node desires an up-to-date routing table, showing the best route to take to any other point in the network. Then, upon receiving packets with particular destinations, rather than repeatedly applying Dijkstra's algorithm, we can simply look up the destination on the node's local routing table.

Single-source shortest-path algorithms address this problem by calculating the shortest paths from the current node to *every* other node in the network topology.

The best-known algorithm for this problem is the Bellman-Ford (or Bellman-Ford-Moore) algorithm (?), which requires worst-case runtime of,

$$O(|V| \cdot |E|). \quad (11.5)$$

Clearly this is more complex than Dijkstra's algorithm for finding a particular shortest-path. But it is more efficient than using brute-force to find the shortest-path between every pair of nodes in the network via $O(|V|^2)$ repeated applications of Dijkstra.

E. Minimum spanning tree

MST algorithms find an MST¹⁸ of some arbitrary graph. Like the shortest-path problem, it has a polynomial-time, deterministic algorithm (i.e it resides in **P**). The first MST algorithm (?) required,

$$O(|E| \log |V|), \quad (11.6)$$

runtime. Numerous variations have since been proposed, with little change to the underlying scaling.

Because MST algorithms are efficient, they play a very useful role in the design of real-world network topologies, where resource minimisation is crucial.

Fig. 14 shows an example of a graph with its MST.

F. Minimum-cost flow

The *minimum-cost flow problem* (?) is that of minimising costs through a network for a specified amount of flow (i.e total bandwidth or throughput), which acts as a constraint on the problem. The definition of 'cost' in this

¹⁸ There may be multiple distinct MSTs for a given graph.

context is compatible with our earlier definition of cost metrics (Def. 1).

This problem can be efficiently solved using linear programming. Specifically, cost metrics along links in series are given by linear combinations of individual link costs. If, in addition, we let our net cost function be linear in the constituent costs then the net cost will also be linear in all the edge weights. This lends itself directly to optimisation via linear programming techniques. Algorithms for linear programming, such as the *simplex* algorithm, have polynomial-time solutions (i.e reside in **P**), and a plethora of software libraries are available for implementing them numerically.

One polynomial-time algorithm, by (?), for solving this problem does so in,

$$O(|V| \log |V|(|E| + |V| \log |V|)), \quad (11.7)$$

time.

G. Maximum flow

The *maximum flow problem* (?) is the seemingly simple goal of – as the name suggests – maximising network flow, without consideration for any of the other cost metrics or attributes associated with the network. This type of problem is relevant when brute bandwidth is the dominant goal.

This problem can be tackled using a number of techniques. In some circumstances, linear programming techniques can be employed. The best-known algorithm is the Ford-Fulkerson algorithm (?), which finds a solution in,

$$O(|E| \cdot c_{\max}), \quad (11.8)$$

runtime, where $|E|$ is the number of links in the network and c_{\max} is the maximum cost present in the network. The algorithm behaves pathologically in some conditions, which can easily be overcome in the context we present here. Using Ford-Fulkerson as a starting point, numerous other more sophisticated algorithms have been developed.

H. Multi-commodity flow

The *multi-commodity flow problem* (?) generalises the previous algorithms to be applicable to multi-user networks. The generalisation is that there may be a number of distinct senders, residing on different nodes, each transmitting to distinct recipients, residing on different nodes. This is the most realistic scenario we are likely to encounter in a real-world quantum internet, where networks will inevitably be shared by many users, residing at different nodes.

Unfortunately the computational complexity of solving this problem is much harder than the previous algorithms in general. Specifically, solving the problem exactly

is **NP**-complete in general. However, in specific circumstances it can be approached using linear programming or polynomial-time approximation schemes.

I. Vehicle routing problem

The vehicle routing problem (VRP) is a multi-user generalisation of the shortest-path problem, where the goal is to minimise total network cost (i.e the sum of all individual users' costs) when there are multiple users sharing the network, each with distinct sources and destinations.

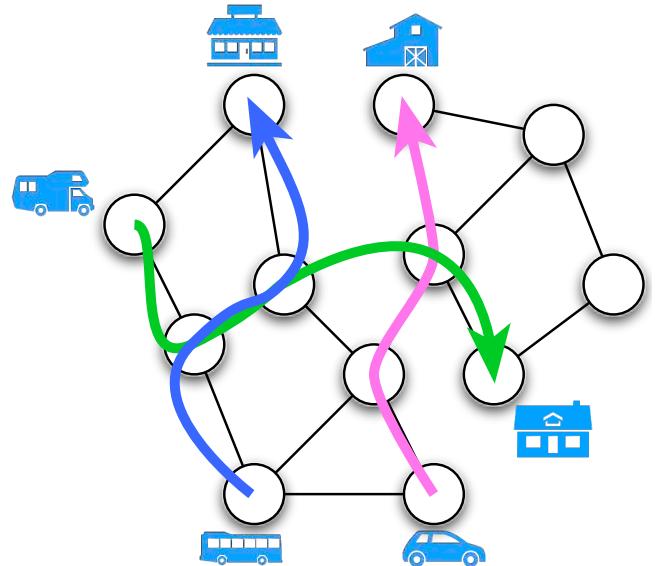


Figure 24: Example of the vehicle routing problem, the multi-user generalisation of the shortest-path problem, where the goal is to minimise the total cost across all users. Intuitively, because whenever individual shortest-paths intersect one must trial different prioritisations, the combinatorics of this grow exponentially with the number of competing users.

Unlike the polynomial-time shortest-path algorithm, exactly solving the VRP is **NP-hard** in general. However, heuristic methods can find approximate suboptimal solutions far more efficiently, and there is a multitude of software packages available for doing so.

The VRP has found widespread use in, for example, the routing of transport networks for delivery companies or public transportation networks (hence the name), and many commercial companies exist, which perform these kinds of optimisations on behalf of transport providers to enhance their efficiency.

It is obvious that this algorithm is directly applicable to multi-user communications networks, which are conceptually identical to transport networks, albeit a bit faster.

A multitude of variations on the VRP exist, accommodating for different types of constraints (or additional flexibilities) in the operation of the network.

J. Vehicle rescheduling problem

The vehicle rescheduling problem (VRSP) generalises the VRP to the case where properties of the network undergo changes dynamically within the course of transmissions over the network. To use the analogy of transport networks, this could entail, for example, a truck breaking down en route to its destination, requiring real-time rescheduling of the other vehicles.

Solving the VRSP exactly is **NP**-complete in general, but as with the VRP, heuristic methods can often be applied, which efficiently find approximate solutions.

In the context of communications over networks, the VRSP has obvious applicability – a quantum internet is likely going to be largely ad hoc in nature, with users coming and going, and many non-deterministic points of failure, requiring ongoing updating of routing decisions if resource allocation is to remain as efficient as possible.

K. Improving network algorithms using quantum computers

Given that we are directing this work at the upcoming quantum era, where quantum computing will become a reality, it is pertinent to ask whether quantum computers might improve the aforementioned network algorithms, some of which are computationally hard problems. Most notably, several of the discussed algorithms are **NP**-complete in general, a complexity class strongly believed to be exponentially complex on classical computers. Can quantum computers help us out here, and improve network resource allocation? Can quantum computers help themselves?

While it is not believed that quantum computers can efficiently solve such **NP**-complete problems, it is known that they can offer a quadratic speedup using Grover's unstructured search algorithm. Specifically, **NP**-complete problems can be treated as satisfiability problems, where we are searching for an input to a classical algorithm that yields a particular output.

To gain a quantum advantage, we treat the classical algorithm as an oracle whose input configurations form an unstructured search space. Then, Grover's algorithm can perform a search over the space of input configurations to find a satisfying solution, with quadratically enhanced runtime.

While a quadratic improvement is far short of the exponential improvement we might hope for, Grover's algorithm is known to be optimal for the unstructured search problem (?). Nonetheless, despite only yielding a quadratic improvement, a quadratic speedup may already be sufficient to significantly improve network resource allocation.



Part IV Quantum networks

"Advances in the technology of telecommunications have proved an unambiguous threat to totalitarian regimes everywhere." — Rupert Murdoch.

We have reviewed some of the key aspects of classical networks, including the real-world implementation of classical networking via the TCP/IP protocol stack, and the essential mathematical foundations for networking theory, including cost vector analysis and routing strategies.

Let us now lay the foundations and some of the key motivations and assumptions we will make in our upcoming discourse on quantum networks, and lay out some of the key differences between classical networks and future quantum ones.

Quantum networks comprise all the same ingredients as classical networks, but with some very important non-classical additions. Nodes can additionally implement quantum computations, quantum-to-classical interfaces (i.e measurements), quantum-to-quantum interfaces (i.e switching data between different physical systems), quantum memories, or any quantum process in general. Many of these are not allowed by the laws of classical physics.

The cost vectors associated with links could include measures that are uniquely quantum, such as fidelity, purity or entanglement measures, none of which are applicable to classical digital data.

As in the classical case, our goal is to find routing strategies that optimise a chosen cost measure. But in the quantum context costs will be constructed entirely

differently owing to the quantum nature of the information being communicated.

We envisage a network with a set of senders and receivers, all residing on a time-dependent network graph as before. Senders have sets of quantum states they wish to communicate. For each state they must choose appropriate strategies, such that the overall cost is optimised, for some appropriate cost measure. Compared to classical resources, equivalent quantum resources are costly and must be used efficiently and frugally. Indeed, the no-cloning theorem imposes the constraint that arbitrary unknown states cannot be replicated at all! This makes resource allocation strategies of utmost importance in the quantum world.

Routing strategies will not always guarantee that packets have immediate access to network bandwidth the moment they demand it. One needs to think about the others too! Inevitably, in shared networks there will sometimes be competition and congestion, forcing some users to wait their turn. For this reason, many quantum networks will require at least some nodes (the ones liable to competition) to have access to quantum memories, such that quantum packets can be buffered for a sufficient duration that they can wait their turn on the shared network resources for which there is high competition. The required lifetime of a quantum memory will then be related to overall network congestion. Of course, quantum memories induce unwanted quantum processes of their own, which need to be factored into cost calculations.

Given that classical networking is decades more advanced than quantum networking, and extremely cheap and reliable in comparison, we will assume that classical resources ‘come for free’, and only quantum resources are of practical interest in terms of their cost. That is, classical communication and computation is a free resource available to mediate the operation of the quantum network. We therefore envisage a *dual network* with two complementary networks operating in parallel and in tandem – the quantum network for communicating quantum data, and a topologically identical classical network operating side-by-side and synchronised with the quantum network, overseeing and mediating the quantum network.

Data packets traversing the network will comprise both quantum and classical fields, which will be separated to utilise the appropriate network, but synchronised such that they arrive at their destination as a single package of joint quantum and classical information to be at the disposal of the recipient.

The motivation for the dual network is to ensure that classical and quantum data that jointly represent packets remain synchronised and subject to the same QoS issues, such as packet collisions and network congestion.

We envisage quantum networks to extend beyond just client/server quantum computation, to include the free trade of any quantum asset. This includes state preparation, measurement, computation, randomness, entangle-

ment, and information. Much like the classical internet, by allowing quantum assets to be exchanged, we can maximise utility, improve economy of scale, and enable new models for commercialisation.

May the games begin.

— *Gladiator in arena consilium capit.*

XII. QUANTUM CHANNELS

LIKE classical channels, quantum channels are inevitably subject to errors. These errors could be an intrinsic part of the system, or induced by interaction with the external environment. The *quantum process* formalism provides an elegant mathematical description for all physically realistic error mechanisms (??). Here we review the quantum process formalism and how it applies to quantum networks. This paves the way for the quantum notion of costs and attributes.

A. Quantum processes

To quantify the operation of nodes and links within our network, we must characterise the evolution they impose upon quantum states passing through them. Quantum processes, also known as *trace-preserving, completely positive maps* (CP-maps) are able to capture all the physical processes relevant to quantum networking, such as: unitary evolution; decoherence; measurement; quantum memory; state preparation; switching; and, indeed entire quantum computations. And they are able to capture physical processes in any degree of freedom, most commonly in the qubit basis, but also, for photons, in the spatio-temporal, photon-number, phase-space, or polarisation degrees of freedom.

Quantum processes are most easily represented using *Kraus operators*, $\{\hat{K}_i\}$,

$$\mathcal{E}(\hat{\rho}) = \sum_i \hat{K}_i \hat{\rho} \hat{K}_i^\dagger, \quad (12.1)$$

where,

$$\sum_i \hat{K}_i^\dagger \hat{K}_i = \hat{\mathbb{I}}, \quad (12.2)$$

for normalisation. Here \mathcal{E} is a super-operator, denoting the action of the process on state $\hat{\rho}$. This is also referred to as the *operator-sum representation*. This representation has the elegant interpretation as the probabilistic application of each of the Kraus operators \hat{K}_i , with probability,

$$p_i = \text{tr}(\hat{K}_i \hat{\rho} \hat{K}_i^\dagger). \quad (12.3)$$

In the ideal case, the two types of evolution of interest are unitary evolution, in which case there is only one Kraus operator, $\hat{K}_1 = \hat{U}$, and projective measurement, where

there is again only one Kraus operator, $\hat{K}_1 = |m\rangle\langle m|$, for measurement outcome m .

Mathematically, quantum processes are equivalent to a state jointly undergoing unitary evolution with an external environment that is not observed (i.e traced out),

$$\mathcal{E}(\hat{\rho}_S) = \text{tr}_E(\hat{U}_{S,E}[\hat{\rho}_S \otimes |0\rangle_E\langle 0|_E]\hat{U}_{S,E}^\dagger), \quad (12.4)$$

where S denotes the primary system to which the process is applied, and E is an auxiliary environment system, as shown in Fig. 25.

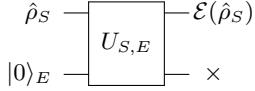


Figure 25: Model for the quantum process formalism, as a system state $\hat{\rho}_S$ undergoing joint unitary evolution with an environment state $|0\rangle_E$, which is subsequently traced out, yielding an arbitrary quantum process $\mathcal{E}(\hat{\rho}_S)$ acting on the primary system. For the most general possible class of quantum processes to be enabled, the dimension of the environment Hilbert space must grow quadratically with the dimension of the primary Hilbert space.

We will require that all our states are normalised,

$$\text{tr}(\hat{\rho}) = 1, \quad (12.5)$$

and that our processes are *trace preserving*. That is, they preserve normalisation,

$$\text{tr}[\mathcal{E}(\hat{\rho})] = 1. \quad (12.6)$$

Multiple consecutive processes may be composed using the notation,

$$\mathcal{E}_n(\dots \mathcal{E}_2(\mathcal{E}_1(\hat{\rho}))) = (\mathcal{E}_n \circ \dots \circ \mathcal{E}_2 \circ \mathcal{E}_1)(\hat{\rho}). \quad (12.7)$$

In general, processes do not commute, i.e $\mathcal{E}_1 \circ \mathcal{E}_2 \neq \mathcal{E}_2 \circ \mathcal{E}_1$. Unless unitary, quantum processes are irreversible, meaning that errors accumulate and cannot be overcome without the overhead of some form of quantum error correction (QEC) (???). The linearity of Eq. (12.1) implies that quantum processes are also linear,

$$\mathcal{E}(\hat{\rho}_1 + \hat{\rho}_2) = \mathcal{E}(\hat{\rho}_1) + \mathcal{E}(\hat{\rho}_2). \quad (12.8)$$

The only limitation faced by the quantum process formalism is that it is described over discrete-time only. To consider continuous-time evolution, *master equations* can be used. These represent the continuous-time evolution of a quantum state as a differential equation in time, combining a usual Hamiltonian term as well as decoherence terms,

$$\frac{d\hat{\rho}}{dt} = -\frac{i}{\hbar}[\hat{H}, \hat{\rho}] + \sum_j (2\hat{L}_j \hat{\rho} \hat{L}_j^\dagger - \{\hat{L}_j^\dagger \hat{L}_j, \hat{\rho}\}), \quad (12.9)$$

where \hat{H} is the Hamiltonian of the isolated system undergoing coherent evolution, and \hat{L}_j are the *Lindblat operators*, capturing the incoherent component of the dynamics (i.e environmental couplings). Here $[\cdot, \cdot]$ and $\{\cdot, \cdot\}$ are the commutator and anti-commutator respectively.

In this work we will only make use of discrete-time quantum processes, since they naturally correspond to the evolution of states between discrete points within a network – we are typically only interested in the process undergone by a state from one end of a link to another, not the continuous-time dynamics of what takes place within them.

B. Quantum process matrices

In general, the Kraus operator representation for quantum processes is not unique – there may be multiple choices of Kraus operators that implement identical physical processes. But if the representation is not unique, how do we compare different quantum processes? To address this, it is common to choose a ‘standard’ basis for representing quantum processes, such that they may be consistently and fairly compared. This requires choosing a basis which is complete for operations on the Hilbert space acted upon by the process.

For example, for a single qubit, the Pauli operators – $\hat{\sigma}_1$ (identity, $\hat{1}$), $\hat{\sigma}_2$ (bit-flip, \hat{X}), $\hat{\sigma}_3$ (bit-phase-flip, \hat{Y}), and $\hat{\sigma}_4$ (phase-flip, \hat{Z}) – are complete for single-qubit operations (\mathbb{C}_2). Therefore by decomposing our Kraus operators into linear combinations of these basis operators we have a standardised representation for single-qubit processes. Formally, for one qubit,

$$\mathcal{E}(\hat{\rho}) = \sum_{i,j=1}^4 \chi_{i,j} \hat{\sigma}_i \hat{\rho} \hat{\sigma}_j^\dagger. \quad (12.10)$$

The Hermitian matrix χ is known as the *process matrix*, from which many other metrics of interest may be directly computed (some of which are discussed in Sec. XV).

Process matrices share many algebraic properties and interpretations in common with density matrices. The diagonal elements can be regarded as the amplitudes associated with applying each of the four Pauli operators, all of which are non-negative, while the off-diagonal elements represent the coherences between them, i.e whether the operations on the diagonal are being applied probabilistically or coherently. For example, a process that simply randomly applies Pauli operators would have a diagonal process matrix in the Pauli basis. But off-diagonal elements would be indicative of applying coherent superpositions of the operators. Like density matrices, the dimensionality of process matrices grows exponentially with the number of qubits in the system being characterised, and for exactly the same conceptual reasons.

For the process to be trace preserving we require,

$$\text{tr}(\chi) = 1. \quad (12.11)$$

We will typically enforce this constraint on our processes. $\text{tr}(\chi) < 1$ implies non-determinism, i.e the process sometimes fails.

As an illustrative example of the interpretation of process matrices, in Fig. 26 we show the process matrix for the CNOT gate, represented in the Pauli basis. The CNOT operator can be expressed in the Pauli operator basis as,

$$\hat{U}_{\text{CNOT}} = \frac{1}{2} (\hat{\mathbb{I}} \otimes \hat{\mathbb{I}} + \hat{\mathbb{I}} \otimes \hat{X} + \hat{Z} \otimes \hat{\mathbb{I}} - \hat{Z} \otimes \hat{X}). \quad (12.12)$$

Then, some density operator evolved under the CNOT gate is simply $\hat{U}_{\text{CNOT}} \hat{\rho} \hat{U}_{\text{CNOT}}^\dagger$. Expanding this out, we obtain a new state comprising 16 terms, each representing the action of some combination of Pauli operators from the left and from the right. The amplitudes of these terms exactly correspond to the 16 non-zero elements of the process matrix shown in Fig. 26.

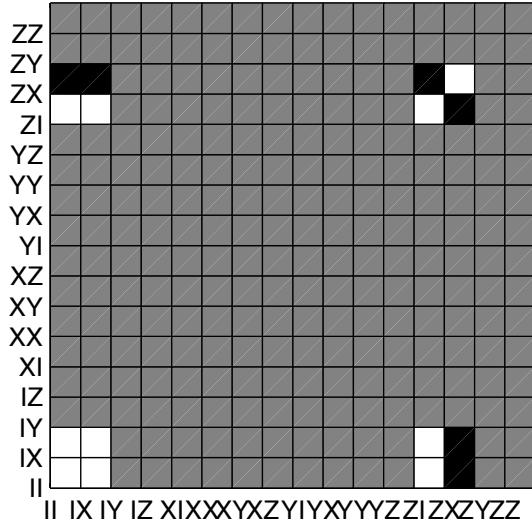


Figure 26: Process matrix for the CNOT gate, expressed in the Pauli basis. Colour coding: grey=0, white=1/4, black=-1/4.

C. Quantum processes in quantum networks

Letting v_i represent the i th node within a route R , the process associated with communication from that node to the next is $\mathcal{E}_{v_i \rightarrow v_{i+1}}$. For the same network used previously, Fig. 27 shows the quantum processes associated with the links in the network. The cumulative process associated with an entire route is therefore,

$$\mathcal{E}_R = \mathcal{E}_{v_{|R|-1} \rightarrow v_{|R|}} \circ \cdots \circ \mathcal{E}_{v_2 \rightarrow v_3} \circ \mathcal{E}_{v_1 \rightarrow v_2}, \quad (12.13)$$

where $|R|$ is the number of nodes in R , and to simplify notation, all v_i are implicitly over the route R .

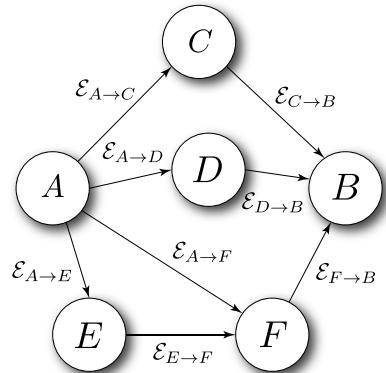


Figure 27: The network from Fig. 7, with the quantum processes associated with each link. The net process associated with a route is given by the composition of each of the processes over the length of the route. For example, the route $R_1 = A \rightarrow C \rightarrow B$ induces the process $\mathcal{E}_{R_1} = \mathcal{E}_{C \rightarrow B} \circ \mathcal{E}_{A \rightarrow C}$.

In general, both nodes and links in a quantum network may implement quantum processes. However, for the purposes of compatibility with the graph-theoretic algorithms described in Sec. XI, we will eliminate node processes by merging them into link processes, such that the processes in the network are described entirely by links. This reduction procedure is straightforward, shown in Fig. 28.

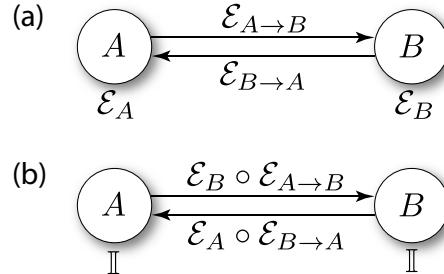


Figure 28: Removing node processes from network graphs on a trivial network with two nodes, A and B . Each node is associated with a quantum process (\mathcal{E}_A and \mathcal{E}_B). Similarly, each link is associated with a process ($\mathcal{E}_{A \rightarrow B}$ and $\mathcal{E}_{B \rightarrow A}$). (a) Representation where the node and link processes are shown explicitly. (b) The node processes are replaced with identity operations by replacing each link process with the composition of the link process and its target node process. Equivalently, the cost of each node process is added to the cost of *every* incoming link and then eliminated. The same may be applied for attributes rather than costs. This procedure requires that all links be directed. If undirected links are present, they may simply be replaced by two directed links, one in each direction, implementing identical quantum processes each way.

D. Characterising quantum states & channels

Given a link implementing some arbitrary quantum process, it is essential that it can be experimentally determined such that network performance may be characterised. For example, if an optical channel is lossy, what is the loss rate? This is crucial when attempting to choose routing strategies that optimise certain cost metrics.

Treating a link or node as an unknown black box, *quantum process tomography* (QPT) (?) is a technique that may be applied to fully characterise the quantum process it implements, reproducing its complete process matrix. QPT has become a standard procedure, demonstrated in numerous architectures, most notably in optics (?).

QPT works in general for processes in any degree of freedom, e.g the qubit degree of freedom. However, it is important to note that full QPT requires statistics across the entire basis over which measurements are defined, which typically grows exponentially with the size of the system. For example, the number of measurement bases required to perform full QPT on n qubits grows exponentially with n .

However, often full process characterisation is not necessary. Instead, knowing particular metrics of interest may suffice. Some of the more noteworthy such metrics will be discussed in Sec. XV. In this instance, much work has been done in the field of *compressed sensing* or *compressed quantum process tomography* (?), in which some process metrics of interest can be experimentally determined using far fewer physical resources (with efficient scaling!) than via a full reconstruction of the process matrix using QPT. As a most trivial example, if the loss associated with a fibre-optic channel is the metric of interest, this can be much more easily determined than by performing full QPT.

On the other hand, however, most quantum channels are designed to accommodate systems with very limited Hilbert space dimensionality per clock-cycle – e.g a fibre-optic link might transmit just one photon at a time – in which case there is no exponentiality to be terribly concerned about (QPT of a single-photon channel is trivial).

Importantly, it is often the case that the quantum process associated with a channel will remain constant over time. The efficiency of a length of fibre, for example, does not change. In this instance, characterising the channel need only be performed once in advance, without requiring ongoing dynamic updating. On the other hand, when communicating with satellites in low Earth orbit it is to be expected that the properties of links will be highly dynamic.

We will now explain QPT in the archetypal context of single-qubit channels, which logically generalises to multiple qubits, and can similarly be generalised to non-qubit systems also.

1. Quantum state tomography

The first stage in QPT is *quantum state tomography* (QST), where the goal is to reconstruct an unknown density matrix via measurements upon multiple copies of the state. QST is based upon the simple observation that the completeness relation for an arbitrary state can be expressed,

$$\hat{\rho} = \sum_i \text{tr}(\hat{E}_i \hat{\rho}) \cdot \hat{E}_i, \quad (12.14)$$

where $\{\hat{E}_i\}$ forms a complete basis for the Hilbert space of $\hat{\rho}$. For a single qubit this decomposition is most often performed in the Pauli basis,

$$\hat{\rho} = \text{tr}(\hat{\rho}) \cdot \hat{\mathbb{1}} + \text{tr}(\hat{X}\hat{\rho}) \cdot \hat{X} + \text{tr}(\hat{Y}\hat{\rho}) \cdot \hat{Y} + \text{tr}(\hat{Z}\hat{\rho}) \cdot \hat{Z}. \quad (12.15)$$

Of course, $\text{tr}(\hat{E}\hat{\rho}) = P(\hat{E}|\hat{\rho})$ is just the expectation value of the measurement operator \hat{E} when measuring $\hat{\rho}$. Thus, measuring the expectation values in each of the four Pauli bases reconstructs $\hat{\rho}$.

This generalises straightforwardly to multi-qubit systems, where we measure all combinations of tensor products of the Pauli operators, the number of which grows exponentially with the number of qubits n , as 4^n . This introduces scalability issues for systems comprising a large number of qubits.

In the case of optical systems, entirely alternate, but equivalent, approaches may be used, such as probing the Wigner function directly using homodyne detection.

2. Quantum process tomography

Now to perform QPT we apply the unknown process to a complete basis of input states $\{\hat{\rho}_i\}$, and perform QST on the output state for each. This yields,

$$\mathcal{E}(\hat{\rho}_j) = \sum_i c_{i,j} \hat{\rho}_i, \quad (12.16)$$

where the sum runs over the basis of states. From QST, all the coefficients $c_{i,j}$ may be determined. Next we define the following decomposition for each of the terms in the sum of Eq. (12.10),

$$\hat{E}_m \hat{\rho}_j \hat{E}_n^\dagger = \sum_k B_{j,k}^{m,n} \hat{\rho}_k, \quad (12.17)$$

where B defines a decomposition in the chosen basis, not dependent on any measurement results. Then we can write,

$$\begin{aligned} \mathcal{E}(\hat{\rho}_j) &= \sum_{m,n} \chi_{m,n} \hat{E}_m \hat{\rho}_j \hat{E}_n^\dagger \\ &= \sum_{m,n} \sum_k \chi_{m,n} B_{j,k}^{m,n} \hat{\rho}_k. \end{aligned} \quad (12.18)$$

Because $\hat{\rho}_k$ form a linearly independent basis, we can write the decomposition,

$$c_{j,k} = \sum_{m,n} \chi_{m,n} B_{j,k}^{m,n}, \quad (12.19)$$

for all j, k . From this, standard linear algebra techniques allow an inversion to obtain,

$$\chi_{m,n} = \sum_{j,k} (B_{j,k}^{m,n})^{-1} c_{j,k}, \quad (12.20)$$

thereby obtaining the full process matrix χ , in the chosen basis.

— *Bulla crustulum.*

XIII. OPTICAL ENCODING OF QUANTUM INFORMATION

WHILE all-optical quantum computing is an unlikely architecture for future scalable quantum computers, it is all but inevitable that optics will play a central role in quantum communications networks. Foremost, this is because photons are ‘flying’ by their very nature and can very easily be transmitted across large distances – it’s quite challenging to transmit a superconducting circuit containing information from Australia to Mozambique in the blink of an eye! Additionally, optical states are, in many cases, relatively easy to prepare, manipulate and measure, and can also be readily interfaced with other physical quantum systems (Sec. XX.A), allowing the transfer of quantum information from optical communications systems to some other architecture better suited to a given task.

Optical systems are very versatile, allowing quantum information to be optically encoded in a number of ways – into single photons, many photons, or even an indeterminate number of photons, and in both discrete or continuous degrees of freedom. Different types of encodings may have very different properties in terms of the errors they are susceptible to (Sec. XIV).

When dealing with single photons, information can be encoded in a number of ways. Most obviously, it can be encoded into the polarisation basis, allowing one qubit of information per photon (i.e horizontal and vertical polarisation represent the logical $|0\rangle$ and $|1\rangle$ states). Or it could be directly encoded into the photon-number basis. However, other degrees of freedom, such as the spectral/temporal degrees of freedom could be employed, encoding information into time- or frequency-bins, with potentially far more levels than a simple polarisation qubit (?). Next we discuss the main methods for optical encoding of quantum information.

A. Single photons

A very attractive feature of single photons is that they undergo very little decoherence, even over large distances – dephasing (Sec. XIV.D) in the polarisation degree of freedom, for example, is negligible in free-space. They are, however, very susceptible to loss, and protocols relying on many single-photon states suffer exponential decay in their success rates as the number of photons is increased (Sec. XIV.C).

We can encode a single qubit into a single photon in the polarisation basis using the horizontal and vertical polarisation degrees of freedom. Equivalently, one can employ ‘dual rail’ encoding, whereby a single photon is placed into a superposition across two spatial modes. Finally, one can use time-bin encoding, whereby discrete windows of time represent logical basis states when occupied by a photon. This leads to the equivalent representations for logical qubits (L),

$$\begin{aligned} |\psi\rangle_{\text{qubit}} &\equiv \alpha|0\rangle_L + \beta|1\rangle_L, \\ |\psi\rangle_{\text{pol}} &\equiv \alpha|H\rangle + \beta|V\rangle, \\ |\psi\rangle_{\text{dual}} &\equiv \alpha|0,1\rangle + \beta|1,0\rangle, \\ |\psi\rangle_{\text{temporal}} &\equiv \alpha|0_t, 1_{t+\tau}\rangle + \beta|1_t, 0_{t+\tau}\rangle, \end{aligned} \quad (13.1)$$

shown graphically in Fig. 29.

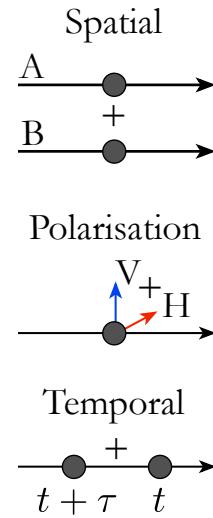


Figure 29: Three approaches to encoding a single qubit using a single photon, via a superposition across two spatial (A and B), polarisation (V and H) or temporal (t and $t + \tau$) modes.

Conversion between polarisation and dual-rail encoding is straightforward and deterministic using standard optical components, as described in Fig. 30.

Note that polarisation encoding requires a single spatial mode per qubit, whereas dual-rail encoding requires two. Polarisation encoding brings with it the advantage that arbitrary single-qubit operations may be implemented using wave-plates, which maintain coherence between the basis states extraordinarily well. In dual-rail encoding, on the

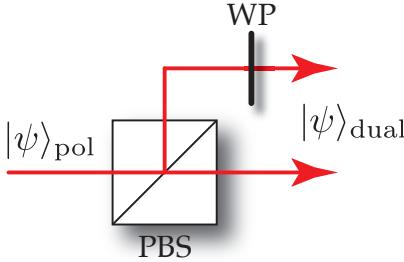


Figure 30: Conversion from single-photon polarisation encoding to dual-rail encoding, using a polarising beamsplitter (PBS) and wave-plate (WP). The PBS separates the polarisation components into two distinct spatial modes. The WP then rotates the polarisation of one of the spatial modes such that it has the same polarisation as the other. Conversion from dual-rail to polarisation encoding is just the reverse of this procedure.

other hand, single-qubit operations are implemented using beamsplitter operations between the two spatial modes, which must be interferometrically stable, since consecutive single-qubit operations yields Mach-Zehnder (MZ) interference (??), to be discussed in detail in Sec. XXII.B.

Single-photon encodings are extremely important, as they form the basis for universal linear optics quantum computing (Sec. XLII.A), BOSON SAMPLING (Sec. XLII.D.1) and quantum walks (Sec. XLII.D.4). They are also the simplest optical states for representing single qubits.

B. Photon-number

Of course, the photon-number degree of freedom needn't be limited to 0 or 1 photons. By fully exploiting the photon-number degree of freedom, we can encode a qudit¹⁹ of arbitrary dimension into a single optical mode,

$$|\psi\rangle_{\text{qudit}} \equiv \sum_{n=0}^{\infty} \alpha_n |n\rangle. \quad (13.2)$$

This may give the impression that a single optical mode has infinite information capacity. Needless to say, this sounds too good to be true, and it is. Loss decoheres photon-number-encoded states exponentially with photon-number, since for large photon-number the probability of a number state retaining its photon-number exponentially asymptotes to zero. So although in principle we can encode an ∞ -level qudit, the moment any non-zero loss is introduced, this exponential dependence destroys the state (Sec. XIV.C).

While photon-number encoding can be useful for communications purposes, it is not very practical for quantum

information processing tasks, since operations between basis states are not energy preserving, with each basis state having energy $E = n\hbar\omega$, where ω is frequency, and \hbar is Planck's constant. Thus, qudit operations would need to be active processes.

C. Spatio-temporal

Completely independent of the photon-number degree of freedom, are the spatio-temporal degrees of freedom, which encode the spatial, temporal, or spectral structure of photons. In the temporal domain, for example, we could define the temporal structure of a single photon as,

$$|\psi\rangle_{\text{temporal}} = \int_{-\infty}^{\infty} \psi(t) \hat{a}^\dagger(t) dt |0\rangle, \quad (13.3)$$

where $\hat{a}^\dagger(t)$ is the time-specific photonic creation operator, and $\psi(t)$ is the temporal distribution function (?). Equivalently, one could take the Fourier transform of the temporal distribution function and represent the same state in the frequency basis,

$$\tilde{\psi}(\omega) = \mathcal{FT}(\psi(t)). \quad (13.4)$$

Likewise, one could employ a similar representation in the transverse spatial degrees of freedom, with spatial distribution function $\psi(x, y)$.

Alternately, we can define *mode operators* (?), which are mathematically equivalent to creation operators, but create photons with a specific temporal envelope,

$$\begin{aligned} \hat{A}_\psi^\dagger &= \int_{-\infty}^{\infty} \psi(t) \hat{a}^\dagger(t) dt, \\ |\psi\rangle_{\text{temporal}} &= \hat{A}_\psi^\dagger |0\rangle. \end{aligned} \quad (13.5)$$

Mode operators commute, inheriting this property directly from photonic creation operators,

$$[\hat{A}_{\psi_1}^\dagger, \hat{A}_{\psi_2}^\dagger] = 0. \quad (13.6)$$

Now by defining an orthonormal basis of temporal distribution functions, $\{\xi_i\}$, such that,

$$\langle 0 | \hat{A}_{\xi_i} \hat{A}_{\xi_j}^\dagger | 0 \rangle = \delta_{i,j}, \quad (13.7)$$

we can encode a qudit of arbitrary dimension into the spatio-temporal degrees of freedom,

$$|\psi\rangle_{\text{qudit}} \equiv \sum_{i=0}^{\infty} \alpha_i \hat{A}_{\xi_i}^\dagger |0\rangle. \quad (13.8)$$

This encoding allows a qudit of arbitrary dimension to be encoded into a single spatial mode. Again, however, summing to infinity is somewhat fanciful, given any physically realistic spatio-temporal error model, such as

¹⁹ A d -level system, as opposed to a qubit's two levels.

an imperfect frequency response in the channel, e.g a bandpass response of an optical fibre or photo-detector.

The spectral basis functions could take the form of any orthonormal basis of complex functions, such as wavelets (?), Hermite functions (shown in Fig. 31), or well-separated functions with finite support.

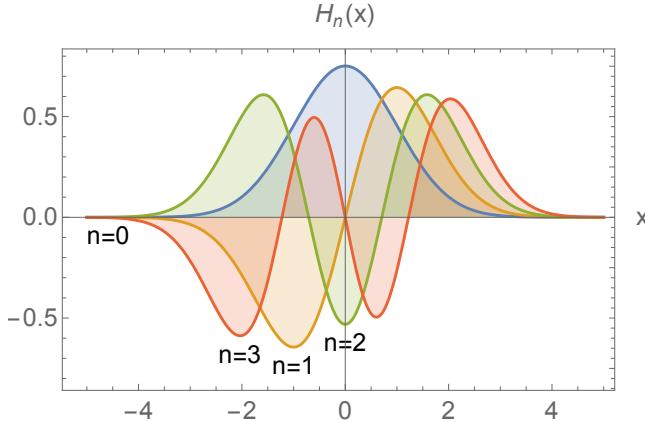


Figure 31: The family of Hermite functions, $H_n(x)$, modulated by Gaussian envelopes, are an example of an orthonormal basis of continuous functions that could be utilised in the spatio-temporal encoding of qudits into photonic wave-packets. Shown here are $n = 0, 1, 2, 3$.

1. Transverse electro-magnetic modes

Of particular interest are transverse electro-magnetic (TEM) modes, which are the two-dimensional eigenfunctions of the electro-magnetic field in the transverse direction of a light field. They are obtained by imposing boundary conditions on the electro-magnetic field, depending on the medium and the geometry of the medium (e.g. optical fibre, free-space, wave-guides).

Specifically, for cylindrical and rectangular boundary conditions, the intensity profile of the TEM modes are given by,

$$I_{p,l}^{\text{cyl}}(\rho, \varphi) = I_0 \rho^l L_p^l(\rho)^2 \cos(l\varphi) e^{-\rho},$$

$$I_{m,n}^{\text{rect}}(x, y, z) = I_0 \left(\frac{\omega_0}{\omega}\right)^2 \left[H_m \left(\frac{\sqrt{2}x}{\omega}\right) e^{-\frac{x^2}{\omega^2}} \right]^2 \left[H_n \left(\frac{\sqrt{2}y}{\omega}\right) e^{-\frac{y^2}{\omega^2}} \right]^2,$$

where,

$$H_n(x) = n! \sum_{m=0}^{\lfloor \frac{n}{2} \rfloor} \frac{(-1)^m}{m!(n-2m)!} (2x)^{n-2m},$$

$$L_n(x) = \sum_{k=0}^n \binom{n}{k} \frac{(-1)^k}{k!} x^k, \quad (13.9)$$

are the Hermite and Laguerre polynomials respectively. These examples are shown in Fig. 32.

The TEM modes are discrete, and denoted TEM_{mn} , where $m, n \in \mathbb{Z}_+$. TEM modes can be prepared and manipulated using holograms in the form of phase-masks.

2. Time-bins

In time-bin encoding we define our basis of modes (whether it be qubits or higher-dimensional qudits) as distinct, non-overlapping time-bins, which are localised wave-packets in the temporal degree of freedom, each

separated from the next by some fixed interval τ . This can be considered a special case of spatio-temporal encoding, where the basis mode functions satisfy the relation,

$$\xi_j(t) = \xi_0(t - j\tau), \quad (13.10)$$

as well as the usual orthonormality constraints. Here τ is sufficiently large, and $\xi_i(t)$ sufficiently temporally localised, that the temporal modes are orthogonal as per Eq. (13.7). The encoding is shown graphically in Fig. 33.

Time-bin encoding arises naturally in architectures where the photon source driving the system is operating at a high repetition rate, R , in which case $\tau = 1/R$. Architectures for optical quantum computing have been described (?), and experimentally demonstrated (?), based entirely on time-bin encoding.

These schemes can be very resource efficient, since a single source operating at high repetition rate can replace an entire bank of distinct sources that would ordinarily be required in spatial architectures. Similarly, a single time-resolved detector, with resolution at least τ , can replace a

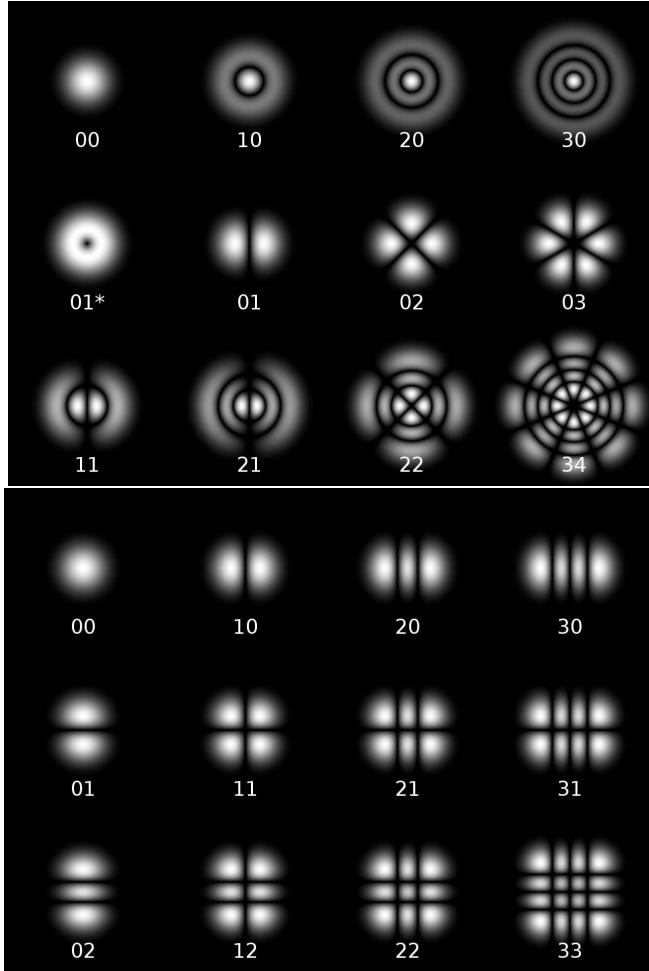


Figure 32: $\text{TEM}_{m,n}$ modes for cylindrical (top) and rectangular (bottom) mode profiles.

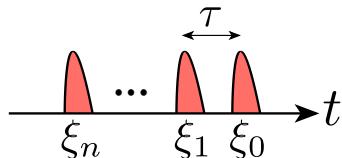


Figure 33: Time-bin encoding on an n -level quantum system. Time-bin are temporally localised with envelopes ξ_j , orthogonal to all others, different only via temporal displacements. The temporal separation between each consecutive time-bin is given by τ .

bank of detectors operating in parallel. And only a single spatial mode is required to store an arbitrary number of qubits/qudits, so long as it is long enough to support the entire pulse-train – at least $2n\tau$ for n qubits.

In the schemes of (??), entire optical quantum computing protocols can be efficiently constructed using only a single source, a single detector, two delay-lines, and three dynamically-controlled beamsplitters, irrespective of the size of the computation, an enormous resource saving compared to traditional spatial encodings. Furthermore, in these schemes, there is only a single point of interfer-

ence, greatly simplifying optical interferometric alignment, which would ordinarily require simultaneously aligning a large number of optical elements, as many as $O(m^2)$ elements for an m -mode network (?).

D. Thermal states

In some quantum protocols, although the inner workings may be quantum mechanical in nature, the inputs and outputs needn't capture any quantum coherence – sometimes *classical* information is sufficient for communications. As discussed above, coherent states are the archetypal example of this, and this is in fact the norm in present-day classical fibre-optic communication, where coherent states prepared via laser diodes are employed.

Another, and even simpler option, is thermal states. These are obtained by fully dephasing a coherent state, retaining the amplitude distribution, while nullifying all the coherence terms,

$$\hat{\rho}_{\text{thermal}}(\alpha) = e^{-|\alpha|^2} \sum_{n=0}^{\infty} \frac{|\alpha|^2}{n!} |n\rangle\langle n|. \quad (13.11)$$

Thermal states can encode classical information into their amplitudes, polarisations, or time-bins, as before. The advantage of this type of encoding is that thermal states are trivial to prepare and measure (e.g ordinary blackbody radiation emits thermal states – this is how a normal incandescent lightbulb emits light). However, they are purely classical states, do not undergo interference with one another, and are therefore useless for, for example, entangling qubits via which-path erasure, any other type of coherent interferometric process, or for representing coherent quantum information such as qubits.

E. Phase-space

When encoding information optically, we needn't restrict ourselves to photon-number states (discrete variables). We also have a lot of flexibility to encode information in phase-space using continuous-variable (CV) states, where phase and amplitude relations encode quantum information (?). In this formalism, rather than expressing states in terms of photonic creation operators, \hat{a}^\dagger , we represent them using phase-space position (\hat{x}) and momentum (\hat{p}) operators.

In phase-space, the most common method for visualising optical states is in terms of quasi-probability functions²⁰, of which there are a multitude. The best known quasi-probability representations are:

²⁰ The term ‘quasi-probability’ arises because in some regimes (for

- P -function: represents a state as a quasi-mixture of coherent states. When the P -function is strictly non-negative, it can be interpreted as a perfect classical mixture of coherent states. However, with any negativity this classical interpretation breaks down, hence ‘quasi’-probability. In general, the P -function representation for a state is not unique.

$$\hat{\rho} = \iint P(\alpha)|\alpha\rangle\langle\alpha|d^2\alpha. \quad (13.12)$$

- Q -function: represents a state in terms of its overlap with the complete set of all coherent states, which form an over-complete basis.

$$Q(\alpha) = \frac{1}{\pi}\langle\alpha|\hat{\rho}|\alpha\rangle. \quad (13.13)$$

- Wigner function: also has a quasi-probability interpretation, and negativity is qualitatively associated with ‘quantumness’. The Wigner function of a state is unique, and isomorphic to the density operator, making it perhaps the most useful phase-space representation for quantum states of light.

$$W(x, p) = \int e^{ips/\hbar} \left\langle x - \frac{s}{2} \right| \hat{\rho} \left| x + \frac{s}{2} \right\rangle ds. \quad (13.14)$$

These representations, whilst entirely equivalent to a photon-number basis representation, are far easier to work with for many types of states. Most notably, Gaussian states are conveniently represented and manipulated using phase-space representations.

1. Coherent states

As the most trivial CV encoding of quantum information, consider coherent states. These are particularly useful since they are pure states, with well defined coherence relationships, and are closely approximated by laser sources, and therefore readily available in the lab.

A coherent state, $|\alpha\rangle$, is parameterised by a single complex parameter, α , given by a phase and amplitude,

$$|\alpha\rangle = e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle. \quad (13.15)$$

Fig. 34 illustrates the phase-space representation for two approximately orthogonal coherent state basis states.

By manipulating these parameters, information can be encoded into coherent states. We could, for example,

example, strictly non-negative P -functions), the function has a true probabilistic interpretation. However this interpretation breaks down for any negativity in the $P(\alpha)$, since negative probabilities have no meaningful classical interpretation.

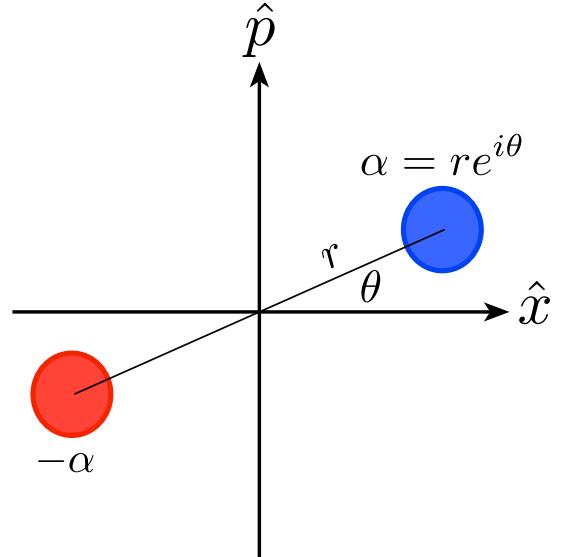


Figure 34: Phase-space representation of a coherent state with complex amplitudes $\pm\alpha$. For sufficiently large amplitude one can approximate orthogonality, enabling a qubit encoding.

define two coherent states of opposite phase to represent qubit basis states,

$$\begin{aligned} |0\rangle &\equiv |\alpha\rangle, \\ |1\rangle &\equiv |-\alpha\rangle. \end{aligned} \quad (13.16)$$

Note, however, that this representation for qubits is only approximate, since coherent states are not orthogonal,

$$\begin{aligned} \langle\alpha|\beta\rangle &= e^{-\frac{1}{2}(|\alpha|^2+|\beta|^2-2\alpha^*\beta)} \\ &\neq \delta(\alpha - \beta), \end{aligned} \quad (13.17)$$

thus the two logical basis states are not perfectly orthogonal,

$$\langle-\alpha|\alpha\rangle = e^{-2|\alpha|^2}, \quad (13.18)$$

which is non-zero for any finite α , whereas for ideal qubits we require $\langle 0|1\rangle = 0$. However, for large α , $|\pm\alpha\rangle$ closely approximate orthogonality, allowing them to be used as qubits.

This representation for qubits using coherent states is easily generalised to qudits by considering coherent states orbiting the origin of phase-space at equal angular intervals of $2\pi/d$, for a d -level qudit. The k th qudit basis state is then,

$$|k\rangle_d = |\alpha e^{ik/d}\rangle, \quad (13.19)$$

for $k = 0, \dots, d-1$, where again the basis states are non-orthogonal, but closely approximate orthogonality for large α . The qudit value k can easily be manipulated using simple phase-shift operators,

$$\hat{\Phi}(\phi) = e^{i\phi\hat{n}}, \quad (13.20)$$

where $\hat{n} = \hat{a}^\dagger \hat{a}$ is the photon-number operator. These phases are trivially implemented in the laboratory as wavelength-scale modulations in optical path length (i.e. a thin piece of glass or other transmissive material with a different refractive index).

Note that despite being pure states, with well-defined coherence, coherent states are considered classical, as they are unable to encode quantum information. That is, the coherence relationships cannot be exploited for the encoding of qubits or qudits.

Coherent states are useful in that they are easy to prepare using modern lasers, including laser diodes, and by turning up the amplitude can be transmitted over long distances, with loss not affecting quantum coherence, only the amplitude (Sec. XIV.C).

Coherent state encoding can be regarded as encoding via the displacement operator,

$$\hat{D}(\alpha) = \exp [\alpha \hat{a}^\dagger - \alpha^* \hat{a}], \quad (13.21)$$

which implements translations in phase-space via the addition of coherent amplitude to a state. Coherent states are simply obtained as displaced vacuum states,

$$\hat{D}(\alpha)|0\rangle = |\alpha\rangle. \quad (13.22)$$

2. Cat states

Another type of CV state, which can in fact encode quantum information, is superpositions of coherent states (colloquially known as ‘cat’ states), with the encoding (?),

$$\begin{aligned} |0\rangle_L &\equiv \mathcal{N}_+ (|\alpha\rangle + |-\alpha\rangle) \\ &= 2\mathcal{N}_+ e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^{2n}}{\sqrt{(2n)!}} |2n\rangle, \\ &= |\text{cat}_+(\alpha)\rangle, \\ |1\rangle_L &\equiv \mathcal{N}_- (|\alpha\rangle - |-\alpha\rangle) \\ &= 2\mathcal{N}_- e^{-\frac{|\alpha|^2}{2}} \sum_{n=0}^{\infty} \frac{\alpha^{2n+1}}{\sqrt{(2n+1)!}} |2n+1\rangle, \\ &= |\text{cat}_-(\alpha)\rangle, \end{aligned} \quad (13.23)$$

where the normalisation factors are,

$$\mathcal{N}_\pm = \frac{1}{\sqrt{2(1 \pm e^{-2|\alpha|^2})}}, \quad (13.24)$$

which arise due to the non-orthogonality of coherent states, Eq. (13.17). These two basis states contain strictly even or odd photon-number terms respectively (i.e. they have well-defined parity), implying that, unlike coherent states, they are always orthogonal, regardless of amplitude,

$$\langle \text{cat}_+(\alpha) | \text{cat}_-(\alpha) \rangle = 0 \quad \forall \alpha, \quad (13.25)$$

making them directly appropriate for qubit encoding, even for weak coherent amplitudes.

Unfortunately, cat states are notoriously difficult to prepare, and extremely sensitive to loss (Sec. XIV.C) and dephasing (Sec. XIV.D). This arises because loss of a single photon flips the parity of the state to an orthogonal one, meaning that as α increases, the state is exponentially more susceptible to decohering into a mixture of the logical basis states.

However, modulo these difficulties, with a resource of cat states at one’s disposal, universal quantum computation may be realised using post-selected linear optics (?).

3. Squeezed states

In the same way that information can be encoded using the displacement operator via coherent states, we can encode information via the single-mode squeezing operator,

$$\hat{S}(\xi) = \exp \left[\frac{1}{2} (\xi^* \hat{a}^2 - \xi \hat{a}^\dagger \hat{a}) \right], \quad (13.26)$$

where $\xi = re^{i\varphi} \in \mathbb{C}$, r is known as the squeezing parameter, which will determine the magnitude of the squeezing, and $\varphi \in [0, 2\pi]$ denotes the axis along which the squeezing is taking place.

Graphically, in terms of their phase-space visualisation, squeezing implements dilations about a given axis. Strongly squeezing the vacuum state along the \hat{x} or \hat{p} directions yields two states that are approximately orthogonal for large squeezing amplitudes, as shown in Fig. 35. Thus, with sufficient squeezing they can be used as a basis for approximating a qubit. This encoding can be exploited for full universal quantum computing, to be discussed in Sec. XLII.E.

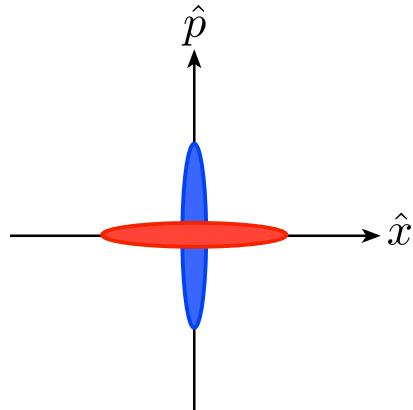


Figure 35: Phase-space representation of two squeezed vacuum states, squeezed in the \hat{x} (blue) and \hat{p} (red) directions. In the limit of large squeezing these states become approximately orthogonal and may therefore approximate the encoding of a qubit.

F. Non-optical encoding

In a non-optical context, the elementary unit of quantum information – the qubit – can be naturally encoded into any system with a natural or engineered two-level structure. This actually encompasses a broad range of possibilities, including, amongst many others:

- Two-level atoms: let two distinct electron energy levels, with long lifetimes, represent the two logical basis states.
- λ -configuration atoms: atoms with two degenerate ground states, which encode the logical qubit, and an additional excited state, which may be transitioned to upon excitation from only one of the ground states. Relaxation from the excited state enables optical coupling via the emitted photon.
- Quantum dots: are essentially artificial atoms, which can be engineered with custom band-structures, allowing two- or higher-level qudits to be easily fabricated.
- Nitrogen-vacancy (NV) centres: are a type of point defect in diamond, which has a very well defined energy level structure that may be utilised to represent qubits.
- Atomic ensembles: encode quantum information similarly to a single atom, except that the excitation is a *collective* one, in superposition across all the atoms in the ensemble.
- Superconducting rings: a superposition of current flow direction in a superconducting ring represents the two logical basis states.
- Trapped ions: qubits are encoded into stable electronic states of electromagnetically trapped ions.

Clearly the non-optical elements in a quantum network must somehow interface with optical states, such that communication is facilitated. This is discussed later in Sec. XX.A.

— *Audaces fortuna juvat.*

XIV. ERRORS IN QUANTUM NETWORKS

As with classical data, quantum data is susceptible to corruption during transmission. However, in addition to all the usual classical error models, quantum information is subject to further uniquely quantum errors. These errors can be represented using the quantum process formalism and fully characterised using QPT (Sec. XII.D). We now briefly discuss several of the dominant errors arising in quantum systems, paying especial attention to error models acting on qubits and optical states, as these are the most relevant in a quantum networking context.

A. Known unitaries

The most trivial error mechanism is when a (potentially multi-qubit) unitary channel (e.g an identity channel for the purposes of quantum memory) actually implements some unitary transformation, \hat{U} , that is not that which is desired. However, the unitary is constant, not varying from trial to trial, and is known, which can be easily determined by performing QPT on the channel. For example, an optical fibre might induce a polarisation rotation on transmitted photons, but the fibre isn't changing and neither is the rotation. If consistently implementing the same known unitary then reversing it is straightforward in most architectures, by applying \hat{U}^\dagger , since $\hat{U}^\dagger \hat{U} = \hat{1}$.

B. Unknown imperfect unitaries

Alternate to known unitaries, the unitary operation implemented by a node/channel may deviate from that which is desired, in an unknown manner, thereby implementing a slightly different operation than that which we intended to engineer. Specifically, the effective unitary can be represented as the ideal unitary, augmented by some deviation matrix,

$$\hat{U}_{\text{effective}} = \hat{U}_{\text{ideal}} + \hat{\Delta}_{\text{error}}, \quad (14.1)$$

where the matrix elements of $\hat{\Delta}_{\text{error}}$ (which is not unitary in general) are unknown, but hopefully small. Since the unknown deviation matrix needn't be constant, it will be a function of random variables, evaluated independently for each trial of the process. Furthermore, since the deviation matrix may vary from trial to trial, QPT cannot be employed to characterise it, unlike unitaries with fixed errors.

C. Loss

Given that quantum communication links will typically be optical, the dominant error mechanism is likely to be loss. We let the *efficiency*, η , of an optical quantum process be the probability that a given photon entering the channel leaves the channel in the desired mode, or probability $1 - \eta$ of being lost. In the case of information encoded into single-photon states, e.g using the polarisation degree of freedom, η corresponds exactly to the success probability of the communication.

When implementing protocols employing post-selection upon detecting all photons, the protocol will be non-deterministic, where loss dictates the protocol's success probability. Specifically, with n photons, each with efficiency η , the net post-selection success probability of the entire device is,

$$P = \eta^n. \quad (14.2)$$

This implies an exponential number of trials,

$$N = \frac{1}{P} = \frac{1}{\eta^n}, \quad (14.3)$$

is required in post-selected protocols. Clearly this exponential scaling is of concern, requiring demanding efficiencies in future large-scale implementations.

1. Model

Formally, let $\mathcal{E}_\eta^{\text{loss}}$ be the loss channel with efficiency η . The channel acting on an initially pure single-photon state, $|1\rangle$, can be modelled as a beamsplitter with transmissivity η acting on the state, where the reflected mode is traced out, shown in Fig. 36. This yields the quantum process,

$$\mathcal{E}_\eta^{\text{loss}}(\hat{\rho}) = \text{tr}_B[\hat{U}_{\text{BS}}(\hat{\rho}_A \otimes |0\rangle_B \langle 0|_B) \hat{U}_{\text{BS}}^\dagger], \quad (14.4)$$

where \hat{U}_{BS} is the beamsplitter operation.

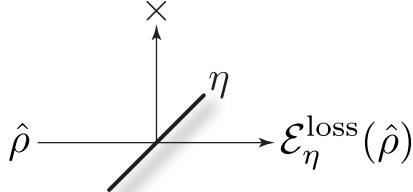


Figure 36: Model for the loss channel. The input state, $\hat{\rho}$, passes through a beamsplitter of transmissivity η , and the reflected mode discarded, yielding the lossy output state $\mathcal{E}_\eta^{\text{loss}}(\hat{\rho})$.

Consecutive loss channels act multiplicatively (in the net efficiency) and commutatively,

$$\mathcal{E}_{\eta_1}^{\text{loss}} \circ \mathcal{E}_{\eta_2}^{\text{loss}} = \mathcal{E}_{\eta_2}^{\text{loss}} \circ \mathcal{E}_{\eta_1}^{\text{loss}} = \mathcal{E}_{\eta_1 \eta_2}^{\text{loss}}. \quad (14.5)$$

2. Linear optics networks

In the special case of linear optics circuits, loss channels have the elegant property that, provided the loss rate is uniform across all modes, they can be commuted through the circuit to the front or back (?). Specifically,

$$(\mathcal{E}_\eta^{\text{loss}})^{\otimes m} \circ \mathcal{E}_U = \mathcal{E}_U \circ (\mathcal{E}_\eta^{\text{loss}})^{\otimes m}, \quad (14.6)$$

where \mathcal{E}_U is a unitary linear optics process, implementing a photon-number-preserving map of the form of Eq. (25.1). This is represented by the circuit diagram shown in Fig. 37. This simplifies the treatment of distinct system inefficiencies (such as source, network and detector inefficiencies) by allowing us to commute them to the beginning or end of the circuit and combine them together into a single net efficiency. In many scenarios, this allows the different system inefficiencies to be dealt with via post-selection.

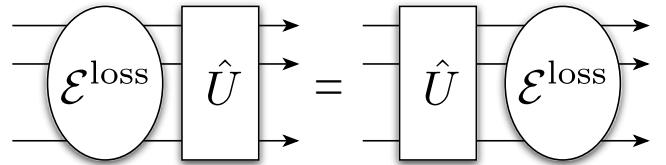


Figure 37: Commutation of a uniform loss channel (i.e identical efficiency on each mode), $\mathcal{E}^{\text{loss}}$, through a passive linear optics network, \hat{U} .

3. Single-photon encoding

In the case of the vacuum and single-photon states, which is most common to qubit encodings, we obtain,

$$\begin{aligned} \mathcal{E}_\eta^{\text{loss}}(|0\rangle\langle 0|) &= |0\rangle\langle 0|, \\ \mathcal{E}_\eta^{\text{loss}}(|1\rangle\langle 1|) &= (1 - \eta)|0\rangle\langle 0| + \eta|1\rangle\langle 1|. \end{aligned} \quad (14.7)$$

This dynamic is of the same form as amplitude damping (Sec. XIV.F).

4. Polarisation & dual-rail encoding

This process would apply equivalently to both horizontal and vertical polarisations. Therefore, via linearity, the loss channel acting on a polarisation-encoded qubit (Sec. XIII.A) yields,

$$\mathcal{E}_\eta^{\text{loss}}(|\psi\rangle_{\text{pol}}\langle\psi|_{\text{pol}}) = (1 - \eta)|0\rangle\langle 0| + \eta|\psi\rangle_{\text{pol}}\langle\psi|_{\text{pol}}. \quad (14.8)$$

The same applies in the context of dual-rail encoding. Note that while this transformation mixes the state in the photon-number degree of freedom, it preserves coherence between the horizontal and vertical single-photon components. Thus, upon successful post-selection, the state is projected back onto the desired qubit state.

5. Photon-number encoding

In the general case of an n -photon Fock state, we obtain,

$$\mathcal{E}_\eta^{\text{loss}}(|n\rangle\langle n|) = \sum_{i=0}^n \binom{n}{i} \eta^i (1 - \eta)^{n-i} |i\rangle\langle i|. \quad (14.9)$$

In the case of higher order photon-number encoding of qudits, as per Eq. (13.2), the probability of an n -photon basis state being maintained scales as η^n . That is, if the highest photon-number term in our qudit is n , that component has an exponentially low probability of being preserved through the loss channel. For this fundamental reason, photon-number encoding does not enable infinite-dimensional qudits to be encoded.

6. Coherent state encoding

Coherent states are one example of states, which are in a sense robust against loss, since a lossy coherent state is another coherent state with lower amplitude, but without any loss in coherence,

$$\mathcal{E}_\eta^{\text{loss}}(|\alpha\rangle\langle\alpha|) = |\eta\alpha\rangle\langle\eta\alpha|. \quad (14.10)$$

This arises because coherent states are eigenstates of the photonic annihilation operator, $\hat{a}|\alpha\rangle = \alpha|\alpha\rangle$.

However, although coherence is maintained under the loss channel, the process is irreversible, since noise-free amplitude amplification is not possible in general (?). Thermal states exhibit the same property, that a loss channel simply yields another thermal state with reduced amplitude, although these exhibit no coherence.

7. Cat state encoding

To the contrary, while cat states (Sec. XIII.E.2) are simple superpositions of coherent states, they are extremely sensitive to loss. This is because cat states have well-defined photon-number parity (strictly even or odd photon-number), and therefore the loss of just a single photon will flip a cat state to an orthogonal one. Since the probability of photon loss occurring increases exponentially with photon-number, large amplitude cat states are exponentially sensitive to loss channels.

Consider a cat state (either even or odd parity), as per Eq. (13.23). Subjecting this to a loss channel yields the mixed state,

$$\begin{aligned} \hat{\rho}_{\text{loss}} &= \mathcal{E}_\eta^{\text{loss}}(|\text{cat}_\pm(\alpha)\rangle\langle\text{cat}_\pm(\alpha)|) \\ &= \mathcal{N}_\pm^{-2}[\mathcal{E}_\eta^{\text{loss}}(|\alpha\rangle\langle\alpha|) + \mathcal{E}_\eta^{\text{loss}}(|-\alpha\rangle\langle-\alpha|)] \\ &\pm \mathcal{E}_\eta^{\text{loss}}(|-\alpha\rangle\langle\alpha|) \pm \mathcal{E}_\eta^{\text{loss}}(|\alpha\rangle\langle-\alpha|)] \\ &= \mathcal{N}_\pm^{-2}[|\eta\alpha\rangle\langle\eta\alpha| + |-\eta\alpha\rangle\langle-\eta\alpha| \\ &\pm e^{-2|(1-\eta)\alpha|^2}(|\eta\alpha\rangle\langle-\eta\alpha| + |-\eta\alpha\rangle\langle\eta\alpha|)]. \end{aligned} \quad (14.11)$$

Now it's evident that the off-diagonal terms (i.e those capturing coherence) accumulate a factor of,

$$C = e^{-2|(1-\eta)\alpha|^2}, \quad (14.12)$$

which rapidly asymptotes to zero for large α with any $\eta < 1$, as shown in Fig. 38.

8. NOON states

Similarly, NOON states (Sec. XXIII.C) undergo complete wave-function collapse if just a single photon is lost to the environment. This is because any single photon reveals complete information about the location of the remaining $N - 1$ photons. Therefore, a NOON state subject

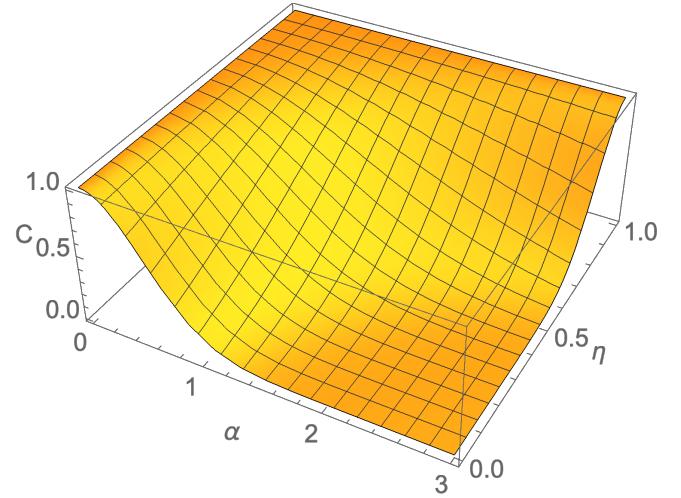


Figure 38: Coherence factor in a cat state subject to decoherence via loss, as a function of efficiency η , and coherent amplitude α .

to loss of a single photon decoheres to the mixture,

$$\hat{\rho} = \frac{1}{2}(|N-1,0\rangle\langle N-1,0| + |0,N-1\rangle\langle 0,N-1|), \quad (14.13)$$

which contains no entanglement.

Because there are N photons in total, the probability of wave-function collapse grows exponentially with photon-number. Specifically, the probability of completely collapsing and losing all entanglement is,

$$P = 1 - \eta^N, \quad (14.14)$$

as shown in Fig. 39.

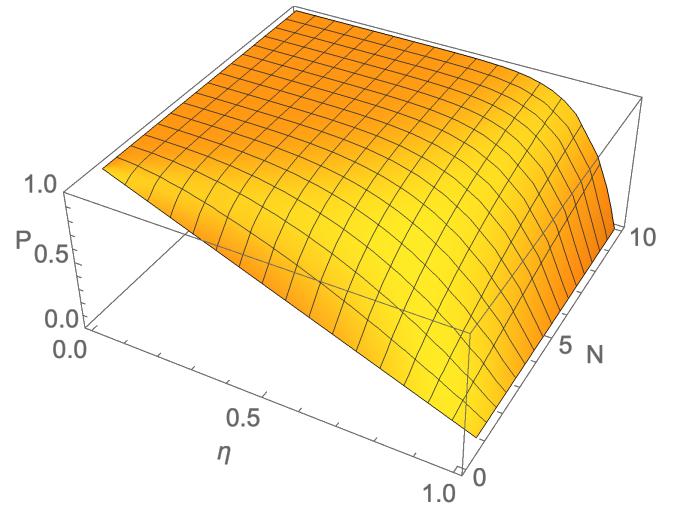


Figure 39: Probability of an N -photon NOON state completely collapsing and losing all entanglement, under a loss channel with efficiency η .

9. Scaling

The scaling of loss over distance d varies depending on the medium through which the light traverses. We will consider two dominant mediums, most relevant to future quantum networking:

- Optical fibre: mode geometry is well-preserved, but optical medium is intrinsically lossy.
- Free-space: mode geometry is subject to dispersion, but the medium is either lossless (in vacuum), or very low-loss (in atmosphere).

When propagating through fibre (or atmosphere, or some other lossy medium) net efficiency scales inverse exponentially as,

$$\eta = O(e^{-\alpha d}), \quad (14.15)$$

where α is a characteristic of the medium²¹.

In free-space on the other hand, where the medium of propagation is vacuum, which is effectively lossless, the effective loss rate is not determined by the medium, but rather by the fact that the spot-size of an optical state is subject to dispersion and grows only quadratically with distance, as shown in Fig. 40. Then when the light is detected, if the spot-size is greater than the detector aperture, the undetected component effectively translates to loss. Thus through free-space the effective efficiency scales as,

$$\eta = O\left(\frac{1}{d^2}\right), \quad (14.16)$$

which is far more favourable than the exponential scaling inherent to lossy mediums. This provides space-based quantum networks with an inherent competitive advantage compared to any form of ground-based network.

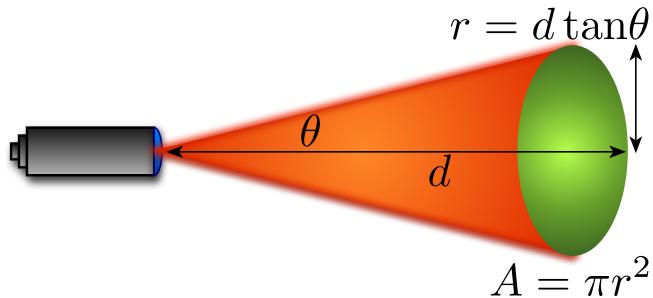


Figure 40: Spot-size of an optical beam grows quadratically with distance from the source. If the beam is then collected via a camera with aperture area A' , any component of A falling outside of A' is effectively lost, yielding an effective loss channel with efficiency $\eta = \frac{A'}{A} = \frac{A'}{\pi d^2 \tan^2 \theta} = O\left(\frac{1}{d^2}\right)$.

²¹ With present-day fibre technology, this characteristic decay rate is on the order of $\alpha = \frac{1}{22\text{km}}$.

Through atmospheric channels we will have both dispersion and distance-dependent loss, yielding an effective loss rate given by the product of the two effects,

$$\eta = O\left(\frac{e^{-\alpha d}}{d^2}\right). \quad (14.17)$$

D. Dephasing

The dephasing error model describes the deterioration of quantum coherence in a state. It does not change the actual amplitudes of the components in the superposition, but rather reduces the state to a mixture of those components. Thus, dephasing can be thought of as destroying quantum information (coherence), while retaining classical information (probability amplitudes).

1. Qubits

In terms of qubits, dephasing is most commonly represented using the Kraus representation,

$$\mathcal{E}_p^{\text{dephasing}}(\hat{\rho}) = p \cdot \hat{\rho} + (1 - p) \cdot \hat{Z} \hat{\rho} \hat{Z}, \quad (14.18)$$

where $\hat{\rho}$ is the state of a single qubit, and \hat{Z} is the Pauli phase-flip operator²². Intuitively this tells us that the dephasing channel creates a mixture of an input state with its phase-flipped self.

An alternate interpretation for the dephasing channel is that it is equivalent to the outside environment measuring $\hat{\rho}$ in the logical (\hat{Z}) basis, but unknown to us, thereby projecting the state onto one basis state or another, yielding a mixture of the two.

Dephasing acting on $\hat{\rho}$ can be very elegantly visualised as simply nullifying the off-diagonal matrix elements, i.e. eliminating coherence terms. Dephasing is a ubiquitous error mechanism and affects all current quantum computing architectures.

Consecutive dephasing channels accumulate into another dephasing channel,

$$\mathcal{E}_{p_1}^{\text{dephasing}} \circ \mathcal{E}_{p_2}^{\text{dephasing}} = \mathcal{E}_{p'}^{\text{dephasing}}, \quad (14.19)$$

where,

$$p' = p_1 p_2 + (1 - p_1)(1 - p_2), \quad (14.20)$$

i.e. the probability that an even number of phase-flips have occurred.

²² Bit-flip and bit-phase-flip channels may be represented similarly by replacing \hat{Z} with \hat{X} or \hat{Y} respectively, although these don't arise as naturally as dephasing in many physical contexts.

As a simple example, consider the $p = 1/2$ dephasing channel acting on the $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ state. Then we have,

$$\begin{aligned}\mathcal{E}_{1/2}^{\text{dephasing}}(|+\rangle\langle+|) &= \frac{1}{2}(|+\rangle\langle+| + \hat{Z}|+\rangle\langle+\hat{Z}|) \\ &= \frac{1}{2}(|+\rangle\langle+| + |-\rangle\langle-|) \\ &= \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \\ &= \frac{\mathbb{1}}{2},\end{aligned}\quad (14.21)$$

is the completely mixed state. That is, the state has completely decohered. Note, however, that this complete decoherence depended on the choice of input state. A computational basis state, on the other hand, would be left unchanged by this channel,

$$\begin{aligned}\mathcal{E}_{1/2}^{\text{dephasing}}(|0\rangle\langle 0|) &= \frac{1}{2}(|0\rangle\langle 0| + \hat{Z}|0\rangle\langle 0|\hat{Z}) \\ &= |0\rangle\langle 0|, \\ \mathcal{E}_{1/2}^{\text{dephasing}}(|1\rangle\langle 1|) &= \frac{1}{2}(|1\rangle\langle 1| + \hat{Z}|1\rangle\langle 1|\hat{Z}) \\ &= |1\rangle\langle 1|.\end{aligned}\quad (14.22)$$

Note that the probability of no dephasing occurring over multiple dephasing channels in series is given by the product of the respective probabilities for the individual channels.

2. T_2 -times

A qubit dephasing channel is often quoted in terms of its T_2 -time, a characteristic time for dephasing to occur under continuous time-evolution. Specifically, the probability of no dephasing occurring scales as,

$$p_{\text{no error}} = e^{-t/T_2}, \quad (14.23)$$

yielding a the equivalent dephasing channel,

$$\mathcal{E}_t^{\text{dephasing}}(\hat{\rho}) = e^{-t/T_2}\hat{\rho} + \frac{1}{2}(1 - e^{-t/T_2})(\hat{\rho} + \hat{Z}\hat{\rho}\hat{Z}), \quad (14.24)$$

as shown in Fig. 41. For a qubit density matrix,

$$\hat{\rho} = \begin{pmatrix} \alpha & \gamma \\ \gamma^* & \beta \end{pmatrix}, \quad (14.25)$$

this is equivalent to adding a factor of e^{-t/T_2} to the two off-diagonal (coherence) elements,

$$\hat{\rho}_t = \begin{pmatrix} \alpha & e^{-t/T_2}\gamma \\ e^{-t/T_2}\gamma^* & \beta \end{pmatrix}. \quad (14.26)$$

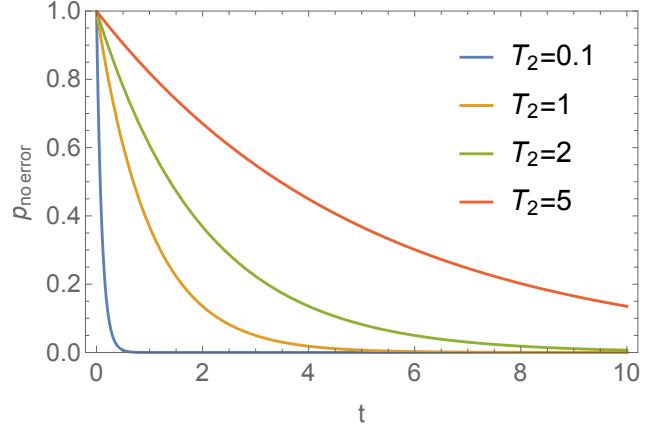


Figure 41: Dephasing under continuous-time evolution, with characteristic decay rate given by the T_2 -time. $p_{\text{no error}}$ is the probability that the state is left unchanged, whereas $1 - p_{\text{no error}}$ is the probability that it is replaced with the completely dephased state.

3. Optical states

The notion of dephasing can be easily generalised to non-qubit states of light, i.e with photon-number $n > 1$. In general, dephasing has the property of mapping a superposition of basis states to a mixture of the same basis states, whilst preserving amplitudes. Thus, for perfect dephasing,

$$\mathcal{E}^{\text{dephasing}} \left(\sum_i \alpha_i |i\rangle \cdot \sum_j \alpha_j^* \langle j| \right) \rightarrow \sum_i |\alpha_i|^2 |i\rangle\langle i|, \quad (14.27)$$

for some arbitrary basis enumerated by i and j . As an example, this process decoheres coherent states into thermal states. For partial dephasing, we can express the channel as creating a mixture over the input state with different phase rotations applied,

$$\mathcal{E}_\phi^{\text{dephasing}}(\hat{\rho}) = \int_0^{2\pi} \phi(\omega) \hat{\Phi}(\omega) \hat{\rho} \hat{\Phi}(\omega)^\dagger d\omega, \quad (14.28)$$

where $\hat{\Phi}(\omega)$ is a phase-shift operator with phase ω , obeying $\hat{\Phi}(\omega)^\dagger = \hat{\Phi}(-\omega)$, and $\phi(\omega)$ is a normalised probability density function characterising the distribution of phase-shifts. In the case of optical states, the phase-shift operators take the form,

$$\hat{\Phi}(\omega) = e^{-i\omega\hat{n}}, \quad (14.29)$$

in the photon-number basis, where $\hat{n} = \hat{a}^\dagger \hat{a}$ is the photon-number operator, satisfying $\hat{n}|n\rangle = n|n\rangle$. With no dephasing, $\phi(\omega) = \delta(\omega)$ and \mathcal{E} reduces to the identity channel. Otherwise, the off-diagonal (coherence) terms in the density operator begin to cancel out, leaving the diagonal (amplitude) terms unchanged. Thus, a perfect dephasing

channel acting on a coherent state yields a thermal state of equal amplitude.

From this definition it can be seen that susceptibility to dephasing increases with photon-number, since the number operator adds a multiplicative factor to the acquired phase-shift,

$$\hat{\Phi}(\omega)|n\rangle = e^{-i\omega n}|n\rangle. \quad (14.30)$$

For number states not in superposition, this corresponds to a simple unimportant global phase, since number states are phase-invariant. However, in superposition this adds relative phases, thereby destroying coherences upon applying the integral from Eq. (14.28).

E. Depolarisation

Depolarisation is a noise model more general than dephasing, that probabilistically replaces a state with the completely mixed state (regardless of the input state). That is, with some probability we lose *all* quantum *and* classical information, i.e both coherences and probability amplitudes. Note that the dephasing channel introduced above only destroys quantum coherence, whilst preserving amplitudes. Formally, the depolarising channel can be expressed as,

$$\mathcal{E}_p^{\text{depolarising}}(\hat{\rho}) = p \cdot \hat{\rho} + (1-p) \cdot \frac{\hat{\mathbb{1}}}{\dim(\hat{\rho})}, \quad (14.31)$$

where $\hat{\mathbb{1}}/\dim(\hat{\rho})$ is the completely mixed state in the d -dimensional Hilbert space.

When acting on qubits, the depolarising channel can equivalently be represented as the action of each of the four Pauli matrices with equal probability, since,

$$\frac{\hat{\mathbb{1}}}{2} = \frac{1}{4}(\hat{\rho} + \hat{X}\hat{\rho}\hat{X} + \hat{Y}\hat{\rho}\hat{Y} + \hat{Z}\hat{\rho}\hat{Z}). \quad (14.32)$$

Thus, both dephasing and depolarisation are examples of Pauli error models.

In the qubit basis (i.e not including loss, for example), the Pauli matrices form a complete basis for quantum operations. Thus, the depolarising channel is the most general qubit error model, since it effectively applies all four Pauli error channels. For this reason, when evaluating fault-tolerance thresholds for QEC codes, thresholds are typically quoted in terms of the depolarising error rate.

Like the dephasing and loss channels, the error probability of multiple channels in series accumulates multiplicatively,

$$\mathcal{E}_{p_1}^{\text{depolarising}} \circ \mathcal{E}_{p_2}^{\text{depolarising}} = \mathcal{E}_{p_1 p_2}^{\text{depolarising}}. \quad (14.33)$$

F. Amplitude damping

An error not so much relevant to optics, but which arises very naturally in some other systems, such as atomic systems or quantum dots, is amplitude damping, also referred

to as a *relaxation channel*. Here the process models the relaxation of a higher energy level, $|1\rangle$, to a lower energy one, $|0\rangle$. The $|0\rangle$ state is assumed to be the ground state and cannot relax any further, but the $|1\rangle$ state can spontaneously relax to the ground state. After complete amplitude damping, any input state will be left in the ground state $|0\rangle$. This model can be thought of as energy dissipating from the qubit system and being measured by the environment, leading to a type of decoherence whereby the input state is probabilistically replaced by the ground state.

The amplitude damping channel is easily represented in the quantum process formalism using two Kraus operators,

$$\begin{aligned} \hat{K}_1 &= |0\rangle\langle 0| + \sqrt{\eta}|1\rangle\langle 1|, \\ \hat{K}_2 &= \sqrt{1-\eta}|0\rangle\langle 1|, \end{aligned} \quad (14.34)$$

where $0 \leq \eta \leq 1$ quantifies the degree of damping ($\eta = 0$ represents complete damping, and $\eta = 1$ represents the identity channel).

The physical intuition is clear upon inspection of the structure of the projectors in the Kraus operators, with \hat{K}_2 representing relaxation from the excited state to the ground state, with probability $1 - \eta$.

In the specific context of optics, the loss channel (Sec. XIV.C) is the equivalent of amplitude damping.

1. T_1 -times

The degree of amplitude damping is often quoted in terms of a channel's T_1 -time, characterising the expected time for the excited state to undergo spontaneous emission and relax to the ground state. Using this parameterisation we can express the amplitude damping channel as,

$$\mathcal{E}_t^{\text{relax}}(\hat{\rho}) = e^{-t/T_1}\hat{\rho} + (1 - e^{-t/T_1})|0\rangle\langle 0|, \quad (14.35)$$

for which the output state is of the form,

$$\hat{\rho}_t = \begin{pmatrix} 1 - (1 - \alpha)e^{-t/T_1} & \gamma e^{-t/T_1} \\ \gamma^* e^{-t/T_1} & \beta e^{-t/T_1} \end{pmatrix}. \quad (14.36)$$

G. Mode-mismatch

Mode-mismatch is an error model unique to optical implementations. For perfect interference to take place between two optical modes, which is necessary to entangle them or perform ideal ‘which-path erasure’²³, the photons

²³ Which-path erasure is the phenomenon whereby a beamsplitter interaction between two modes makes processes associated with those two modes indistinguishable, thereby projecting them into a superposition state of both possibilities. This is most commonly used to entangle distinct photon-emitting systems. This is discussed in detail in Sec. XLII.F.

in those modes must be perfectly indistinguishable, i.e they must exhibit identical spatio-temporal structure (?) and must be pure states.

This phenomenon arises very naturally whenever optical path-lengths are not perfectly aligned, or there is imperfect spatial mode-overlap between optical modes interfering at beamsplitters. Furthermore, even if optical networks are perfect, photon distinguishability may arise during state preparation, since no two photon sources are absolutely identical – engineering photon sources is a precise business and no two are ever exactly alike.

In real-world experiments, the most common form of mode-mismatch is temporal mode-mismatch, whereby the timing of different photons are not perfectly synchronised, yielding temporal distinguishability, thereby reduced quantum interference. This type of error is easily introduced via mismatched path lengths in an experiment, or incorrectly accounted for changes in refractive index. This is easily represented mathematically via translations in the temporal distribution functions (Sec. XIII.C) of photons,

$$\psi(t) \rightarrow \psi(t - \Delta_t), \quad (14.37)$$

for temporal mismatch Δ_t . Of course, this logically generalises to other degrees of freedom, such as spatial mode-mismatch, in which case a translation of the following form would take place,

$$\psi(x, y) \rightarrow \psi(x - \Delta_x, y - \Delta_y), \quad (14.38)$$

where x and y are the two transverse spatial dimensions perpendicular to the direction of propagation.

The Hong-Ou-Mandel (HOM) (?) *visibility* is a direct measure of the indistinguishability of two photons based on their interference fringes. Specifically, interference fringes are reduced as the photons become more distinguishable. Once completely distinguishable, they obey classical statistics.

Let us consider this in detail. Consider the two-mode, two-photon state,

$$|\psi_{\text{in}}\rangle = \hat{A}_{\psi_1}^\dagger \hat{B}_{\psi_2}^\dagger |0\rangle, \quad (14.39)$$

where \hat{A}^\dagger and \hat{B}^\dagger denote the mode operators for two spatial modes, with respective temporal distribution functions ψ_1 and ψ_2 . Evolving this through a 50:50 (Hadamard) beam splitter yields,

$$\begin{aligned} |\psi_{\text{out}}\rangle &= \hat{U}|\psi_{\text{in}}\rangle \\ &= \frac{1}{2} [\hat{A}_{\psi_1}^\dagger + \hat{B}_{\psi_1}^\dagger] [\hat{A}_{\psi_2}^\dagger - \hat{B}_{\psi_2}^\dagger] |0\rangle \\ &= \frac{1}{2} [\hat{A}_{\psi_1}^\dagger \hat{A}_{\psi_2}^\dagger - \hat{A}_{\psi_1}^\dagger \hat{B}_{\psi_2}^\dagger + \hat{A}_{\psi_2}^\dagger \hat{B}_{\psi_1}^\dagger - \hat{B}_{\psi_1}^\dagger \hat{B}_{\psi_2}^\dagger] |0\rangle. \end{aligned} \quad (14.40)$$

Post-selecting upon detecting a coincidence event (i.e one photon per mode), the conditional state is projected onto,

$$|\psi_{\text{cond}}\rangle = \frac{1}{2} [\hat{A}_{\psi_1}^\dagger \hat{B}_{\psi_2}^\dagger - \hat{A}_{\psi_2}^\dagger \hat{B}_{\psi_1}^\dagger] |0\rangle. \quad (14.41)$$

The probability of this coincidence event occurring is then given by the normalisation of the residual state,

$$\begin{aligned} P_{\text{coincidence}} &= |\langle \psi_{\text{cond}} | \psi_{\text{cond}} \rangle|^2 \\ &= \frac{1}{2} - \frac{1}{2} \left| \int_{-\infty}^{\infty} \psi_1(t) \psi_2^*(t) dt \right|^2. \end{aligned} \quad (14.42)$$

Now if we let both input photons have identical temporal structure, ψ , but with a time-delay τ between them, this reduces to,

$$P_{\text{coincidence}} = \frac{1}{2} - \frac{1}{2} \left| \int_{-\infty}^{\infty} \psi(t) \psi^*(t - \tau) dt \right|^2. \quad (14.43)$$

It is clear upon inspection that when $\tau = 0$, the coincidence probability $P_{\text{coincidence}} = 0$, and we observe perfect photon bunching at the output (quantum statistics). On the other hand, as $\tau \rightarrow \pm\infty$, the photons become completely distinguishable, and we reduce to classical statistics, whereby $P_{\text{coincidence}} = 1/2$. In the intermediate regime, there will be a monotonic tradeoff between distinguishability (determined by $|\tau|$) and the coincidence probability. As an example, if we let the temporal distribution function be a normal Gaussian distribution,

$$\psi(t) = \frac{1}{\sqrt[4]{2\pi}} e^{-\frac{t^2}{4}}, \quad (14.44)$$

then,

$$P_{\text{coincidence}} = \frac{1}{2} - \frac{1}{2} e^{-\frac{\tau^2}{8}}, \quad (14.45)$$

which is shown in Fig. 42. Thus, experimentally measuring $P_{\text{coincidence}}$ directly determines the degree of photon distinguishability.

In the above representation of mode-mismatch as a temporal or spatial translation, the process is entirely coherent, and could in principle be reversed if the translation were known (which might easily be established using tomographic characterisation techniques). Of course, such translations could occur incoherently also. In particular, ‘time-jitter’ is where this process occurs incoherently, and the photons are subject to probabilistic temporal displacements. In this instance, a pure single-photon state would evolve into a mixture of states subject to different displacements. Since the mode-mismatch is now probabilistic, it is not reversible. The state of a single photon subject to time-jitter would be of the form,

$$\hat{\rho}_{\text{jitter}} = \int_{-\infty}^{\infty} p_{\text{jitter}}(\Delta_t) |\psi - \Delta_t\rangle \langle \psi - \Delta_t| d\Delta_t, \quad (14.46)$$

where $p_{\text{jitter}}(\Delta_t)$ characterises the classical probability distribution of the temporal displacement. Time-jitter is particularly natural in heralded spontaneous parametric down-conversion (SPDC) sources (Sec. XIII.B), where imprecision in the measurement time of the heralding

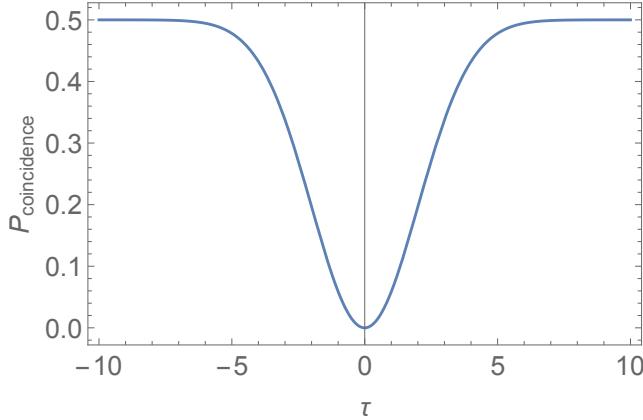


Figure 42: Hong-Ou-Mandel dip for two photons with normal Gaussian temporal distribution functions, and temporal offset τ between them. τ effectively characterises the degree of photon distinguishability, where $\tau = 0$ represents complete indistinguishability (quantum statistics), and $\tau \rightarrow \pm\infty$ represents complete distinguishability (classical statistics). Thus, performing this experiment and measuring $P_{\text{coincidence}}$ can be used to characterise the degree of photon distinguishability.

mode projects that temporal uncertainty onto the heralded state. For this reason, much time is being invested into engineering SPDC sources with separable output photons, such that pathological behaviour of the detection of the heralding photon does not project the heralded photon onto a mixed state. Time-jitter is a major consideration in all present-day single-photon source technologies.

When considering mode-mismatch, there are two general regimes for how it manifests itself in an optical system. The first is when the interference taking place is between distinct, independent photons, i.e HOM interference (or its equivalent generalisations to higher-photon-number systems). The second is when multiple paths followed by a given photon interfere it with itself, i.e Mach-Zehnder (MZ) interference. The former only requires mode-matching on the scale of the photons' wave-packets, whereas the latter requires interferometric stability on the order of the photons' wavelength, a far more demanding requirement. This is discussed in greater detail in Sec. XXII.

Mode-mismatch has been studied extensively in the context of linear optics quantum computing (LOQC), introduced in Sec. XLII.A. In particular, it was shown that in the cluster state formalism (Sec. XL.B), mode-mismatch in a fusion gate is equivalent to a dephasing error model, where the dephasing rate is related to the degree of photon distinguishability (i.e visibility) (?). More generally, the operation of entangling gates (????) and BOSONSAMPLING (Sec. XLII.D.1) (?) have been considered, and explicit error models derived.

H. Dispersion

Dispersion is the phenomenon of frequency-dependent velocity of light in a given medium. These effects can be very diverse, but can always be expressed in the mode-operator representation using an appropriate transformation in the temporal or spectral wave-function,

$$f_{\text{disp}} : \tilde{\psi}(\omega) \rightarrow \tilde{\psi}(\omega)' . \quad (14.47)$$

I. Spectral filtering

In Sec. XIV.C we discussed the loss channel, whereby with some fixed probability photons are lost to the environment. In reality, this process is often not uniform, but frequency-dependent, resulting in spectral filtering effects. For example, optical fibres are typically designed to operate with a particular optical frequency in mind, and will attenuate frequencies outside a given range, implementing, for example, low-pass, high-pass or band-pass spectral filtering.

Because spectral filtering can be regarded as frequency-dependent loss, it can be modelled in the same way as per the loss channel, but using a frequency-dependent beamsplitter with transmissivity $\eta_f(\omega)$, which models the frequency response of the channel. The model is shown in Fig. 43.

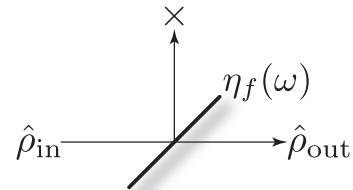


Figure 43: Model for the spectral filtering channel. The input state, $\hat{\rho}_{\text{in}}$, passes through a beamsplitter of frequency-dependent transmissivity $\eta_f(\omega)$, and the reflected mode discarded, yielding the lossy output state $\hat{\rho}_{\text{out}} = \mathcal{E}_{\eta_f}^{\text{filter}}(\hat{\rho}_{\text{in}})$.

This channel has the effect of modulating the spectral distribution function of a photonic mode operator \hat{A}_ψ^\dagger to $\hat{A}_{\psi'}^\dagger$, where,

$$\psi'(\omega) = \sqrt{\eta_f(\omega)} \psi(\omega) . \quad (14.48)$$

Note that unless $\eta_f(\omega) = 1 \forall \psi(\omega) \neq 0$, the new distribution function $\psi'(\omega)$ will not be normalised, where the normalisation reflects the loss probability,

$$p_{\text{loss}} = 1 - \int_{-\infty}^{\infty} \eta_f(\omega) |\psi(\omega)|^2 d\omega . \quad (14.49)$$

J. Phase-space

In Sec. XXV.B we introduce the displacement and squeezing operations, two non-linear operations which

are important ingredients in CV quantum information processing schemes. Of course, such processes are subject to errors.

In the case of the displacement operation, which is implemented by mixing a state with a coherent state on a beamsplitter, errors in the amplitude of the coherent state or in the beamsplitter reflectivity will introduce an offset in the displacement amplitude. Thus, instead of implementing $\hat{D}(\alpha)$, we might over- or under-displace the state, implementing,

$$\hat{D}(\Delta)\hat{D}(\alpha) \propto \hat{D}(\alpha + \Delta), \quad (14.50)$$

for some error Δ .

In the case of the squeezing operation, we might similarly have uncertainty in the squeezing parameter, thus implementing $\hat{S}(\xi + \Delta)$ instead of $\hat{S}(\xi)$.

— *Igne natura renovatur integra.*

XV. QUANTUM COST VECTOR ANALYSIS

As with the classical case in Sec. IX.B, there will be costs associated with the links and nodes in a network – nothing is free! In the quantum case, all the usual classical costs are valid, but there are some very important additions of far greater relevance to most quantum applications. Classical digital data is discretised, resulting in data transmission highly robust against noise. In a quantum setting this is necessarily not the case, as the coefficients in quantum superpositions are continuous, meaning that errors accumulate during transmission and states will inevitably deteriorate, unlike digital states. This requires a rethinking of appropriate cost metrics.

A. Costs

We now briefly introduce some of the key measures for quantifying the quality of quantum communications links, and how they may be expressed as metrics with meaningful operational interpretations. Many of the measures typically employed for characterising quantum systems are not true metrics (i.e costs), but in many cases can be converted to metrics, or used meaningfully as attributes instead.

1. Efficiency

The efficiency measure introduced previously is multiplicative, so for consecutive lossy channels the net efficiency is,

$$\eta_{\text{net}} = \prod_i \eta_i, \quad (15.1)$$

where η_i is the efficiency of the i th channel. Intuitively, this is simply telling us that if a photon passes through a channel with success probability η_1 , followed by another with η_2 , the total success probability is $\eta_1\eta_2$.

When employing single-photon encoding of qubits (e.g using the polarisation degree of freedom), there are three basis states of interest: a single photon horizontally polarised ($|H\rangle$); a single photon vertically polarised ($|V\rangle$); and, the vacuum state ($|0\rangle$). The effect of the loss channel on this type of state is to map $|H\rangle$ and $|V\rangle$ to $|0\rangle$ with probability $1 - \eta$, while doing nothing to $|0\rangle$. Note that because the loss process affects both logical basis states ($|H\rangle$ and $|V\rangle$) identically, its action is invariant under unitary operations in the logical (i.e polarisation) basis space.

2. Spectral filtering

Because spectral filtering can be regarded as a frequency-dependent loss channel, its associated cost can be treated in the same manner, except that rather than keeping track of a single efficiency η , we track a frequency response function $\eta_f(\omega)$, with the same multiplicative property (on a per-frequency basis),

$$\eta_f^{(\text{net})}(\omega) = \prod_i \eta_f^{(i)}(\omega). \quad (15.2)$$

If we are keeping track of the frequency response, the usual efficiency metric can be made redundant and absorbed into the frequency response function as a uniform response,

$$\eta_f(\omega) = \eta \forall \omega. \quad (15.3)$$

3. Decoherence

The dephasing and depolarising channels, given by Eqs. (14.18 & 14.31), also behave multiplicatively. If p_i is the probability that the state passing through the i th channel in series does not undergo the error process, then the probability of the state passing though the entire series without error is simply,

$$p_{\text{net}} = \prod_i p_i, \quad (15.4)$$

exhibiting the same multiplicative behaviour as the loss channel. The same observation applies to any of the other Pauli error channels.

4. Mode-mismatch

In Sec. XIV.G we introduced a simple model for temporal mode-mismatch as a displacement in the temporal

wave-function of photons propagating through a channel. Clearly, such a process is cumulative – a temporal displacement of Δ_1 followed by another of Δ_2 yields a net displacement of $\Delta_1 + \Delta_2$. Thus, for a chain of such channels we simply accumulate a net temporal displacement of,

$$\Delta_{\text{net}} = \sum_i \Delta_i. \quad (15.5)$$

For an incoherent mode-mismatching process, such as time-jitter, an upper bound on the accumulated mismatch may be obtained by summing the maximum temporal displacements at each step.

5. Distance measures

The fidelity of two states directly quantifies how close they are to one another in a geometric sense, i.e on the Bloch sphere (?), or, in the context of a state passing through a quantum channel, a measure of how well the state is preserved.

The fidelity between two states is defined as,

$$\mathcal{F}(\hat{\rho}_1, \hat{\rho}_2) = \text{tr} \left(\sqrt{\hat{\rho}_1^{1/2} \cdot \hat{\rho}_2 \cdot \hat{\rho}_1^{1/2}} \right), \quad (15.6)$$

where,

$$\begin{aligned} \mathcal{F}(\hat{\rho}_1, \hat{\rho}_2) &= \mathcal{F}(\hat{\rho}_2, \hat{\rho}_1), \\ 0 \leq \mathcal{F}(\hat{\rho}_1, \hat{\rho}_2) &\leq 1. \end{aligned} \quad (15.7)$$

$\mathcal{F}(\hat{\rho}_1, \hat{\rho}_2) = 1$ iff the states are equal, and $\mathcal{F}(\hat{\rho}_1, \hat{\rho}_2) = 0$ iff they are orthogonal. In the case where one of the states is a pure state, this simplifies to,

$$\mathcal{F}(\hat{\rho}_1, |\psi_2\rangle) = \langle \psi_2 | \hat{\rho}_1 | \psi_2 \rangle, \quad (15.8)$$

and when both states are pure to simply,

$$\mathcal{F}(|\psi_1\rangle, |\psi_2\rangle) = |\langle \psi_1 | \psi_2 \rangle|^2. \quad (15.9)$$

The fidelity is invariant under a common unitary applied to both states,

$$\mathcal{F}(\hat{\rho}_1, \hat{\rho}_2) = \mathcal{F}(\hat{U}\hat{\rho}_1\hat{U}^\dagger, \hat{U}\hat{\rho}_2\hat{U}^\dagger). \quad (15.10)$$

We define the fidelity of two processes, the process fidelity (?), to be the fidelity between two identical copies of a state that have been evolved under each of those processes, minimised over all possible states. That is, it provides a lower bound on the fidelity between identical states evolved under the two processes. In the context of networking, where quality must be guaranteed, this definition is more appropriate than, say, the average case fidelity. Specifically,

$$\mathcal{F}(\mathcal{E}_1, \mathcal{E}_2) = \text{tr} \left(\sqrt{\chi_1^{1/2} \cdot \chi_2 \cdot \chi_1^{1/2}} \right), \quad (15.11)$$

where χ_1 and χ_2 are the process matrices for \mathcal{E}_1 and \mathcal{E}_2 .

The fidelity of two processes is invariant under a common unitary applied to both channels before or after the process. Specifically,

$$\begin{aligned} \mathcal{F}(\mathcal{E}_1, \mathcal{E}_2) &= \mathcal{F}(\mathcal{E}_U \circ \mathcal{E}_1, \mathcal{E}_U \circ \mathcal{E}_2) \\ &= \mathcal{F}(\mathcal{E}_1 \circ \mathcal{E}_U, \mathcal{E}_2 \circ \mathcal{E}_U), \end{aligned} \quad (15.12)$$

where \mathcal{E}_U is an arbitrary unitary process.

In the special case of an identity channel, $\hat{\mathbb{1}}$, which is of special interest in many communications scenarios, we employ the shorthand,

$$\mathcal{F}(\mathcal{E}) = \mathcal{F}(\mathcal{E}, \hat{\mathbb{1}}) = \min_{\hat{\rho}} [\mathcal{F}(\hat{\rho}, \mathcal{E}(\hat{\rho}))]. \quad (15.13)$$

By definition $\mathcal{F}(\mathcal{E}) = 1$ iff $\mathcal{E} = \hat{\mathbb{1}}$.

A lower bound on the process fidelity of multiple processes in series is multiplicative,

$$\begin{aligned} \mathcal{F}(\mathcal{E}_2 \circ \mathcal{E}_1, \mathcal{E}_3) &\geq \mathcal{F}(\mathcal{E}_2, \mathcal{E}_3) \cdot \mathcal{F}(\mathcal{E}_1, \mathcal{E}_3), \\ \mathcal{F}(\mathcal{E}_2 \circ \mathcal{E}_1) &\geq \mathcal{F}(\mathcal{E}_2) \cdot \mathcal{F}(\mathcal{E}_1). \end{aligned} \quad (15.14)$$

Generalising to a sequence of n processes in series yields,

$$\mathcal{F}(\mathcal{E}_n \circ \dots \circ \mathcal{E}_1) \geq \prod_{i=1}^n \mathcal{F}(\mathcal{E}_i). \quad (15.15)$$

An alternate measure for the distance between two quantum states is the trace-norm distance, defined as,

$$\begin{aligned} D(\hat{\rho}_1, \hat{\rho}_2) &= \frac{1}{2} \|\hat{\rho}_1 - \hat{\rho}_2\|_1 \\ &= \frac{1}{2} \sum_i |\lambda_i|, \end{aligned} \quad (15.16)$$

where λ_i are the eigenvalues of $\hat{\rho}_1 - \hat{\rho}_2$. Like the fidelity, the trace-norm distance is invariant under unitary transformation. Furthermore, it is contractive under the action of quantum processes,

$$D(\mathcal{E}(\hat{\rho}_1), \mathcal{E}(\hat{\rho}_2)) \leq D(\hat{\rho}_1, \hat{\rho}_2). \quad (15.17)$$

The trace-norm distance relates to the fidelity according to the following bounds,

$$1 - F(\hat{\rho}_1, \hat{\rho}_2) \leq D(\hat{\rho}_1, \hat{\rho}_2) \leq \sqrt{1 - F(\hat{\rho}_1, \hat{\rho}_2)^2}. \quad (15.18)$$

6. Purity

The purity of a state that was initially pure quantifies how well quantum coherence was maintained during evolution, equivalently how well superpositions are maintained. The purity is defined as,

$$\mathcal{P}(\hat{\rho}) = \text{tr}(\hat{\rho}^2), \quad (15.19)$$

where,

$$\frac{1}{\dim(\hat{\rho})} \leq \mathcal{P}(\hat{\rho}) \leq 1. \quad (15.20)$$

We have $\mathcal{P}(\hat{\rho}) = 1$ iff $\hat{\rho} = |\psi\rangle\langle\psi|$ is a pure state, and $\mathcal{P}(\hat{\rho}) = 1/\dim(\hat{\rho})$ iff $\hat{\rho} = \hat{1}/\dim(\hat{\rho})$ is the maximally mixed state.

The purity is invariant under unitary operations,

$$\mathcal{P}(\hat{\rho}) = \mathcal{P}(\hat{U}\hat{\rho}\hat{U}^\dagger). \quad (15.21)$$

The purity of a process is defined analogously to the fidelity of a process,

$$\mathcal{P}(\mathcal{E}) = \text{tr}(\chi^2), \quad (15.22)$$

and as with the fidelity, a lower bound on the purity of multiple processes in series is multiplicative,

$$\mathcal{P}(\mathcal{E}_2 \circ \mathcal{E}_1) \geq \mathcal{P}(\mathcal{E}_2) \cdot \mathcal{P}(\mathcal{E}_1). \quad (15.23)$$

If the channel implements a unitary operation then necessarily $\mathcal{P}(\mathcal{E}) = 1$.

Like the process fidelity, the purity of a quantum process is invariant under unitary operations,

$$\begin{aligned} \mathcal{P}(\mathcal{E}) &= \mathcal{P}(\mathcal{E}_U \circ \mathcal{E}) \\ &= \mathcal{P}(\mathcal{E} \circ \mathcal{E}_U). \end{aligned} \quad (15.24)$$

Generalising to a sequence of n processes in series yields,

$$\mathcal{P}(\mathcal{E}_n \circ \dots \circ \mathcal{E}_1) \geq \prod_{i=1}^n \mathcal{P}(\mathcal{E}_i). \quad (15.25)$$

7. Entanglement

When distributing entanglement between separate nodes, metrics quantifying bipartite entanglement are relevant. For pure bipartite states $|\psi\rangle_{A,B}$, the purity of one of the reduced subsystems directly quantifies the degree of entanglement between them,

$$\begin{aligned} \mathcal{M}(|\psi\rangle_{A,B}) &= \mathcal{P}(\text{tr}_A(|\psi\rangle_{A,B})) \\ &= \mathcal{P}(\text{tr}_B(|\psi\rangle_{A,B})), \end{aligned} \quad (15.26)$$

The entanglement between two systems is invariant under local unitaries,

$$\mathcal{M}(|\psi\rangle_{A,B}) = \mathcal{M}([\hat{U}_A \otimes \hat{U}_B]|\psi\rangle_{A,B}). \quad (15.27)$$

8. Phase-space

Displacements in phase-space accumulate additively, up to a phase-factor. Specifically, the composition of two displacements is given by,

$$\hat{D}(\alpha)\hat{D}(\beta) = e^{\frac{1}{2}(\alpha\beta^* - \alpha^*\beta)}\hat{D}(\alpha + \beta). \quad (15.28)$$

Thus, the composition of a chain of unwanted or uncertain displacements yields, up to phase, a displacement with amplitude given by the sum of the individual displacement amplitudes.

Similarly, from the definition of the squeezing operator,

$$\hat{S}(\xi) = \exp\left[\frac{1}{2}(\xi^*\hat{a}^2 - \xi\hat{a}^{+2})\right], \quad (15.29)$$

it is evident that squeezing accumulates additively as well,

$$\hat{S}(\xi_1)\hat{S}(\xi_2) = \hat{S}(\xi_1 + \xi_2). \quad (15.30)$$

9. Channel capacity

The measures considered until now have quantified the preservation of quantum states. Alternately, one might consider information theoretic measures, which quantify the number of bits/qubits transmitted by a link, i.e the number of bits/qubits in common before and after the channel. This is an extremely powerful tool as it upper bounds the amount of information the receiver can extract from the transmitter under *any* measurement scheme, very useful in a cryptographic context, where we want security to be attack-independent (Sec. XXXII).

Suppose $\hat{\rho}_{A,B}$ is shared between Alice and Bob, from which they would like to extract maximal correlations (i.e joint information). They measure using local measurement bases $\{\Lambda_A^x\}$ and $\{\Lambda_B^x\}$, yielding random variables A and B . Recall from Sec. ?? that their quantum mutual information of the channel is defined as,

$$I_c(\hat{\rho}_{A,B}) = \max_{\rho, \{\Lambda_A^x\}, \{\Lambda_B^x\}} I(A; B). \quad (15.31)$$

We are then lead to the definition of the *classical channel capacity* of a quantum channel \mathcal{E} ,

$$\mathcal{C}(\mathcal{E}) = \lim_{n \rightarrow \infty} \frac{I_c(\mathcal{E}^{\otimes n})}{n}, \quad (15.32)$$

which can be interpreted as the maximum bitrate of the channel, per use of the channel, in the limit of an infinite number of uses, where entangled between multiple applications is allowed in general.

For a noisy channel, if the input contains product states as well as separable measurements, the classical capacity is equal to Eq. (15.31), that is, the single-use and the asymptotic classical capacities are equal.

If we also allow joint measurement (but not entanglement between input states), the classical channel capacity can be expressed as,

$$\chi(\mathcal{E}) = \max_{p_i, \hat{\rho}_i} S(\hat{\rho}) - \sum_i p_i S(\hat{\rho}_i), \quad (15.33)$$

which is known as the Holevo bound.

However, if entangled input states are allowed in conjunction with joint measurement setting, the Holevo

bound no longer holds, and the asymptotic classical channel capacity,

$$\chi(\mathcal{E}) < \mathcal{C}(\mathcal{E}). \quad (15.34)$$

It has been shown that using entangled input states in conjunction with joint measurements can increase the amount of classical information which can be transmitted over a noisy quantum channel. In this case,

$$C_{\text{ent}} = \lim_{n \rightarrow \infty} \frac{1}{n} \chi(\mathcal{E}^{\otimes n}). \quad (15.35)$$

The private classical capacity $\mathcal{P}(\mathcal{E})$ of a quantum channel \mathcal{E} describes the maximum rate at which the channel is able to send classical information through the channel reliably and privately, that is, the trusted parties do not want the environment to have access to their classical correlations.

Here Alice prepares the quantum state ρ_x^A according to the probability distribution $\{p_X(x)\}$. The expected density operator of the ensemble is of the form,

$$\rho^{XA} \equiv \sum_x p_X(x) |x\rangle\langle x| \otimes \rho_x^A. \quad (15.36)$$

The difference between the amount of classical correlation that she can establish with Bob, minus the correlation obtainable by Eve is given by,

$$I(X; B)_\rho - I(X; E)_\rho, \quad (15.37)$$

which leads to the definition of the *private information of a quantum channel*, where Alice maximises over all of her inputs,

$$I_{\mathcal{P}} = \max_{\rho^{XA}} \{I(X; B)_\rho - I(X; E)_\rho\}. \quad (15.38)$$

The private classical capacity is defined as,

$$\mathcal{P}(\mathcal{E}) = \lim_{n \rightarrow \infty} \frac{1}{n} I_{\mathcal{P}}(\mathcal{E}^{\otimes n}) \quad (15.39)$$

The last capacity in regards to classical communication over quantum channel is the *entanglement-assisted classical capacity*, $\mathcal{C}_E(\mathcal{E})$. This quantity measures the classical information that can be transmitted through the channel, if Alice and Bob possess shared entanglement prior to the transmission,

$$\mathcal{C}_E(\mathcal{E}) = \max_{p_i, \hat{\rho}_i} I(A : B). \quad (15.40)$$

The difference between $\mathcal{C}(\mathcal{E})$ and $\mathcal{C}_E(\mathcal{E})$ is that $\mathcal{C}_E(\mathcal{E})$ is equal to the maximised quantum mutual information (i.e additivity holds), and is equal to the single-use version of $\mathcal{C}_E(\mathcal{E})$. This implies that shared entanglement does not change the additivity of quantum mutual information.

a. *Quantum channel capacities* The quantum coherent information exhibits much of the same mathematical structure as the classical mutual information. And analogously,

$$\mathcal{I}_Q(\mathcal{E}) = \max_{\hat{\rho}} I_{\text{coh}}(\hat{\rho}, \mathcal{E}). \quad (15.41)$$

The *quantum channel capacity* is then analogously defined as,

$$Q(\mathcal{E}) = \lim_{n \rightarrow \infty} \frac{\mathcal{I}_Q(\mathcal{E}^{\otimes n})}{n}. \quad (15.42)$$

The most important distinction between quantum mutual information and quantum coherent information is that the mutual information is always additive, but the quantum coherent information is not.

Analytic solutions to Eq. (15.42) are notoriously difficult to calculate. However, once net dephasing or depolarisation rates have been calculated across a route, the *single-use* quantities can be calculated, which can serve as a lower bound to the ultimately achievable rates, if a number representing *cost*, rather than an analytic solution, is all we need.

A summary of the different measure of classical and quantum capacities are given in Tab. III.

We now discuss the quantum capacity for the erasure channel and the amplitude damping channel, both of which the quantum capacities can be calculated.

We recall that the erasure channel acts on a state as follows is,

$$\hat{\rho} \rightarrow (1 - \eta)\hat{\rho} + \eta|e\rangle\langle e| \quad (15.43)$$

where η is the probability of losing the state and $|e\rangle\langle e|$ is orthogonal to ρ . Given a system of dimension d_A , For $\eta \in [1/2, 1]$, the quantum capacity is zero. For $\eta \in [0, 1/2]$, the quantum capacity is $(1 - 2\eta)\log d_A$.

For an amplitude damping channel with damping parameter γ , the Kraus operators can be written as,

$$\begin{aligned} A_0 &= |0\rangle\langle 0| + \sqrt{1 - \gamma}|1\rangle\langle 1|, \\ A_1 &= \sqrt{\gamma}|1\rangle\langle 1|. \end{aligned} \quad (15.44)$$

For $\gamma \in [0, 1/2]$, the quantum capacity is equal to,

$$\max_{p \in [0, 1]} h_2[(1 - \gamma)p] - h_2[\gamma_p], \quad (15.45)$$

where h_2 represents the binary entropy function. For $\gamma \in [1/2, 1]$, the quantum capacity is equal to zero.

Two of the most surprising discoveries of quantum Shannon theory is the *superadditivity* of coherent information for the depolarising channel, and *superactivation* of quantum capacity, where two channels whose individual capacities are zero can be combined to make a channel with non-zero capacity.

Capacity	Type of information	Correlation measure	Capacity formula
Classical	Classical information	Holevo information	Eq. (15.33)
Private	Classical information	Private information	See (?)
Entanglement Assisted Classical	Classical information	Quantum mutual information	See (?)
Quantum	Quantum information	Quantum Coherent Information	Eq. (15.42)

Table III: Measure of classical and quantum channel capacities. Taken from (?).

1. Superadditivity of coherent information For qubits, the depolarising channel transmits its input with probability $(1 - p)$ and replaces it with the maximally mixed state $\hat{1}/2$ with probability p ,

$$\rho \rightarrow (1 - p)\rho + p\frac{\hat{1}}{2}. \quad (15.46)$$

It is known that its coherent information is strictly super-additive when p is large.

$$5I_Q(\mathcal{E}) < 5I_Q(\mathcal{E}^{\otimes 5}). \quad (15.47)$$

It can be shown that the state which maximises the coherent information is the Bell state $\frac{1}{\sqrt{2}}(|00\rangle_{AA_1} + |11\rangle_{AA_1})$, and the one-use coherent information is,

$$I_Q = 1 + \left(1 - \frac{3p}{4}\right) \log \left[1 - \frac{3p}{4}\right] + \frac{3p}{4} \log \left[\frac{p}{4}\right]. \quad (15.48)$$

At $p = 0.25$, Eq. (15.48) is equal to zero, but if we calculate the coherent information for a five-qubit repetition code,

$$\begin{aligned} \frac{1}{\sqrt{2}}(|00\rangle_{AA_1} + |11\rangle_{AA_1}) &\rightarrow \\ \frac{1}{\sqrt{2}}(|000000\rangle_{AA_1A_2A_3A_4A_5} + |111111\rangle_{AA_1A_2A_3A_4A_5}), \end{aligned} \quad (15.49)$$

at p just above 0.25, the quantity of interest is positive (?). This demonstrates that the quantum capacity of the depolarising channel is superadditive. Despite the fact that the channel looks simple, the best strategy for achieving the quantum capacity of the depolarising channel remains very poorly understood (?).

2. Superactivation Suppose that Alice is connected to Bob by two quantum channel $\mathcal{E}_1, \mathcal{E}_2$ both with zero quantum capacity. Intuitively one would expect that Alice would not be able to transmit quantum information over the channel $\mathcal{E}_1 \otimes \mathcal{E}_2$. However, examples of two zero-capacity channels are known to have non-zero joint quantum capacity. The phenomenon is known as *superactivation*.

One of the channels is a 50% erasure channel (which we know to have zero quantum capacity from the non-cloning theorem), and the other is known as an entanglement-binding channel (?).

This phenomenon has important implications for quantum information transmission. It implies that if there are other seemingly useless channels available, one could transmit more quantum information than if the channels were used alone. It also implies that the coherent information is strongly non-additive, and there is much to be understood about reliable communication rates over quantum channels.

10. Latency

Aside from the actual information content of a transmitted quantum state, the latency associated with its transmission is a key consideration in many time-critical applications.

By defining the latency of a link/node as the time between receipt of a quantum state and its retransmission, the total latency of a route is simply the sum of all the individual node and link latencies across the route,

$$\mathcal{L}(R) = \sum_{i \in R} \mathcal{L}_i, \quad (15.50)$$

where \mathcal{L}_i is the latency associated with the i th link in route R .

11. Dollars

Not to be overlooked is the actual dollar cost of communicating information. It is unlikely that Alice and Bob outright own the entire infrastructure of particular routes. Rather, different links and nodes are likely to be owned by different operators (particularly in ad hoc networks), who are most likely going to charge users for bandwidth in their network (quantum networks won't be cheap). Clearly dollar costs are additive over the links and nodes within routes,

$$\mathcal{C}(R) = \sum_{i \in R} \mathcal{C}_i, \quad (15.51)$$

where \mathcal{C}_i is the dollar cost of utilising the i th link in route R .

B. Costs as distance metrics

Def. 1 defines the properties of a cost metric in the classical context. We now wish to consider this in the quantum context.

If we consider a lossy photonic channel for example, efficiencies (η) are multiplicative – for a route $v_1 \rightarrow v_2 \rightarrow v_3$, the net efficiency is given by the product of the individual efficiencies,

$$\eta_{v_1 \rightarrow v_2 \rightarrow v_3} = \eta_{v_1 \rightarrow v_2} \eta_{v_2 \rightarrow v_3}. \quad (15.52)$$

This is multiplicative rather than additive, clearly not satisfying our definition for a cost metric. However, multiplicative metrics such as this can easily be made additive by shifting to a logarithmic scale, since

$$\log(\eta_{v_1 \rightarrow v_2 \rightarrow v_3}) = \log(\eta_{v_1 \rightarrow v_2}) + \log(\eta_{v_2 \rightarrow v_3}), \quad (15.53)$$

which now has a legitimate interpretation as a distance. The same applies to, for example, frequency response functions, which are equivalent to frequency-dependent loss.

In general, for a series of links $v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n$ characterised by multiplicative measure m , the equivalent cost metric is,

$$c_{v_1 \rightarrow v_2 \rightarrow \dots \rightarrow v_n} = - \sum_{i=1}^{n-1} \log(m_{v_i \rightarrow v_{i+1}}). \quad (15.54)$$

We have assumed that $0 \leq m \leq 1$, where $m = 0$ represents complete failure, and $m = 1$ represents ideal operation.

With these properties, the costs in our graph have an elegant interpretation. In the case of perfect operation, $m = 1$, the cost is $c = 0$, creating an ideal direct link between neighbouring nodes at no cost. On the other hand for complete failure, $m = 0$, the cost metric is $c = \infty$, effectively removing the link from the network and prohibiting pathfinding algorithms from following that route altogether.

Such a logarithmic scale is particularly convenient when a cost metric over links accumulates on a per physical distance basis, in which case the cost metric is simply the physical length of the link multiplied by the metric per unit distance. For example, if a fibre channel implements loss at 3dB/km, the loss over 10km is $10 \times 3\text{dB}$.

Note that lower bounds on fidelity, purity, efficiency and dephasing are all multiplicative on a scale of 0 to 1, and thus their logarithms may be regarded as cost metrics. Spatio-temporal mode-mismatch, latency, dollar cost and displacements are clearly automatically metrics as they are additive.

A dephasing channel can be easily converted to a distance metric as follows. First we reparameterise the dephasing channel into,

$$\begin{aligned} \mathcal{E}(\hat{\rho}) &= p\hat{\rho} + (1-p)\hat{Z}\hat{\rho}\hat{Z} \\ &= (2p-1)\hat{\rho} + (1-p)(\hat{Z}\hat{\rho}\hat{Z} + \hat{\rho}). \end{aligned} \quad (15.55)$$

Now $2p-1$ is the probability that the state is not dephased and $1-p$ is the probability that the state is replaced with the completely dephased state. Therefore

the probability of multiple applications of the channel ($\mathcal{E}_n \circ \dots \circ \mathcal{E}_1$) not dephasing the state scales multiplicatively as²⁴,

$$p_{\text{no error}} = \prod_i (2p_i - 1), \quad (15.56)$$

which is additive in a logarithmic scale as before,

$$\log(p_{\text{no error}}) = \sum_i \log(2p_i - 1), \quad (15.57)$$

which acts as a distance metric. This approach can similarly be applied to other Pauli channels.

In the case of mutual information and channel capacity, it makes most sense to consider the number of bits that are lost by a channel, rather than the number communicated, since then we have a measure with quasi-metric properties. Specifically, let the number of bits lost by a channel be the difference between the number of bits in the input state and the channel capacity,

$$B_{\text{lost}}(\mathcal{E}, \hat{\rho}) = S(\hat{\rho}) - C(\mathcal{E}). \quad (15.58)$$

Then there are two cases to consider – upper and lower bounds on accumulated lost bits.

The best-case scenario is that subsequent channels lose the same bits, giving us a lower bound on the number of lost bits as the maximum number of bits lost by the constituent channels,

$$B_{\text{lower}}(\mathcal{E}_2 \circ \mathcal{E}_1, \hat{\rho}) = \max[B_{\text{lost}}(\mathcal{E}_1, \hat{\rho}), B_{\text{lost}}(\mathcal{E}_2, \hat{\rho})]. \quad (15.59)$$

Alternately, each subsequent channel could lose a different set of bits, in which case the number of lost bits accumulates additively,

$$B_{\text{upper}}(\mathcal{E}_2 \circ \mathcal{E}_1, \hat{\rho}) = B_{\text{lost}}(\mathcal{E}_1, \hat{\rho}) + B_{\text{lost}}(\mathcal{E}_2, \hat{\rho}). \quad (15.60)$$

Then, the number of actual bits lost is bounded from above and below as,

$$B_{\text{lower}} \leq B_{\text{lost}} \leq B_{\text{upper}}. \quad (15.61)$$

C. Non-trivial node operations

Thus far we have considered how to accumulate cost metrics across routes through a network, where each link is subject to some quantum process obeying our notion of a cost metric. But what happens when the links are

²⁴ With this parameterisation, in the limit of many applications of the dephasing channel, an input state asymptotes to the completely dephased state, $\lim_{n \rightarrow \infty} \mathcal{E}^n(\hat{\rho}) = \frac{1}{2}(\hat{Z}\hat{\rho}\hat{Z} + \hat{\rho})$. Thus, $2p-1$ can be regarded as a discretised parameterisation of a system's T_2 -time.

interspersed with nodes that may be doing more than just simple switching?

A more general scenario to consider is where the nodes are not restricted to routing, but can additionally implement arbitrary unitary operations. This substantially broadens the class of networks under consideration, to encompass nodes capable of doing everything from straightforward routing to entire quantum computations.

All of the examples for cost metrics we introduced in Sec. XV have the property that they are invariant under unitary operations. Therefore the costs along a route may simply be accumulated as before, summing up the edge weights, without needing any special treatment for node operations, provided they are unitary. For non-unitary node processes, we can merge them into their neighbouring link processes as before (see Fig. 28).

D. Negative cost vectors

When we initially introduced cost vector analysis in the classical context (Sec. IX.B) we insisted that costs be positive by definition. However, in the quantum scenario we will loosen this demand since negative costs arise quite naturally in the context of operations that *improve* quantum data. Specifically this arises naturally when nodes implement operations such as entanglement purification or quantum error correction, to be discussed in detail in Sec. XVII. In that case making what would otherwise be a routing detour can yield net benefit, and so the cost vector analysis must take these negative costs into consideration and give them the favourable treatment they deserve.

— *Flectere si nequeo superos, acheronta movebo.*

XVI. QUANTUM TRANSMISSION CONTROL PROTOCOL (QTCP)

IN the classical world, TCP is employed for data transmission and routing. Next we present a simple toy model for a proposed quantum equivalent – the Quantum Transmission Control Protocol (QTCP). The stack is described in detail in Sec. XVI.A.

We emphasise that this toy model is not intended to be a proposal suited to immediate implementation, solving all the problems of quantum communication in the most effective way. Rather, we simply aim to construct a sketch of the data structures and algorithms that might form a basis for future, more well-considered real-world implementations. Alternately, it is plausible that a QTCP protocol may never reach the light of day at all, instead being made redundant by networks built entirely on entanglement distribution, discussed in detail in Secs. XXIX & XXVIII. The answer to this question is difficult to foresee, largely depending on the future requirements of

quantum communication protocols.

The goal of QTCP is to abstract away the low-level physical operation of a quantum network to create a virtual interface between Alice and Bob, allowing direct access to data as if it were held locally, in much the same way that high-level services like classical FTP facilitate interaction with remote data as though it were a local asset, blind to the intermediate networking.

The design goals of our elementary toy model are simply to capture the quintessential feature requirements of real-world protocols, and a sketch for their implementation. The designs we present should not be interpreted as a final proposal, but merely as laying a foundation of ideas to build upon.

We consider the scenario where Alice (or a set of Alices) is in possession of some quantum state, which she wishes to communicate to Bob (Bobs), with the aim of optimising some arbitrary cost measure. Bob is no guru and doesn't want to concern himself with how the state was communicated from Alice to himself – his only concern is that he receives it and that it satisfies quality constraints he and Alice have agreed upon.

QTCP is the joint software/hardware stack that facilitates these objectives. QTCP begins by logically separating different levels of network functionality into distinct layers of abstraction. This includes primarily:

- Encapsulation of data into packets of quantum information (Secs. XVI.A.1 & XVI.A.2).
- Cost vector analysis (Sec. IX.B).
- Routing decisions (Secs. XVI.A.3 & XIX).
- Reconstruction of communicated quantum states upon receipt (Sec. XVI.A.5).
- Enforcement of quality of service requirements and error correction (Sec. XVI.A.5).
- Providing a high-level virtual interface between end-users, which abstracts away low-level operations (Sec. XVI.A.7).

All the while, Alice and Bob, as end-users, ought to be as blind as possible to the lower-level layers, instead only directly interfacing with the highest layer of abstraction, that which provides the virtual interface between end-users.

— *O derint dum metuant.*

A. QTCP protocol stack

As with classical networking, our protocols for quantum networks will be separated into distinct layers, each performing a specific set of tasks with different levels of abstraction.

The structure of the protocol stack for QTCP is shown in Fig. 44. In summary, the layers in the protocol stack are, beginning from the lowest level:

- DATA (MESSAGE): Raw data ('payload') Alice wishes to transmit to Bob. Comprises both QUANTUM DATA and CLASSICAL DATA.
- PACKET: Decomposition of DATA into blocks (PACKET DATA), and associated classical PACKET HEADERS containing metadata (e.g routing information).
- STRATEGY: Construct PACKET routing strategies based on a cost optimisation algorithm.
- TRANSPORT: Physical routing of PACKETS to arrive at their destination, based upon metadata contained in PACKET HEADER. Perform collision detection during transit.
- RECONSTRUCTION: Reconstruct DATA from received PACKETS.
- QUALITY OF SERVICE (QoS): Apply QEC and determine whether QoS requirements have been satisfied.
- SERVICES & APPLICATIONS: High-level interface to DATA presented to Bob's services and applications. The interface abstracts away lower levels of the protocol stack, presenting Bob with only DATA and its associated metadata.

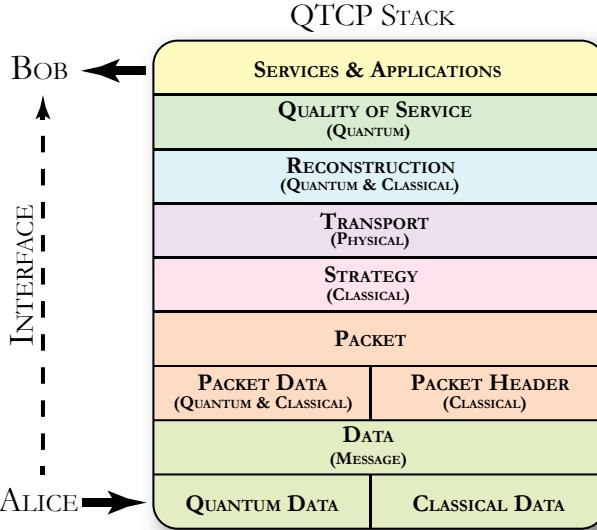


Figure 44: Protocol stack for QTCP. The protocol stack mediates communication of quantum data from Alice to Bob using the shown layers of abstraction. The end goal is to provide Bob a virtual interface to Alice's transmitted data, while remaining oblivious to the underlying protocol.

Next we describe the operation of these layers in detail.

1. Data (Message)

At the lowest level of the protocol we have the raw DATA Alice wishes to communicate to Bob. DATA is allowed to comprise both QUANTUM DATA and CLASSICAL DATA components, and may contain one or the other, or both.

a. *Quantum data* QUANTUM DATA is allowed to be an arbitrary quantum state. It could be a pure or mixed state, of arbitrary (but predetermined) dimension, or even a subsystem of a larger external state (i.e entangled with another system). We stress that it needn't be expressed using a conventional qubit representation, in the way digital data is necessarily represented using bits. Keep in mind that the quantum internet isn't just there to communicate qubit data streams. Rather, it is intended to act as generally as possible, such that essentially arbitrary *quantum assets* can be exchanged. These needn't be restricted to any particular type of encoding, such as those discussed in Sec. XIII. For example, in addition to something 'standard' like polarisation-encoded qubits in single photons, one network user might like to share an exotic CV state of light, like a cat state, with his mate whose cat died. Indeed, multiple types of state encoding might be encapsulated within a single PACKET. The QTCP acts only as an abstract interface for quantum networking, but is completely blind as to what the underlying data in the network is. QTCP is only concerned with getting that state from Alice to Bob.

b. *Classical data* CLASSICAL DATA is a purely classical state with no coherence (i.e a diagonal density matrix), which may be represented as a classical bit-string. We very intentionally segregate the CLASSICAL and QUANTUM components of DATA, since the classical network is expected to be cheaper and more reliable than the quantum network operating in parallel to it. The CLASSICAL DATA could, for example, provide nodes with classical instructions on what quantum computations to perform on the QUANTUM DATA.

2. Packet

DATA is transmitted as PACKETS, much in the same way as conventional TCP. The DATA is decomposed into three components: QUANTUM DATA, CLASSICAL DATA, and PACKET HEADER.

We can express the state of an entire PACKET as,

$$\hat{\rho}_{\text{packet}}(i) = \hat{\rho}_{\text{quantum}}(i) \oplus \hat{\rho}_{\text{classical}}(i) \oplus \hat{\rho}_{\text{header}}(i), \quad (16.1)$$

where i denotes the i th packet. Here $\hat{\rho}_{\text{quantum}}$ ($\hat{\rho}_{\text{classical}}$) is a block of QUANTUM BLOCK SIZE qubits (CLASSICAL

BLOCK SIZE bits) taken from the user's QUANTUM DATA (CLASSICAL DATA), while $\hat{\rho}_{\text{header}}$ is the PACKET's classical PACKET HEADER. As discussed earlier, since $\hat{\rho}_{\text{quantum}}$ is quantum, and $\hat{\rho}_{\text{classical}}$ and $\hat{\rho}_{\text{header}}$ are classical, they needn't be transmitted together over the same quantum network. Instead all CLASSICAL DATA and PACKET HEADER could be transmitted over a classical network operating in parallel to and synchronised with the quantum network, which carries the QUANTUM DATA.

Note that while one can always measure classical data without disturbance, this is not the case with quantum data, where measurements cause wave-function collapse. Thus, while Alice is always able to know $\hat{\rho}_{\text{classical}}$ and $\hat{\rho}_{\text{header}}$, she may or may not know $\hat{\rho}_{\text{quantum}}$. Clearly if she prepared the state herself, she would (hopefully) know what she was doing. But in general, quantum networks could be used for far less trivial networking, where Alice is, for example, an intermediary in a distributed quantum computation. In this instance, Alice is unlikely to know what her QUANTUM DATA is.

a. *Packet data* Comprises blocks of both QUANTUM DATA and CLASSICAL DATA, of sizes QUANTUM BLOCK SIZE and CLASSICAL BLOCK SIZE respectively. PACKET DATA requires that DATA be decomposed into distinct subsystems which are independently transmitted by QTCP. For ease of exposition, we will restrict ourselves to the case where the DATA is encoded into a stream of qubits (QUANTUM DATA) and bits (CLASSICAL DATA). But of course other encodings could be used. Also bear in mind that any quantum (classical) information can be encoded into qubits (bits), and once represented as such, the decomposition of DATA into PACKETS arises very naturally and intuitively. The PACKET's QUANTUM DATA is what is transmitted via the quantum channels, while the CLASSICAL DATA is communicated via classical channels.

b. *Packet header* PACKET HEADER is purely classical and needn't be transmitted over the costly quantum network, instead being transmitted over a complementary classical network, running in parallel to, and synchronised with the quantum network. PACKET HEADER contains no information content from DATA, instead comprising only metadata relevant to the higher levels of the protocol stack. In particular, PACKET HEADER contains the following fields:

- HEADER SIZE: The number of bits in the PACKET HEADER.
- MESSAGE ID: A unique identifier for the complete message to which this PACKET belongs. This field mitigates ambiguity as to which MESSAGE this packet belongs when performing RECONSTRUCTION.

- LIFETIME: How long the PACKET has been in existence for, i.e since it was initially sent by the SENDER. This is used by strategies to prevent collisions.
- SENDER: A unique node identifier for the sender (Alice).
- RECIPIENT: A unique node identifier for the recipient (Bob).
- ORDER: To which block taken from DATA does this PACKET DATA belong? This is extremely important in networks where PACKETS may arrive out of order. The ORDER field forms the basis for the RECONSTRUCTION layer.
- QUANTUM BLOCK SIZE: The number of qubits contained in the QUANTUM component of PACKET DATA. This is important for the RECONSTRUCTION layer.
- CLASSICAL BLOCK SIZE: The number of bits contained in the CLASSICAL component of PACKET DATA, also important for the RECONSTRUCTION layer.
- ROUTING QUEUE: A first in, first out (FIFO) queue of node identifiers, tracing out the entire route for the PACKET to follow, in chronological order from the next node to visit all the way to the RECIPIENT.
- COSTS: A tuple characterising all the accumulated costs of the PACKET at the current stage in the route. These are treated as accumulators that are incremented appropriately after each step, since costs are additive.
- ATTRIBUTES: A tuple characterising all the non-COST properties associated with the PACKET. Examples include: the PRIORITY of routing a PACKET to its destination; suggesting a preferred routing STRATEGY; or, indicating whether or not a RESEND UNTIL SUCCESS protocol may be applied to this PACKET.
- PADDING: Null data to pad the joint CLASSICAL PACKET DATA and PACKET HEADER fields to be of the same length as the QUANTUM PACKET DATA. This ensures that the components of the PACKET traversing the quantum and classical channels remain in perfect tandem – bit for qubit. In Sec. XVI.A.4 we show that this facilitates collision detection without the need to measure quantum states.
- CHECKSUM: A regular checksum of the entire CLASSICAL component of the PACKET, including both CLASSICAL PACKET DATA and PACKET HEADER.

This also forms a part of the collision detection protocol.

One might question why PACKET HEADERS tally accumulated costs when we ought to already know all the costs, since these were employed by the algorithm for choosing strategies in the first case. In the ideal case where all strategies are determined *a priori* and are implemented as intended, this is certainly valid. However, for generality we retain this option since more realistic networks may require dynamically updating strategies during the course of propagation, in which case dynamically tallying costs is appropriate.

3. Strategy

Based on the PACKET HEADERS of all users sharing the network, choose routing strategies to optimise cost metrics. The notion of strategies is introduced in Sec. IX.E, and a detailed discussion of example strategies is presented in Sec. XIX.

Once routings have been determined for all PACKETS, the ROUTING QUEUES in their PACKET HEADERS are initialised accordingly by pushing the sequence of node identifiers tracing out the desired route.

In the case of dynamic, time-dependent strategies, which can be updated within the duration of transmissions, the ROUTING QUEUES may need to be updated. A change in a PACKET’s route simply requires flushing the queue and pushing new node identifiers for each of the nodes in the new route, in chronological order.

The STRATEGY layer is responsible for evaluating the net cost function f_{cost} from Eq. (9.3), which accounts for both COSTS and ATTRIBUTES to calculate a single effective cost measure that may be employed in routing decisions.

4. Transport

The TRANSPORT layer is responsible for actual routing at the physical level, making direct decisions as to what to do with a PACKET at each step, based upon the metadata contained in PACKET HEADER, most notably the ROUTING QUEUE, which specifies the full route a PACKET is destined to follow. It is also responsible for keeping track of costs that accumulate over their route.

Additionally, the TRANSPORT layer is responsible for collision detection, whereby multiple packets being transmitted simultaneously over a network interfere with one another, corrupting the data. In classical networking, collision detection is straightforward using checksums. But the usual classical approach breaks down in the quantum setting. In Sec. XVI.B we discuss in detail collision detection in QTCP.

The pseudo-code algorithm implemented by the TRANSPORT layer, including collision detection, is shown in Alg. III.

```

function Transport(Packet):
    1. nextNode = Packet.RoutingQueue.Pop()
    2. Packet.PhysicallySendTo(nextNode)
    3. Packet.WaitUntilArrivesAt(nextNode)
    4. checksum = Hash(Packet.Header +
        Packet.ClassicalData)
    5. if(checksum ≠ Packet.Header.Checksum) {
        6.   Packet.Sender.Notify(FAILURE)
        7.   Packet.Recipient.Notify(FAILURE)
        8.   Packet.Discard()
        9.   □
    10. }
    11. Packet.Costs += IncomingLink.Costs
    12. Packet.Attributes.Update()
    13. if(Packet.RoutingQueue.Length = 0) {
        14.   Return(Packet)
        15. }
    16. □

```

Algorithm III: Algorithm implemented by the TRANSPORT layer of QTCP for each PACKET. The `Attributes.Update()` function is left undefined. This is where arbitrary ATTRIBUTE dynamics may take place.

5. Reconstruction

The RECONSTRUCTION layer only serves one purpose – to chronologically reorder the received PACKETS based on the ORDER field in their PACKET HEADERS. This stage is only performed by Bob – the final recipient – and not at any intermediate stage. In general this will require Bob to have a quantum memory, able to hold all PACKET DATA for a sufficient duration as to enable an arbitrary permutation of PACKETS to be applied, reproducing the correct chronological order. The algorithm for this is shown in Alg. IV.

```

function Reconstruction(Packets):
    1. Packets.WaitUntilAllReceived()
    2. message =
        Packets.SortByOrderAscending().data
    3. Packets.Receiver.Notify(message)
    4. □

```

Algorithm IV: The goal of the RECONSTRUCTION layer, is to take a collection of received PACKETS and reassemble them into the MESSAGE.

6. Quality of service

In classical networking theory, error detection and correction is an important element of networking protocols. Communication links may be unreliable, or subject to external noise, which users must be able to detect so as to guarantee the quality of their data.

Classically, error detection is typically performed using checksums (hash functions), which generate a short digest of a packet's data that can be recalculated upon arrival to verify integrity. The checksum can be included in the header component of each packet, allowing the remainder of the protocol to remain unchanged.

In the quantum context the elegant notion of checksums is complicated by the fact that calculating a hash function of a quantum state would necessarily entangle the state with the output hash. This would have the undesired effect of causing measurement of the checksum to collapse the quantum state of the data, thereby altering it in an uncontrollable way.

As an alternative to checksums, we could borrow the notion of quantum error correction (QEC) and fault-tolerance from quantum computing theory. Here we encode a quantum state into a (polynomially) larger Hilbert space. *Syndrome measurements* on some of the states in this larger space allow us to both detect and correct universal error models, such as depolarisation or dephasing, provided that error rates are within the fault-tolerance threshold of the code being employed.

Thus, enforcing QoS is fundamentally different for quantum data packets than for classical ones, requiring entirely different techniques. Given how large a field this has become in its own right, we dedicate Sec. XVII entirely to protocols suited to the implementation of quantum QoS requirements in quantum networks.

7. Services & applications

Having communicated all the PACKET DATA from Alice to Bob, performed RECONSTRUCTION, and applied QoS protocols, Bob ought to have $\hat{\rho}_{\text{data}}$ to a good approximation. The quality of Bob's received state can be inferred directly from the COSTS vector contained in the PACKET HEADERS. The final state and its associated quality metrics (COSTS and QEC outcomes) may then be provided to Bob as a software interface for end use.

— *Divide et impera.*

B. Collision handling & classical errors

In classical networking, protocols such as Ethernet allow users to simply broadcast data at their leisure and rely on *collision detection* to detect when the broadcasts of multiple users have interfered, signalling that both users

ought to retransmit following backoff, to minimise the chances of another collision occurring. While this RESEND UNTIL SUCCESS approach has certainly proven to be effective in classical networking, in a quantum setting the rules of the game are entirely different.

First, collision detection necessarily requires measuring a communications channel to test whether data has been corrupted. Classical networks typically do this by transmitting a checksum with the data, which is recalculated upon arrival for comparison. This raises the obvious problem that quantum measurements are destructive, which means that testing the integrity of our data destroys it in the process – the last feature we'd like our network to exhibit! However, to overcome this, in Sec. XVI.A.4 we describe a protocol based on the dual classical/quantum network that allows collision detection without measuring quantum states.

Second, collision detection is not always even allowed at all. If one of Alice's packets was entangled with another (i.e she was communicating an entangled system, where the different subsystems resided in different packets), she would not be able to simply retransmit an identical copy of the corrupted packet, since the entanglement with the other system would have been lost and there are no local operations she can do to recover it.

Alternately, Alice might be a part of a distributed computation, where she didn't prepare the data in the first place. In this instance, the no-cloning theorem²⁵ implies that she cannot, in general, learn what the quantum state was, and therefore would be unable to make a second transmission attempt.

From Alice's point of view, RESEND UNTIL SUCCESS would clearly work if she was preparing a known state, separable from the other packets. However, collisions on the network caused by her reckless resending would likely corrupt the communications between other parties, leaving them rather ticked off at her.

²⁵ The no-cloning theorem states that it is impossible to take a completely general, unknown quantum state and prepare an identical copy of it. That is, there is no quantum process implementing the map $\mathcal{E}(\hat{\rho} \otimes |0\rangle\langle 0|) \rightarrow \hat{\rho} \otimes \hat{\rho}$ for unknown $\hat{\rho}$. Note that this result applies when $\hat{\rho}$ is arbitrary and unknown, but in general needn't apply if restricted to a particular basis, subject to other constraints, or we have *a priori* knowledge of the state. *Proof:* To see this, consider the unitary evolution $\hat{U}|\psi\rangle|e\rangle = e^{i\alpha(\psi,e)}|\psi\rangle|\psi\rangle$, for some normalised ancillary state $|e\rangle$, where α denotes an irrelevant global phase factor. We have $\langle\psi|\phi\rangle\langle e|e\rangle = \langle\psi|\langle e|\hat{U}^\dagger\hat{U}|\phi\rangle|e\rangle = e^{-i[\alpha(\psi,e)-\alpha(\phi,e)]}\langle\psi|\phi\rangle^2$. Since $|e\rangle$ is normalised, this implies $|\langle\psi|\phi\rangle| = |\langle\psi|\phi\rangle|^2$, which can only be true if $|\langle\psi|\phi\rangle| = 0$ or $|\langle\psi|\phi\rangle| = 1$, which does not hold in general for arbitrary states $|\psi\rangle$ and $|\phi\rangle$. However, clearly if $|\psi\rangle$ and $|\phi\rangle$ are restricted to being equal or orthogonal then this constraint can be satisfied with an appropriate choice of \hat{U} , since this corresponds to classical cloning. This proof applies to unitary evolution. What about generic quantum processes? In that case we merely need to assume that the ancillary state $|e\rangle$ acts on a joint primary/environment system, and the same argument applies.

There are therefore two main approaches to dealing with collisions. First, central planning of all routing could be employed, precisely scheduling all routes *a priori* so as to entirely eliminate any possibility for collisions. Second, if all users in the network were communicating data where packet loss could be tolerated, they could all mutually agree to use the RESEND UNTIL SUCCESS protocol. This would not require a central authority, and be highly desirable for ad hoc networks. It is important to stress, however, that the latter requires unanimity amongst network participants to function, and the restriction to known, separable states is a major limitation that would prohibit many important uses for quantum networks, such as distributed quantum computation. However, both these approaches are entirely valid in their appropriate context.

In classical TCP all components of data packets are classical and are kept together throughout every stage of transmission. In the quantum case we will instead have a mixture of both quantum and classical data. As mentioned, we will assume that classical communication and computation resources ‘come for free’ (or are at least cheap compared to quantum resources), so there will be a clear disambiguation as to what data is quantum or classical within packets.

As discussed, quantum collision detection is complicated by the fact that measuring quantum data to determine whether it has been corrupted disturbs the quantum state. We address this problem by taking advantage of the duality of the quantum/classical network, discussed in Part. IV. Because the quantum and classical components of the PACKETS are synchronised and of equal length (thanks to the PADDING field of the PACKET HEADER), and because the same applies to all other PACKETS on the network, a collision in the quantum data necessarily implies a collision in the classical data, and vice versa. Therefore, by applying regular classical collision detection techniques based on checksums (recall PACKET HEADER contains a CHECKSUM field), we can infer collisions in quantum data without actually measuring it. We refer to this as *indirect collision detection*. This guarantees us the ability to detect when a collision has occurred, in which case both quantum *and* classical data are corrupted, or has not occurred, in which case both quantum and classical data are uncorrupted and the quantum data remains unmeasured. Collision detection is incorporated into the pseudo-code implementation of the TRANSPORT layer shown in Alg. III.

This algorithm could, depending upon implementation, be executed locally on the node currently hosting the PACKET, or it could be delegated to a central authority, but with the overhead of additional classical communication.

In instances where corruption or loss of packets cannot be tolerated, a more proactive approach may be applied. To preempt the risk of packet collision, one could introduce *probe packets* – packets containing only classical data,

that query a route ahead to negotiate channel usage for the following proper packet, thereby avoiding collisions. Of course, if an upcoming node is unable to guarantee channel capacity immediately, the packet may need to be stored in quantum memory until the channel is available. Thus, it is important to accommodate for this by ensuring that quantum memory is available in nodes preceding links/nodes where collisions are not guaranteed to be mitigated immediately. Quantum memory will be discussed in Sec. XXVI.

— *Esse est percipi.*

C. Multi-packet operations

To do

Thus far we have considered nodes which implement single-packet operations. However, many real-world quantum protocols (to be presented in Part. V) will require multi-packet operations, which must be carefully combined in some well-defined way. For example, for state teleportation (Sec. XXVII.C) we must combine a data packet with half of a Bell pair (Sec. XXIII.G). For entanglement swapping (Sec. XXVII.E) – the essential component in quantum repeater networks (Sec. XXIX) – two halves of distinct Bell pairs undergo a joint entangling measurement.

To accommodate such multi-packet operations, we must add some modifications to the QTCP protocol to handle:

- Packet synchronisation: multiple packets subject to joint operations must be time-synchronised upon implementation of the joint operation. This might involve holding one packet in a quantum memory whilst waiting upon another. Thus, the time required to implement the joint operation is bottlenecked by the packet slowest to arrive.
- Bookkeeping of costs and attributes: when multiple packets are combined under some operation, how do error costs and attributes accumulate?
- Partner packet tracking: packets must keep track of the partner packets with which they are intended to undergo joint operations with. This may be achieved with the addition of PARTNERS fields in the packet headers, containing an array of their MESSAGE ID’s.
- Operation signalling: the recipient node of multiple packets undergoing a joint operation must be informed as to what joint operations should be applied. This may be implemented via the addition of a WHAT OPERATION field to the header.

At a high level, these may be implemented in the TRANSPORT layer as outlined in Alg. V, where we have intentionally left the functions for combining costs and attributes unspecified. In many instances, the

`CombineCosts()` function could be as simple as a summation of its arguments.

```
function Transport.MultiPacketOperation(packets):
    1. for(packet∈packets) {
        2. packet.WaitFor()
        3. packet.HoldInMemory()
    4. }

    5. output = SomeOperation(packets.WhatOperation,
        packets)

    6. if(output.OperationSuccess = true) {
        7. packets.Sender.Notify(SUCCESS)
        8. output.Costs =
            CombineCosts(packets.Costs)
        9. output.Attributes =
            CombineAttributes(packets.Attributes)
    10. return(output)
    11. } else {
        12. packets.Sender.Notify(FAILURE)
    13. }
    14. □
```

Algorithm V: High-level structure of transport layer implementation of multi-packet operations.

1. Combining costs & attributes

Let us now turn our attention to the details of how the `CombineCosts()` and `CombineAttributes()` functions might be implemented. In Sec. XV we presented a number of examples of prominent costs and attributes that might arise in quantum networks.

To illustrate the basic idea behind this, we now turn our attention to several examples for combining costs and attributes. Some details may vary between specific quantum protocols, although the following typically apply in most general cases.

These ideas logically extend all the way from simple state teleportation to more complicated multi-packet protocols, all the way up to fully distributed quantum computations involving large numbers of qubits and nodes (Sec. XLIII.B).

To do

a. Efficiency To do

b. Decoherence To do

c. Latency In state teleportation there are three channels side-by-side, each exhibiting their own distinct latencies.

```
function CombineCosts.Efficiency(packets):
    1. costs = packets.Costs.Efficiency
    2. netCost = sum(costs)
    3. return(netCost)
    4. □
```

Algorithm VI: Algorithm for combining efficiency metrics in a multi-packet protocol.

```
function CombineCosts.Decoherence(packets):
    1. maxDecoherence =
        max(packets.Costs.Decoherence)
    2. return(maxDecoherence)
    3. □
```

Algorithm VII: Transport layer algorithm for combining decoherence rates in a multi-packet protocol.

However, as they are in parallel, the net latency will not be the accumulation of the individual latencies, but rather determined by the largest of those three latencies. That is, the slowest qubit acts as a bottleneck. This is shown in Alg. VIII.

```
function CombineCosts.Latency(packets):
    1. maxLatency = max(packets.Costs.Latency)
    2. return(maxLatency)
    3. □
```

Algorithm VIII: Transport layer algorithm for combining latencies in a multi-packet protocol.

To do

d. Dollars Dollars are perhaps the simplest of costs to account for, as they are simply additive – the dollar cost of a protocol is the sum of the dollar cost of each component within it. This is shown in Alg. IX.

```
function CombineCosts.Dollars(packets):
    1. costs = packets.Costs.Dollars
    2. netCost = sum(costs)
    3. return(netCost)
    4. □
```

Algorithm IX: Transport layer algorithm for combining dollar costs in packet teleportation.

To do

— *Capax infiniti.*

D. Extensibility of QTCP

In Sec. IX.B we introduced the notion of the *costs* and *attributes* of links in a classical network, and in Sec. XV generalised these notions to the quantum case. In Sec. XVI.A.2.b we described the header format for quantum data packets.

The COSTS and ATTRIBUTES fields within the HEADER are very powerful data structures, implemented as ordered sets of arbitrary dimension, comprising arbitrary data fields. The intention here is to allow QTCP to be extensible into the future, with the flexible addition of new data structures into the protocol. These can be custom designed to, in conjunction with appropriate routing strategies and cost functions, influence the operation of QTCP completely arbitrarily, and easily implement entirely different quantum networking paradigms than presented here.

COSTS naturally capture characteristics of the network that accumulate additively along routes, whereas ATTRIBUTES capture any other characteristics that aren't additive. A network needn't have both costs *and* attributes. It may have one or the other, or both, but not neither, since there must be some measure by which to judge routes, even if via a very trivial measure.

To do!

— *Cedere nescio.*

E. Scalability of QTCP

Discussion of how QTCP protocol/policies might be designed to change with scale. E.g for a small LAN the policies/metrics/attributes might be completely different than for an intercontinental uplink. Discuss how this scalability might be implemented, with examples.

“Beware of bugs in the above code. I have only proved it correct, not tried it.” — Donald Knuth.

XVII. QUALITY OF SERVICE

As discussed previously in Sec. XVI.A.6, quality of service is a major consideration in any quantum network. If we are transmitting quantum states over a quantum channel, there will inevitably be deterioration in the form of decoherence, loss, and other undesirable effects we wish to mitigate. In this section we review some of the essential quantum error correction (QEC) techniques that can be used to achieve this.

Different error correcting codes have different error correcting power, and different resource overheads, which must be taken into consideration. The appropriate choice of code will largely be determined by the final SERVICES & APPLICATIONS layer. For some simple quantum com-

munications protocols, simple error correction (or even no error correction) may suffice. A full-fledged quantum computation on the other hand will require error rates within a fault-tolerance threshold.

A. Entanglement purification

As we will later see in Sec. XXVIII, Bell pairs are a ubiquitous resource for many quantum information processing applications, and warrant special treatment in their own right.

Entanglement purification (or distillation) is a process by which two imperfect distributed Bell pairs can be condensed into a single Bell pair of higher quality (as measured by a fidelity metric, Sec. XV.A.5).

More concretely, we distribute two Bell pairs of fidelity F between Alice and Bob. Let us assume a bit-flip error model in this instance²⁶, in which case the imperfect Bell pairs are of the form,

$$\begin{aligned}\hat{\rho}_{A,B}^F &= F|\Phi^+\rangle\langle\Phi^+| + (1-F)\hat{X}_1|\Phi^+\rangle\langle\Phi^+|\hat{X}_1 \\ &= F|\Phi^+\rangle\langle\Phi^+| + (1-F)|\Psi^+\rangle\langle\Psi^+|,\end{aligned}\quad (17.1)$$

That is, we have a mixture of,

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B), \quad (17.2)$$

with probability F , and,

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + |1\rangle_A|0\rangle_B), \quad (17.3)$$

with probability $1 - F$. The goal is to ‘filter’ out the latter of these. Alice and Bob then perform their entanglement purification protocol – operations to reduce them to a single Bell pair of higher fidelity,

$$\mathcal{E}_{\text{purif}}(\hat{\rho}_{A,B}^{F_{\text{in}}} \otimes \hat{\rho}_{A,B}^{F_{\text{in}}}) \rightarrow \hat{\rho}_{A,B}^{F_{\text{out}}}, \quad (17.4)$$

where,

$$F_{\text{out}} > F_{\text{in}}. \quad (17.5)$$

The purification process $\mathcal{E}_{\text{purif}}$ is realised very simply by Alice and Bob both performing parity measurements on their two qubits, whose measurement outcomes indicate whether the two qubits were equal (+) or different (-) – their parity. These are easily implemented using CNOT gates, using the construction shown in Fig. 45, which implement the projectors,

$$\begin{aligned}\hat{\Pi}^+ &= |0,0\rangle\langle 0,0| + |1,1\rangle\langle 1,1|, \\ \hat{\Pi}^- &= |0,1\rangle\langle 0,1| + |1,0\rangle\langle 1,0|.\end{aligned}\quad (17.6)$$

²⁶ The same protocol can be employed to correct against phase-flip errors by operating in a Hadamard-rotated basis.

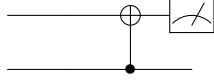


Figure 45: Construction of a parity measurement circuit using a single CNOT gate and a single-qubit measurement. The measurement outcome indicates whether the two input qubits were the same ($|0\rangle|0\rangle$, $|1\rangle|1\rangle$) or different ($|0\rangle|1\rangle$, $|1\rangle|0\rangle$). See Sec. XXIV.D for a simple linear optics construction for this operation.

They complete the protocol by classically communicating their measurement outcomes, and performing appropriate local corrections.

Two such parity projections project a 4-qubit state down to a 2-qubit state. The full circuit for this entanglement purification protocol is shown in Fig. 46.

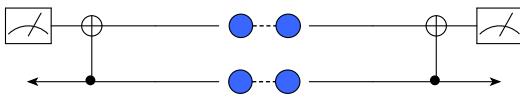


Figure 46: Protocol for entanglement purification of two noisy Bell pairs (blue) into a single less noisy Bell pair. Not shown are trivial single-qubit local corrections that may need to be made, depending upon measurement outcomes.

Let us illustrate the operation of the circuit by example of different possible input states.

Suppose there were no errors, the ideal case. Then the input state was,

$$\begin{aligned} |\psi\rangle_{\text{in}} &= |\Phi^+\rangle_{1,2}|\Phi^+\rangle_{3,4} \\ &= \frac{1}{2}(|0\rangle_1|0\rangle_2 + |1\rangle_1|1\rangle_2)(|0\rangle_3|0\rangle_4 + |1\rangle_3|1\rangle_4). \end{aligned} \quad (17.7)$$

Upon parity measurement we obtain,

$$\begin{aligned} \hat{\Pi}_{1,3}^+|\Phi^+\rangle_{1,2}|\Phi^+\rangle_{3,4} &= |\Phi^+\rangle_{2,4}, \\ \hat{\Pi}_{1,3}^-|\Phi^+\rangle_{1,2}|\Phi^+\rangle_{3,4} &= \hat{X}_2|\Phi^+\rangle_{2,4}, \end{aligned} \quad (17.8)$$

and we obtain the desired output Bell pair (with an \hat{X} correction in the case of the odd-parity outcome).

Note that if both qubits from one Bell pair are subject to bit-flips, this is equivalent to no error, since,

$$\hat{X}_1\hat{X}_2|\Phi^+\rangle_{1,2} = |\Phi^+\rangle_{1,2}. \quad (17.9)$$

Now suppose there was a bit-flip error on the first qubit of the first pair. Now the input state is,

$$\begin{aligned} |\psi\rangle_{\text{in}} &= \hat{X}_1|\Phi^+\rangle_{1,2}|\Phi^+\rangle_{3,4} \\ &= \frac{1}{2}(|1\rangle_1|0\rangle_2 + |0\rangle_1|1\rangle_2)(|0\rangle_3|0\rangle_4 + |1\rangle_3|1\rangle_4). \end{aligned} \quad (17.10)$$

Now upon parity measurement we obtain,

$$\begin{aligned} \hat{\Pi}_{1,3}^+ \hat{X}_1 |\Phi^+\rangle_{1,2} |\Phi^+\rangle_{3,4} &= \hat{X}_2 |\Phi^+\rangle_{2,4}, \\ \hat{\Pi}_{1,3}^- \hat{X}_1 |\Phi^+\rangle_{1,2} |\Phi^+\rangle_{3,4} &= |\Phi^+\rangle_{2,4}, \end{aligned} \quad (17.11)$$

and following correction we achieve a bit-flipped Bell pair.

Finally, with both Bell pairs subject to a single bit-flip,

$$\begin{aligned} |\psi\rangle_{\text{in}} &= \hat{X}_1 \hat{X}_3 |\Phi^+\rangle_{1,2} |\Phi^+\rangle_{3,4} \\ &= \frac{1}{2}(|1\rangle_1|0\rangle_2 + |0\rangle_1|1\rangle_2)(|1\rangle_3|0\rangle_4 + |0\rangle_3|1\rangle_4). \end{aligned} \quad (17.12)$$

Upon performing the parity measurement,

$$\begin{aligned} \hat{\Pi}_{1,3}^+ \hat{X}_1 \hat{X}_3 |\Phi^+\rangle_{1,2} |\Phi^+\rangle_{3,4} &= |\Phi^+\rangle_{2,4}, \\ \hat{\Pi}_{1,3}^- \hat{X}_1 \hat{X}_3 |\Phi^+\rangle_{1,2} |\Phi^+\rangle_{3,4} &= \hat{X}_2 |\Phi^+\rangle_{2,4}, \end{aligned} \quad (17.13)$$

recovering the expected Bell pair, without an erroneous bit-flip.

Evidently, the output Bell pair is subject to an unexpected bit-flip iff one of the Bell pairs was subject to a bit-flip error. This yields the output fidelity,

$$F_{\text{out}} = \frac{F_{\text{in}}^2}{F_{\text{in}}^2 + (1 - F_{\text{in}})^2}, \quad (17.14)$$

and thus the purification protocol improves fidelity provided that $F_{\text{out}} > F_{\text{in}}$, which implies the input fidelity must satisfy $F_{\text{in}} > 1/2$.

B. 3-qubit code

The 3-qubit bit-flip code is traditionally used as a basic introduction to the concept of Quantum Error Correction. However, the 3-qubit code *does not* represent a full quantum code. This is due to the fact that the code cannot simultaneously correct for both X -errors and Z -errors, which is required for correcting errors for an arbitrary error mapping on a single qubit. This code is a quantum analogue of the classical repetition code.

The 3-qubit code encodes a single logical qubit into three physical qubits with the property that it can correct for a single X -error. The two logical basis states $|0\rangle_L$ and $|1\rangle_L$ are defined as,

$$|0\rangle_L = |000\rangle, \quad |1\rangle_L = |111\rangle, \quad (17.15)$$

such that an arbitrary single qubit state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is mapped to,

$$\begin{aligned} \alpha|0\rangle + \beta|1\rangle &\rightarrow \alpha|0\rangle_L + \beta|1\rangle_L \\ &= \alpha|000\rangle + \beta|111\rangle = |\psi\rangle_L. \end{aligned} \quad (17.16)$$

As three individual X -errors are required to switch between logical states, $|0\rangle_L \leftrightarrow |1\rangle_L$, any single X -error, for example if the encoded state is $|0\rangle_L$. results in a erred state that is closer to $|0\rangle_L$ than $|1\rangle_L$. The distance between two codeword states, d , defines the number of errors that can be corrected, t , as, $t = \lfloor(d-1)/2\rfloor$. In this case, $d = 3$, hence $t = 1$.

How do we detect and correct errors without directly measuring or obtaining information about the logical state,

which would collapse the computational wave-function and destroy the encoded information? Two additional ancilla qubits, which are used to extract *syndrome* information (information regarding possible errors) are introduced.

Assuming all gate operations are perfect and the only place where the qubits are susceptible to error is the region between encoding and correction, correction proceeds by introducing two ancilla qubits and performing a sequence of CNOT gates, which checks the parity of the three qubits. Algorithm X summarizes the state of the whole system, for each possible error, just prior to measurement.

Error Location	Final State, $ \text{data}\rangle \text{ancilla}\rangle$
No Error	$\alpha 000\rangle 00\rangle + \beta 111\rangle 00\rangle$
Qubit 1	$\alpha 100\rangle 11\rangle + \beta 011\rangle 11\rangle$
Qubit 2	$\alpha 010\rangle 10\rangle + \beta 101\rangle 10\rangle$
Qubit 3	$\alpha 001\rangle 01\rangle + \beta 110\rangle 01\rangle$

Algorithm X: Final state of the five qubit system prior to the syndrome measurement for no error or a single X error on one of the qubits. The last two qubits represent the state of the ancilla. Note that each possible error will result in a unique measurement result (syndrome) of the ancilla qubits. This allows for a X correction gate to be applied to the data block which is classically controlled from the syndrome result. At no point during correction do we learn anything about α or β .

For each possible situation, either no error or a single bit-flip error, the ancilla qubits are flipped to a unique state based on the parity of the data block. These qubits are then measured to obtain the classical *syndrome* result. The result of the measurement will then dictate if an X correction gate needs to be applied to a specific qubit, i.e.

Ancilla Measurement: $ 00\rangle$,	Collapsed State: $\alpha 000\rangle + \beta 111\rangle$	\therefore Clean State
Ancilla Measurement: $ 01\rangle$,	Collapsed State: $\alpha 001\rangle + \beta 110\rangle$	\therefore Bit Flip on Qubit 3
Ancilla Measurement: $ 10\rangle$,	Collapsed State: $\alpha 010\rangle + \beta 101\rangle$	\therefore Bit Flip on Qubit 2
Ancilla Measurement: $ 11\rangle$,	Collapsed State: $\alpha 100\rangle + \beta 011\rangle$	\therefore Bit Flip on Qubit 1

(17.17)

Provided that only a single error has occurred, the data block is restored.

This code will only work if a maximum of one error occurs. If two X errors occur, then the syndrome result becomes ambiguous. For example, if an X error occurs on both qubits one and two, then the syndrome result will be $|01\rangle$. This will cause us to mis-correct by applying an X gate to qubit 3. Therefore, two errors will induce a logical bit flip and causes the code to fail. This flow is illustrated in Algorithm XI.

C. 9-qubit code

The nine qubit error correcting code was first developed by Shor (?) in 1995 and is based on the 3-qubit repetition code. The Shor code can correct a logical qubit from one discrete bit flip, one discrete phase flip or one of each on

any of the nine physical qubits and is therefore sufficient to correct for any continuous linear combination of errors on a single qubit. The two basis states for the code are,

$$\begin{aligned} |0\rangle_L &= \frac{1}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle) \\ |1\rangle_L &= \frac{1}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle) \end{aligned} \quad (17.21)$$

Correction for X -errors occurs by treating each block of three qubits in the code identically to the three qubit code [Algorithm XI]. Phase errors (Z -errors) are corrected by examining the sign differences between the three blocks.

The nine qubit code is effectively a concatenation of two bit-flip codes. By combining three bit-flip encoded qubits into a second level encoded qubit, the lower level codes can correct for a single bit-flip error in any single block, while the upper level code corrects for a single

function ThreeQubitCode($|\psi\rangle$):

1. Using two CNOT gates, redundantly encode the logical single-qubit state,

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (17.18)$$

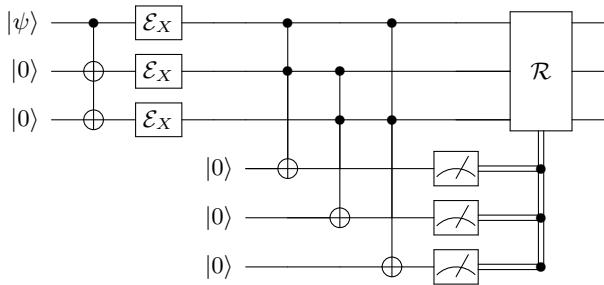
into the 3-qubit state,

$$\begin{aligned} |\psi\rangle_R &= \hat{\text{CNOT}}_{1,2}\hat{\text{CNOT}}_{1,3}|\psi\rangle|00\rangle \\ &= \alpha|000\rangle + \beta|111\rangle. \end{aligned} \quad (17.19)$$

2. Independently apply bit-flip channels \mathcal{E}_X to each of the three encoded qubits.
3. If exactly one bit-flip operation was applied in total, the three possible erroneous encoded states are,

$$\begin{aligned} |\psi\rangle_1 &= \hat{X}_1|\psi\rangle_R = \alpha|001\rangle + \beta|110\rangle, \\ |\psi\rangle_2 &= \hat{X}_2|\psi\rangle_R = \alpha|010\rangle + \beta|101\rangle, \\ |\psi\rangle_3 &= \hat{X}_3|\psi\rangle_R = \alpha|100\rangle + \beta|011\rangle. \end{aligned} \quad (17.20)$$

4. Determine the parity of each of the three pairs of encoded qubits.
5. Assuming at most a single bit-flip operation has occurred on the encoded state, the three parity outcomes uniquely determine which encoded qubit the bit-flip was applied to.
6. Apply classically-controlled bit-flip recovery operations, \mathcal{R} , to correct the encoded state, recovering $|\psi\rangle_R$.
7. Apply the inverse of the encoding operation to recover $|\psi\rangle$.
8. \square



Algorithm XI: 3-qubit code for protecting against at most a single logical bit-flip error. The doubly-controlled CNOT gates represent parity measurements, $n_3 = n_1 \oplus n_2$, where n_i represents the value of the i th qubit. In a Hadamard-rotated basis, the same circuit may be employed to protect against a single phase-flip error. And by concatenating the two we obtain a 9-qubit code protecting against a single depolarising error (i.e joint bit-flip/phase-flip), which is a universal single-qubit error model.

function NineQubitCode($|\psi\rangle$):

1. Using eight CNOT gates and three Hadamards, redundantly encode the logical single-qubit state,

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (17.22)$$

into the 9-qubit state,

$$\begin{aligned} |\psi\rangle|00000000\rangle &\rightarrow |\psi\rangle_R \\ &= \frac{\alpha}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle) \\ &+ \frac{\beta}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle) \end{aligned} \quad (17.23)$$

2. Independently apply both bit-flip channels \mathcal{E}_X and phase-flip channels \mathcal{E}_Z to each of the three encoded qubits.
3. The nine qubit code is degenerate, meaning that multiple single qubit errors have the same effect on the code-state. For all possible single qubit phase errors, there are only three possible changes to the logic state.

$$\begin{aligned} |\psi\rangle_1 &= \hat{Z}_{1,2,3}|\psi\rangle_R \\ &= \frac{\alpha}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle + |111\rangle)(|000\rangle + |111\rangle) \\ &+ \frac{\beta}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle - |111\rangle)(|000\rangle - |111\rangle), \\ |\psi\rangle_2 &= \hat{Z}_{4,5,6}|\psi\rangle_R \\ &= \frac{\alpha}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle - |111\rangle)(|000\rangle + |111\rangle) \\ &+ \frac{\beta}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle + |111\rangle)(|000\rangle - |111\rangle), \\ |\psi\rangle_3 &= \hat{Z}_{7,8,9}|\psi\rangle_R \\ &= \frac{\alpha}{\sqrt{8}}(|000\rangle + |111\rangle)(|000\rangle + |111\rangle)(|000\rangle - |111\rangle) \\ &+ \frac{\beta}{\sqrt{8}}(|000\rangle - |111\rangle)(|000\rangle - |111\rangle)(|000\rangle + |111\rangle). \end{aligned} \quad (17.24)$$

4. Correct all X -errors by applying the three qubit correction protocol to each of the three blocks.
5. To check for Z -errors, compare the \pm signs between blocks by first coupling all qubits in blocks one and two via CNOT gates to one ancilla and all qubits in blocks two and three to a second ancilla.
6. To correct for Z -errors, apply a physical Z -gate to any of the physical qubits in a block identified by the ancilla measurements (for a degenerate code, the correction gate does not need to be identical to the original error).
7. \square

Algorithm XII: 9-qubit code for protecting against at most a single bit- and/or phase flip error. The code is a concatenation of three blocks of three bit flip codes. Bit flips are corrected within blocks of three qubits, while phase flips are corrected by comparing blocks.

phase flip error in any single block.

Note that a phase flip on *any* one qubit in a block of three has the same effect, this is why the 9-qubit code is referred to as a degenerate code. In other error correcting codes, such as the 5- or 7-qubit codes (??), there is a one-to-one mapping between correctable errors and unique states, in degenerate codes such as this, the mapping is not unique. If we know which block the error occurs it does not matter which qubit we apply the correction operator.

Even if a bit *and* phase error occurs on the same qubit (i.e. a Y -error), the X correction circuit will detect and correct for bit flips while the Z correction circuit will detect and correct for phase flips. The X -error correction has the ability to correct for up to three individual bit flips (provided each bit flip occurs in a different block of three). However, in general the 9-qubit code is only a single error correcting code as it cannot handle more than a single bit flip error if they occur in certain locations.

D. Stabiliser codes

So far we have presented error correcting codes from the perspective of their state representations and their preparation and correction circuits. This is a rather inefficient method for describing the codes as the state representations and circuits clearly differ from code to code. The majority of error correcting codes that are used within the literature are members of a class known as stabilizer codes. Stabilizer codes are very useful to work with. The general formalism applies broadly and there exists general rules to construct preparation circuits, correction circuits and fault-tolerant logical gate operations once the stabilizer structure of the code is specified.

The stabilizer formalism which was first introduced by Daniel Gottesman (?) uses essentially the Heisenberg representation for quantum mechanics which describes quantum states in terms of operators rather than the basis states themselves. An arbitrary state $|\psi\rangle$ is defined to be stabilized by some operator, K , if it is a +1 eigenstate of K , i.e.

$$K|\psi\rangle = |\psi\rangle. \quad (17.25)$$

For example, the single qubit state $|0\rangle$ is stabilized by the operator $K = \sigma_z$, i.e.

$$\sigma_z|0\rangle = |0\rangle \quad (17.26)$$

Defining multi-qubit states with respect to this formalism relies on the group structure of multi-qubit operators.

Within the group of all possible, single qubit operators, there exists a subgroup, denoted the Pauli group, \mathcal{P} , which contains the following elements,

$$\mathcal{P} = \{\pm\sigma_I, \pm i\sigma_I, \pm\sigma_x, \pm i\sigma_x, \pm\sigma_y, \pm i\sigma_y, \pm\sigma_z, \pm i\sigma_z\}. \quad (17.27)$$

It is easy to check that these matrices form a group under multiplication through the commutation and anti-commutation rules for the Pauli set, $\{\sigma_i\} = \{\sigma_x, \sigma_y, \sigma_z\}$,

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k, \quad \{\sigma_i, \sigma_j\} = 2\delta_{ij}, \quad (17.28)$$

where,

$$\epsilon_{ijk} = \begin{cases} +1 & \text{for } (i, j, k) \in \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\} \\ -1 & \text{for } (i, j, k) \in \{(1, 3, 2), (3, 2, 1), (2, 1, 3)\} \\ 0 & \text{for } i = j, j = k, \text{ or } k = i \end{cases} \quad (17.29)$$

and

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j. \end{cases} \quad (17.30)$$

The Pauli group extends over N -qubits by simply taking the N fold tensor product of \mathcal{P} , i.e.

$$\begin{aligned} \mathcal{P}_N &= \mathcal{P}^{\otimes N} \\ &= \{\pm\sigma_I, \pm i\sigma_I, \pm\sigma_x, \pm i\sigma_x, \pm\sigma_y, \pm i\sigma_y, \pm\sigma_z, \pm i\sigma_z\}^{\otimes N}. \end{aligned} \quad (17.31)$$

An N -qubit stabilizer state, $|\psi\rangle_N$ is then defined via an N -element Abelian subgroup, \mathcal{G} , of the N -qubit Pauli group, in which $|\psi\rangle_N$ is a +1 eigenstate of each element,

$$\begin{aligned} \mathcal{G} &= \\ &\{G_i \mid G_i|\psi\rangle = |\psi\rangle, [G_i, G_j] = 0 \forall (i, j)\} \subset \mathcal{P}_N. \end{aligned} \quad (17.32)$$

Given this definition, the state $|\psi\rangle_N$ can be equivalently defined either through the state vector representation *or* by specifying the stabilizer set, \mathcal{G} .

Many extremely useful multi-qubit states are stabilizer states, including two-qubit Bell states, Greenberger-Horne-Zeilinger (GHZ) states (??), Cluster states (??) and codeword states for QEC. As an example, consider a three qubit GHZ state, defined as,

$$|\text{GHZ}\rangle_3 = \frac{|000\rangle + |111\rangle}{\sqrt{2}}. \quad (17.33)$$

This state can be expressed via any three linearly independent elements of the $|\text{GHZ}\rangle_3$ stabilizer group for example,

$$\begin{aligned} G_1 &= \sigma_x \otimes \sigma_x \otimes \sigma_x \equiv XXX, \\ G_2 &= \sigma_z \otimes \sigma_z \otimes \sigma_I \equiv ZZZ, \\ G_3 &= \sigma_I \otimes \sigma_z \otimes \sigma_z \equiv IZZ. \end{aligned} \quad (17.34)$$

where the right-hand side of each equation is the short-hand representation of stabilizers. Note that these three operators form an Abelian group [Eq. 17.32] as,

$$\begin{aligned} |\psi\rangle &= G_i G_j |\psi\rangle - G_j G_i |\psi\rangle \\ &= |\psi\rangle - |\psi\rangle = 0, \quad \forall [i, j, |\psi\rangle]. \end{aligned} \quad (17.35)$$

Similarly, the four orthogonal Bell states,

$$\begin{aligned} |\Phi^\pm\rangle &= \frac{|00\rangle \pm |11\rangle}{\sqrt{2}}, \\ |\Psi^\pm\rangle &= \frac{|01\rangle \pm |10\rangle}{\sqrt{2}}, \end{aligned} \quad (17.36)$$

are stabilized by the operators, $G_1 = (-1)^a XX$, and $G_2 = (-1)^b ZZ$. Where $[a, b] \in \{0, 1\}$ and each of the four Bell states correspond to the four unique pairs, $\{\Phi^+, \Psi^+, \Phi^-, \Psi^-\} = \{[0, 0], [0, 1], [1, 0], [1, 1]\}$.

XVIII. QEC WITH STABILIZER CODES

The use of the stabilizer formalism to describe quantum error correction codes is extremely useful since it allows

for easy synthesis of correction circuits and also clearly shows how logical operations can be performed directly on encoded data. As an introduction we will focus on arguably the most well known quantum code, the 7-qubit Steane code, first proposed in 1996 (?).

The 7-qubit code represents a full quantum code that encodes seven physical qubits into one logical qubit, with the ability to correct for a single X and/or Z error. The $|0\rangle_L$ and $|1\rangle_L$ basis states are defined as,

$$\begin{aligned} |0\rangle_L &= \frac{1}{\sqrt{8}}(|0000000\rangle + |1010101\rangle + |0110011\rangle + |1100110\rangle + |0001111\rangle + |1011010\rangle + |0111100\rangle + |1101001\rangle), \\ |1\rangle_L &= \frac{1}{\sqrt{8}}(|1111111\rangle + |0101010\rangle + |1001100\rangle + |0011001\rangle + |1110000\rangle + |0100101\rangle + |1000011\rangle + |0010110\rangle). \end{aligned} \quad (18.1)$$

The stabilizer set for the 7-qubit code is fully specified by the six operators,

$$\begin{aligned} K^1 &= IIIXXX, & K^2 &= XIXIXX, \\ K^3 &= IXXIIXX, & K^4 &= IIIZZZZ \\ K^5 &= ZIZIZIZ, & K^6 &= IZZIIZZ. \end{aligned} \quad (18.2)$$

As the 7-qubit codeword states are specified by only six stabilizers, the code contains two basis states, which are the logical states. With a final operator, $K^7 = ZZZZZZZ = Z^{\otimes 7}$ fixing the state to one of the codewords, $K^7|0\rangle_L = |0\rangle_L$ and $K^7|1\rangle_L = -|1\rangle_L$. The 7-qubit code is defined as a $[[n, k, d]] = [[7, 1, 3]]$ quantum code, where $n = 7$ physical qubits encode $k = 1$ logical qubit with a distance between basis states $d = 3$, correcting $t = (d - 1)/2 = 1$ error. Notice that the stabilizer set separates into X and Z sectors which defines the code as a Calderbank-Shor-Steane (CSS) code. CSS codes are extremely useful since they allow for straightforward logical gate operations to be applied directly to the encoded data and are easy to derive from classical codes.

Although the 7-qubit code is the most well known Stabilizer code, there are other stabilizer codes which encode multiple logical qubits and correct for more errors (?). The downside to these larger codes is that they require more physical qubits and more complicated error correction circuits. Tables XIII and XIV shows the stabilizer structure of two other well known codes, the 9-qubit code (?) which we have examined and the 5-qubit code (?) which represents the smallest possible quantum code that corrects

for a single error.

K^1	Z	Z	I						
K^2	Z	I	Z	I	I	I	I	I	I
K^3	I	I	I	Z	Z	I	I	I	I
K^4	I	I	I	Z	I	Z	I	I	I
K^5	I	I	I	I	I	I	Z	Z	I
K^6	I	I	I	I	I	I	Z	I	Z
K^7	X	X	X	X	X	X	I	I	I
K^8	X	X	X	I	I	I	X	X	X

Algorithm XIII: The eight Stabilizers for the 9-qubit Shor code, encoding nine physical qubits into one logical qubit to correct for a single X and/or Z error.

1. Performing error correction

Error correction using stabilizer codes is a straightforward extension of state preparation. Consider an arbitrary single qubit state that has been encoded,

$$\alpha|0\rangle + \beta|1\rangle \rightarrow \alpha|0\rangle_L + \beta|1\rangle_L = |\psi\rangle_L. \quad (18.3)$$

Now assume that an error occurs on one (or multiple) qubits which is described via the operator E , where E is a combination of X and/or Z errors over the N physical qubits of the logical state. By definition of stabilizer codes, $K^i|\psi\rangle_L = |\psi\rangle_L$, $i \in [1, \dots, N - k]$, for a code encoding k

K^1	X	Z	Z	X	I
K^2	I	X	Z	Z	X
K^3	X	I	X	Z	Z
K^4	Z	X	I	X	Z

Algorithm XIV: The Four Stabilizers for the [[5,1,3]] quantum code, encoding five physical qubits into one logical qubit to correct for a single X and/or Z error. Unlike the 7- and 9-qubit codes, the [[5,1,3]] code is a non-CSS code, since the stabilizer set does not separate into X and Z sectors.

logical qubits. Hence the errred state, $E|\psi\rangle_L$, satisfies,

$$K^i E |\psi\rangle_L = (-1)^m E K^i |\psi\rangle_L = (-1)^m E |\psi\rangle_L. \quad (18.4)$$

where m is defined as, $m = 0$, if $[E, K^i] = 0$ and $m = 1$, if $\{E, K^i\} = 0$. Therefore, if the error operator commutes with the stabilizer, the state remains a +1 eigenstate of K^i , if the error operator anti-commutes with the stabilizer then the logical state is flips to now be a -1 eigenstate of K^i .

Each of the code stabilizers are sequentially measured. Since a error free state is already a +1 eigenstate of all the stabilizers, any error which anti-commutes with a stabilizer will flip the eigenstate and consequently the parity measurement will return a result of $|1\rangle$.

Taking the [[7, 1, 3]] code as an example, if the error operator is $E = X_i$, where $i = (1, \dots, 7)$, representing a bit-flip on any *one* of the 7 physical qubits, then regardless of the location, E will anti-commute with a unique combination of (K^4, K^5, K^6) . Hence the classical results of measuring these three operators will indicate if and where a single X error has occurred. Similarly, if $E = Z_i$, then the error operator will anti-commute with a unique combination of, (K^1, K^2, K^3) . Consequently, the first three stabilizers for the [[7, 1, 3]] code correspond to Z sector correction while the second three stabilizers correspond to X sector correction. Note, that correction for Pauli Y errors are also taken care of by correcting in the X and Z sector since a Y error on a single qubit is equivalent to both an X and Z error on the same qubit, i.e. $Y = iXZ$.

A. Surface codes

Discuss genus and how this adds logical qubits.

If the intention is to perform quantum computations using cluster states shared over a network, QEC and fault-tolerance *must* be taken into consideration, or catastrophic algorithmic failure will inevitably follow – the cluster state model is no different from the circuit model in this respect. Fault-tolerance theory places hard thresholds on the amount of noise (typically depolarising errors and loss) qubits may be subject to in order for fault-tolerance to be possible and computations to succeed. This places strict QoS (Sec. XVI.A.6) constraints on the

network, which can be accommodated for using the usual depolarising and efficiency cost metrics when developing networking strategies and link performance requirements.

It has been shown that fault-tolerance is possible within the cluster state model (??) using variations of conventional QEC codes. However, more importantly, from cluster states certain *topological QEC codes* (?) can be readily constructed. This implements a form of QEC-encoded measurement-based quantum computing protocol, where the computation proceeds in a measurement-based fashion, but is ‘natively’ fault-tolerant.

These codes have been shown to have very favourable fault-tolerance thresholds in terms of both depolarising noise and loss (??), as well as frugal resource overhead compared to traditional concatenated codes. Additionally, loss- and gate-failure-tolerant codes, uniquely applicable to the cluster state model, have been described, with very favourable loss thresholds (????).

Importantly, topological codes do not require joint measurements across the entire graph state, instead requiring only operations localised to small regions within the graph. Thanks to this, computation using such topological codes can remain distributed without requiring the entire state to be held locally by a particular host, or requiring full access to the entire state by any particular user. As with cluster states, a large graph can be stitched together from a patchwork of smaller neighbouring graphs.

The most common topological code, which we will use here as an example, is the toric code, which resides on a lattice graph over the surface of a torus²⁷. As with cluster states (Sec. XL.B), the toric code is most easily visualised in the stabiliser formalism, and the toric code is an example of a stabiliser code (Sec. XVII.D). Consider a rectangular sub-graph of the torus. We place a qubit on each edge (not vertex) of the graph. Now we define two sets of stabiliser operators: *star* (\hat{S}_+) and *plaquette* (\hat{S}_\square) operators,

$$\begin{aligned} \hat{S}_+(v) &= \prod_{i \in e(v)} \hat{X}_i, \\ \hat{S}_\square(p) &= \prod_{i \in e(p)} \hat{Z}_i, \end{aligned} \quad (18.5)$$

where $e(v)$ are the edges neighbouring vertex v , and $e(p)$ are the edges surrounding plaquette p . By definition, the toric code state, $|\psi\rangle_{\text{toric}}$, satisfies the stabiliser relations,

$$\begin{aligned} \hat{S}_+(v)|\psi\rangle_{\text{toric}} &= |\psi\rangle_{\text{toric}} \forall v, \\ \hat{S}_\square(p)|\psi\rangle_{\text{toric}} &= |\psi\rangle_{\text{toric}} \forall p. \end{aligned} \quad (18.6)$$

The stabilisers can be visualised graphically as per Fig. 47.

²⁷ As with cluster states, this graph needn't (but could) correspond to a network graph.

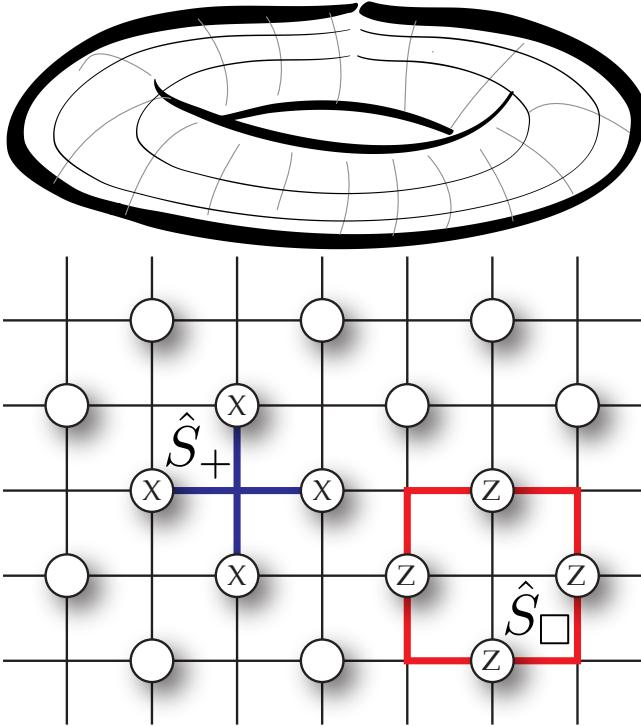


Figure 47: Graph representation of the toric QEC code, and its associated stabilisers. The graph resides on a torus (top), while the physical qubits reside on the embedded lattice (bottom). The star and plaquette stabilisers across all vertices, jointly specify the state of the graph up to two missing degrees of freedom, which encode a single logical qubit. Thus, a logical qubit is encoded jointly across the entire graph, not at any specific vertex. Logical operations are performed via operations following topological paths through the lattice (not shown). The graph may be distributed across multiple hosts for distributed quantum computation.

Unlike the cluster state stabilisers from Eq. (40.8), these stabilisers are insufficient to fully characterise a unique quantum state. Rather, there are two unspecified degrees of freedom, which allows for a single qubit to be represented. Modifications of the topology, in the form of holes in the lattice (the genus of the topology), allow larger numbers of qubits to be encoded. Logical operations are implemented by performing sequences of local gates and measurements across topologies over the surface.

The important feature to note is that logical qubits encoded into the toric code do not reside locally at any of the physical qubits in the topology. Rather, they reside jointly across the entire graph, which, like cluster states, might be partitioned across multiple hosts, enabling distributed computation.

Having defined the toric code as such, QEC proceeds in a similar manner to any other stabiliser code – we measure all the stabilisers, yielding a syndrome, from which we can determine geometrically where errors took place in the graph, which can subsequently be corrected (if below threshold).

The simplest example of error detection is the scenario where a single bit-flip (\hat{X}) error has occurred in the graph. Now exactly two plaquette stabilisers (the ones upon which the respective qubit acts) will yield the -1 measurement outcome, instead of the expected $+1$ outcome. These two stabilisers will necessarily be neighbouring ones, overlapping at the qubit where the error took place. Thus, using this geometric reasoning, we are able to identify the location of the single \hat{X} error and subsequently correct it. On the other hand, if there were too many errors, it is possible they could conspire against us to create ambiguity in the geometric argument for the location of the errors. The QEC procedure is illustrated in Fig. 47.

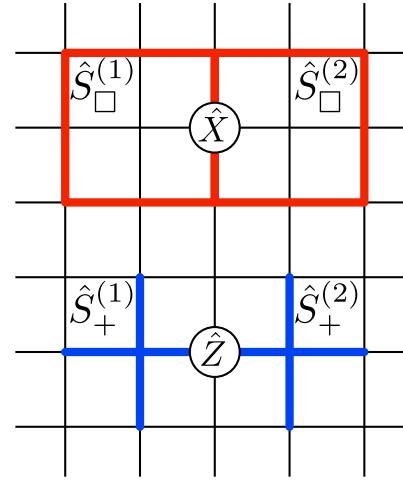


Figure 48: QEC of a single bit-flip (top) and phase-flip (bottom) error via measuring two overlapping stabilisers. If there is at most one bit- or phase-flip error, as per this graphic, then upon measuring all the star and plaquette stabilisers, only those whose qubits are affected by the error channel will yield the -1 measurement outcome, the others all yielding $+1$. Such an error will affect exactly two stabilisers, whose qubit in common must have been the one affected by the respective error, which can subsequently be corrected since the type of the error and its location are known. Thus, in this example, all of the $\hat{S}_+^{(1)}$, $\hat{S}_+^{(2)}$, $\hat{S}_\square^{(1)}$ and $\hat{S}_\square^{(2)}$ stabilisers will yield -1 outcomes for the shown \hat{X} and \hat{Z} errors, whereas any of the remaining (not shown) $\hat{S}_k^{(j)}$ will yield $+1$ outcomes, which uniquely specifies where those errors took place.

Importantly, the stabilisers are all defined over geometrically localised neighbourhood regions, and do not require long-range measurements, making this type of code suitable to distributed models for quantum computation, much like cluster states.

Logical operations similarly have geometric interpretations. Most simply, logical Pauli- \hat{X} and $-\hat{Z}$ gates may be implemented by applying chains of local Pauli operations along topological paths, specifically closed loops around the two different axes of the torus (Fig. 49). These topologies may be deformed, provided that they are topologically equivalent to the desired closed loops, hence the name ‘topological code’. With more qubits and more

elaborate gates, similar topological definitions exist for implementing logical gates on encoded qubits.

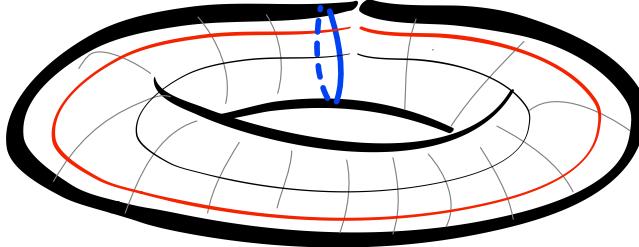


Figure 49: Implementation of logical \hat{X} (**which colour?**) and $-\hat{Z}$ (**which colour?**) operations on a toric code qubit via the application of a chain of local Paulis. The chains must circumnavigate the torus around either axis. Any path topologically equivalent to such a circumnavigation of the torus is sufficient, enabling paths to bypass defects in the lattice.

What about the actual computation? Can this still be distributed when we do the topological gates etc?

Figures for thresholds FT and QEC

Explain stabiliser measurement and how this yields error syndromes

Explain logical operations on code

Explain how number of logical qubits can be increased by changing topology

How to convert cluster states to topological codes, or any other way of preparing them. Add discussion of how to perform gates.

Figure for both examples of this – under and over threshold!

Missing degree of freedom in stabilisers defines qubit. How does topology (genus) relate to number of logical qubits?

B. Unitary error averaging

Aside from the *active* QoS techniques discussed until now, there is also a recently described *passive* technique that requires no feedforward to operate. This technique, called *unitary error averaging* (?), is formulated specifically in the linear optical context. It is an open question whether it generalises to other models, such as conventional quantum circuits.

Suppose we desire to implement some linear optics unitary network, \hat{U}_{target} , but our fabrication techniques are imperfect and we instead implement a close approximation to it. How can we overcome this?

(?) showed that by using an optical fanout operation, which splits a set of optical modes equally across multiple sets of modes in a type of redundant encoding²⁸, and

passing each set through an independently manufactured copy of the imperfect unitary, \tilde{U}_i , upon performing a fan-in to recombine the modes, and post-selection on detecting all photons in the desired output modes, the output state will behave as if it had evolved through a network given by the arithmetic average of each of the imperfect copies of \hat{U}_{target} ,

$$\hat{U}_{\text{av}} = \frac{1}{M} \sum_{i=1}^M \tilde{U}_i. \quad (18.7)$$

A schematic for the optical circuit is shown in Fig. 50.

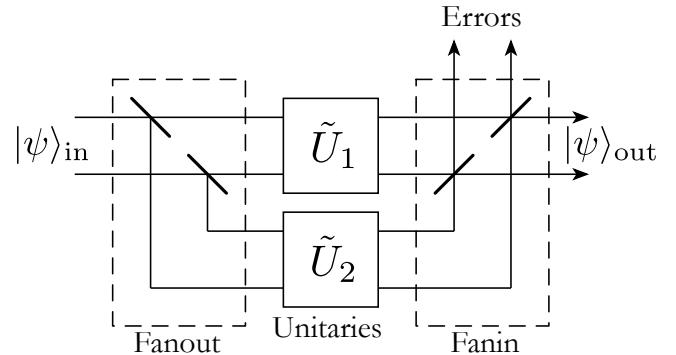


Figure 50: Protocol for error correction via unitary error averaging. Here our computation resides across $N = 2$ modes, and there are $M = 2$ redundant copies of the manufactured unitary, which straightforwardly generalises to arbitrary M and N . A fanout operation splits the wave amplitudes from the set of N input modes uniformly across M sets of N modes. Each set of N evolves through an independent copy of the manufactured imperfect unitary, \tilde{U}_i . A reverse fanout operation recollects the bundles of modes. Upon post-selecting upon detecting all photons within the first N modes, and none in the other ancillary modes, the state is projected into the subspace as if it had been acted upon by the arithmetic average of the imperfect unitaries. If photons are detected in the error modes, our logical state has effectively suffered photon loss and has been corrupted.

Note that the protocol doesn't require that the M instances of the unitary be held locally. They operate independently in parallel and could therefore be distributed, with the independent copies held by entirely different parties in distant locations (Sec. XLIII.B.5).

The protocol is necessarily non-deterministic, since sometimes photons will be measured in the error modes, signalling that the output state has been projected onto an erroneous subspace. Thus, this is a heralded error correction scheme. When post-selection succeeds we have effectively evolved via the averaged operator, which statistically ought to more closely approximate \hat{U}_{target} than individual \tilde{U}_i .

²⁸ The optical fanout operation effectively maps single-photon states

to W-states (Sec. XXIII.F).

C. Qubit loss codes

More severe than a simple, say, dephasing channel is qubit loss, whereby an entire physical qubit is lost to the environment, effectively tracing it out of the system. Depending on physical architecture, there are two ways in which qubit loss might manifest itself:

- Located errors: We know that a loss occurred and which physical qubit was affected.
- Unlocated errors: We do not know that a physical qubit was lost.

These manifestations arise naturally in different architectural contexts. For example, it is relatively easy to look at an atomic qubit to see whether it's still where it should be, whereas observing a photonic qubit typically requires destroying it, making the former vulnerable to located errors and the latter to unlocated errors.

It turns out that the distinction between these two modes of error can have a significant effect on the error correction thresholds of QECs. Specifically, a located error effectively gives us an additional classical bit of information diagnosing that an error took place, which we are not privy to for unlocated errors. This additional bit of information can be exploited by QEC codes to improve thresholds for located errors.

In (?) a derivation of a generalisation of the quantum Hamming bound for non-degenerate²⁹ QEC codes was presented, where the number of located and unlocated errors were treated as independent parameters. They found that for non-degenerate codes the Hamming bound allows exactly twice as many located as unlocated errors. Specifically, for t_l located errors and t_u unlocated errors in an n -qubit non-degenerate code encoding k logical qubits, QEC is possible provided that the number of errors satisfied,

$$\sum_{i=0}^{t_u + \lfloor t_l/2 \rfloor} \binom{n}{i} 3^i 2^k \leq 2^n. \quad (18.8)$$

Subsequently, using surface codes (?) arrived at the similar conclusion that QEC codes are more robust against located than unlocated errors.

D. Gate failure codes

Quantum gates, especially 2-qubit entangling ones in many architectures, can have some non-zero chance of

failure. In linear optics this gate non-determinism is an inherent feature of entangling gates like the CNOT. In fact it is provable that such gates cannot be implemented deterministically. With a large number of such gates in a circuit, the probability of them all working drops exponentially with the number of non-deterministic gates, thereby undermining efficiency. Can we overcome this?

In Sec. XLII.B we discuss techniques for performing scalable linear optics quantum computing with non-deterministic gates, within the cluster state formalism. The idea is to construct micro-clusters with redundant dangling bonds, which facilitate multiple bonding attempts to merge smaller clusters into larger ones. Once all the entangling operations have been performed, all that remains is to perform a sequence of single-qubit unitaries and measurements, both of which are deterministic in-principle.

This same approach could be logically generalised to any physical architecture where entangling gates are non-deterministic, but single-qubit operations are deterministic.

E. Decoherence-free subspaces

The QoS techniques discussed previously were based on the notion of performing measurements on quantum systems to project them into subspaces devoid of errors. For example, in the 3-qubit code, measurement of the syndrome qubits projects the encoded state into a subspace where there was either no error, or in which an error occurred whose location is known and may therefore be corrected.

An alternate approach is to encode quantum information into Hilbert spaces which are invariant under a given error model. Such spaces are referred to as *decoherence-free subspaces* (DFSs). In this instance we assume the error model is known, for example a dephasing channel, such that we can choose the appropriate DFS.

To illustrate this idea we will consider encoding a single logical qubit into two physical qubits. The error model we will encode against is a collective Z -rotation, where the two physical qubits are subject to perfectly correlated Z errors. This arises naturally in the context of, say, atomic qubits subject to the same external electromagnetic field, and therefore accumulate the associated phase errors in tandem.

A single-qubit Z -rotation of angle θ on the Bloch sphere is given by,

$$\hat{Z}(\theta) = e^{i\frac{\theta}{2}\hat{Z}} = \begin{pmatrix} e^{i\frac{\theta}{2}} & 0 \\ 0 & e^{-i\frac{\theta}{2}} \end{pmatrix}. \quad (18.9)$$

where \hat{Z} is the usual Pauli phase-flip operator (Sec. XL.A).

²⁹ In a non-degenerate code, all correctable errors map the codeword to orthogonal states. For degenerate codes multiple distinct errors can map the codeword to the same state. The Hamming bound exploits the non-degeneracy using a counting or ‘pigeonholing’ argument to count correctable errors.

This operates on the physical basis states as,

$$\begin{aligned}\hat{Z}(\theta)|0\rangle &\rightarrow e^{i\frac{\theta}{2}}|0\rangle, \\ \hat{Z}(\theta)|1\rangle &\rightarrow e^{-i\frac{\theta}{2}}|1\rangle.\end{aligned}\quad (18.10)$$

Now we employ the encoding for logical basis states,

$$\begin{aligned}|0\rangle_L &\equiv |0\rangle_1 \otimes |1\rangle_2, \\ |1\rangle_L &\equiv |1\rangle_1 \otimes |0\rangle_2,\end{aligned}\quad (18.11)$$

Note that both logical basis states are invariant under a common Z -rotation (the other two physical basis states, $|0\rangle_1 \otimes |0\rangle_2$ and $|1\rangle_1 \otimes |1\rangle_2$ do not observe this property),

$$\begin{aligned}\hat{Z}_1(\theta)\hat{Z}_2(\theta)|0\rangle_L &= |0\rangle_L, \\ \hat{Z}_1(\theta)\hat{Z}_2(\theta)|1\rangle_L &= |1\rangle_L.\end{aligned}\quad (18.12)$$

Thus, when acting on an arbitrary linear combination of these basis states (i.e a logical qubit),

$$|\psi\rangle_L = \alpha|0\rangle_L + \beta|1\rangle_L,\quad (18.13)$$

via linearity the logical qubit must also be invariant under the collective error,

$$\hat{Z}_1(\theta)\hat{Z}_2(\theta)|\psi\rangle_L = |\psi\rangle_L.\quad (18.14)$$

This type of DFS encoding therefore protects a logical qubit against arbitrary, unknown but correlated Z -rotations.

The same principle can be logically extended to many other correlated error models. For example, operating in a rotated basis (under a Hadamard transform), one could similarly protect against correlated X -rotations using the encoding,

$$\begin{aligned}|0\rangle_L &\equiv |-1\rangle_1 \otimes |+2\rangle, \\ |1\rangle_L &\equiv |+1\rangle_1 \otimes |-2\rangle,\end{aligned}\quad (18.15)$$

where $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$.

This idea, although very simple, is very powerful, since it is a completely passive form of error correction, requiring no syndrome measurements, feedforward or correction operations. It also arises quite naturally in some systems where undesired external fields act roughly uniformly across the physical qubits within a system.

F. Dynamical decoupling

An alternate mechanism by which errors could be introduced into our system is via coupling to an external environment (for example via an electromagnetic field) introducing a persistent evolution of our qubits, which is slow-moving compared to the rate at which the implemented computation is evolving the system. We can

model this as a joint system/environment Hamiltonian of the form,

$$\hat{H}_{\text{total}} = \lambda_{\text{comp}}\hat{H}_{\text{comp}} + \lambda_{\text{env}}\hat{H}_{\text{env}} + \lambda_{\text{int}}\hat{H}_{\text{int}},\quad (18.16)$$

where the different components represent, in order, the Hamiltonians of the: total joint system; quantum computer (or system of interest); environment; interaction between system and environment. We are specifically operating in the regime where $\lambda_{\text{int}} \ll \lambda_{\text{comp}}$, such that the computation is the dominant term in the evolution and the environmental coupling can be treated as a small perturbation from the desired evolution.

The goal of dynamical decoupling is to minimise the influence of the system/environment interaction term, \hat{H}_{int} , by manipulating the system in such a way that this term continuously cancels itself out over time.

Let us illustrate how this can be achieved using a simple example, whereby a single-qubit system couples with an environment which introduces a slow, unknown phase evolution. Discretising time, we can write this phase evolution as a Pauli Z -rotation on the Bloch sphere,

$$\hat{Z}(\theta) = e^{i\frac{\theta}{2}\hat{Z}} = \begin{pmatrix} e^{i\frac{\theta}{2}} & 0 \\ 0 & e^{-i\frac{\theta}{2}} \end{pmatrix},\quad (18.17)$$

for some unknown, but small θ . Next we observe that, quite obviously, a Z -rotation of angle θ can be trivially undone by applying another Z -rotation of angle $-\theta$, since,

$$\hat{Z}(-\theta)\hat{Z}(\theta) = \hat{1}.\quad (18.18)$$

Therefore, if the system/environment coupling introduced an evolution of $\hat{Z}(\theta)$ in the previous unit of time, our goal is to manipulate it into implementing $\hat{Z}(-\theta)$ during the next one. Alas, the environment is beyond our control and we cannot directly order it to reverse direction. We do, however, have complete control over our qubit system, so we will achieve the same outcome by flipping the direction of the Bloch sphere underneath the environments foot, allow it to take a step forward, before flipping it back.

How do we achieve this flip in the Bloch sphere? Simply by using the following identity from the algebra of the Pauli matrices,

$$\hat{X}\hat{Z}(\theta)\hat{X} = \hat{Z}(-\theta).\quad (18.19)$$

That is, applying a bit-flip to a qubit, followed by an arbitrary phase-rotation, followed by another bit-flip is equivalent to having taken the same phase-rotation in the reverse direction. Effectively we are tricking the environment into time-reversal!

Now if we proceed for two time-steps, once bit-flipped and another not, we have,

$$\underbrace{\hat{Z}(\theta)}_{\text{forward}} \cdot \underbrace{[\hat{X}\hat{Z}(\theta)\hat{X}]}_{\text{reverse}} = \hat{1},\quad (18.20)$$

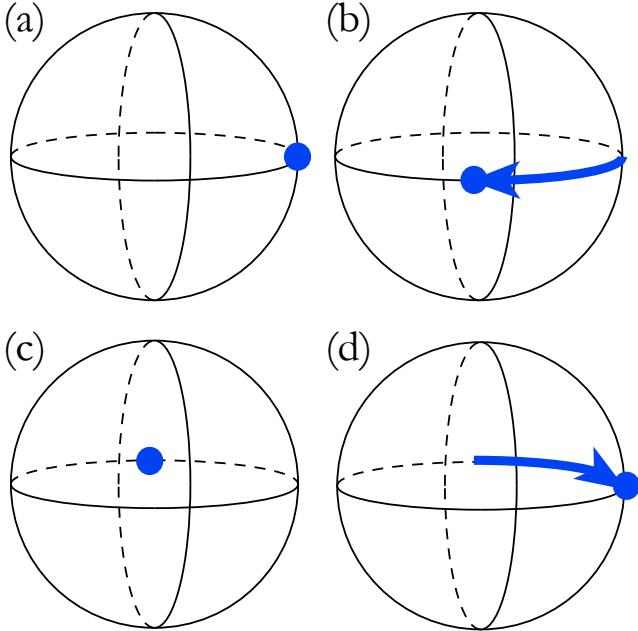


Figure 51: Bloch sphere representation of a sequence to dynamically decouple an unknown \hat{Z} rotation from a qubit. (a) Blue dot represents the initial state. (b) The state undergoes an unknown $\hat{Z}(\theta)$ rotation. (c) A bit-flip (\hat{X}) flips the Bloch sphere upside down. (d) Waiting for the same time as before such that the same unknown \hat{Z} -rotation takes place, the state ends up in its initial state, independent of the \hat{Z} -rotation angle θ , but assuming it was the same on both occasions in (b) and (d).

and the unknown phase-rotation has been eliminated. The sequence is illustrated on the Bloch sphere in Fig. 51.

The coupling to the external environment and the qubit state needn't be constant over time, but could be time-varying. In this instance, it is essential that the decoupling control sequence of bit-flip operations be much faster than the rate at which both the phase-rotations and computational operation are changing, such that consecutive unknown phase-rotations are approximately equal and thereby almost perfectly cancel. This rapid control sequence is sometimes referred to as ‘bang-bang’ control, since we are repeatedly implementing bit-flips at a fast rate.

As described above, we have merely error corrected a quantum memory. Of course we wish to implement far more sophisticated evolutions. This requires breaking down the computational evolution (given by \hat{H}_{comp}) into a large number of small, discrete steps. These are interspersed with our bang-bang control sequence so as to continuously remove any phase-errors accumulated during the course of the computation.

Dynamical decoupling extends to all manner of error models, beyond the simple Z -error model presented above. They are therefore a very powerful tool in error correction. However, unfortunately they are only naturally suited to continuous evolutions governed by Hamiltonians of the

form shown in Eq. (18.16), not to the more common discretised models such as the circuit and cluster state models.

References

G. Continuous-variables

First CV QEC protocol is the direct analogue of the the qubit redundancy codes. However, the noise models do not correspond to what usually occur in Gaussian systems, namely, thermal noise and loss.

It has been proven that QEC of Gaussian noise on Gaussian states, using only Gaussian operations is impossible (?). It is closely related to entanglement distillation, discusses above. Several approaches have since been developed.

This no-go theorem does not apply if the initial states are non-Gaussian. One can also encode qubits using non-Gaussian continuous-variable states (?). Then error correction can be implemented using Gaussian operations, resulting in a Gaussian state. This protocol is known to be fault tolerant, however, the threshold requirements are quite stringent, such as needing the photon number in the squeezed state to be around 20 (?).

Another approach is to error correct on Gaussian states using non-Gaussian operations. Such protocols have been proposed (?) and demonstrated (?) for entanglement distillation. It has been shown that CV teleportation and distillation based on noiseless amplification can be combined to error correct Gaussian states against Gaussian noise (?), implementable using linear optics and photon counting. However, it is unclear whether such protocols can be made fault tolerant.

— *Ex luna scientia.*

XIX. ROUTING STRATEGIES

In Sec. IX.B we introduced the notion of network costs, strategies for allocating network resources in Sec. IX.D, and a general formalism for optimising strategies so as to minimise costs in Sec. IX.E. In this section we present some meaningful example strategies and associated pseudo-code fragments, illustrating the implementation of various aspects of strategies of practical real-world interest.

A. Single user

Let us begin our discussion of strategies by considering the simplest case of just a single user on the network. Consider the graph shown in Fig. 8. This is the same example used earlier, but now the edges have been weighted by some arbitrary cost metric. There are four routes from A to B . All have cost $c = 3$ except the route indicated by the

red arrow, which has cost $c = 2$. Clearly the latter is optimal in terms of cost minimisation, and any shortest-path algorithm applied between A and B will accurately come to that conclusion. Thus, single-user networks are very trivial to optimise, and there is no distinction between LOCAL and GLOBAL strategies.

The very trivial algorithm for this route finding is shown in Alg. XV, where the `ShortestPath()` function could be any of the existing, well-known shortest path algorithms (Sec. XI.B).

```

function Strategy.SingleUser(Packets):
    1. for(packet in Packets) {
        2. currentNode = packet.RoutingQueue.Pop()
        3. shortestRoute =
           ShortestPath(currentNode, packet.Recipient)
        4. packet.RoutingQueue.Flush()
        5. packet.RoutingQueue.Push(shortestRoute)
    6. }
    7.  $\square$ 
```

Algorithm XV: For a single user, a simple shortest-path algorithm necessarily finds the optimal route, as there is no potential for packet collisions or competition for network resources.

B. Multiple users

Next consider the more complex network shown in Fig. 52. We consider two sender/receiver pairs, $A_1 \rightarrow B_1$ and $A_2 \rightarrow B_2$. The available routes connecting both pairs overlap, creating competition for network resources.

Let us assume there are just two properties of interest when deciding strategies – cost in dollars (which may differ for different links), and availability (i.e how many states can the channel handle at once). Let c_1 be the dollar cost, and a_1 be the amount of available channel capacity. Our network is very primitive and each channel can only accommodate one state at a time. Thus, we let $a_1 = 1$ for all links, except for the one common to both R_1 and R_3 , ($R_1 \cap R_3$), which we invest more heavily into, since both routes are going to be wanting to use this link.

To define our net cost measure, we combine c_1 and a_1 according to,

$$\mathcal{S} : f_{\text{net}}(\vec{c}) = \begin{cases} c_1, & \text{if } a_1 > 0 \\ \infty, & \text{if } a_1 = 0 \end{cases}. \quad (19.1)$$

That is, provided bandwidth is available, the link will have the dollar cost c_1 . If no bandwidth is available, the cost is infinite, thereby removing the respective link from the graph.

Next the cost metrics are updated by the strategy \mathcal{S} following each communication. In this instance this simply

decrements the bandwidth attribute for the links that were utilised,

$$\mathcal{S} : a_1 \rightarrow a_1 - 1. \quad (19.2)$$

Suppose the strategy optimises the $A_1 \rightarrow B_1$ route first, yielding R_3 , before moving onto the $A_2 \rightarrow B_2$ route. In this case, the reduction of the bandwidth attribute signals that the cheapest route R_2 is no longer available to be utilised simultaneously to R_3 , and must therefore wait its turn on the following clock-cycle. Alternately, the strategy could employ R_1 for $A_2 \rightarrow B_2$, in which case their common link with capacity for two states would eliminate the competition between the two communications, allowing both to take place simultaneously. Thus, there is a tradeoff: for $A_2 \rightarrow B_2$, we could achieve a net cost of $c(A_2 \rightarrow B_2) = 5$, requiring 2 clock-cycles; or we could achieve simultaneous communication at the expense of increasing cost to $c(A_2 \rightarrow B_2) = 6$. This indicates that when choosing strategies, we must carefully define its goals.

Suppose net cost, rather than clock-cycles, was the key measure of interest. Then choosing the routes R_1 and R_3 would be the optimal choice. An optimal GLOBAL optimisation would recognise this. However, a LOCAL optimisation, based on choosing shortest-paths one-by-one for each sender/receiver pair, may or may not choose the optimal routes, depending on the order in which the decisions were made.

Suppose the $A_2 \rightarrow B_2$ route were optimised first. We would choose R_2 . Then there would be a traffic jam on the $A_1 \rightarrow B_1$ route, and it would necessarily have to wait its turn. In a time-critical application, where waiting is intolerable, this effectively renders the network useless to the first sender/receiver pair.

If, however, the $A_1 \rightarrow B_1$ were optimised first, choosing R_3 , then R_2 would be prohibited once the bandwidth attributes were updated, and the second best option, R_1 , would be chosen. Now both communications could take place simultaneously. So we see that the outcomes of LOCAL optimisations needn't always be consistent or unique. Rather, they can be highly dependent upon circumstantial issues, such as the arbitrary order in which routes are chosen for optimisation.

Generalising this to any number of users is a straightforward extension to the route optimisation problem, incurring a higher computational overhead due to the increased optimisation complexity.

In the upcoming sections we discuss multi-user strategies in more detail. None of these are true GLOBAL strategies, but nonetheless address some of the problems facing LOCAL strategies mentioned above.

Truly GLOBAL strategies could employ, for example, the vehicle routing problem (Sec. XI.I) or vehicle rescheduling problem (Sec. XI.J) algorithms. However, both of these are NP-hard in general. Thus, the approximation heuristics to be discussed in the following sections are highly applicable.

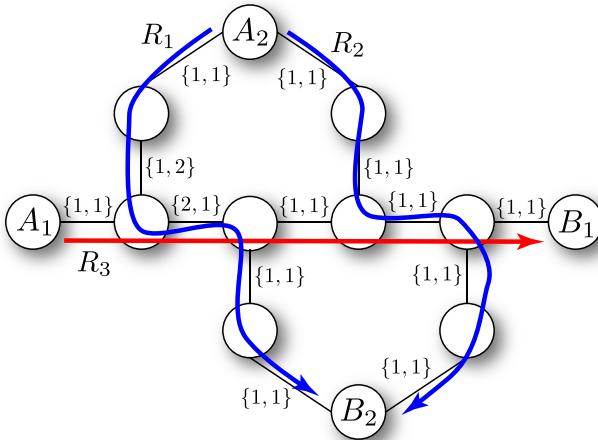


Figure 52: A simple network with two competing pairs of senders and receivers, $A_1 \rightarrow B_1$ and $A_2 \rightarrow B_2$. Edges are labelled by $\{b, d\}$, where b is the bandwidth attribute of the link (i.e. number of states that can be communicated simultaneously), and d is the cost metric associated with the link, e.g. loss in dB. (blue line) R_1 and R_2 are two routes from $A_2 \rightarrow B_2$. Either of these routes could be declared optimal, depending on the choice of cost function. For a trivial additive cost function, R_2 would be declared optimal. (red lines) R_3 is the optimal route from $A_1 \rightarrow B_1$.

C. Round robin

Perhaps the simplest and most elegant multi-user scheduling strategy is to borrow from the idea of time-division multiplexing for preemptive multitasking employed by UNIX operating systems. Here we simply put all live packets in a list, and go through the list, one-by-one, giving each packet an equal time-share of network resources, independent of costs. The algorithm for this is shown in Alg. XVI.

```

function Strategy.RoundRobin(Packets):
    1. for(packet ∈ Packets) {
        2. currentNode = packet.RoutingQueue.Pop()
        3. shortestRoute =
            ShortestPath(currentNode, packet.Recipient)
        4. packet.RoutingQueue.Flush()
        5. packet.RoutingQueue.Push(shortestRoute)
    }
    6. □

```

Algorithm XVI: In the ROUND ROBIN strategy we simply iterate through the list of active packets, with no regard for any metrics, or conflicts between them. Rather, we strive for perfect time-sharing equality, and every packet entirely ignores the actions of all other packets, performing a completely selfish optimisation strategy.

The ROUND ROBIN strategy can be considered base skeleton code for more sophisticated algorithms to build upon, simply by reordering the packet queue.

While such a protocol clearly ensures scheduling that gives all packets attention, it is the perfect example of an algorithm subject to the resource allocation imbalance discussed in Sec. XIX.B. Specifically, the routes being followed by some packets may systematically receive more favourable treatment than others, based on the arbitrary ordering of the list of packets. Also, equal timesharing fails to accommodate for the fact that some routes are inherently more costly than others and deserve a greater share of network resources.

D. Data priority

Are all men created equal? No. Some packets may inherently be more important than others, and ought to receive priority when allocating network resources. A simple variation on the ROUND ROBIN strategy is to, before iterating through the list of packets, order them according to a PRIORITY attribute. Thus, when applying a shortest-path algorithm, it is deemed most important to minimise the costs of the more important packets first.

This is trivially achieved by taking the existing ROUND ROBIN strategy, and first ordering the packet list by their priority attributes, i.e by inserting a new line 1, `Packets.SortByPriority()`.

E. Randomisation

The imbalance issue facing the ROUND ROBIN strategy (Sec. XIX.C) may be most trivially addressed using randomisation of the strategy, such that routes are optimised in an order chosen randomly each time. This would allow the different sender/receiver pairs to have equal access to network resources, when averaged over many network uses.

This is also a straightforward variation of the ROUND ROBIN strategy, achieved by first randomising the list of packets before the other stages, i.e insert a new line 1, `Packets.RandomizeOrder()`.

F. Cost priority

The RANDOM strategy overcomes one key problem facing any LOCAL optimisation strategy. But it is nonetheless merely a mild variation on the ROUND ROBIN strategy, guaranteeing equal time-share to each sender/receiver pair. But does equal time-sharing actually represent the best allocation of resources?

It isn't just the order in which routes are chosen, which creates imbalance between users. The costs and attributes of the routes themselves is inevitably biased more in favour of some users than others. To accommodate this we now introduce the COST PRIORITY strategy. Here, rather than prioritising packets on a random basis, or according

to a fixed, predetermined priority attribute, we do so according to their net accumulated cost. Those who have accumulated the highest cost will subsequently be treated with highest priority. This strategy effectively introduces a negative feedback loop into resource allocation, creating a self-regulating (and hopefully stable!) time-multiplexed packet-switched network. The pseudo-code for the COST PRIORITY strategy is shown in Alg. XVII.

```

function Strategy.CostPriority(Packets):
    1. packetsAndCosts = []
    2. for(packet $\in$ Packets) {
        3. cost = costFunction(packet)
        4. packetsAndCosts.Append([packet, cost])
    5. }
    6. sorted =
        SortByCostDescending(packetsAndCosts)
    7. for(packet $\in$ sorted) {
        8. currentNode = packet.RoutingQueue.Pop()
        9. shortestRoute =
            ShortestPath(currentNode, packet.Recipient)
        10. packet.RoutingQueue.Flush()
        11. packet.RoutingQueue.Push(shortestRoute)
    12. }
    13.  $\square$ 
```

Algorithm XVII: The COST PRIORITY strategy scheduling algorithm that gives highest routing priority to PACKETS with the highest accumulated cost (i.e which have suffered the most). The as-yet undefined `costFunction()`, which refers to f_{cost} from Eq. (9.3), is where the details of the priority decisions are made, which could be entirely arbitrary. In this example, the shortest route is recalculated at each step, based on the expectation that network metrics are dynamic.

This is an example of a GREEDY optimisation algorithm, which attempts to optimise routing by always optimising the most desperate packets first, in descending order down to the least. It is well-known that GREEDY algorithms often do not find global optima. Nonetheless, this approach improves on the previous multi-user protocols.

Let us consider a simple example scenario. Imagine we begin with an ordinary network graph, with edges weighted by costs and attributes. For generality, we will additionally assume the available network resources are very dynamic and unpredictable. The costs associated with links are at the whim of market forces we do not understand (do we ever?). And, for the sake of example, and to make matters worse, the links have been very unreliable lately, and are routinely dropping in and out – ‘blackouts’. This effectively rules out *a priori* route optimisation, requiring something dynamic.

There are many users on the network, with many active packets at any give time, but because of the constant oscillations in network resources, some packets have received

second-class treatment, and through neglect accumulated an unfair share of state degradation. This simple toy model is, at least qualitatively, something that could arise quite naturally in networks with constrained or unreliable resources.

Let us define an example COST PRIORITY strategy using the following:

- LATENCY cost: How long has the packet has been in transit for? This is actually a very general cost metric, since any other cost metric measured in units per time will be directly proportional to this metric. That is, loss, fidelity, purity, efficiency, and so on, all mirror this metric when expressed on a decibel scale. Of course, the same strategy could have easily been applied to any other cost metric.
- BLACKOUT attribute: Is our unreliable link actually working right now? A given link will have probability p_{op} of being operational at any given time, chosen independently for each link at each clock-cycle. The `Attributes.Update()` function from Alg. III is responsible for implementing this.
- COST FUNCTION (f_{cost} , `costFunction()` in Alg. XVII): The strategy must make sensible decisions based upon only the above two parameters. Because of the previously mentioned generality of the LATENCY metric, we would like the net cost to directly reflect this metric, but only of course, if the link is operational. If it is not, then that link must be ruled out entirely by assigning it an infinite cost. Thus, we simply choose,

$$\mathcal{S} : f_{\text{cost}}(c, a) = \begin{cases} c, & \text{if } a = \text{True} \\ \infty, & \text{if } a = \text{False} \end{cases}. \quad (19.3)$$

Note that different packets could be associated with different net cost functions, f_{net} , to accommodate for the different QoS requirements of different users and messages.

In other words, the net cost is taken directly from the underlying cost metric, and modulated by an attribute, yielding a net cost for each packet, which is used to determine which packets receive priority.

This provides us with a simple illustration of how costs and attributes can compliment one another to yield meaningful strategies, that improve network performance over naïve, but well-intentioned, time-sharing approaches.

G. All or nothing

In some cases, end-user applications may have strict QoS constraints associated with any data they receive. For example, in a time-critical enterprise, say high-frequency trading, receiving information a millisecond too late is

worthless, and it would be best to discard the out of date information to free up bandwidth for the next round of information. Alternately, if the fidelity of a state is required to strictly fall within a fault-tolerance threshold, it will be useless if the threshold is violated. In such a context, the STRATEGY will apply hard boundaries on QoS metrics, discarding anything violating it, after which some other STRATEGY is applied. The algorithm is summarised in Alg. XVIII.

```

function Strategy.AllOrNothing(Packets,
threshold):
    1. for(packet∈Packets) {
        2. cost = packet.costFunction()
        3. if(cost ≥ threshold) {
            4. packet.Sender.Notify(FAILURE)
            5. packet.Recipient.Notify(FAILURE)
            6. packet.Discard()
        7. }
    8. }
    9. Strategy.SomeOtherStrategy(Packets)
10. □

```

Algorithm XVIII: The ALL OR NOTHING strategy. If the net cost of a packet exceeds a certain `threshold`, it is discarded outright, and the sender and recipient notified.

H. Optimal flow

In Sec. IX.C we introduced flow networks as a means for analysing networks where maximising network flow (throughput) is the primary objective. Formulating our quantum networks in this manner is extremely convenient since, combined with our existing definitions for cost metrics and attributes, we can easily exploit a plethora of known results from flow network theory.

As an example of how load allocation might be applied in a simple network, consider again the network shown in Fig. 8, where the edge weights are regular cost metrics (not capacities). Alice wishes to send two packets to Bob, simultaneously if possible. Clearly she would transmit her first packet over the $A \rightarrow F \rightarrow B$ route, since this has lowest cost. But let us assume that every link has a maximum capacity of one packet per unit time. In this case Alice will be unable to send her second packet via the same route and must instead resort to using $A \rightarrow C \rightarrow B$ or $A \rightarrow D \rightarrow B$. The optimisation is straightforward in this instance. However, in general these types of optimisations are somewhat more involved.

These scenarios are handled by flow network optimisation algorithms, of which there are many. We discuss a few of the most relevant ones for our purposes in Sec. XI. Note that these algorithms are GLOBAL optimisation al-

gorithms, requiring complete knowledge of the status of the entire network to perform the optimisation.

The routing strategy is very straightforward, shown in Alg. XIX, since the GLOBAL flow-optimisation algorithm completely specifies the entire configuration of routes through the network.

```

function Strategy.OptimalFlow(Packets):
    1. routes = Packets.OptimalFlowRoutes()
    2. for(packet∈Packets) {
        3. packet.RoutingQueue.Flush()
        4. packet.RoutingQueue.Push(routes[packet])
    5. }
    6. □

```

Algorithm XIX: A generic optimal flow routing strategy. PACKETS is the array of all packets that ought to be transmitted simultaneously, which are collectively optimised using some flow optimisation algorithm before undergoing transport.

— *Disce quasi semper victurus vive quasi cras moritrus.*

XX. INTERCONNECTING & INTERFACING QUANTUM NETWORKS

ANY global-scale network will inevitably comprise participants choosing to go about things their own way. The physical architecture and medium may vary from one subnetwork to the next, as may the QTCP policies they adopt. The key then is to construct efficient *interconnects* between different levels of the network hierarchy, each of which may subscribe to their own QTCP policies and cross between different physical mediums. Note that the QTCP protocol presented here does not enforce any particular networking policies, but rather provides a high-level framework that can be customised essentially arbitrarily.

For example, the cost metrics and attributes employed at the intercontinental level would most certainly be very different to those in a small LAN. A small LAN might be running applications whereby they can easily reproduce packets and thereby tolerate packet loss. But for a warehouse-scale commercial quantum computing enterprise, responsible for performing one stage of a distributed quantum computation, the loss of a single packet could be extremely costly, requiring the entire computation to be performed completely from scratch due to no-cloning and no-measurement limitations, something that may not come cheaply.

Such interconnects will typically comprise a combination of:

- Packet switching: such that packets can be arbitrarily switched between the different levels of the network hierarchy.

- Physical interface: interconnect may be switching between different media. Such physical interfaces have costs associated with them. For example, coupling between free-space and fibre is typically very lossy. Sec. XX.A discusses optical interfacing with matter qubits, and Sec. XLII.F discusses hybrid architectures, where optics mediates entanglement generation between matter qubits.
- Quantum memory: such that data can be buffered while it awaits its turn at being switched between networks, as different networks may have different loads and operate at different clock-rates. This is discussed in Sec. XXVI.
- Packet format conversion: different levels of the network hierarchy may be employing entirely different cost metrics, attributes, and cost functions, requiring packet headers to be reformatted upon switching between networks.

The packet switching and quantum memory are implemented as quantum processes at nodes, using the usual quantum process formalism. The physical interface between different mediums, if there is one, could be very diverse, encompassing many types of physical systems, but can always be characterised using the quantum process formalism. Packet headers, which contain all formatting, cost, and routing information are represented entirely classically and communicated entirely by the classical network. Thus, this operation also takes place at nodes, but no quantum processes are taking place.

A. Optical interfacing

Unless the entire pipeline of quantum operations through the course of a protocol is all-optical, there will be a need to exchange information between physical systems, for example via light-matter interactions (?). We will now discuss optical interfacing with some of the significant types of matter systems, such that their intercommunication can be optically mediated over the network.

1. Two-level systems

The archetypal interface is that between a photonic qubit in the $\{|0\rangle, |1\rangle\}$ photon-number basis, and a two-level matter qubit in the $|g\rangle$ (ground) and $|e\rangle$ (excited) state basis. The logical qubit is defined as,

$$\begin{aligned} |0\rangle_L &\equiv |g\rangle, \\ |1\rangle_L &\equiv |e\rangle. \end{aligned} \quad (20.1)$$

. Examples include atoms in cavities, NV centres, and engineered quantum dots.

In the case of a photon interacting with a two-level matter qubit, the interface can be expressed via the Jaynes-Cummings interaction Hamiltonian of the form,

$$\hat{H}_{\text{int}} = \hbar\chi(\hat{a}\hat{\sigma}^+ + \hat{a}^\dagger\hat{\sigma}^-), \quad (20.2)$$

where \hat{a} (\hat{a}^\dagger) is the photonic annihilation (creation) operator, $\hat{\sigma}^\pm$ are the Pauli spin-flip operators, and χ is the interaction strength. The interpretation of this Hamiltonian is very clear upon inspection – the annihilation (creation) of a photon is associated with the excitation (relaxation) of the two-level matter system, thereby directly coherently exchanging quantum information between the two systems, as shown in Fig. 53.

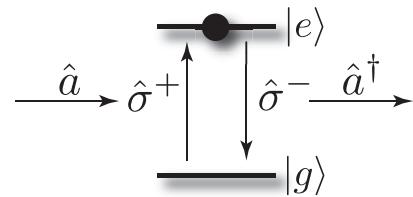


Figure 53: Light-matter interfacing between a single-photon state (\hat{a}, \hat{a}^\dagger) and a two-level matter qubit ($|g\rangle, |e\rangle$). The absorption (emission) of a photon is associated with the excitation (relaxation) of the matter qubit ($\hat{\sigma}^\pm$).

2. λ -configuration systems

Alternately, one can easily optically interface with a λ -configuration system, as shown in Fig. 54. Here there are two degenerate ground states representing the logical qubit basis states ($|0\rangle_L \equiv |\uparrow\rangle, |1\rangle_L \equiv |\downarrow\rangle$), one of which may undergo a transition to an excited state, $|e\rangle$. By pumping the system to the excited state and waiting for a coherent relaxation, the emitted photon may be used to couple the qubit state of the λ -configuration to an optical mode, mapping the qubit value of the matter qubit to a photon-number representation.

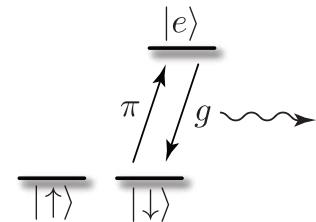


Figure 54: Light-matter interfacing between an optical mode and a λ -configuration system. The two degenerate ground states represent the logical qubit ($|0\rangle_L \equiv |\uparrow\rangle, |1\rangle_L \equiv |\downarrow\rangle$), only one of which may undergo transition to the excited state $|e\rangle$. Upon pumping the $|\downarrow\rangle \rightarrow |e\rangle$ transition with a π -pulse, a relaxation back to the ground state maps the logical qubit value to photon-number.

3. Atomic ensembles

In addition to single atoms with well-defined electronic structure, atomic ensembles (??) can be used, whereby the absorption of a photon creates a *collective excitation* – a superposition of a single excitation across all the atoms in the ensemble. Specifically, excitations are represented using collective excitation operators,

$$\hat{S}^\dagger = \frac{1}{\sqrt{N}} \sum_{i=1}^N \hat{S}_i^\dagger, \quad (20.3)$$

where,

$$\hat{S}_i^\dagger = |e\rangle_i \langle g|_i, \quad (20.4)$$

is the excitation operator for the i th atom in the ensemble, $|g\rangle_i$ and $|e\rangle_i$ are the ground and excited states for the i th particle, and there are N atoms. The state of a single collective excitation is then given by,

$$|\psi_{\text{collective}}\rangle = \hat{S}^\dagger |g\rangle^{\otimes N}. \quad (20.5)$$

Atomic ensembles are essentially well-engineered clouds of atomic gasses, trapped in a glass container, coupled to an optical mode. Atomic ensembles have been demonstrated with extremely long coherence lifetimes (T_2 -times on the order of milliseconds), operating at room temperatures (a very attractive feature on its own). They exhibit *collective enhancement* in their coupling to the optical mode – the optical coupling strength is amplified by a factor quadratic in N compared to single-atom optical coupling, mitigating the need for a cavity.

The collective excitations exhibit the same general mathematical structure as single-atom excitations – the absorption (emission) of a single photon is associated with a single collective excitation (relaxation), albeit with the favourable collective enhancement in the coupling strength.

To couple with a polarisation-encoded photonic qubit, a PBS can be employed to spatially separate the horizontal and vertical modes, each of which couples to a separate atomic ensemble, which jointly represent the logical qubit, as shown in Fig. 55.

Atomic ensembles have been proposed as quantum memories, given their long coherence lifetimes. Additionally, a protocol for universal cluster state quantum computation (Sec. XL.B) based upon atomic ensemble qubits has been described (?).

Essentially, the long coherence lifetimes of collective excitations owes to the fact that the excitation is effectively encoded as a W-state (Sec. XXIII.F), an equal superposition of a single excitation across many (N) atoms, of the

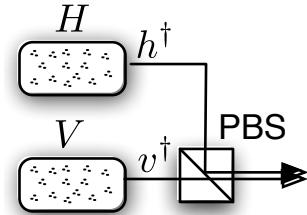


Figure 55: Coupling a polarisation-encoded photonic qubit to a pair of atomic ensembles, each of which corresponds to one of the qubit's logical basis states ($|0\rangle$ or $|1\rangle$). The horizontal and vertical components of the photonic qubit are spatially separated using a PBS, which subsequently independently interface with distinct atomic ensembles via collective excitation.

form,

$$\begin{aligned} |\psi_W^{(N)}\rangle = & \frac{1}{\sqrt{N}}(|e, g, g, \dots\rangle \\ & + |g, e, g, \dots\rangle \\ & + |g, g, e, \dots\rangle \\ & + \dots \\ & + |g, g, \dots, e\rangle). \end{aligned} \quad (20.6)$$

W-states are favourable from a decoherence perspective as tracing out a single particle has minimal impact on the coherence of the residual state, which preserves most entanglement, with this robustness growing with the number of particles. This is in stark contrast to GHZ states, which completely decohere under the loss of just a single particle.

Specifically, if $|\psi_W^{(N)}\rangle$ is the N -particle W-state (collective excitation), tracing out a single particle yields,

$$\begin{aligned} \hat{\rho}_{\text{tr}} &= \text{tr}_1(|\psi_W^{(N)}\rangle \langle \psi_W^{(N)}|) \\ &= \left(1 - \frac{1}{N}\right) |\psi_W^{(N-1)}\rangle \langle \psi_W^{(N-1)}| + \frac{1}{N} (|g\rangle \langle g|)^{\otimes (N-1)}, \end{aligned} \quad (20.7)$$

which for $N \gg 1$ approaches the pure state $|\psi_W^{(N-1)}\rangle$, i.e a W-state with one fewer particles.

4. Superconducting qubits

— by Chandrashekhar Radhakrishnan.

In the context of superconducting qubits (Sec. XLII.I), the energy difference between the energy levels being utilised to encode the qubit is extremely small. Therefore photons coupled to these transitions sit in the microwave regime, whose wavelength lies in the range $\lambda \sim 100\mu\text{m}-1\text{m}$. x

Information transfer between distinct superconducting qubits is achieved using a resonator, which acts as a quantum data bus. A simple resonator is an LC circuit,

which can support only one frequency mode, but a waveguide resonator can support multiple modes. In general, the transmission line circuits used in non-linear quantum electric circuits are in the form of coplanar waveguides. These waveguides are engineered to handle a particular set of frequencies, and produce transmission lines with tuneable frequency. Tuneable resonators are very important in quantum optics, and are useful in implementing controllable coupling between different quantum elements, and also in shaping photon wave-packets.

In cavity quantum electrodynamics (QED) the interaction of a natural atom with an optical photon in the visible wavelength regime is considered. Similarly, the interaction between quantum non-linear electrical circuits and microwave photons are investigated in circuit QED. The coupling-strength between a natural atom and visible light photon is fixed, where atoms couple weakly with photons (?). Meanwhile, the coupling-strength between a superconducting qubit and a microwave can be manipulated by engineering the parameters of the qubit and resonator, yielding strong and ultra-strong coupling between qubits and photons (?) (see Fig. 56). Furthermore, the coupling between an atom and a photon can be tuned dynamically during the course of an experiment. Several quantum optics components such as mirrors, beamsplitters, circulators and switches can also be designed based on quantum electric circuits.

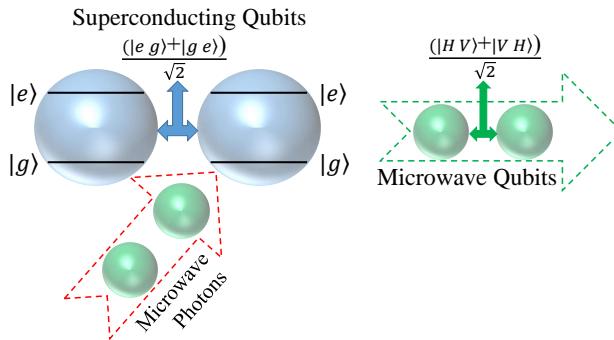


Figure 56: Schematic sketch of the interaction between superconducting qubits and microwave qubits. The superconducting qubits are in the entangled state $\frac{1}{\sqrt{2}}(|e, g\rangle + |g, e\rangle)$, where $|g\rangle$ and $|e\rangle$ are the ground and excited states of the superconducting qubits. These are subjected to interact with the flying qubits, the microwave qubit (red dashed arrow). The photons become entangled and the output state of the photons are $\frac{1}{\sqrt{2}}(|H, V\rangle + |V, H\rangle)$, where $|H\rangle$ and $|V\rangle$ are the horizontal and vertical polarisation modes respectively.

Due to scalability requirements, superconducting qubits are the most widely used qubit implementation used today. The energy-level spacing in superconducting qubits lie in the microwave regime. Hence to control and transfer information from a superconducting qubit one might need to use microwave photons. In principle, we can use microwave photons to transfer information from one node

to another, but such a transmission process is extremely lossy. Also, such processes have very demanding technical requirements, like the design of specialised Niobium waveguides, maintained at extremely low temperatures. Hence it isn't feasible to use microwave photons for the long-distance transfer of quantum information. Meanwhile, it is well known that photons in the visible spectrum can be transmitted easily using optical fibres, with favourable efficiency.

To convert microwave photons to optical photons we can use a quantum transducer. A sketch of a typical design for a quantum transducer is shown in Fig. 57 through a flowchart diagram. The quantum computer is made up of superconducting qubits, which feed quantum information to microwave qubits. The microwave qubits are then interfaced to optical qubits in the visible spectrum through a 3-level quantum system, which can couple at both microwave and optical frequencies. These steps should be reversible. Hence it should be possible to convert the optical qubits back into microwave qubits at the receiving end.



Figure 57: Block diagram for the quantum transducer. The quantum computer comprising superconducting qubits couples with the microwave qubit. The microwave qubit and the photonic qubit are coupled by a three-level system in which the energy difference between the levels correspond to both microwave and visible optical frequencies. The thick arrows denote the forward process, required to convert the microwave to optical frequencies, and the dashed arrows represent the reverse process, required at the receiving end to convert the optical frequency to a microwave frequency.

There are several proposals for quantum transducers in existence and they can be classified into two major classes:

- Opto-mechanical (??????).
- Spin-ensembles (?).

The opto-mechanical quantum transducer (see Fig. 58), as the name suggests, combines optical components with a nano-mechanical resonator and converts the microwave photon into a phonon (acoustic) mode. The acoustic mode is then transmitted via waveguides. Since we need to fabricate waveguides with very high precision to transmit phonon modes, this scheme is not suitable for communicating between two distant quantum computers. A coupling of the phonon mode to the optical photon mode in the visible region was suggested to enable long-distance transfer of quantum information.

The Hamiltonian of an opto-mechanical quantum transducer, which converts a microwave photon to optical photon using an intermediate nano-mechanical resonator

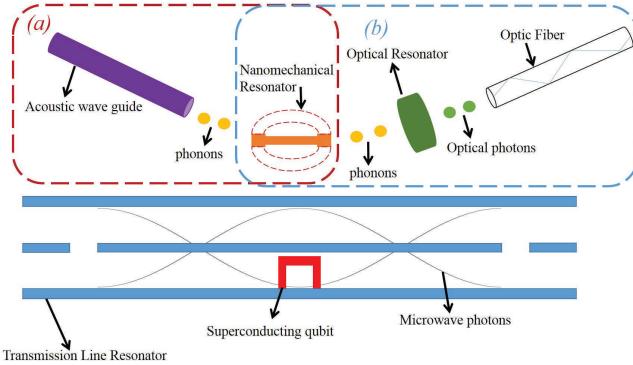


Figure 58: Scheme for an opto-mechanics-based quantum transducer. There are two possible ways of converting microwaves. The final steps for the process in which acoustic modes are transported using waveguides is enclosed by a brown coloured box with dashed outlines labeled (a). Similarly the blue coloured box labeled (b) represents the scheme where phonons are converted to optical photons which are then transmitted via optical fibres. The initial few steps consisting of the superconducting qubit, transmission line resonator and microwaves are common to both processes.

reads,

$$\hat{H} = \hbar\omega_1 \hat{a}_1^\dagger \hat{a}_1 + \hbar\omega_2 \hat{a}_2^\dagger \hat{a}_2 + \hbar\Omega \hat{b}^\dagger \hat{b} + \hbar g (\hat{b} + \hat{b}^\dagger)(\hat{a}_2^\dagger \hat{a}_1 + \hat{a}_1^\dagger \hat{a}_2), \quad (20.8)$$

where ω_1 (ω_2) is the frequency of the microwave (optical) photon, and Ω is phonon frequency. The operators \hat{a}_1 (\hat{a}_1^\dagger) and \hat{a}_2 (\hat{a}_2^\dagger) denote the annihilation (creation) operators corresponding to the microwave and optical photons respectively. Meanwhile, \hat{b}^\dagger (\hat{b}) denotes the phonon creation (annihilation) operator corresponding to the phonons. The factor g is the coupling-strength between the microwave, phonon and photon modes. This design for a quantum transducer is widely preferred, since the optical photons in the visible spectrum can be transmitted over long distances using fibre optics. But the scheme requires two intermediate conversions, each of which reduces overall efficiency.

The spin-ensemble-based quantum transducer (see Fig. 59) is an alternative to the opto-mechanical quantum transducer. In this scheme an ensemble of spins interact with microwave qubits via magnetic dipole coupling, while the superconducting qubits interact via electric dipole coupling with the microwave coupling. The Hamiltonian for such a system is,

$$\hat{H} = \hat{H}_{\text{mw}} + \hat{H}_{\text{spin}} + \hat{H}_{\text{opt}}, \quad (20.9)$$

where,

$$\begin{aligned} \hat{H}_{\text{mw}} &= \hbar\omega_{\text{sq}} \hat{\sigma}_{\text{sq}}^\dagger \hat{\sigma}_{\text{sq}} + \hbar\omega_\mu \hat{a}^\dagger \hat{a} + \hbar g_\mu (\hat{a}^\dagger \hat{\sigma}_{\text{sq}} + \hat{a} \hat{\sigma}_{\text{sq}}^\dagger), \\ \hat{H}_{\text{spin}} &= \hbar g_s (\hat{\sigma}_{\text{ba}}^\dagger \hat{\sigma}_{\text{ba}} + \hat{\sigma}_{\text{bs}}^\dagger \hat{\sigma}_{\text{bs}}), \\ \hat{H}_{\text{opt}} &= \hbar g_{ab} (\hat{\sigma}_{\text{ba}}^\dagger \hat{c} + \hat{c}^\dagger \hat{\sigma}_{\text{ba}}). \end{aligned} \quad (20.10)$$

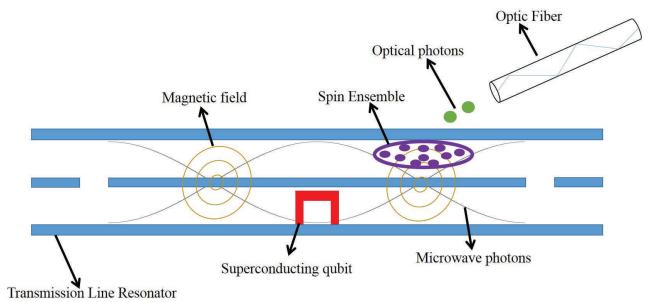


Figure 59: The spin-ensemble-based quantum transducer. Microwave photons are converted to optical photons via a spin-ensemble.

The factor ω_{sq} is the frequency of the superconducting qubit, and $\hat{\sigma}_{\text{sq}}^\dagger$ and $\hat{\sigma}_{\text{sq}}$ are the raising and lowering operators corresponding to the superconducting qubits. The frequency of the microwaves is given by ω_μ , and the \hat{a}^\dagger and \hat{a} are the creation and annihilation operators for the microwave photons. The factor g_s is the coupling strength between the various levels of the spin, and $\hat{\sigma}_{\text{ba}}$ and $\hat{\sigma}_{\text{bs}}$ are the spin operators corresponding to the transition between the level a , b and s . Finally, the coupling strength of the spin interaction with the photon is denoted by g_{ab} , and \hat{c}^\dagger (\hat{c}) is the creation (annihilation) operator corresponding for the photon. Again this is a two-step process, which is in addition beset with the problem of inhomogenous line broadening. The design of an experimental, high-fidelity quantum transducer is still an open and ongoing challenge in the field of quantum technology.

— *Vir sapit qui pauca loquitur.*

XXI. OPTICAL ROUTERS

PERHAPS the most fundamental building block in any network is routers, devices which switch data packets between multiple inputs and outputs so as to relay them to a destination. Indeed, in many real-world networks, many nodes will purely implement routing, and nothing more elaborate such as computations or other end-user protocols, to be discussed in Part. V.

We now discuss the implementation of optical routers, beginning with the simplest two-port switch, upon which we build to construct more general and powerful routers.

There are many parameters of interest characterising the operation of optical routes. We will introduce the terminology convention that:

- *Ports:* number of input and output optical modes in a device.
- *Channels:* number of simultaneous communications streams running in parallel through the device.
- *Optical depth:* number of primitive optical elements/devices an optical path traverses through

the course of its trajectory from input to output.

- *Directionality*: whether information is transferred in one (unidirectional) or two (bidirectional) directions.
- *Switching time*: time for the switch to be reconfigured from one state to another.
- *Delay time*: time taken by signal to reach the output line of the switch from input.
- *Throughput*: maximum data rate that can flow through the switch.
- *Switching energy*: energy input required for activating and deactivating the switch.
- *Power dissipation*: power dissipated during the process of switching.
- *Insertion loss*: loss in signal power when the switch is connected.
- *Crosstalk*: coupling to other optical modes.

A summary of the routing devices we consider, and their associated resource requirements, is provided in Tab. III.

Of course, real-world routers will not only switch optical paths, but also implement some (probably undesired) quantum processes across those paths, such as a loss channel or temporal mode-mismatch. Thus, proper analysis of optical router performance in quantum networks requires treating them as legitimate nodes in the network graph, with associated costs and attributes, as per the QTCP framework.

A. Mechanical switches

Most obviously, optical switching could be performed mechanically, by physically displacing fibre endpoints, directing them towards different routes³⁰. Such switches have found use in other areas, but are not particularly appropriate for quantum information processing applications, as they are extremely slow compared to electro- or acousto-optic technologies. Certainly, mechanical switching would not be applicable to optical fast-feedforward, such as that required by optical quantum computing, on the order of nanoseconds.

A second disadvantage of mechanical switches is that the introduction of moving parts into quantum optics protocols makes optical stabilisation extremely challenging. The mechanical control required to preserve wavelength-level coherence, for example, is effectively ruled out by moving mechanical parts.

³⁰ Remember, the telephone network used to be mechanically routed by human switchboard operators, manually routing point-to-point connections!

B. Interferometric switches

— by Rohit Ramakrishnan.

Interferometric routers are based on the principle that the evolution implemented by interferometers are in general highly dependent on the phase relationships within them. This reduces the seemingly uphill task of high-speed, dynamic switching between modes to the problem of implementing dynamically-controllable phases. Thankfully there are a number of techniques for implementing such phase-switching. We will discuss these phase-modulation techniques, before moving onto combining them into more complex routing systems.

A phase modulator is a classically-controlled device that lets us tune the local phase accumulated by an optical path, ideally over the full range of $\{0, 2\pi\}$. These may be implemented in several ways:

1. Electro-optic modulators

Electro-optics modulators (EOMs) are based on anisotropic materials, in which the refractive index changes according to an applied electric field. There are two primary variations on this:

- Pockel's effect: a linear electro-optic effect, where the refractive index change is proportional to the applied electric field.
- Kerr's effect: a quadratic electro-optic effect, where the refractive index change is proportional to the square of the applied electric field.

These changes in refractive index are typically small, such that the effects are significant over propagation distances larger than the light's wavelength. For example, in a material where the refractive index increases by 10^{-4} , an optical wave propagating a distance of 10^{-4} wavelengths will acquire a phase-shift of 2π .

The refractive index of an electro-optic medium is a function $n(E)$ of the applied electric field E . This function varies only slightly with E , such that using a Taylor series expansion about $E = 0$ we obtain,

$$n(E) = n + a_1 E + \frac{1}{2} a_2 E^2 + \dots \quad (21.1)$$

In a Pockel's medium this relation becomes (after approximating and simplifying),

$$n(E) = n - \frac{1}{2} \chi n^3 E, \quad (21.2)$$

where χ is called the Pockel's coefficient or linear electro-optic coefficient. Typical values of χ lie in the range $10^{-12} - 10^{-10} \text{ mV}^{-1}$. The most common crystals used as the medium for Pockel's cells are $\text{NH}_4\text{H}_2\text{PO}_4$ (ADP), KH_2PO_4 (KDP), LiNbO_3 , LiTaO_3 , and CdTe .

Device	Resource requirements	Optical depth
Two-channel two-port switch	$N_{\text{bs}} = 2, N_{\text{ps}} = 1$	$d = 1$
Linear n -port multiplexer	$N_s = n - 1$	$1 \leq d \leq n - 1$
Pyramid n -port multiplexer	$N_s = n - 1$	$d = \log_2 n$
Single-channel multi-port switch (linear)	$N_s = 2n - 3$	$2 \leq d \leq 2n - 3$
Single-channel multi-port switch (pyramid)	$N_s = 2n - 3$	$d = 2 \log_2 n - 1$
Multi-channel multi-port switch	$N_s = \left[\frac{n^2}{2} \right] - n + 1$	$\left[\frac{n}{2} \right] \leq d \leq n - 1$
Crossbar switch	$N_s = n^2$	$1 \leq d \leq 2n - 1$

Table III: Summary of different primitives for constructing optical routers. $n, N_{\text{bs}}, N_{\text{ps}}$ and N_s are the number of input/output ports, beamsplitters, phase-shifters, and two-port switches respectively. d is the optical depth (in units of number of two-port switches). Since all of the multi-port devices are constructed from two-port switches, in all cases $N_{\text{bs}} = 2N_s$ and $N_{\text{ps}} = N_s$.

In a centrosymmetric material or Kerr's medium this relation becomes (again after approximating and simplifying),

$$n(E) = n - \frac{1}{2}\xi n^3 E^2, \quad (21.3)$$

where ξ is the Kerr's coefficient or the quadratic electro-optic coefficient. Typical values of ξ lies in the range $10^{-18} - 10^{-14} \text{ m}^2 \text{ V}^{-2}$.

The refractive index profiles as a function of applied electric field strength for Kerr's and Pockel's mediums are shown in Fig. 60.

Light transmitted through a transparent plate with controllable refractive index undergoes a controllable phase-shift. This plate can be used as an optical phase modulator.

Consider light traversing a Pockels cell of length L to which an electric field E is applied. The phase-shift undergone is given by,

$$\phi \approx \phi_0 - \pi \frac{\chi n^3 E L}{\lambda_0}, \quad (21.4)$$

where,

$$\phi_0 = \frac{2\pi n L}{\lambda_0}. \quad (21.5)$$

If the electric field generated by applying a voltage V across the faces of the cell of dimension d is,

$$E = \frac{V}{d}, \quad (21.6)$$

then,

$$\phi = \phi_0 - \pi \frac{V}{V_\pi}, \quad (21.7)$$

where,

$$V_\pi = \frac{d\lambda_0}{L\chi n^3}, \quad (21.8)$$

is the half-wave voltage, the voltage at which the phase-shift changes by π .

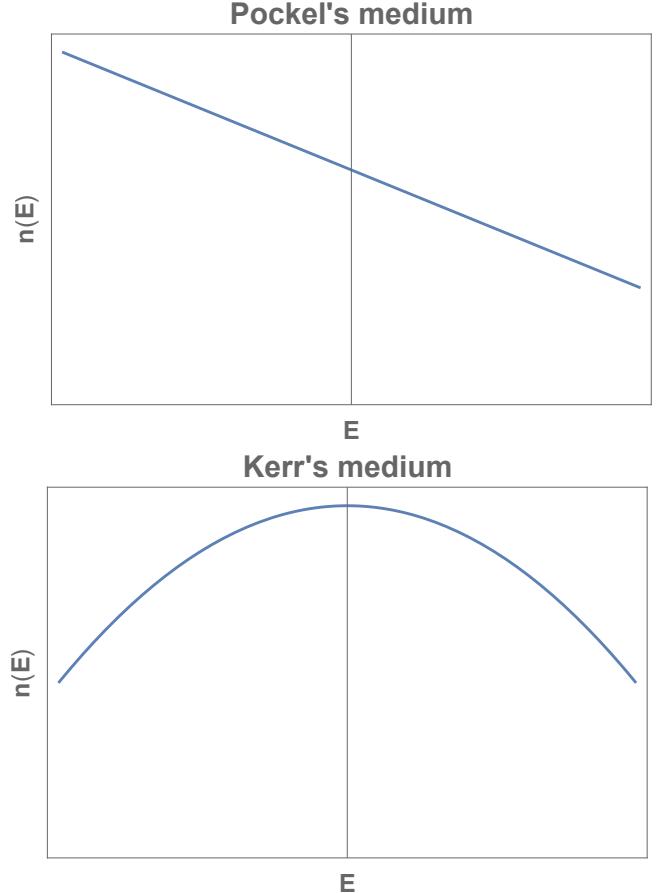


Figure 60: Dependence of refractive index on electric field in: Pockel's medium, exhibiting linear electric field dependence; Kerr's medium, exhibiting quadratic electric field dependence. Graphs express qualitative behaviour only, hence no numbers are provided.

The electric field is applied either perpendicular (transverse modulators) or parallel (longitudinal modulators) to the direction of the propagation light. The value of the electro-optic coefficient χ depends on the directions of propagation and the applied field. The speed of operation is limited by the capacitive effects and the transit time of the signal through the material.

State of the art electro-optic modulators are integrated

optic devices based on LiNbO_3 , in which materials like titanium are used to increase the refractive index. The typical operation speed is above 100GHz. Light signals can be coupled in and out using optical fibres.

2. Acousto-optic modulators

Sound, or acoustic waves, are vibrations that travel through a medium with a velocity characteristic of the medium. This can create perturbations in the refractive index of the optical medium, thus modifying the velocity of light passing through the medium. Thus sound can be used to modify the effect of the medium on light. That is, sound can control the direction of propagation of light. This acousto-optic effect is used to make a variety of devices like optical modulators, switches, deflectors, filters, isolators, frequency shifters and spectrum analysers. This is shown in Fig. 61.

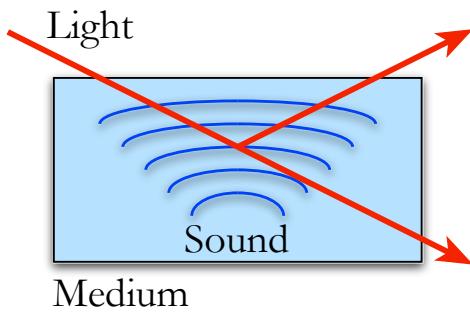


Figure 61: Acousto-optic modulators as a classically-controlled optical switch. The light signal is refracted depending on the applied sound wave.

According to quantum theory, a light wave of angular frequency ω and wave-vector k is a stream of photons each with energy $\hbar\omega$ and angular momentum $\hbar k$. Additionally, acoustic waves with frequency Ω and wave-vector q are a stream of phonons each with energy $\hbar\Omega$ and momentum $\hbar q$. When light and sound interact, a photon combines with a phonon to generate a new photon with energy and wave-vector subject to energy and momentum conservation laws,

$$\begin{aligned}\hbar\omega_r &= \hbar\omega + \hbar\Omega, \\ \hbar k_r &= \hbar k + \hbar q.\end{aligned}\quad (21.9)$$

The associated energy conservation diagram is shown in Fig. 62.

Since the intensity of the reflected light is proportional to the intensity of the sound (provided the intensity of sound is low), the intensity of reflected light can be varied proportionally by using an electrically controlled acoustic transducer. This device can be used as a linear modulator of light.

When the acoustic power increases beyond a certain threshold level, total reflection of light occurs whereby

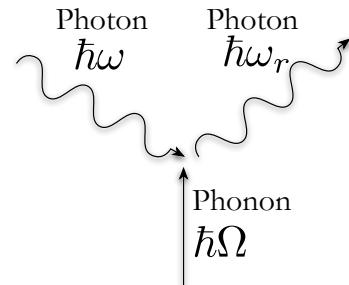


Figure 62: Energy diagram for an acousto-optic modulator, based on Eq. (21.9). Energy and momentum must be conserved from the incident photon of energy $\hbar\omega$ and phonon of energy $\hbar\Omega$, yielding a scattered photon of energy $\hbar\omega_r$.

the modulator behaves as an optical switch. By switching the sound on and off, the reflected light can be turned on and off, yielding an acoustically-controlled switch.

3. Magneto-optic modulators

In the presence of a static magnetic field, certain materials act as polarisation rotators, known as the Faraday effect. The angle of rotation is proportional to distance and the rotary power ρ (angle per unit length), which is proportional to the component B of the magnetic flux density in the direction of wave propagation,

$$\rho = VB, \quad (21.10)$$

where V is known as the Verdet constant, which is a function of wavelength λ_0 .

Examples of materials that exhibit the Faraday effect include glass, Yttrium-iron-garnet (YIG), Terbium-gallium-garnet (TGG) and Terbium-aluminium-garnet (TbAlG).

A simple form of magneto-optic modulator comprises a parallel-sided disk of material placed in a small coil. An alternating current in the coil provides a magnetic field normal to the plane of the disk. The material becomes magnetised in this direction and light propagating through the disk undergoes a polarisation rotation about its plane of polarisation. The modulation of the angle of the plane of polarisation induced by the alternating current may be converted to amplitude modulation by subsequently passing the beam through a polariser.

C. Two-channel two-port switches

The elementary primitive switch from which more complicated routers may be constructed is the two-channel two-port switch. This switch may be constructed from a Mach-Zehnder interferometer, with a classically-controlled phase-shifter in one arm. By switching the phase to either $\phi = 0$ or $\phi = \pi$, the MZ may be tuned to implement either an identity or swap operation respectively. This is shown in Fig. 63.

In the upcoming diagrams we present, arrows are used to indicate the time-ordering of the flow of data. However, it should be noted that a MZ interferometer is reversible and therefore bidirectional, and so too are all of the more complex routers based upon them.

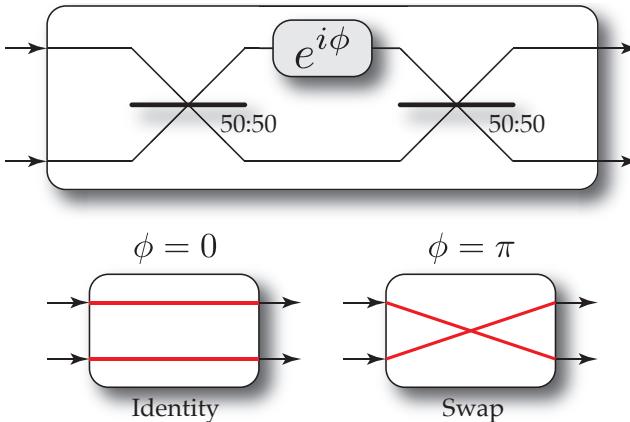


Figure 63: (top) A two-channel two-port switch has two inputs and two outputs, implementing either an identity or swap operation between them. This may be constructed using a Mach-Zehnder interferometer with a variable, classically-controlled phase-shift, $e^{i\phi}$, in one of the arms, which may be implemented using an acousto-optic or electro-optic modulator (AOM or EOM). The phase-shift is allowed to be either $\phi = 0$ for an identity channel (bottom left) or $\phi = \pi$ for a swap operation (bottom right). Because the switch is based on MZ interference, this technique only applies to optical states which undergo MZ interference. The total resource requirements are two 50:50 beam splitters and a single phase-shifter.

Because the two-port switch is based upon MZ interference, it will only function for optical states subject to such MZ interference. Thus, single-photons and coherent states are applicable, whereas thermal states, for example, are not.

D. Multiplexers & demultiplexers

From the two-port switch, which implements a controlled permutation of two optical modes, we can construct multi-port multiplexers and demultiplexers, which controllably route a single input port to one of n multiple output ports, or vice versa.

There are two main architectures that may be employed for implementing such multiplexers/demultiplexers. The first is to use a linear cascade of two-port switches, shown in Fig. 64. The second is to use a pyramid cascade, shown in Fig. 65. Both layouts require,

$$N_s = n - 1, \quad (21.11)$$

two-port switches to implement. However, they differ in one important respect. In the linear multiplexer, different routes experience different optical depth, ranging from $d = 1$ (for the first port) to $d = n - 1$ (for the final port).

This will lead to asymmetry in accumulated errors. In the pyramid multiplexer, on the other hand, all optical paths have the same optical depth, $d = \log_2 n$, yielding completely symmetric operation.

The differing optical depths of linear and pyramid multiplexers lend themselves naturally to different applications. Suppose that in a network a single input-to-output route through a multiplexer is used far more often than the others. In that case, utilising a linear multiplexer will minimise average optical depth since that route can be designated to the first output port, which has an optical depth of only $d = 1$. On the other hand, in a very balanced network, in which all optical routes are used roughly uniformly, the average case logarithmic optical depth of the pyramid multiplexer outperforms the average case linear optical depth of the linear multiplexer.

Note that the logarithmic optical depth of the pyramid configuration grows less quickly than the linear average optical depth of the linear configuration. Thus, on average, optical paths pass through fewer optical elements in the pyramid configuration, reducing average accumulated error rates when using noisy optical elements. This, in conjunction with the pyramid's perfect symmetry, makes the pyramid multiplexer configuration generally most favourable.

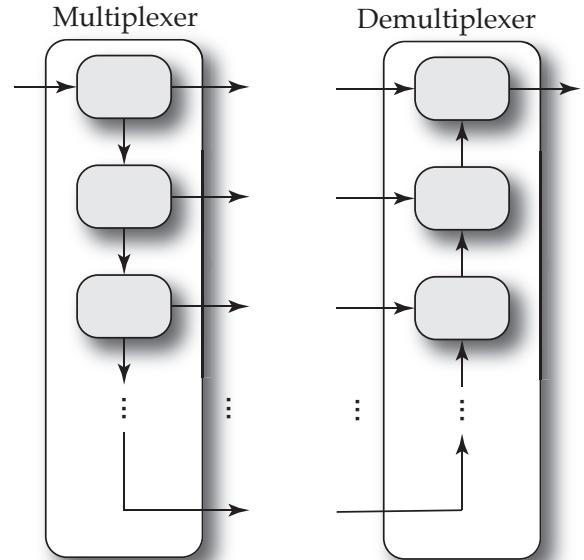


Figure 64: Linear multiplexers (left) and demultiplexers (right) may be constructed from a linear chain of two-port switches (grey boxes), cascading into one another. These switch a single optical channel between n ports. The total resource requirements are $n - 1$ two-port switches. The optical depth ranges from 1 (for the first port) to $n - 1$ (for the final port).

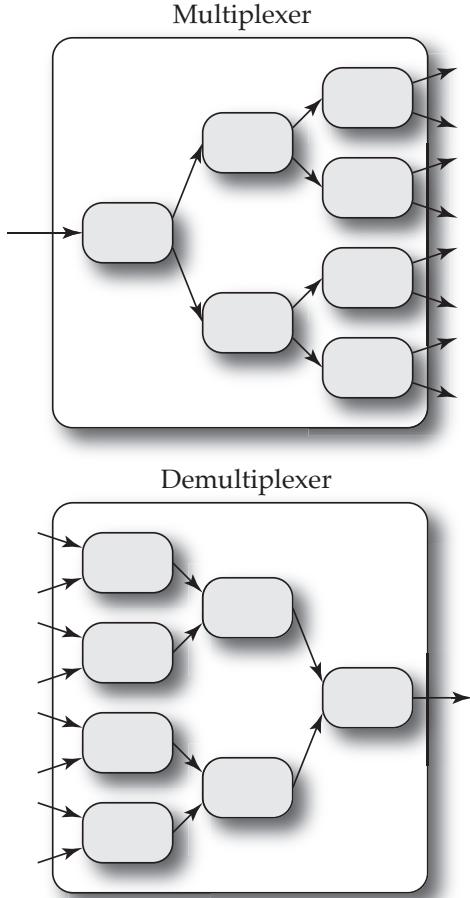


Figure 65: Pyramid multiplexers (top) and demultiplexers (bottom) decompose the multiplexing into a binary tree-structure of two-port switches (grey boxes), shown here for the case of $n = 8$ ports. For n ports, all optical paths observe an optical depth of $d = \log_2(n)$ two-port switches, of which there are $n - 1$ in total.

E. Single-channel multi-port switches

The multiplexers and demultiplexers route between one port and n ports. In the more general and useful case, we wish to route between n inputs and n outputs. If we only require one active channel at a given time, such a router may be trivially constructed from an n -port multiplexer connected to an n -port demultiplexer, as shown in Fig. 66. Here, the demultiplexer chooses one of the input modes to route to its single output, which then feeds into the multiplexer to fan it out to the desired output. The multiplexers/demultiplexers could be implemented using either of the aforementioned layouts, yielding a total resource count of,

$$N_s = 2n - 3, \quad (21.12)$$

two-port switches³¹.

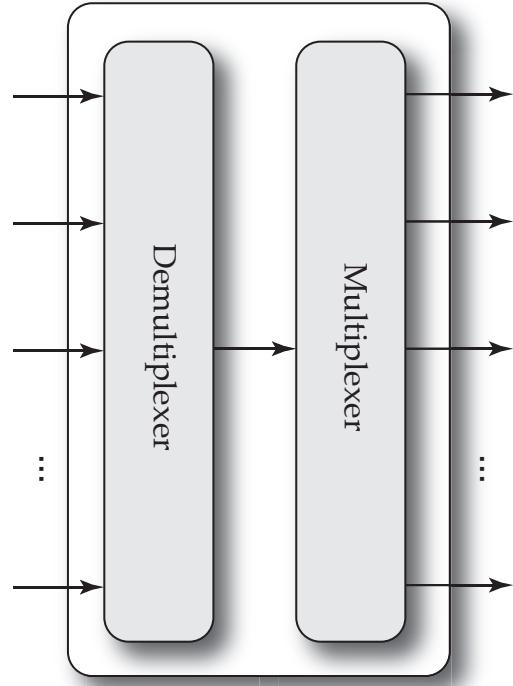


Figure 66: A single-channel multi-port switch may be constructed by demultiplexing the n input ports to a single port, routing the desired input channel to that port, before multiplexing it back out to the desired output port. This allows an arbitrary input to be routed to an arbitrary output, but only one channel at a time. This requires $2n - 3$ two-port switches in total.

F. Multi-channel multi-port switches

The single-channel multi-port switch enables switching between an arbitrary number of input/output ports, but suffers that it can only route a single channel at a time. The most general scenario to consider is multi-channel multi-port switching, which implements an arbitrary permutation between n inputs and n outputs. That is, all n ports may be routing active channels, enabling simultaneous routing of multiple data-flows.

Such a switch may be constructed from a staggered, rectangular lattice of two-port switches, as shown in Fig. 67. It is easy to see upon inspection that optical paths exist between every input/output pair of ports. The total

³¹ Note that the multiplexer and demultiplexer each require $2(n - 1)$ two-port switches, but one of the central ones adjoining the multiplexer and demultiplexer is redundant and may be eliminated, reducing the number of two-port switches to $2n - 3$.

resource count for this device is,

$$N_s = \left\lceil \frac{n^2}{2} \right\rceil - n + 1, \quad (21.13)$$

two-port switches.

The operation implemented by this device can therefore be expressed as,

$$\begin{pmatrix} \hat{b}_1^\dagger \\ \hat{b}_2^\dagger \\ \vdots \\ \hat{b}_m^\dagger \end{pmatrix} = \hat{\sigma} \cdot \begin{pmatrix} \hat{a}_1^\dagger \\ \hat{a}_2^\dagger \\ \vdots \\ \hat{a}_m^\dagger \end{pmatrix}, \quad (21.14)$$

where $\hat{\sigma} \in S_m$ is an arbitrary element of the symmetric group (i.e a permutation matrix), and \hat{a}_i^\dagger (\hat{b}_i^\dagger) are the input (output) photonic creation operators.

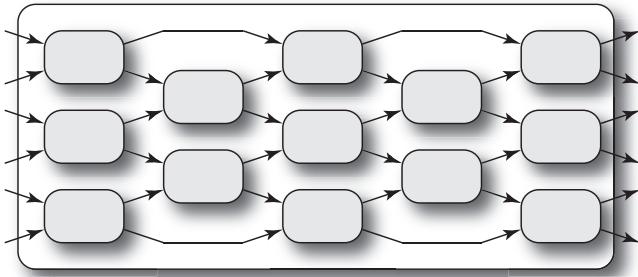


Figure 67: A completely general multi-channel multi-port switch may be constructed using a staggered grid of two-port switches (grey boxes), shown here for $n = 6$ ports. This allows the implementation of an arbitrary permutation between input and output ports, enabling all n channels to be simultaneously utilised and routed across distinct input-to-output routes. This requires $\left\lceil \frac{n^2}{2} \right\rceil - n + 1$ two-port switches in total. Optical depth is approximately equal across all input-to-output paths.

Note that this decomposition is more favourable than the completely general Reck *et al.* decomposition presented in Fig. 86(a), since the circuit is balanced, with (almost!) identical optical depths across all input-to-output paths.

G. Crossbar switches

A general multi-port switching architecture, that gained popularity in the early days of channel-switched telecommunications networks, is the crossbar architecture, whereby n inputs are mapped to n outputs via a binary permutation matrix, which controls a lattice of 2×2 switches. The general layout of the architecture is shown in Fig. 68, and an example of a routing sequence corresponding to a particular binary control matrix is shown in Fig. 69.

Clearly the scheme requires n^2 two-port switches to implement arbitrary $n \times n$ mode permutations. The main

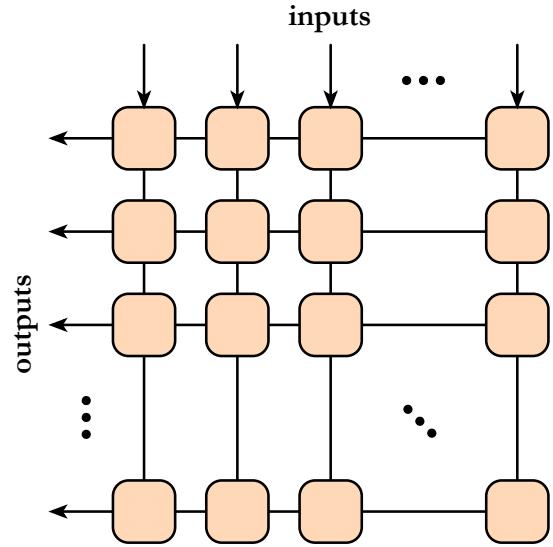


Figure 68: Crossbar architecture for multi-port switching. Each orange box represents a 2×2 switch, of any physical implementation. The switching sequence of the constituent two-port switches is defined by a binary $n \times n$ permutation matrix, whose elements determine whether a given two-port switch flips modes or doesn't.

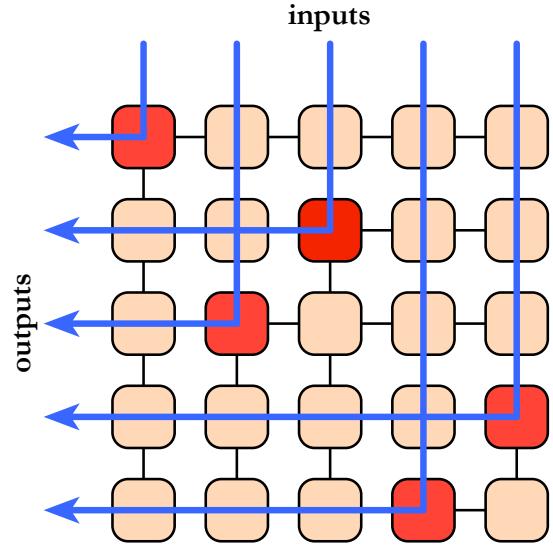


Figure 69: Example switching configuration for a 5×5 crossbar switch. The switch colours represent whether the respective 2×2 switch is set to flip (red) the modes or not (orange).

disadvantage of this scheme is that in general different paths within a given permutation experience differing optical depths, ranging from 1 (best case) to $2n - 1$ (worst case).

— *Semper inops quicunque cupit.*

XXII. OPTICAL STABILITY IN QUANTUM NETWORKS

GIVEN that communications links in quantum networks are expected to be optical, an issue of central importance is optical stability when signals from remote sources interfere or interact with local quantum states. For example, in an entanglement swapping protocol (Sec. XXVII.E) forming a part of a quantum repeater network (Sec. XXIX), if the entangling operation between the remotely prepared qubits suffers errors, so too will the prepared distributed entangled state.

If we consider the simplest scenario of employing a PBS (Sec. XIV.D) to implement the entangling operation in the polarisation degree of freedom, photon distinguishability in the form of mode-mismatch (Sec. XIV.G) will undermine quantum interference, thereby reducing the entangling power of the gate. Similar observations apply to many other protocols involving entangling measurements, or multi-photon interference more generally.

In present-day laboratories, mode-mismatch and photon-distinguishability can be controlled with exceptionally high fidelity. However, in the networking context this is likely to not be so easy, since perfectly aligning states emanating over long-distance communications channels, which we do not have exquisite control over in a well-controlled laboratory setting, is going to be a somewhat unpredictable and time-varying technological challenge.

Such processes are likely to arise in a multitude of ways, including, but certainly not limited to:

- Optical fibre: slight variations in temperatures induce refractive index changes, or changes in physical dimension, resulting in temporal displacements of optical wave-packets.
- Satellite: precise knowledge of the distance to a rapidly moving target, at the scale of photon wave-packets, is an extremely daunting prospect.
- Free-space (including via satellite): unpredictable temperature and pressure fluctuations in the atmosphere cause unpredictable variations in the speed of light.

For these inevitable reasons, it is important to understand the susceptibility of different network protocols to optical stability.

There are two dominant forms of photonic interference that must be considered, each with quite distinct behaviours under the influence of optical instability. These are:

- Hong-Ou-Mandel (HOM) interference (Sec. XXII.C): interference between two distinct photons at a beamsplitter.
- Mach-Zehnder (MZ) interference (Sec. XXII.B): self-interference of a single-photon traversing multiple paths in superposition within an interferometer.

A. Photon wave-packets

Before describing optical interference in detail, we must first formalise a definition for the optical wave-packets we will be dealing with. We will assume wave-packets with Gaussian temporal envelope of width σ (the coherence length), frequency-shifted by some carrier frequency ω_0 (the wavelength).

The temporal distribution function is then,

$$\psi(t) = \sqrt[4]{\frac{2}{\sigma\pi}} e^{-\frac{t^2}{\sigma^2} - i\omega_0 t}, \quad (22.1)$$

with associated mode-operator \hat{A}_ψ^\dagger (Sec. XIII.C). This wave-packet is normalised such that,

$$|\langle 0 | \hat{A}_\psi \hat{A}_\psi^\dagger | 0 \rangle|^2 = \int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1. \quad (22.2)$$

Of course the temporal envelope needn't be Gaussian in general, and could take any other form, subject to normalisation. In Fig. 70 we illustrate the two main features of this representation: the temporal envelope, and the underlying carrier frequency that it modulates.

In real-world scenarios we are likely to encounter carrier frequencies sufficiently large that oscillations at the carrier frequency level are far more rapid than that of the temporal envelope. For this simple reason, it is to be expected that interference dependent only on σ will be far more robust against temporal instability than interference dependent on ω_0 .

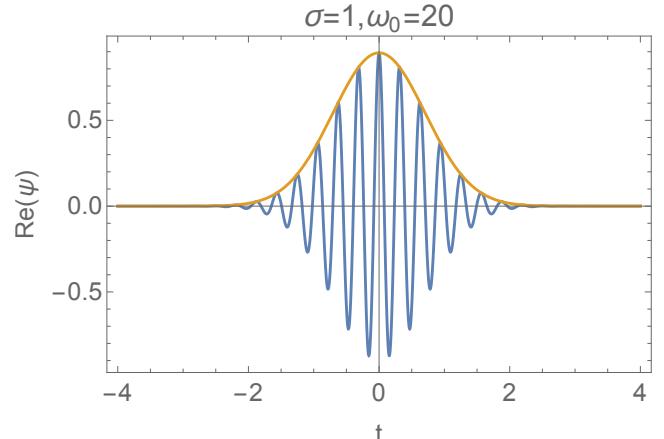


Figure 70: A photonic wave-packet of the form of Eq. (22.1), with Gaussian temporal envelope of width σ (orange), shifted by a carrier frequency ω_0 (blue).

B. Mach-Zehnder interference

Mach-Zehnder (MZ) interference is the interference of a photon or coherent state with itself in a two-mode interferometer constructed from two 50:50 beamsplitters

in series, as shown in Fig. 71(top). This is MZ interference in its simplest form, which can of course be generalised to more complex networks involving self-interference across multiple optical paths.

Within the interferometer is a time-delay, τ , which acts as a temporal mismatch between the two optical paths.

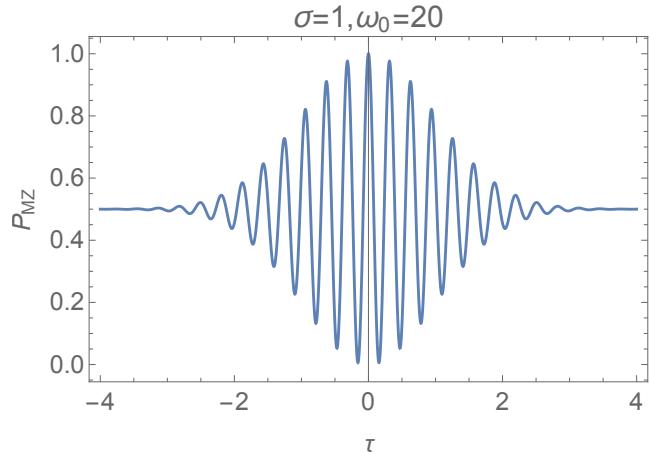
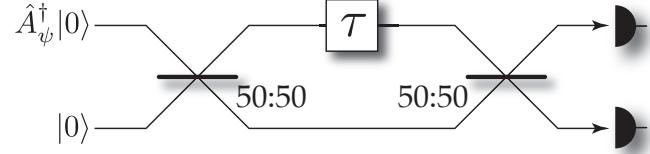


Figure 71: Mach-Zehnder self-interference of a single photon. (top) Layout of the interferometer. The photon is subject to a time-delay τ within only the upper arm of a balanced interferometer, comprising two 50:50 beamsplitters. (bottom) Interference fringe $P_{\text{MZ}}(\tau)$ as a function of the time-delay. The interference is sensitive at the scale of the photon wavelength, (blue) in Fig. 70.

Let us calculate explicitly the evolution of a single-photon through this device, beginning with a photon described by mode operator \hat{A}_ψ^\dagger (Sec. XIII.C), with the temporal distribution function from Eq. (22.1). We have,

$$\begin{aligned} |\psi_{\text{in}}\rangle &= \hat{A}_\psi^\dagger |0, 0\rangle \\ &\xrightarrow[\text{BS}]{} \frac{1}{\sqrt{2}} [\hat{A}_\psi^\dagger + \hat{B}_\psi^\dagger] |0, 0\rangle \\ &\xrightarrow[\tau]{} \frac{1}{\sqrt{2}} [\hat{A}_{\psi-\tau}^\dagger + \hat{B}_{\psi-\tau}^\dagger] |0, 0\rangle \\ &\xrightarrow[\text{BS}]{} \frac{1}{2} [\hat{A}_{\psi-\tau}^\dagger + \hat{B}_{\psi-\tau}^\dagger + \hat{A}_\psi^\dagger - \hat{B}_\psi^\dagger] |0, 0\rangle \\ &\xrightarrow[\text{PS}]{} \frac{1}{2} [\hat{A}_{\psi-\tau}^\dagger + \hat{A}_\psi^\dagger] |0, 0\rangle \\ &= \frac{1}{2} \int_{-\infty}^{\infty} [\psi(t) + \psi(t-\tau)] \hat{a}^\dagger(t) dt, \end{aligned} \quad (22.3)$$

where BS denotes the evolution implemented by a 50:50 beamsplitter, and PS denotes post-selecting upon detecting a single-photon in the first output mode.

We now characterise the operation of the device in terms of the probability of detecting the photon in the

first output mode,

$$\begin{aligned} P_{\text{MZ}}(\tau) &= \frac{1}{4} \int_{-\infty}^{\infty} |\psi(t) + \psi(t-\tau)|^2 dt \\ &= \frac{1}{2} \left[1 + e^{-\frac{\tau^2}{2\sigma^2}} \cos(\omega_0 \tau) \right]. \end{aligned} \quad (22.4)$$

These dynamics are shown in Fig. 71(bottom). There are two key features in the behaviour of $P_{\text{MZ}}(\tau)$. First, there is a slowly varying Gaussian term. Second, the Gaussian term modulates a rapidly oscillating sinusoidal term associated with the carrier frequency. This implies that τ on the order of the photon's wavelength dominates the measurement dynamics, making it extremely sensitive to temporal instability.

C. Hong-Ou-Mandel interference

In Hong-Ou-Mandel (HOM) interference, there is no self-interference as per MZ, but rather interference between two independent but indistinguishable photons. The interference takes place at a single 50:50 beamsplitter, with a temporal delay in one input mode modelling temporal instability. The model is shown in Fig. 72(top).

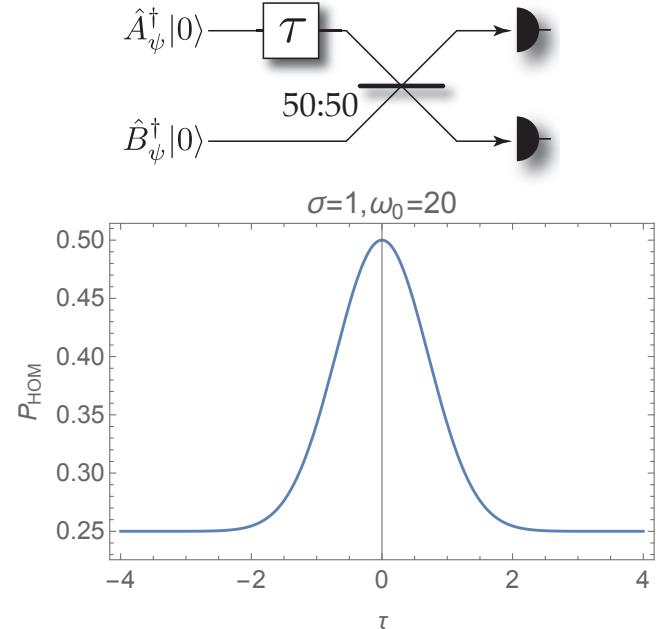


Figure 72: Hong-Ou-Mandel interference between two independent photons, A and B . (top) Layout of the interferometer. Two photons given by mode operators \hat{A}^\dagger and \hat{B}^\dagger (Sec. XIII.C), both with temporal distribution function $\psi(t)$, interfere at a 50:50 beamsplitter, where mode A is first subject to a time-delay τ . (bottom) Interference fringe $P_{\text{HOM}}(\tau)$ as a function of the time-delay. The fringe is only sensitive at the scale of the wave-packet envelope, (orange) in Fig. 70.

Performing the same evaluation of the evolution of the

system as before, we obtain,

$$\begin{aligned}
|\psi_{\text{in}}\rangle &= \hat{A}_\psi^\dagger \hat{B}_\psi^\dagger |0,0\rangle \\
&\xrightarrow{\tau} \hat{A}_{\psi-\tau}^\dagger \hat{B}_\psi^\dagger |0,0\rangle \\
&\xrightarrow[\text{BS}]{\frac{1}{2}} [\hat{A}_{\psi-\tau}^\dagger + \hat{B}_{\psi-\tau}^\dagger][\hat{A}_\psi^\dagger - \hat{B}_\psi^\dagger] |0,0\rangle \\
&\xrightarrow[\text{PS}]{\frac{1}{2}} \hat{A}_\psi^\dagger \hat{A}_{\psi-\tau}^\dagger |0,0\rangle \\
&= \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(t)\psi(t'-\tau) \hat{a}^\dagger(t)\hat{a}^\dagger(t') dt dt' |0,0\rangle.
\end{aligned} \tag{22.5}$$

We then characterise the operation of the device in terms of the probability of detecting both photons in the first output mode (photon bunching),

$$\begin{aligned}
P_{\text{HOM}}(\tau) &= \frac{1}{4} \left[1 + \left| \int_{-\infty}^{\infty} \psi(t)\psi(t-\tau)^* dt \right|^2 \right] \\
&= \frac{1}{4} \left[1 + e^{-\frac{\tau^2}{\sigma^2}} \right].
\end{aligned} \tag{22.6}$$

These dynamics are shown in Fig. 72(bottom). Now, unlike MZ interference, we observe no dependence on the carrier frequency and its associated rapidly oscillating terms. Rather, operation depends only on the temporal envelope, which exists over a far larger time-scale.

Importantly, unlike MZ interference, HOM interference is not applicable to coherent states, which do not entangle or enter into superposition at beamsplitters. The photon bunching effect is unique to single-photons.

The intuition behind the HOM-dip phenomenon is as follows. We know that for identical, indistinguishable photons, an input photon-pair evolves as,

$$|1,1\rangle \xrightarrow[\text{BS}]{} \frac{1}{\sqrt{2}}(|2,0\rangle - |0,2\rangle), \tag{22.7}$$

yielding perfect photon-bunching. This bunching effect arises from quantum mechanical interference between the photons. Next imagine that the two photons arrived a long time apart from one another, so long that their wave-packets do not overlap at all. In that instance, the photons do not ‘see’ one another and no quantum interference takes place. Instead, rather than a two-photon quantum interference experiment, we effectively have two independent instances of single-photon experiments, given by,

$$\begin{aligned}
|1,0\rangle &\xrightarrow[\text{BS}]{} \frac{1}{\sqrt{2}}(|1,0\rangle + |0,1\rangle), \\
|0,1\rangle &\xrightarrow[\text{BS}]{} \frac{1}{\sqrt{2}}(|1,0\rangle - |0,1\rangle).
\end{aligned} \tag{22.8}$$

Note that each of these independent instances obeys the classical statistics of a 50/50 distribution. Combining the two instances using classical probability theory, we now observe a 50% chance of measuring a coincidence, as opposed to the 0% chance for true HOM interference.

D. HOM vs MZ interference

Let us now examine the implications of these different types of interference. The key observation was that MZ is far more sensitive to temporal mismatch than HOM, the former at the scale of the photons’ wavelength, the latter at the scale of their temporal envelope, which is far larger.

This leads to the immediate conclusion that network protocols relying on HOM interference will be far more robust against temporal instability than those relying on MZ interference. Realistically, it is to be expected that the latter might be impossibly challenging in many contexts, as wavelength-scale stabilisation over long distances seems implausible.

In Tab. IV we summarise the network protocols discussed in Part. V in terms of the types of interference they rely upon. This creates a picture of which are more realistic from a near-term engineering perspective.

Protocol	Interference type
Cluster state measurement	None
Quantum anonymous broadcasting	None
QKD (BB84, E91)	None
Quantum memory	None
Quantum process tomography	None
Quantum state tomography	None
Random number generation	None
Separable measurements	None
Separable state preparation	None
Optical interfacing	None
Cluster state preparation (fusion gates)	HOM
Entanglement purification	HOM
Entanglement swapping	HOM
Matter qubit entangling operations	HOM
Partial Bell state measurements	HOM
Quantum gate teleportation	HOM
Quantum state teleportation	HOM
Superdense coding	HOM
BOSONSAMPLING	MZ
General linear optics networks	MZ
Universal LOQC (KLM)	MZ
Quantum metrology	MZ
Quantum walks	MZ

Table IV: Summary of the major quantum network protocols presented in Part. V, and their required type of interference.

E. Optical stabilisation

To do!

Discussion of both static and dynamic stabilisation

— *Dum inter homines sumus, colamus humanitatem.*



Part V

Protocols for the quantum internet

THESE are countless applications for the long-distance communication and processing of quantum data. We will outline some of the most notable examples. Broadly, we will begin with discussion of *low-level protocols* that form the primitives upon which other protocols are built. We will then progressively move towards *high-level protocols*, culminating with full *cloud quantum computing*.

Much of the recent experimental progress in quantum technology has been in the area of low-level protocols, although demonstrations of higher-level protocols are rapidly accelerating.

We keep in mind that although throughout this presentation we have been very quantum computing-centric, quantum computing is not the *only* quantum resource worth communicating. In the same way that *digital assets* encompass a broad range of digital systems and information, any aspect of a quantum system – from a state, to an operation, storage, to a measurement, or anything else – could be treated as a *quantum asset*, which, for generality, we would like our quantum networks to be able to handle.

At the lowest physical level, quantum protocols have in common that they involve state preparation, evolution, and measurement as the fundamental primitives upon which more complex protocols are constructed. We consider these primitive resources in detail, before building upon them to consider some of the major elementary quantum protocols that implement tasks of practical interest. All of those discussed here have been subject to extensive experimental investigation and demonstration, which will be summarised in Part. XI. We treat full quantum computation separately in Secs. XL & XLII, as this is such an involved topic in its own right.

We will employ circuit model diagrams when describ-

ing some protocols. The unfamiliar reader may refer to Sec. XL.A for a very brief introduction to quantum circuits.

Throughout this section the material will be optics-heavy, and not include discussion of some purely non-optical architectures, based on the reasonable assumption that networked quantum protocols will be optically mediated.

XXIII. STATE PREPARATION

THE first step in any quantum protocol involves the preparation of some kind of quantum state. Some quantum states are easy and cheap to prepare. Others are complex and costly. Thus, the most fundamental quantum asset that a quantum network must handle is the preparation and communication of quantum states.

A state prepared by Bob and sent to Alice might be prepared in isolation, or it might be entangled with a much larger system held by Charlie, that Alice does not have full access to. In that case, it would be impossible for Alice to prepare the state on her own, unless she were to first establish a relationship with Charlie. Alternately, maybe Alice just isn't very well-resourced, and can't do much on her own. The ability to let someone else prepare her desired quantum states for her would be highly appreciated.

Given the emphasis on quantum optics in quantum networking, it should be noted that optical quantum state engineering has broad applications, but can be very challenging in general. Single-photon state engineering, for example, finds ubiquitous applications in quantum information processing protocols, and has become commonplace. Most notably, linear optics quantum computing (Sec. XLII.A), and some quantum metrology protocols (Sec. XXVII.H) rely on single-photon state preparation. ‘Push-button’ (i.e on-demand) single-photon sources would be a prized asset to many undergraduate experimentalists, were they able to afford them. But with access to the quantum internet, they could purchase single photons from another better-resourced lab, with QoS constraints guaranteed by QTCP.

The QTCP protocol is ideally suited to facilitating this kind of transaction. With the use of efficiency and purity cost metrics, QoS guarantees could be established for the efficiency and purity of a licensed single-photon source. In the case of single photons, the dephasing metric is irrelevant, since photon-number states are phase-invariant. This is an elegant example of the value of the versatility of having the QTCP protocol track multiple cost metrics for quantum packets, since different metrics will be of relevance to different messages. Were the message a coherent state, $|\alpha\rangle$, on the other hand, dephasing would be of utmost importance, whereas loss would be less critical, as lossy coherent states remain as coherent states and

retain their coherence.

We see that even the most basic primitive in quantum technologies – state preparation – already brings with it much to take into consideration when designing quantum networks. However, the QTCP protocols we described earlier are versatile enough to be capable of mediating their distribution across quantum networks, whilst providing QoS guarantees.

A. Coherent states

Coherent states (Sec. XIII.E.1), although not strictly *quantum* states, nonetheless find broad applications in quantum protocols, for example as the pump for SPDC sources (Sec. XXIII.B), or as a phase-reference for homodyne detection (Sec. XXIV.C). Coherent states are rather trivial to prepare, as they are closely approximated by laser sources. Despite their triviality, high quality lasers can nonetheless become very expensive, large, and inaccessible to the not-so-well-resourced end-user. It is not uncommon for laser sources in contemporary labs to be valued in the \\$100k's.

B. Single-photons

Single-photon sources (Sec. XIII.A) (?) are of particular interest, as a foundational building block in many optical quantum information processing applications, such as linear optics quantum computing (Sec. XLII.A) and quantum key distribution (Sec. XXXVII.A).

The most common approach to preparing single-photon states is via heralded SPDC (??), whereby a coherent pump source is down-converted into two-mode photon-pairs via a second-order non-linear crystal with interaction Hamiltonian of the form,

$$\hat{H}_{\text{SPDC}} = \xi(\hat{a}_p \hat{a}_s^\dagger \hat{a}_i^\dagger + \hat{a}_p^\dagger \hat{a}_s \hat{a}_i), \quad (23.1)$$

where ξ is the interaction strength, and \hat{a}_p , \hat{a}_s and \hat{a}_i are the photonic annihilation operators for the pump (input), and *signal* and *idler* (output) modes respectively. This has the clear intuitive interpretation as the coherent exchange of photon-pairs in the output modes with photons in the coherent pump.

Specifically, a two-mode SPDC state takes the form,

$$|\psi\rangle_{\text{SPDC}} = \sqrt{1 - \chi^2} \sum_{n=0}^{\infty} \chi^n |n\rangle_s |n\rangle_i, \quad (23.2)$$

where χ is the squeezing parameter, a function of the pump-power and properties of the crystal. The layout is shown in Fig. 73.

Applying the single-photon projector, $|1\rangle\langle 1|$, to the first mode yields the single-photon state in the other, up to normalisation, which reflects the inherent non-determinism.

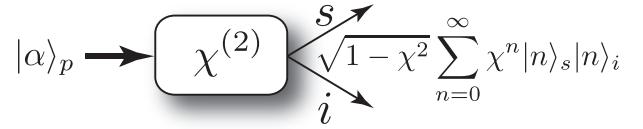


Figure 73: Layout of an SPDC single-photon source. A second-order non-linear crystal is pumped with a coherent state (i.e laser source), yielding a two-mode output state with perfect photon-number correlation between the two modes. Then, post-selecting upon detecting a single photon in one mode in principle guarantees a single photon in the other.

The preparation success probability is derived from the amplitude of the $n = 1$ term as,

$$P_{\text{prep}} = \chi^2(1 - \chi^2), \quad (23.3)$$

assuming ideal photo-detection. Thus, the perfect photon-number correlation enables heralded preparation of states with exactly one photon in principle.

Transitioning from heralded state preparation to quasi-deterministic state preparation may then be achieved by operating a bank of such sources in parallel, and multiplexing their outputs, such that when all sources are triggered simultaneously, if any one succeeds, the respective single photon is routed to the desired output mode, as shown in Fig. 74.

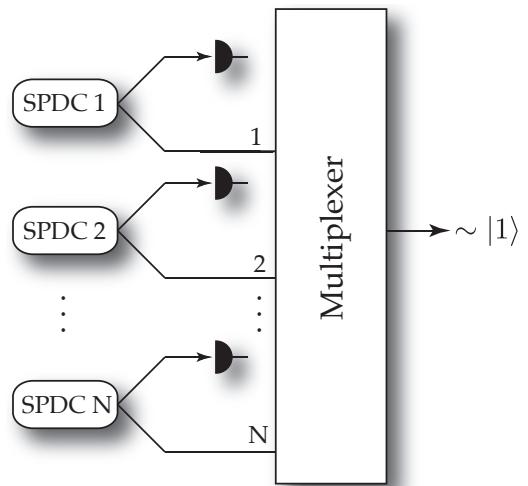


Figure 74: Quasi-deterministic single-photon state preparation using N -fold multiplexing of heralded SPDC sources (or any other non-deterministic, but heralded source). All N SPDC sources are triggered simultaneously. The heralding detectors feedforward to the multiplexer, which routes a successfully heralded single-photon state (if there is one) to the output mode. With a sufficiently large bank of sources in parallel, the probability of successfully preparing a single-photon state approaches unity.

The success probability of the multiplexed source exponentially asymptotes to unity as the number of in-parallel sources increases,

$$P_{\text{success}} = 1 - (1 - P_{\text{prep}})^N, \quad (23.4)$$

where there are N sources in parallel. This relationship is shown in Fig. 75 for sources with varying heralding probabilities. This principle could also obviously be applied to any other type of non-deterministic, but heralded source.

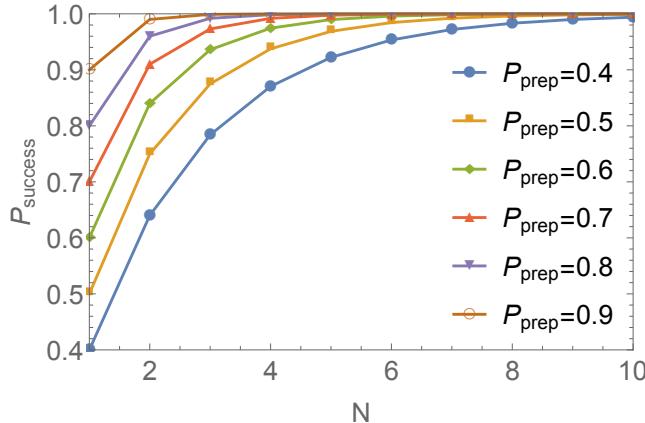


Figure 75: Single-photon state preparation probability, P_{success} , using N -fold multiplexing, where the individual heralded sources have heralding probability P_{prep} . P_{success} always exponentially asymptotes to unity with increasing N , for any $P_{\text{prep}} > 0$.

The multiplexing approach needn't be restricted to the spatial domain, but could also be equivalently implemented in the temporal domain (???), as shown in Fig. 76.

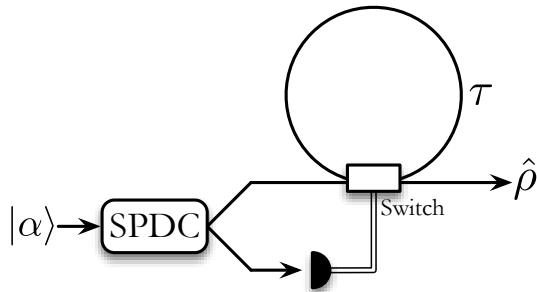


Figure 76: Multiplexed single-photon state preparation in the temporal domain. An SPDC source operating at high repetition rate, with time-bin separation τ , enters a fibre-loop with an in/out coupling switch classically controlled by the heralding outcomes. The fibre-loop acts as a quantum memory, keeping the most recent successfully heralded time-bin in memory until the procedure terminates. The output is a pulse-train where the last time-bin closely approximates a single-photon.

Of course, operating a large bank of sources in parallel, along with the associated multiplexing, which requires nanosecond-scale fast-feedforward, is experimentally costly (both in physical size, complexity and dollars), making outsourcing of this technology potentially highly desirable.

This description is purely in the photon-number basis. However, as discussed in Sec. XIII.C, photons also have

spatio-temporal characteristics. This strongly affects state preparation when using heralded SPDC, particularly state purity, and much effort has been invested into engineering the spectral structure of SPDC states so as to maximise purity and indistinguishability (??). Specifically, we wish to engineer the photon-pairs to be spectrally separable, such that the heralded photon remains spectrally pure even if the heralding photon was measured with undesirable spectral characteristics (e.g finite resolution).

SPDC is relatively cheap, and widely used, but nonetheless might be out of reach for many end-users, particularly when the previously discussed multiplexing techniques are employed to boost heralding efficiencies. It is quickly being superseded by superior technologies in cutting-edge labs, such as quantum dot sources, which have deterministic, push-button potential (??). Techniques based on cavity quantum electrodynamics (QED) (?) and molecular fluorescence (?) have also been demonstrated. However, such sources are very much in their developmental stages, and relatively expensive.

Generally speaking, a push-button photon source could be constructed from any two-level system, comprising a ground state, $|g\rangle$, and an excited state, $|e\rangle$, with short lifetime, whereby relaxation via the $|e\rangle \rightarrow |g\rangle$ transition emits a photon. Then, pumping the system to excite it to the $|e\rangle$ state, and waiting for spontaneous decay yields a single photon.

C. NOON states

So-called NOON states, path-number entangled two-mode states of the form,

$$|\psi\rangle_{\text{NOON}} = \frac{1}{\sqrt{2}}(|N, 0\rangle + |0, N\rangle), \quad (23.5)$$

may be exploited to perform Heisenberg-limited quantum metrology (Sec. XXVII.H), allowing extremely precise phase measurement with large photon-number N (?).

The extreme sensitivity of NOON states owes to the N -fold enhancement in phase-dependence associated with the N -photon component of the superposition. Specifically, if a phase ϕ is present in only the second mode, the state evolves to,

$$e^{i\phi\hat{n}}|\psi\rangle_{\text{NOON}} = \frac{1}{\sqrt{2}}(|N, 0\rangle + e^{i\phi N}|0, N\rangle), \quad (23.6)$$

where $\hat{n} = \hat{a}^\dagger \hat{a}$ is the photon-number operator associated with the second mode. The phase enhancement arises because $\hat{n}|N\rangle = N|N\rangle$. Then, a simple interferometric procedure is able to extract this enhanced phase-dependence as an observable.

However, these states are notoriously difficult and technologically challenging to prepare, and can only be prepared non-deterministically using linear optics (???). If a remote server had the capacity to prepare such states,

they would be in high demand across the globe. Hindering this, NOON states are very fragile creatures. First, they exhibit exponentially increased susceptibility to loss – loss of just a single photon completely decoheres the state, rendering it useless for metrological purposes. Second, the large photon-number, N , amplifies unwanted dephasing by a factor of N , as discussed in Sec. XIV.D. These considerations can be readily accommodated for in the QTCP protocol by tracking dephasing and loss metrics of the packets encapsulating the NOON states.

D. Cluster states

In addition to the simple single- or two-mode states discussed above, an entire universal quantum computation can be performed using the ‘cluster state’ measurement-based model for quantum computation (explained in detail in Sec. XL.B). Here state preparation can not only be outsourced, but distributed, with different hosts preparing different geometric parts of the state, which are then ‘stitched together’.

The beauty of this type of state is that there is a natural separation between state preparation and computation, with the preparation stage being far more technologically challenging than the computation stage. Thus, Alice might ask better-resourced Bob to prepare a cluster state and send it to her, at which point she implements the computation herself using only simple single-qubit measurement operations.

A detailed discussion of optical cluster state preparation is presented in Sec. XLII.B, with a focus on protocols employing non-deterministic gates.

E. Greenberger-Horne-Zeilinger states

Another class of states is Greenberger-Horne-Zeilinger (GHZ) states (?), which are maximally-entangled states across an arbitrary number of qubits, n , of the form,

$$|\psi\rangle_{\text{GHZ}}^{(n)} = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}). \quad (23.7)$$

GHZ states are useful for various quantum information processing applications, including quantum anonymous broadcasting (Sec. XXXVII.D). These states are particularly susceptible to loss, since the loss of a single qubit completely decoheres the state into a perfect mixture of the $|0\rangle^{\otimes(n-1)}$ and $|1\rangle^{\otimes(n-1)}$ states, with complete loss of entanglement and coherence,

$$\begin{aligned} \hat{\rho}_{\text{GHZ}}^{\text{loss}} &= \text{tr}_1(|\psi\rangle_{\text{GHZ}}^{(n)}\langle\psi|_{\text{GHZ}}^{(n)}) \\ &= \frac{1}{2}(|0\rangle^{\otimes(n-1)}\langle 0|^{\otimes(n-1)} + |1\rangle^{\otimes(n-1)}\langle 1|^{\otimes(n-1)}). \end{aligned} \quad (23.8)$$

A simple linear optics circuit for the preparation of 3-qubit polarisation-encoded GHZ states is shown in Fig. 77.

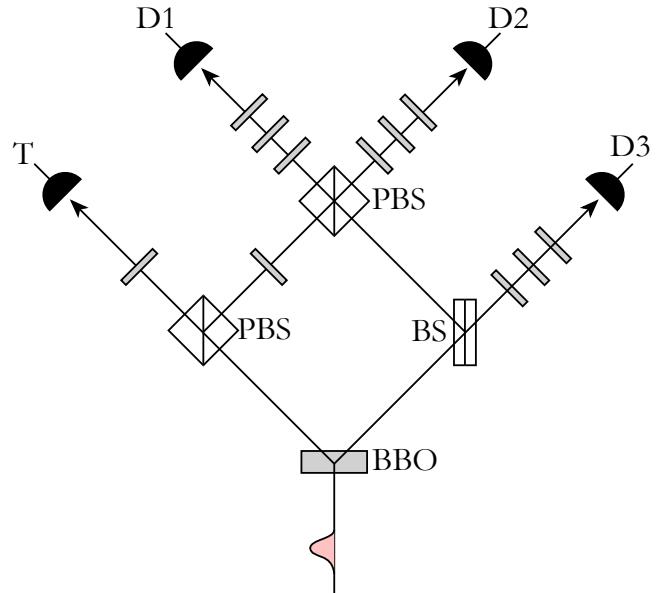


Figure 77: Linear optics circuit for the non-deterministic preparation of 3-qubit polarisation-encoded GHZ states (?). A BBO SPDC source is pumped into the double excitation regime. Following evolution through the linear optics network, measurement of a photon in the tigger mode (T) heralds the preparation of a GHZ state across modes $D1$, $D2$ and $D3$.

F. W-states

W-states are a class of non-maximally-entangled states across an arbitrary number of qubits/modes. They are given by the equal superposition of a single excitation/‘1’ across all n qubits. Thus, there are n terms in the superposition, of the form,

$$\begin{aligned} |\psi\rangle_{\text{W}}^{(n)} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{a}_i^\dagger |0\rangle^{\otimes n} \\ &= \frac{1}{\sqrt{n}}(|1, 0, 0, 0, \dots\rangle + |0, 1, 0, 0, \dots\rangle \\ &\quad + |0, 0, 1, 0, \dots\rangle + |0, 0, 0, 1, \dots\rangle + \dots). \end{aligned} \quad (23.9)$$

W-states are a class of states distinct from GHZ states (for any $n \geq 3$). Unlike GHZ states, W-states preserve some entanglement when a single qubit is traced out of the system, giving them a degree of robustness against qubit loss. This property is discussed in further detail in Secs. XX.A.3 & XVIII.B, where this property is exploited for error correction purposes.

Using linear optics, W-states are amongst the most trivial entangled states to prepare, using a simple fanout operation. An n -mode W-state can be prepared directly using n beamsplitters in a linear cascade, as shown in Fig. 78(a), where the beamsplitter reflectivities are chosen so as to create the uniform superposition, given by the

pattern,

$$\begin{aligned}
\eta_1^2 &= 1 - \frac{1}{n}, \\
\eta_1^2 \eta_2^2 &= \frac{1}{n}, \\
\eta_1^2 \eta_3^2 (1 - \eta_2^2) &= \frac{1}{n}, \\
\eta_1^2 \eta_4^2 (1 - \eta_2^2)(1 - \eta_3^2) &= \frac{1}{n}, \\
\eta_1^2 \eta_i^2 \prod_{j=2}^{i-1} (1 - \eta_j^2) &= \frac{1}{n}, \\
\eta_n^2 &= 1. \tag{23.10}
\end{aligned}$$

Alternately, they can be made from a pyramid fanout network of $\eta = 1/\sqrt{2}$ beamsplitters, as shown in Fig. 78(b).

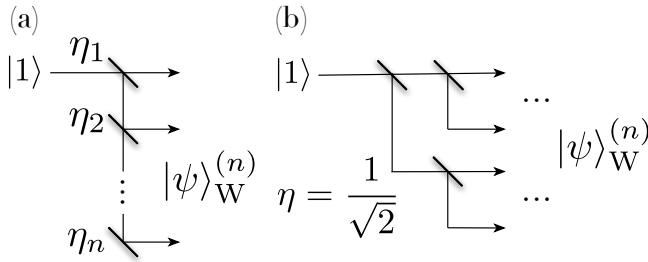


Figure 78: Preparation of an n -mode W-state using a single-photon source and a (a) linear cascade of beamsplitters, (b) a pyramid network of beamsplitters. The latter has the advantage that paths are balanced, in the sense that all routes pass through the same number of optical elements. Additionally the pyramid network has only $O(\log n)$ depth, compared to the worst-case $O(n)$ depth of the linear cascade. This is a relevant consideration when optical elements are lossy or induce errors.

More generally, an entire basis of W-states³² is accessible using linear optics networks described by unitaries where all matrix elements have amplitude $1/\sqrt{n}$, differing only via local phases. This includes the generalised Hadamard transform, and quantum Fourier transform.

G. Bell states

Bell states, also known as Einstein-Podolsky-Rosen (EPR) pairs (?), which are maximally-entangled 2-qubit states, are particularly useful for many applications, including quantum teleportation (Sec. XXVII.C), cluster state preparation (Sec. XL.B), and entanglement swapping (Sec. XXVII.E). Bell states are the special case of 2-qubit cluster states, or equivalently, 2-qubit GHZ states,

$$|\Phi^+\rangle = |\psi\rangle_{\text{GHZ}}^{(2)}. \tag{23.11}$$

³² We refer to a ‘basis’ of W-states as being a set of states that are all equivalent to a standard W-state up to local phases.

Bell pairs may be directly prepared as the two-mode output from a type-II³³ SPDC source, or using non-deterministic linear optics from single-photon sources.

There are four Bell states³⁴, defined as,

$$\begin{aligned}
|\Phi^\pm\rangle &= \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B \pm |1\rangle_A|1\rangle_B), \\
|\Psi^\pm\rangle &= \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B \pm |1\rangle_A|0\rangle_B), \tag{23.12}
\end{aligned}$$

which are locally equivalent to one another via the application of Pauli operators, and may therefore be transformed to one another without classical or quantum communication between the two parties. Specifically,

$$|\Phi^+\rangle = \hat{Z}|\Phi^-\rangle = \hat{X}|\Psi^+\rangle = \hat{X}\hat{Z}|\Psi^-\rangle, \tag{23.13}$$

where \hat{X} and \hat{Z} could apply to either qubit, up to global phase.

In Sec. XXVIII we present the case that these states are so useful on their own that one might be justified in building entire quantum networks based purely upon the distribution of Bell pairs. This is the basis for *quantum repeater networks*, which will be discussed in Sec. XXIX.

H. Cat states

Cat states (Sec. XIII.E.2) – superpositions of coherent states – are extremely difficult to prepare, most easily via non-linear processes. However, they are very useful for optical quantum computation and for the study of macroscopic quantum systems, when using large coherent amplitudes. Because of the difficulty of their preparation, the ability to outsource it would be very valuable.

However, it must be cautioned that cat states decohere very readily, since their well-defined parity implies decoherence upon loss or measurement of just a single photon. This makes QoS considerations particularly pertinent.

I. Squeezed states

Of particular interest to metrology and CV quantum computing in particular, are squeezed states, states which have been longitudinally distorted in phase-space. In the metrological context, squeezed states enable sub-shot-noise limited metrology (?), thereby outperforming any classical protocol, using, for example, coherent states.

³³ In type-II SPDC the photon-pair is polarisation-entangled, directly preparing a Bell pair in the polarisation basis. In type-I SPDC both photons have the same polarisation, yielding only photon-number correlation, but no polarisation entanglement.

³⁴ $|\Psi^-\rangle$ is also referred to as a *singlet* state, and $|\Psi^+\rangle$ as a *triplet* state.

Mathematically, squeezing is represented using the squeezing operator introduced in Eq. (42.50). Experimentally, such states are prepared using non-linear crystals. It is intuitively obvious that linear optics alone cannot prepare such states, owing to the non-linear terms in the definition of the operator, which do not preserve photon-number, and therefore cannot be passive.

Of particular interest are squeezed coherent states, $\hat{S}(\xi)|\alpha\rangle$, which are minimum uncertainty states, saturating the Heisenberg uncertainty relation. A special case of this is squeezed vacuum states, $\hat{S}(\xi)|0\rangle$, which are even-parity states (i.e containing strictly even photon-number terms). This implies that, like cat states, they are very vulnerable to decoherence for the same reason.

J. Matter qubits

In addition to preparing optical states, they can be used to mediate the preparation of systems comprising matter qubits, using which-path erasure techniques (Sec. XLII.F & Fig. 148) or light-matter interactions (Sec. XX.A & Fig. 53). These techniques are very versatile, and apply to many different matter qubit systems, such as two-level atoms, nitrogen-vacancy centres, quantum dots, and atomic ensembles.

This is useful when matter-based architectures are more scalable or technologically simpler than all-optical architectures (Sec. XLII.A), and particularly for quantum memory (Sec. XXVI), when using matter qubits with long lifetimes.

XXIV. MEASUREMENT

As a last (and possibly also intermediate) step in any quantum protocol is the measurement of quantum states. State measurement is, in the most general context, essentially state preparation in reverse, and brings with it many of the same challenges.

Different detection schemes bring with them their own (potentially substantial) costs and technological challenges. State of the art micro-pillar photo-detectors (?), at the time of writing, cost on the order of \$100k's, and require a sophisticated laboratory setup. Clearly this type of infrastructure is inaccessible to many players, and borrowing or licensing access to such equipment over a quantum network would pave the way for broader accessibility to state of the art technology.

Each type of state being measured, in combination with the nature of the detection scheme, brings with it their own limitations. Specifications of interest include dead-time, speed (relevant when implementing feedforward), and spatio-spectral filtering characteristics.

These represent significant technological challenges, which are costly to overcome, necessitating outsourcing

over the quantum internet to become economically viable on a large scale. However, the QTCP protocol is able to accommodate error metrics and attributes covering all the above error models, enabling reliable, predictable QoS for outsourced quantum measurement.

With the ability to perform measurements over a complete basis for the respective system, QST, and consequently QPT (Sec. XII.D), can also be outsourced, as both these protocols are built entirely upon determining measurement expectation values in some known basis.

A. Photo-detection

Perhaps the most useful, and ubiquitous, type of optical state measurement is photo-detection (?), where we would like to count photon-number. Broadly, there are two main classes of photo-detectors – *number-resolved* and *non-number-resolved* (or ‘bucket’ detectors). These behave exactly as the names suggest, with the former typically being more expensive and technologically demanding than the latter.

1. Mathematical representation

A completely general detector can be modelled as a POVM,

$$\hat{\Pi}_m = \sum_{n=0}^{\infty} P(m|n)|n\rangle\langle n|, \quad (24.1)$$

where $P(m|n)$ is the conditional probability of measuring m photons given n incident photons. The POVM is fully characterised by the conditional probabilities, which must be inferred from the specifics of the architecture. Alternatively, a quantum process formalism can be constructed as,

$$\mathcal{E}_m(\hat{\rho}) = \sum_{n=0}^{\infty} P(m|n)\hat{E}_n\hat{\rho}\hat{E}_n^\dagger, \quad (24.2)$$

where $\hat{E}_n = \hat{E}_n^\dagger = |n\rangle\langle n|$ are the Kraus operators.

These mathematical representations very conveniently reduce the characterisation and representation of photo-detectors to calculating the matrix of conditional probabilities, $P(m|n)$. This readily allows various experimental effects and imperfections to be accommodated.

2. Avalanche photo-diodes

The most common form of photo-detection is using avalanche photo-diodes (APDs), which are cheap but non-number-resolving. Here, a single photon excites an electron into the conduction band at a semiconductor junction, enabling a detectable current flow. However,

a single excitation triggers an ‘avalanche’ of further excitation making the magnitude of the detected current essentially unrelated to exact photon number.

3. Superconducting photo-detectors

More recently, superconducting detectors have been adopted, as they have the potential for number-resolution. Here a superconductor is kept just below its critical temperature, and the absorption of a photon is enough to heat the superconductor above the critical temperature, creating a detectable change in resistance across the superconductor. This is shown in Fig. 79. Despite their superior performance, however, superconducting detectors are very expensive (for obvious reasons) and only accessible to well-resourced labs. However, unlike APDs they can be photon-number-resolving.

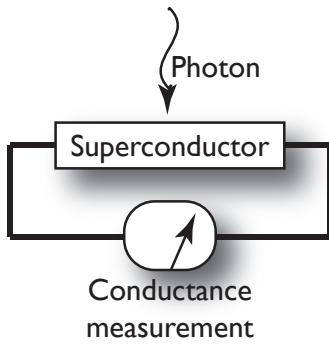


Figure 79: A superconducting photo-detector. A superconductor is held at just below its critical temperature. The absorption of a photon is sufficient to heat it above the critical temperature, yielding a detectable change in conductance across the device.

Is this the same as transition edge sensor (TES)

4. Quantum dot photo-detectors

To do

5. Photo-diodes

To do

6. Experimental issues

The key parameter of interest in a photo-detector, in addition to whether or not it is number-resolving, is its efficiency, η – the probability that a given incident photon will trigger the detector. For most applications, the goal is to maximise η . As one might expect, there is a direct tradeoff between η and cost, with very high-efficiency

detectors often economically out of reach for many experimentalists. Also of interest is the ‘dark-count’ rate – the rate at which the detector falsely clicks in the absence of photons. However, this is often ignored as modern detectors typically exhibit very low dark-count-rates.

Mathematically, the measurement operators for inefficient number-resolved detection are,

$$\hat{\Pi}_n = \eta^n \sum_{m=n}^{\infty} \binom{m}{n} (1-\eta)^{m-n} |m\rangle\langle m|, \quad (24.3)$$

for measurement outcome n , in the photon-number basis. And for non-number-resolved detection,

$$\begin{aligned} \hat{\Pi}_0 &= \sum_{m=0}^{\infty} (1-\eta)^m |m\rangle\langle m|, \\ \hat{\Pi}_{>0} &= \hat{\mathbb{1}} - \hat{\Pi}_0. \end{aligned} \quad (24.4)$$

Thus, inefficiency results in projection onto the wrong photon-number, making measurement outcomes incorrect.

In addition to their operation in the photon-number basis, photo-detectors exhibit spatio-temporal characteristics, which affect their operation in quantum information processing protocols (?). For example, imperfect spectral response can undermine photonic interference, affecting which-path erasure protocols, such as Bell state projection (Sec. XXIV.D). However, in many cases this can be improved upon using spectral filtering or time-gating techniques, also at the expense of experimental complexity and resource overhead.

Furthermore, photo-detectors are subject to ‘dead-time’, which renders them inactive for a finite recovery period following a detection event. This is of especial importance in time-bin-encoded schemes (Sec. XIII.C.2), where detectors must resolve photons over very short timescales on the order of nanoseconds. Dead-time can be modelled as a time-dependent efficiency of the form,

$$\eta(t) = \begin{cases} 0, & t < \tau_{dt} \\ \eta_{ss}, & t \geq \tau_{dt} \end{cases}, \quad (24.5)$$

where t is time, τ_{dt} is the detector’s dead time, and η_{ss} is the detector’s steady-state efficiency (i.e when not dead).

Photo-detectors of all types are inevitably subject to ‘dark-counts’, whereby thermal noise, either within the detector or coupled from the noisy external environment, triggers non-existent detection events. The distribution follows exactly that of the thermal state photon-number distribution (Sec. XIII.D). Thus, the probability of n dark-counts occurring is,

$$p_{dc}(n) = e^{-|\alpha|^2} \frac{|\alpha|^n}{n!}, \quad (24.6)$$

where α is a parameterisation of the temperature of the environmental noise. Fortunately, modern detector technology is able to keep dark-count rates very low, making

this far less of an issue than the aforementioned ones, loss being the dominant.

Finally, all photo-detection techniques are subject to some degree of ‘time-jitter’ – noise in the detector’s reported time of detection. This can be extremely important in the context of temporal mode-matching, where post-selection upon detection events in an extremely narrow time-window effectively enforces temporal indistinguishability.

B. Multiplexed photo-detection

Number-resolved detectors are the more challenging ones to experimentally realise. However, using multiplexing techniques, non-number-resolved detectors can be used to closely approximate number-resolution (????), at the expense of an (efficient) overhead in the complexity of the experiment, which comes at a cost.

Specifically, there is a direct tradeoff between the confidence in photon-number outcomes, and experimental overhead. The idea behind this is simple. We spread out an n -photon state evenly across a large number of modes, m , and detect each one independently using a non-number-resolved photo-detector. If $m \gg n$, it is unlikely that more than a single photon will reach any given detector. Thus, by summing the total number of clicks across all detectors, we closely approximate the true photon-number. This multiplexing can be performed in the spatial- or temporal-domains, shown in Fig. 80, and has been a widely employed technique in laboratories without access to expensive number-resolved detectors.

Mathematically, we are interested in the probability $P(n_{\text{meas}} = n_{\text{inc}})$ that the measured number of photons (n_{meas}) matches the actual number of incident photons (n_{inc}). The structure of this expression will vary enormously depending on the details of the architecture (e.g multi-port interferometer vs fibre-loop). However, (?) presented a very general mathematical formalism applicable to all architectural variants. The simplest case to consider is the multi-port interferometer, owing to its perfect symmetry. The probability is simply the probability that no output mode from the multi-port contain multiple photons. A quick calculation yields,

$$P(n_{\text{meas}} = n_{\text{inc}}) = \frac{\eta^n m!}{m^n (m-n)!}, \quad (24.7)$$

for efficiency η , not accounting for other lesser errors such as dark-counts. This is shown in Fig. 81. For perfect efficiency, $\eta = 1$, this probability always approaches unity in the limit of a large number of modes,

$$\lim_{m \rightarrow \infty} P(n_{\text{meas}} = n_{\text{inc}}) = 1. \quad (24.8)$$

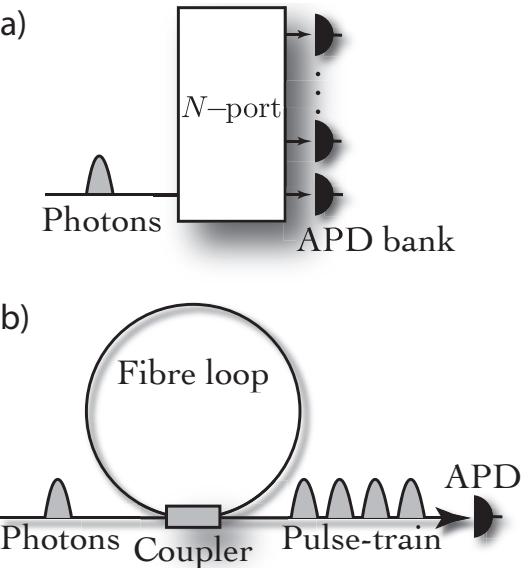


Figure 80: Multiplexed number-resolved photo-detection using non-number-resolved photo-detectors. The principle is to spread out (as uniformly as possible) a multi-photon state across a large number of modes, sufficiently large that it is unlikely that more than one photon will be present in any given mode. Then, the sum of the number of clicks at each mode closely approximates the incident photon-number. (a) In the spatial domain. (b) In the temporal domain. The advantage of employing the temporally multiplexed architecture is that only a single detector is required, unlike the multiple independent detectors required in the spatially multiplexed scheme. However this requires that the dead-time of the detector is less than the round-trip time of the fibre loop. An alternate, but conceptually equivalent approach is to spatially disperse the optical field across a charge-coupled device (CCD), much like that found in a regular digital camera, except with single-photon resolution per-pixel. This achieves an effectively very large number of optical modes.

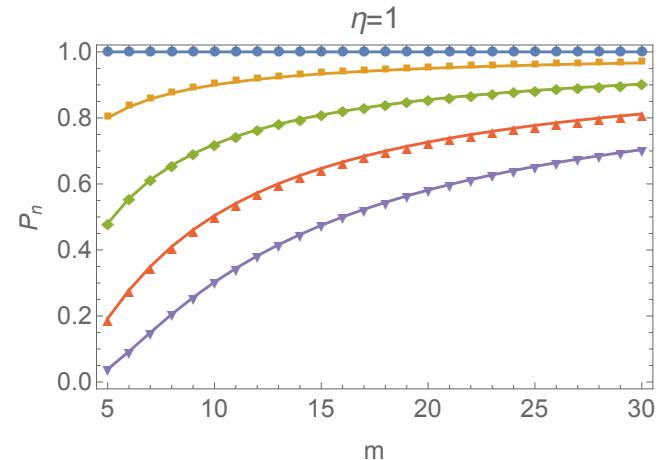


Figure 81: Probability of a balanced multiplexed detector inferring the correct number of photons with n incident photons across m modes. The detectors have perfect efficiency, $\eta = 1$, and all other errors are ignored. The number of incident photons ranges from $n = 1$ (top) to $n = 5$ (bottom). All curves asymptotically approach $P_n = 1$ for large m .

C. Homodyne detection

A homodyne detector interferes a state with a coherent state on a beamsplitter, which acts as a phase-reference, before photo-detecting both output modes and taking the difference in the photon count-rates (Fig. 82). This effectively allows us to observe ‘beating’ effects between the signal and reference probe.

By sweeping through the amplitude and phase of the reference beam, we are able to directly sample points in phase-space, allowing the Wigner function – which is isomorphic to the density operator – of an unknown state to be fully reconstructed.

Equivalently, homodyne detection can directly sample the position (\hat{x}) and momentum (\hat{p}) operators, or arbitrary linear combinations of the two.

The operation of homodyne measurement is most easily visualised in phase-space, where it can be regarded as integrating along an infinite line with arbitrary rotation determined by the phase-reference.

This measurement technique is typically applied to CV states rather than photon-number states. While conceptually straightforward, preparing the reference beam requires a coherent source, which can become costly (Sec. XXIII.A).

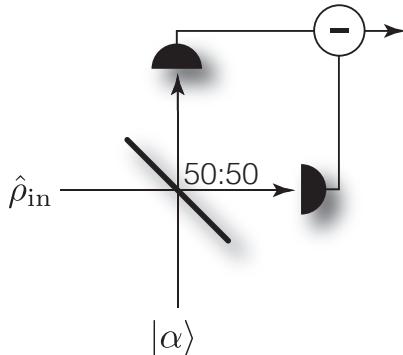


Figure 82: Homodyne detection of an unknown optical state $\hat{\rho}_{\text{in}}$, by mixing it with a reference coherent state $|\alpha\rangle$ on a 50:50 beamsplitter, and taking the difference in the photo-detection rates at the output modes. By sweeping through the phase and amplitude of $|\alpha\rangle$, we can directly sample the Wigner function of $\hat{\rho}_{\text{in}}$, allowing its full reconstruction.

D. Bell state & parity measurements

For the purposes of which-path erasure, essential for optical cluster state (Sec. XL.B) preparation and quantum teleportation (Sec. XXVII.C), Bell state measurements [i.e projections onto the Bell basis given in Eq. (23.12)], or equivalently parity measurements, are important.

To realise this, there are two primary options. The first is to use a CNOT gate, for example an LOQC gate (Sec. XLII.A). The second is to perform a *partial* Bell

state projection using a polarising beamsplitter (PBS) – a beamsplitter which completely transmits vertical polarisation, and completely reflects horizontal polarisation (?).

In the Heisenberg picture, the transformation of the photonic creation operators implemented by a PBS is,

$$\begin{aligned}\hat{h}_1^\dagger &\rightarrow \hat{h}_2^\dagger, \\ \hat{h}_2^\dagger &\rightarrow \hat{h}_1^\dagger, \\ \hat{v}_1^\dagger &\rightarrow \hat{v}_1^\dagger, \\ \hat{v}_2^\dagger &\rightarrow \hat{v}_2^\dagger,\end{aligned}\quad (24.9)$$

where \hat{h}_i^\dagger (\hat{v}_i^\dagger) are the horizontal (vertical) creation operators for the i th mode. The measurement projectors implemented by the PBS, when both modes are measured in the diagonal (+/-) basis³⁵, are then,

$$\begin{aligned}\hat{\Pi}_{\text{Bell}}^+ &= |H, H\rangle\langle H, H| + |V, V\rangle\langle V, V|, \\ \hat{\Pi}_{\text{Bell}}^- &= |H, H\rangle\langle H, H| - |V, V\rangle\langle V, V|, \\ \hat{\Pi}_{\text{HV}} &= |H, V\rangle\langle H, V|, \\ \hat{\Pi}_{\text{VH}} &= |V, H\rangle\langle V, H|,\end{aligned}\quad (24.10)$$

where the former two represent successful projection onto the Bell basis, and the latter two represent failures, effectively measuring both modes in the H/V basis. This approach is described in Fig. 83.

Technically, $\hat{\Pi}_{\text{Bell}}^\pm$ are not Bell measurements, but rather projections onto the even parity subspace. A true Bell projection would implement $|\Phi^\pm\rangle\langle\Phi^\pm|$. However, in an optical context the two terms are often used interchangeably, since they exhibit effectively the same behaviour, given that the detection process is destructive.

Bell projections using CNOT gates can be implemented with arbitrarily high success probability in principle. However, in most scenarios of interest (such as cluster state preparation and entanglement purification) Bell projection using a PBS succeeds with probability of 1/2, since a PBS is only able to uniquely distinguish two of the four Bell states. To its advantage, such ‘partial’ Bell measurements only require high HOM visibility, avoiding the need for the interferometric stability inherent internally within LOQC CNOT gates.

While partial Bell state projection using a PBS is relatively straightforward, LOQC CNOT gates (which are very desirable owing to their near-determinism) are very

³⁵ By measuring in the diagonal basis we erase information about whether photons were horizontally or vertically polarised, thereby projecting onto the coherent subspace of both possibilities. Such a diagonal basis measurement may be implemented using a waveplate to perform a Hadamard polarisation rotation, followed by another PBS, separating the horizontal and vertical components, which are then independently measured via regular photo-detection.

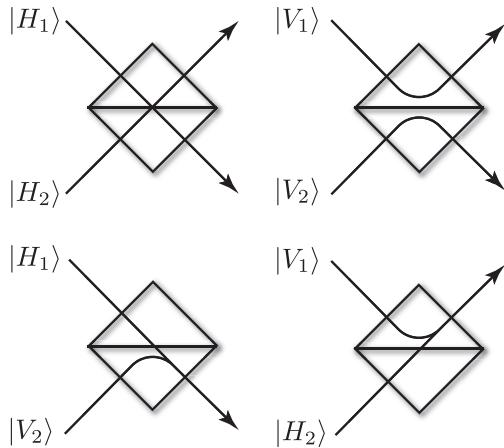


Figure 83: Partial Bell state projection using a polarising beamsplitter (PBS). The PBS completely transmits horizontally polarised light, whilst completely reflecting vertically polarised light. Shown are the four possible two-photon input states, and the respective trajectories followed by the photons. To complete the partial Bell projection we measure the output modes in the diagonal basis, $|\pm\rangle = \frac{1}{\sqrt{2}}(|H\rangle \pm |V\rangle)$, such that horizontally and vertically polarised photons cannot be distinguished. If the input state was $|H, H\rangle$ or $|V, V\rangle$, we would measure one photon in each output mode (both transmitted or both reflected). Since the detectors cannot distinguish $|H\rangle$ from $|V\rangle$, this effectively projects us onto the coherent superposition of both possibilities ('which-path erasure'), implementing the measurement projector $\hat{\Pi}_{\text{Bell}}^{\pm} = |H, H\rangle\langle H, H| \pm |V, V\rangle\langle V, V|$. If, on the other hand, we measure two photons at one output mode, we know with certainty what the polarisations of both incident photons were and we probabilistically implement one of the projectors $\hat{\Pi}_{\text{HV}} = |H, V\rangle\langle H, V|$ or $\hat{\Pi}_{\text{VH}} = |V, H\rangle\langle V, H|$, effectively performing polarisation-resolved detection upon both modes, which equates to a \hat{Z} measurement on the logical qubits. The practical outcome of this is that, when using a PBS to prepare cluster states, with probability 1/2 we are able to successfully fuse two smaller cluster states together into a larger one, and with probability 1/2 we fail to do so, instead removing two qubits from the clusters.

technologically challenging, with drastic resource overheads, particularly for high success probability. Thus, outsourcing them to the cloud may be very economically efficient.

Talk about CV equivalent

E. Matter qubits

Many non-optical systems can be indirectly measured by first entangling optical states with the matter qubits and then measuring the optical state. Because of the entanglement, projective measurement on the optical state teleports the measurement onto the matter qubit.

In Fig. 148 we illustrate a scheme for entangling two λ -configuration atoms using which-path erasure. Consider just one of these qubits in isolation. If a π -pulse is applied to the atom, the $|\downarrow\rangle$ state is excited to the $|e\rangle$ state, after

which, upon relaxation, it emits a photon. Thus, upon measurement, the presence or absence of a photon directly indicates whether the qubit was in the $|\uparrow\rangle$ or $|\downarrow\rangle$ state.

The attractive feature of this is that although the matter qubit is stationary, its indirect measurement via optical coupling may be performed over arbitrary distances across the optical network, allowing the measurement stage to be outsourced. This includes entangling measurements, useful for, for example, cluster state preparation (Sec. XL.B).

F. Quantum non-demolition measurement

When it comes to optical systems, measurement is typically *destructive*, i.e the photon (or other optical state) is destroyed in the process of measurement. An APD, for example, converts a photon to an electrical current and then the photon is gone. Can we measure the logical value of optical states without destroying them?

The answer is yes, via *quantum non-demolition measurements* (QND). The central idea here is to entangle a system which mustn't be destroyed with an ancillary system that we can happily afford to lose. Because the two systems are correlated, a destructive measurement on the ancillary state yields an effective measurement on the primary system, but without destroying it.

The simplest example to illustrate this is by considering the measurement of a single qubit,

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle. \quad (24.11)$$

We begin by introducing an additional ancillary qubit in the logical $|0\rangle$ state, which we maximally entangle to the primary system with a CNOT gate. This yields two redundantly encoded qubits,

$$\text{CNOT}|\psi\rangle|0\rangle = \alpha|0\rangle|0\rangle + \beta|1\rangle|1\rangle, \quad (24.12)$$

characterised by the same amplitudes as the original qubit. The ancilla is then measured (destructively), which reads out the original state of the primary system to which it was correlated, but which has not been destroyed and may freely continue on its journey.

Unfortunately, the required CNOT gate is cumbersome in optical architectures. But it can be done, and in Sec. XLII.A we describe in detail how such gates can be constructed.

The quantum circuit implementation for QND is shown in Fig. 84.

G. Weak measurement

The measurements considered until now have been *projective* measurements. These completely collapse the wave-function of a state onto an eigenstate of the measurement operator, and also give us perfect information about

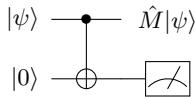


Figure 84: Quantum non-demolition measurement of a single qubit. \hat{M} is the measurement operator applied during measurement to the ancillary state, which may be destructive. The primary qubit has now been projected onto the measurement outcome \hat{M} , but has not been destroyed.

the associated eigenvalue (i.e measurement outcome). But in some scenarios this might be too aggressive – could we instead just extract *some* information about the system, and only *partially* collapse it. This is achieved using *weak measurement*.

Consider the circuit shown in Fig. 85. This is identical to the previous circuit for QND, but with the addition of the arbitrary rotation on the ancilla qubit prior to the CNOT gate.

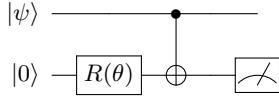


Figure 85: Quantum weak measurement. The choice of rotation on the ancillary qubit determines how much information about the state is extracted from the measurement, and also how much the state is disturbed by the measurement.

When $\theta = 0$ this circuit implements an ordinary projective measurement, projecting onto the eigenbasis of the measurement operator. We'll refer to this as a *strong measurement* – maximum information and maximum disturbance.

On the other hand, consider the case where $\hat{R}(\theta) = \hat{H}$. This transforms the ancillary $|0\rangle$ state into the $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ state. However, the $|+\rangle$ state is an eigenstate of the \hat{X} operator that resides in the CNOT gate, and is therefore invariant under the action of the CNOT. Now the measurement outcome will be completely randomised and independent of $|\psi\rangle$, revealing no information about $|\psi\rangle$ whatsoever. But simultaneously, $|\psi\rangle$ will be unaffected by the measurement since it was not entangled with the ancilla, remaining separable. We have effectively implemented no measurement at all.

However, we can choose θ to be any value in between these two extremities – *weak measurement* – where the strength of the measurement, how much information is extracted from it, and how much it disturbs the state $|\psi\rangle$, will be a function of the rotation θ . The rule is, the more information our measurement extracts, the more we disturb the state.

Add some equations.

XXV. EVOLUTION

THE evolution of optical states represents an extremely broad category of quantum operations, including: passive linear optics; post-selected linear optics; non-linear optics; and, light-matter interactions. Clearly, the items in this list present technological challenges, inaccessible to many users.

The error models in the evolution of optical states are largely accounted for by those discussed in Sec. XIV.

A. Linear optics

Linear optics networks (?) implement unitary linear maps on the photonic creation operators, of the form,

$$\hat{U}\hat{a}_i^\dagger\hat{U}^\dagger \rightarrow \sum_{j=1}^m U_{i,j}\hat{a}_j^\dagger, \quad (25.1)$$

where \hat{a}_i^\dagger is the photonic creation operator on the i th of the m modes, and U may be any $SU(m)$ matrix. It was shown by (?) that arbitrary transformations of this form may be decomposed into $O(m^2)$ linear optical elements (beamsplitters and phase-shifters), enabling efficient construction of arbitrary linear transformations. Furthermore, the algorithm for determining the decomposition of such transformations has polynomial classical runtime (i.e residing in **P**). Note that the original Reck *et al.* decomposition is not unique, and various other topologies of optical elements also enable universality, each with their own implementational advantages and disadvantages.

Each individual beamsplitter in such a decomposition is an arbitrary $SU(2)$ matrix acting on two photonic creation operators, \hat{a}^\dagger and \hat{b}^\dagger ,

$$\begin{pmatrix} \hat{a}_{\text{out}}^\dagger \\ \hat{b}_{\text{out}}^\dagger \end{pmatrix} = \begin{pmatrix} e^{i\phi_1}\sqrt{\eta} & e^{i\phi_2}\sqrt{1-\eta} \\ e^{-i\phi_2}\sqrt{1-\eta} & -e^{-i\phi_1}\sqrt{\eta} \end{pmatrix} \begin{pmatrix} \hat{a}_{\text{in}}^\dagger \\ \hat{b}_{\text{in}}^\dagger \end{pmatrix}, \quad (25.2)$$

where $0 \leq \eta \leq 1$ is the reflectivity, and $0 \leq \phi_1, \phi_2 \leq 2\pi$ determine the phase relationships.

When operating in the polarisation basis, wave-plates enable the same transformation as beamsplitters do in dual-rail encoding. The phase-shifters implement the unitary operation,

$$\hat{\Phi}(\phi) = e^{-i\phi\hat{n}}, \quad (25.3)$$

or equivalently in the Heisenberg picture,

$$\hat{U}\hat{a}^\dagger\hat{U}^\dagger \rightarrow e^{-i\phi\hat{a}^\dagger}, \quad (25.4)$$

for phase-shift ϕ .

These linear optics evolutions are most commonly implemented using either:

- Bulk optics: discrete optical elements are arranged on an optical table.

- Time-bin architectures: time-bin encoded qubits (Sec. XIII.C.2) evolve through delay lines and interfere at a single central optical component.
- Integrated waveguides: all passive components are etched into a chip. When optical modes are brought physically close together, evanescent coupling allows photons to coherently hop between modes. This gives rise to evolution described by the coupled oscillator Hamiltonian, given in Eq. (42.41), where the coupling coefficients are dictated by the proximity and geometry of the waveguides.

These three main contenders are illustrated in Fig. 86.

B. Non-linear optics

Aside from the linear transformations described above, which are passive and photon-number-preserving, various active, non-linear interactions are also of interest to optical quantum information processing. The most prominent of these are primarily considered as transformations in phase-space (Sec. XIII.E), using, for example, the Wigner function representation.

The most well-known non-linear transformation is the displacement operation, which translates the Wigner function by some arbitrary amplitude in phase-space, whilst preserving all other features of the phase-space representation. This is described by the unitary operator,

$$\hat{D}(\alpha) = \exp [\alpha \hat{a}^\dagger - \alpha^* \hat{a}], \quad (25.5)$$

where α is the displacement amplitude. This transformation is easily implemented by mixing a state on a low-reflectivity beamsplitter with a coherent state of some arbitrary complex amplitude, which determines the displacement amplitude. In the special case of a displacement operator acting on the vacuum state, we obtain a coherent state of equal amplitude, $\hat{D}(\alpha)|0\rangle = |\alpha\rangle$.

Another common non-linear transformation is squeezing, discussed in Sec. XXIII.I. This implements the unitary operator,

$$\hat{S}(\xi) = \exp \left[\frac{1}{2} (\xi^* \hat{a}^2 - \xi \hat{a}^{\dagger 2}) \right], \quad (25.6)$$

where ξ is the squeezing parameter, which has the effect of applying a dilation of some arbitrary factor along a particular axis in phase-space.

Thus, jointly, the displacement and squeezing operators enable arbitrary translations and dilations in phase-space. These operations form the basis for CV quantum computing schemes, to be discussed in more detail in Sec. XLII.E.

C. Non-optical systems

There are countless non-optical systems applicable to quantum information processing applications. For exam-

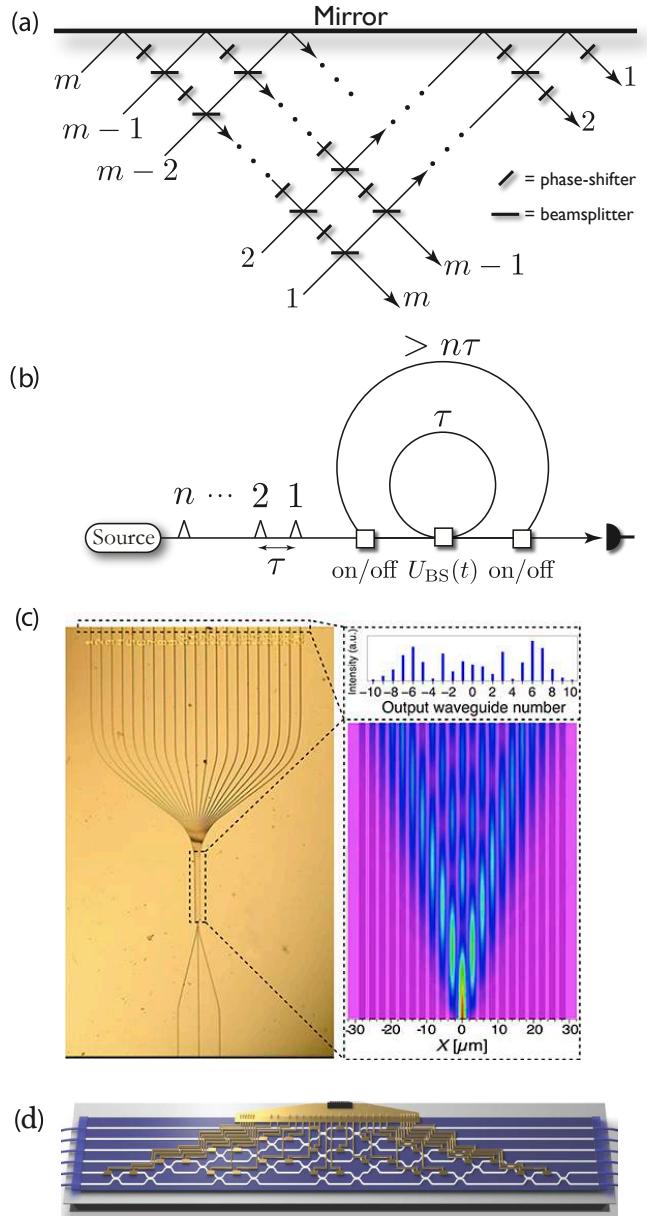


Figure 86: The three primary approaches for implementing linear optics transformations. (a) Bulk optics, where each optical mode is spatially encoded, and the linear transformation is decomposed into a discrete array of beam splitters and phase-shifters, an appropriate choice of which enables arbitrary linear optics transformations to be implemented. (b) Time-bin encoding, where each optical mode is designated a distinct time-bin. Fibre-loops meeting at dynamically reconfigurable beam splitters enable arbitrary linear transformations to be implemented. (c) Integrated wave-guide chips, where evanescent coupling between neighbouring wave-guides within a chip facilitates interference between modes (graphic courtesy of Alberto Peruzzo (?)). (d) Integrated wave-guide chip with electrically controllable phase-shifters, implementing a programmable, universal 6×6 linear optics network (graphic courtesy of Jeremy O'Brien (?)).

ple, quantum computing schemes have been described using:

- Two-level.
- λ -configuration atoms.
- Superconducting rings.
- Ion traps.
- Atomic ensembles.
- Countless more...

From a networking perspective, we are not terribly interested in the inner workings of all these schemes, as we are reasonably confident that optics will be mediating networking, even if other aspects of the protocol are non-optical. Thus, we will not go into great detail about the evolution of non-optical systems. Instead, for our purposes, the relevant issue is interfacing between optical and non-optical systems, such that networking protocols between them may be implemented. Optical interfacing is discussed in detail in Sec. XX.A.

XXVI. QUANTUM MEMORY

Discuss atomic ensembles, two-level systems, lambda systems, 3-level systems with two ground states (better since no decay).

A final building block, that will be essential in many networks, is quantum memory, which simply delays a packet by some fixed amount of time, ideally implementing an identity channel ($\hat{1}$) in the non-temporal degrees of freedom. This will be required when, for example, quantum data packets reach a network bottleneck, and face one of two options: wait, or be discarded. As discussed earlier, discarding quantum packets is often a highly undesirable enterprise, as they often cannot be easily recreated, most notably when entangled with other systems. Quantum memory is an essential component in quantum repeater networks, to be discussed in Sec. XXIX.

A. Network graph representation

Quantum memory is modelled in our network graph representation as per Fig. 87, via a self-loop implementing a process that delays packets. Ideally, the associated process should implement the identity operation in all degrees of freedom, except the temporal one, affecting only the LIFETIME metric of the packets passing through it, incrementing it by the duration of the quantum memory.

Note that this is not directly compatible with conventional shortest-path algorithms, which ignore self-loops. One approach is to modify our strategy optimisation algorithms to accommodate self-loops. Alternately, we could

construct a ‘virtual’ graph, obtained by adding additional nodes to the network, with connections determined by ‘unravelling’ the self-loops. For example, in Fig. 87, we could eliminate the self-loop, and instead replace B with multiple redundant nodes in series between A and C , each associated with their own latency cost.

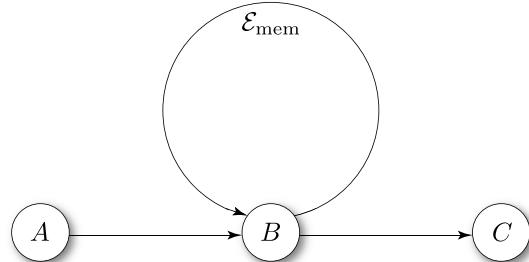


Figure 87: Simple model for a quantum memory via a self-loop that passes through a memory process, E_{mem} . Ideally, E_{mem} does not affect any of the costs or attributes of states passing through the link, except for the LATENCY cost, which is incremented according to the duration of the memory.

B. Physical implementation

At the physical level, there are two main approaches we could use to put optical states into memory. The first is simply to employ optical delay lines (shown in Fig. 88), either in free-space or in fibre. The second is to interface the state with a non-optical system with a long coherence lifetime, which holds the information content until needed before being out-coupled. This can be achieved using, for example, the light-matter interfacing techniques discussed in Sec. XX.A.

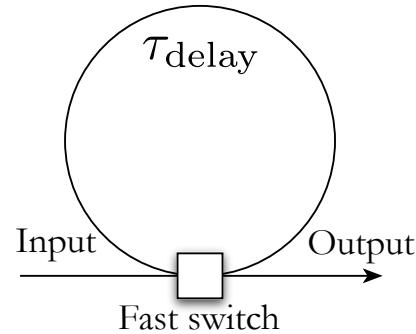


Figure 88: Simple architecture for a controllable optical quantum memory using an optical delay line. A length of optical fibre (or free-space), with roundtrip time τ_{delay} , couples to a fast controllable switch, with switching time $\tau_{\text{switch}} \ll \tau_{\text{delay}}$. The switch allows the optical state to be coupled into the loop, maintained there for an arbitrary number of roundtrips, and then coupled out. The storage time of the memory is restricted to being integer multiples of the roundtrip time, $\tau_{\text{storage}} = n \cdot \tau_{\text{delay}}$, where $n \in \mathbb{Z}_+$ is the number of roundtrips. If the efficiency of the fibre-loop is η , the effective efficiency of the quantum memory after readout is $\eta_{\text{eff}} = \eta^n$.

The former is experimentally straightforward, but plagued by loss, and is only suitable over short timescales, on the order of nanoseconds. The latter is more experimentally challenging, but can achieve longer storage times, limited by the lifetime (T_1 - and T_2 -times) of the non-optical system. For some physical systems, this can be very high, on the order of milliseconds for atomic ensemble qubits (??), for example, which is typically adequate for the purposes of waiting out network bottlenecks.

C. Error correction

Since quantum memories are subject to errors that accumulate with time, long-life quantum memories will necessarily require error correction mechanisms to preserve qubit states held within them.

To achieve this, standard QEC codes (Sec. XVI.A.6) can be employed to encode a number of logical qubits into a larger number of physical qubits, which are held in quantum memory and undergo active error correction. Any of the previously discussed QEC techniques are applicable to this.

An alternate approach is to use W-state encoding to implement unitary error averaging (Sec. XVIII.B), where the unitaries are single-qubit channels (??). The protocol is shown in Fig. 89. We begin by using a QFT fanout operation to encode a dual-rail encoded qubit across N optical modes. Each optical mode then feeds into a solid-state qubit (e.g a two-level atom), which ideally implement an identity channel but are inevitably subject to usual error processes such as dephasing and amplitude damping (characterised by T_1 - and T_2 -times). Finally, the inverse fanout operation is applied and success of the protocol is defined as there being exactly one photon between the first two ('success') output modes. If a photon is detected in any of the other ('failure') modes the state is presumed erroneous and discarded.

For a logical qubit of the form,

$$\begin{aligned} |\psi\rangle_L &= \alpha|0\rangle + \beta|1\rangle, \\ \alpha &= \cos\left(\frac{\theta}{2}\right), \\ \beta &= e^{i\phi} \sin\left(\frac{\theta}{2}\right), \end{aligned} \quad (26.1)$$

the heralding success probability, and associated fidelity of the heralded events is given by,

$$\begin{aligned} P_H &= 1 - e^{-t/T_1} + e^{-t/T_1}(e^{-t/T_2} + \frac{2}{N}(1 - e^{-t/T_2})), \\ F_H &= e^{-t/T_1} \frac{e^{-t/T_2} + (1 - e^{-t/T_2})\left(\frac{2+\sin^2\theta}{2N}\right)}{e^{-t/T_2} + \frac{2}{N}(1 - e^{-t/T_2})}. \end{aligned} \quad (26.2)$$

Fig. 90 illustrates the fidelity of an error-corrected logical qubit using W-state encoding.

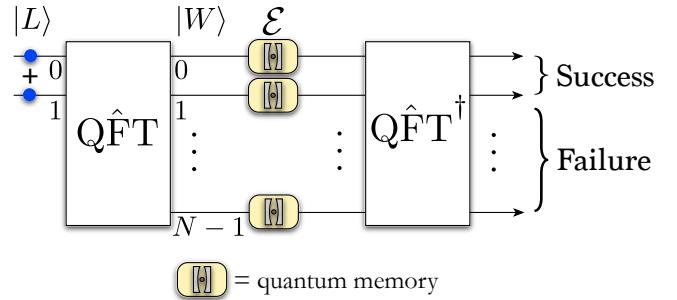


Figure 89: Quantum memory based on W-state encoding. Rails represent optical modes, and the input logical qubit state, $|L\rangle$, is represented using dual-rail encoding. The QFT fanout operation maps each of the two input basis states to one of two orthogonal N -qubit W-states, $|W\rangle$, differing only by their phase relationships. These are then fed into a bank of N quantum memories. To readout the memories we convert back to optical encoding and decode using the inverse QFT operation. Success of the protocol is defined as there being exactly one photon between the first two 'success' output modes. If a photon leaks into any of the other 'failure' output modes we discard the state and assume it was erroneous. Thus error detection is heralded via the presence or absence of a photon in the 'failure' modes.

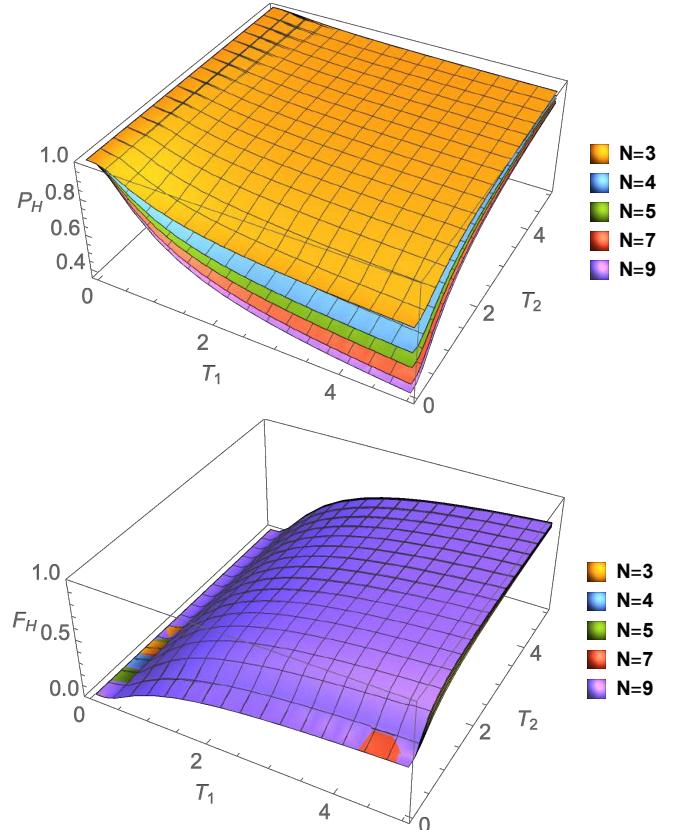


Figure 90: Heralding probability (P_H) and associated error-corrected fidelity (F_H) for a single photonic qubit stored in a W-state quantum error correction circuit with decoherence times T_1 (amplitude damping) and T_2 (dephasing). We have chosen $\theta = 0$ (representing worst-case behaviour), storage time $t = 1$, and several levels of encoding N .

XXVII. HIGH-LEVEL PROTOCOLS

BUILDING upon the aforementioned primitive protocols for quantum networking, we can construct a plethora of higher-level protocols that implement more powerful end-user applications. These high-level protocols are ubiquitous in quantum information processing and form building blocks for even more powerful architectures, such as full cloud quantum computing, to be discussed in Sec. XLIII.

A. Random number generation

“God does not play dice!” — Albert Einstein.

Perhaps the simplest quantum information processing task is that of perfect random number generation. True random numbers have widespread applications in cryptography, Monte-Carlo simulations, and any type of randomised (e.g BPP) algorithm.

Classical random number generators are actually deterministic, following the laws of classical physics, but so difficult to predict that we accept them to be as good as random. But for some applications this isn't enough, and we must make sure that no correlations of any type exist between different random numbers, or between the random numbers and their environment.

This can be achieved in many different ways quantum mechanically. Ultimately, they are all based on the Heisenberg uncertainty principle, that certain quantum mechanical measurements yield uncertainty. The procedure is shown in Alg. XX, and a simple optical implementation is shown in Fig. 91.

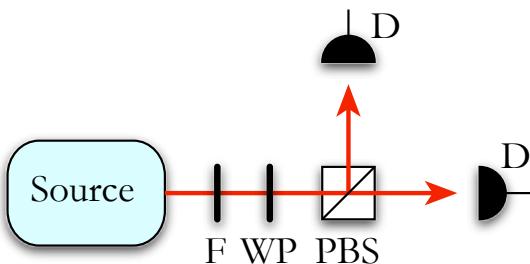


Figure 91: A simple optical implementation for a quantum binary random number generator. The source prepares single photons, which pass through a polarisation filter (F) to retain only horizontally polarised photons, $|H\rangle$. A wave-plate (WP) then rotates the polarisation into the diagonal basis, yielding a stream of $|+\rangle$ polarised photons. These are then separated into horizontal and vertical components via a polarising beamsplitter (PBS), each of which are independently photo-detected (D). If correctly implemented, the two detectors click one at a time, each with 50% probability, and the random stream is inherently non-deterministic, guaranteed by the non-determinism of quantum mechanics. Random numbers are produced at a rate of 1 bit per photo-detection.

function RandomBit():

1. Prepare the equal superposition state,

$$|\psi\rangle_{in} = |+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle). \quad (27.1)$$

2. Most commonly this is in the polarisation basis,

$$\begin{aligned} |H\rangle &\equiv |0\rangle, \\ |V\rangle &\equiv |1\rangle. \end{aligned} \quad (27.2)$$

3. Measure the state in the logical basis, with measurement projectors,

$$\begin{aligned} \hat{\Pi}_0 &= |0\rangle\langle 0|, \\ \hat{\Pi}_1 &= |1\rangle\langle 1|. \end{aligned} \quad (27.3)$$

4. The measurement outcomes occur with probabilities,

$$\begin{aligned} P_0 &= |\langle 0|+\rangle|^2 = \frac{1}{2}, \\ P_1 &= |\langle 1|+\rangle|^2 = \frac{1}{2}, \end{aligned} \quad (27.4)$$

following a uniform, random, binary distribution.

5. Repeat for as many random bits as are required.

6. \square

Algorithm XX: Procedure for the generation of random bit-strings. Assuming the device is perfectly implementing this procedure, we will measure a perfect random 50/50 distribution between the two measurement outcomes. Note that the procedure requires no quantum interference, and no entanglement. Only single-qubit state preparation and measurement are required. Thus, a single-photon source, wave-plate, polarisation filter, and photo-detector are sufficient for its realisation. Favourably, if the detector is inefficient it simply reduces the bit-rate, but does not compromise the randomness of the distribution. The scheme is very robust, in the sense that there are no temporal synchronisation or mode-matching requirements.

The cynics amongst us might question the non-determinism of the laws of Nature, and ask whether quantum random numbers really are truly random (in the sense of non-determinism), or whether they also are just too hard to predict that we treat them as effectively random. The answer to this is that it has been proven that quantum mechanics is inconsistent with ‘hidden variable theories’ (?), i.e that there is an underlying, but inaccessible determinism in the world, which is guiding quantum measurements in a completely deterministic manner. This disproof effectively validates the notion of quantum mechanical perfect random number generation.

Consider the scenario where a client needs a stream of true random numbers for use in her Monte-Carlo simulation algorithm or as a secret-key for her email encryption.

She has limited quantum resources herself, so she outsources it to her better-equipped mate. Depending on her own resource limitations and potential security considerations, her friend could either: (1) implement the full protocol described above, providing her with a classical random bitstream; or (2) only take care of photon generation, providing her with a perpetual source of high-quality photons for her to measure herself using a simple photo-detector. (1) and (2) would both be suitable if the intention was to apply the source of randomness to a Monte-Carlo simulation. But in a cryptographic scenario, where the randomness is being used for key generation, clearly Alice could not outsource the measurement stage without revealing her secret-key. In this instance, Bob can act as the provider of photons, while Alice does the measurements herself so as to keep her random bit-string secret.

For cryptographic purposes, there are far more stringent constraints placed upon our random number generator than for use in say a Monte-Carlo computer simulation – cryptographic random number generation. In this context, the demands placed upon the amount of bias or correlations in the random number stream are very stringent. An enormous amount of research has been invested into this topic, and sophisticated statistical tests have been developed for establishing crypto-worthiness of a random number stream. The fact that quantum random numbers, via the inherent non-determinism of quantum mechanics, do not obey any hidden variable theory, implies there is intrinsically no underlying ‘seed’ to the random number stream, which reveals the entire deterministic sequence. This is unlike any classical generator, where there always is such a seed, but it is simply taken to be too hard to determine.

This scenario is an obvious example of where a UDP-like SEND-AND-FORGET protocol may be viable. Unlike most other applications, Bob is broadcasting a stream of identical, pure quantum states, that are not entangled with any peripheral system, and are easily replicated, with no correlations between distinct photons. Thus, if any particular photon fails to reach Alice, it matters not, as she can simply await the next one emanating from Bob’s bombardment of photons (the ‘shotgun’ approach). There are no QoS requirements.

B. Entanglement purification

Entangled states, most notably Bell pairs (Sec. XXIII.G), play a central role in many quantum technologies. These maximally entangled states are easily represented optically using polarisation encoding of single photons, and can be non-deterministically prepared directly using SPDC (Sec. XXIII.B), or post-selected linear optics.

Bell pairs are the basis for building cluster states

(Sec. XL.B), some quantum cryptography protocols (Sec. XXXVII.A), and quantum teleportation (Sec. XXVII.C), to name just a few applications. Therefore distributing entangled states with the highest entanglement metrics is extremely important. In short, entanglement can be considered a valuable quantum resource (discussed in detail in Sec. XXVIII), upon which many other protocols may be built.

Suppose Alice and Bob share an entangled pair. Quantum mechanics, specifically the very definition of entanglement itself, prohibits local operations performed by Alice and Bob from increasing the level of entanglement. However, if Alice and Bob share multiple pairs, they can perform an operation known as *entanglement purification* or *entanglement distillation*, whereby two lower-fidelity entangled pairs are consumed and projected onto a single entangled pair with higher fidelity (??). Such protocols will be extremely useful in protocols where achieving the highest possible degree of entanglement is paramount, for example when error thresholds must be achieved for the purpose of error-correction and fault-tolerance (?).

Taking two polarisation-encoded photonic Bell pairs, say $|\Psi^+\rangle$, and subjecting them to a dephasing error model (Sec. XIV.D) yields a mixed state of the form,

$$\hat{\rho}_{\text{in}} = F|\Psi^+\rangle\langle\Psi^+| + (1 - F)|\Psi^-\rangle\langle\Psi^-|, \quad (27.5)$$

where F is the entanglement fidelity, which is a function of the dephasing rate. Note that $|\Psi^+\rangle$ and $|\Psi^-\rangle$ are related by local Pauli phase-flip operations (\hat{Z}) applied to either qubit,

$$|\Psi^-\rangle = \hat{Z}_A|\Psi^+\rangle = \hat{Z}_B|\Psi^+\rangle. \quad (27.6)$$

A linear optics entanglement purification protocol can be simply implemented using two polarising beamsplitters PBSs (??). Alice uses one PBS to interfere the photons from her side of each of the photon pairs, measuring one output only, which implements a non-deterministic, partial Bell state projection. Bob does the same on his side. What’s left is one photon in Alice’s hands and one in Bob’s. When successful, they will now be sharing a single entangled pair of higher Bell state fidelity than the two starting states. The protocol is shown in Fig. 92.

Note that when using PBSs to perform the Bell projections, the protocol is necessarily non-deterministic, since PBSs are only able to distinguish two of the four Bell states. Thus, each PBS has a success probability of $1/2$. And there are two PBSs per instance of the protocol, therefore the net success probability is $1/4$. When concatenated, n applications of the protocol thus has an exponentially low success probability of $1/4^n$. This could be overcome using deterministic CNOT gates (Sec. XLII.A), but these are challenging using linear optics.

Furthermore, the protocol consumes two Bell pairs upon each trial, only one quarter of which are successful. Thus, on average, 8 Bell pairs are consumed for every

purified Bell pair prepared, and the expected number of Bell pairs required to perform n iterations of entanglement purification grows exponentially as 8^n .

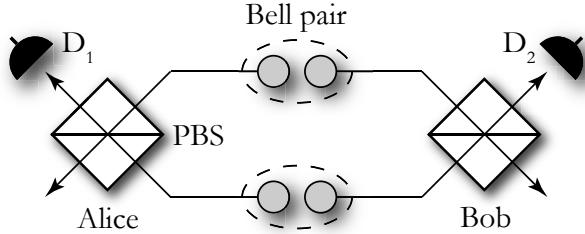


Figure 92: Elementary entanglement purification using linear optics. Two Bell pairs are distributed between Alice and Bob, each of which has been subject to a dephasing error model. Alice and Bob perform Bell measurements on their two qubits using a PBS and polarisation-resolved photo-detection (D_1 and D_2). Upon successful Bell state projection (Bell measurements are necessarily non-deterministic using linear optics), Alice and Bob will share a single Bell pair with higher fidelity than the two input pairs.

Specifically, the relationship between the input (F_{in}) and output (F_{out}) fidelities of the protocol is,

$$F_{\text{out}} = \frac{F_{\text{in}}^2}{F_{\text{in}}^2 + (1 - F_{\text{in}})^2}. \quad (27.7)$$

This input/output relationship is shown in Fig. 93.

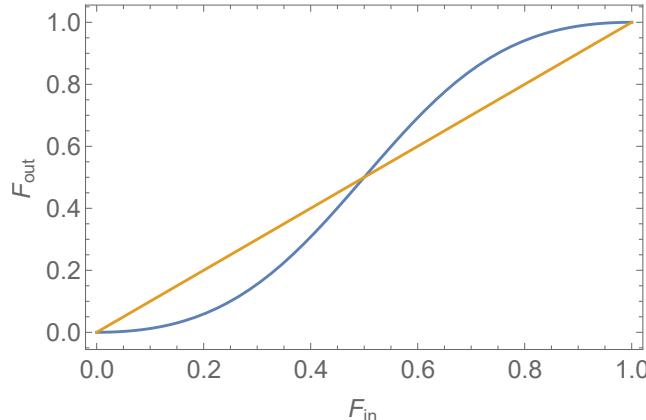


Figure 93: Entanglement purification of two polarisation-encoded, photonic Bell pairs. F_{in} (F_{out}) are the input (output) fidelities of the Bell pairs. The protocol consumes two Bell pairs for every one purified pair. The straight line represents the break-even point in terms of state fidelity, above which the protocol enhances fidelity, and below which reduces it. This places a strict bound on the fidelity of Bell pairs reaching the purifier. This equates to route cost, if measured by the fidelity metric, stipulating network performance requirements. This threshold requirement presents an example of where an ALL OR NOTHING strategy might be appropriate.

Note that there is a break-even point, above which the protocol strictly increases fidelity, and below which strictly decreases it. This occurs at $F_{\text{in}} = 1/2$. Provided

pairs can be communicated above this fidelity threshold, bootstrapped application of the protocol could be employed to boost entanglement fidelity asymptotically close to unity (but with exponential resource overhead, since each operation non-deterministically consumes two pairs to produce one). But below this threshold it is impossible to recover any more entanglement than we started with. This provides an example of an application where the protocol being implemented dictates strict requirements on network cost metrics. Specifically, assuming perfect Bell pairs to begin with, the routes by which they are communicated must strictly ensure entanglement fidelities of at least $F = 1/2$ upon reaching their destination. Here, a type of ALL OR NOTHING networking strategy would be applicable – if the fidelity requirement is not met, the state cannot be purified and might as well be thrown away to make way for other traffic.

A theoretical analysis of this protocol has been performed, accounting for mode-mismatch (Sec. XIV.G) in the protocol (?), where it was found that mode-mismatch shifts the break-even point upwards, and lowers the maximum value of F_{out} – with more mode-mismatch, a higher starting fidelity is required to break even, and we achieve a lower, sub-unity output fidelity. In this case, a cost function that combines the dephasing and mode-mismatch metrics of the network will be required.

Importantly, this protocol is based on partial Bell state measurement, and therefore does not require interferometric stability, only high HOM visibility, thus making stabilisation comparatively easy over long distances.

Entanglement purification can also be performed using physical encodings other than single photons. For example, this has been demonstrated using Gaussian CV quantum states (?).

C. Quantum state teleportation

Quantum state teleportation (?) is an essential ingredient in many higher-level protocols. It forms the basis of cluster state quantum computing (Sec. XL.B), some QEC codes, the KLM linear optics quantum computing scheme (Sec. XLII.A), and can act as a mediator for long-range transmission of quantum states, amongst others.

In the standard teleportation protocol, Alice begins with a single qubit,

$$|\phi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad (27.8)$$

which she would like to teleport to Bob. Importantly, no quantum communication between the two is allowed, since obviously this would make the problem trivial. However, classical communication is allowed (and turns out to be necessary), and furthermore they share an entangled Bell pair as a resource. Thus, Alice begins with two qubits, and Bob begins with one – his half of the entangled pair onto

which Alice's state ought to be teleported. The initial state is therefore,

$$\begin{aligned} |\psi\rangle_{\text{in}} &= |\phi\rangle_{A_1} |\Psi^+\rangle_{A_2, B} \\ &= \frac{1}{\sqrt{2}} (\alpha|0\rangle_{A_1} + \beta|1\rangle_{A_1}) (|0\rangle_{A_2}|1\rangle_B + |1\rangle_{A_2}|0\rangle_B). \end{aligned} \quad (27.9)$$

The first step of the protocol is for Alice to perform a 2-qubit entangling measurement on her two qubits, projecting onto the Bell basis, Eq. (23.12). She obtains one of four measurement outcomes. For illustration, suppose she measures the $|\Psi^+\rangle$ outcome. Then the projected state is,

$$\begin{aligned} |\psi\rangle_{\text{proj}}^{\Psi^+} &= \langle\Psi^+|_{A_1, A_2} |\psi\rangle_{\text{in}} \\ &= \frac{1}{\sqrt{2}} \langle\Psi^+|_{A_1, A_2} |\psi\rangle_{A_1} (|0\rangle_{A_2}|1\rangle_B + |1\rangle_{A_2}|0\rangle_B) \\ &= \frac{1}{2} (\langle 0|_{A_1} \langle 1|_{A_2} + \langle 1|_{A_1} \langle 0|_{A_2}) \\ &\quad \cdot (\alpha|0\rangle_{A_1} + \beta|1\rangle_{A_1}) (|0\rangle_{A_2}|1\rangle_B + |1\rangle_{A_2}|0\rangle_B) \\ &= \frac{1}{2} (\alpha|0\rangle_B + \beta|1\rangle_B) \\ &= \frac{1}{2} |\phi\rangle_B, \end{aligned} \quad (27.10)$$

which is Alice's initial state. For all four possible Bell measurement outcomes we have,

$$\begin{aligned} |\psi\rangle_{\text{proj}}^{\Psi^+} &= \frac{1}{2} (\alpha|0\rangle_B + \beta|1\rangle_B) \\ &= \frac{1}{2} |\phi\rangle_B, \\ |\psi\rangle_{\text{proj}}^{\Psi^-} &= \frac{1}{2} (\alpha|0\rangle_B - \beta|1\rangle_B) \\ &= \frac{1}{2} \hat{Z} |\phi\rangle_B, \\ |\psi\rangle_{\text{proj}}^{\Phi^+} &= \frac{1}{2} (\alpha|1\rangle_B + \beta|0\rangle_B) \\ &= \frac{1}{2} \hat{X} |\phi\rangle_B, \\ |\psi\rangle_{\text{proj}}^{\Phi^-} &= \frac{1}{2} (\alpha|1\rangle_B - \beta|0\rangle_B) \\ &= \frac{1}{2} \hat{X} \hat{Z} |\phi\rangle_B, \end{aligned} \quad (27.11)$$

which are all locally equivalent to $|\phi\rangle$ under Pauli gates, and can be corrected by Bob, given communication of the classical Bell measurement outcome provided by Alice. The full protocol is described in Alg. XXI.

In general, the protocol is deterministic, although using PBSs to perform partial Bell measurements, the success probability is at most 1/2.

The question now is what error metrics apply and how do they accumulate in the teleportation protocol. The answer is straightforward – the final teleported qubit accumulates all local Pauli errors (e.g. dephasing or depolarisation) associated with Alice's input state as well

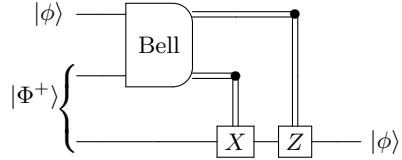
function StateTeleportation($|\phi\rangle_{A_1}$, $|\Phi^+\rangle_{A_2, B}$):

1. Alice prepares the state $|\phi\rangle_{A_1}$, which she would like to teleport to Bob.
2. Alice and Bob share the Bell pair $|\Phi^+\rangle_{A_2, B}$.
3. Alice performs a Bell state projection between qubits A_1 and A_2 .
4. Alice communicates the classical measurement outcome to Bob – one of four outcomes.
5. Bob applies an appropriate local correction to his qubit – some combination of the Pauli operators \hat{X} and \hat{Z} – according to the classical measurement outcome:

$$\begin{aligned} |\Psi^+\rangle \langle \Psi^+| &\rightarrow \hat{1}, \\ |\Psi^-\rangle \langle \Psi^-| &\rightarrow \hat{Z}, \\ |\Phi^+\rangle \langle \Phi^+| &\rightarrow \hat{X}, \\ |\Phi^-\rangle \langle \Phi^-| &\rightarrow \hat{Z} \hat{X}. \end{aligned} \quad (27.12)$$

6. Bob is left with the state $|\phi\rangle_B$.

7. \square



Algorithm XXI: Quantum state teleportation of a single qubit.

as any that acted upon the shared Bell pair. That is, the errors get teleported along with the state being teleported, plus any errors on the Bell pair.

In the case of loss, loss of either of Alice's qubits will immediately be detected when she performs her Bell measurement. Thus, loss becomes a located error, and the knowledge of the error allows the associated packet to be discarded, and the sender and recipient notified. On the other hand, loss of Bob's qubit will behave no differently than loss acting on an ordinary qubit channel.

Thus, in terms of Pauli errors, no special treatment is required by the QTCP protocol – it is almost as if the teleportation protocol weren't there. And in terms of loss, the Bell state projection diagnoses lost qubits, allowing appropriate action to be taken, which is actually better than if the error were undiagnosed. These are often referred as *located* and *unlocated* errors.

The total resources required to teleport a single-qubit state are:

1. The qubit to be teleported.

2. A shared Bell pair.
3. A 2-qubit entangling measurement in the Bell basis.
4. The transmission of two classical bits.
5. Two classically-controlled Pauli gates for correction.

This is more costly than sending the qubit directly over a quantum channel, but may be the only approach if a direct link is not available. In the context of an internet where entanglement distribution is treated as the fundamental resource (Sec. XXVIII), state teleportation is the natural approach for communicating quantum states, since no quantum communication of any kind is required once the two parties have a shared Bell pair between them.

The important feature of this protocol to note is that there is no direct quantum communication between Alice and Bob, only a classical communications channel. Rather, the Bell pair mediates the transfer of quantum information, despite there being no direct quantum channel between Alice and Bob.

Relying on teleportation rather than direct quantum communication makes frugal use of quantum channels, since there is no need for direct quantum routes between every pair of nodes in the network. Instead, each node need only have a direct one-way quantum link with the central authority responsible for entanglement distribution, thereby significantly reducing the complexity of the topology of the quantum network.

The Bell state measurement can be implemented either using a CNOT gate, or as a non-deterministic partial Bell state measurement using a PBS (Sec. XXIV.D), both of which are non-deterministic using purely linear optics.

The above describes quantum state teleportation at the level of single qubits. However, when dealing with more general QTCP packets, which may have multi-qubit payloads, we may wish to teleport an entire packet. This is implemented as a simple extension of the above procedure – we simply implement n multiple independent teleportation protocols to all of the packet’s n constituent qubits. Via linearity, although the teleportation protocols are being applied independently to each qubit, the net packet teleportation operation will preserve their joint state, including entanglement between them. Note, however, that if the qubit state teleportation protocols are individually non-deterministic with success probability p_{teleport} , the net success probability for the teleportation of the entire packet scales inverse exponentially with n , as p_{teleport}^n .

1. Open-destinations

In the standard quantum state teleportation protocol there is a single sender and a single receiver. A generalisation of the protocol is *open-destination quantum state*

teleportation, whereby there is still just one sender, but any number of potential recipients. At the time of transmission, the sender does not specify the recipient, but rather wishes to ‘broadcast’ the state to *all* recipients, such that any *one* of them can subsequently read out the state. Note that only a single recipient may actually perform the readout, since multiple readouts would violate the no-cloning theorem. However, the key new feature introduced by this variant of the protocol is that the final choice of which recipient performs the readout needn’t be known in advance, but can be decided at an arbitrary later stage, well after the sender has completed their side of the protocol.

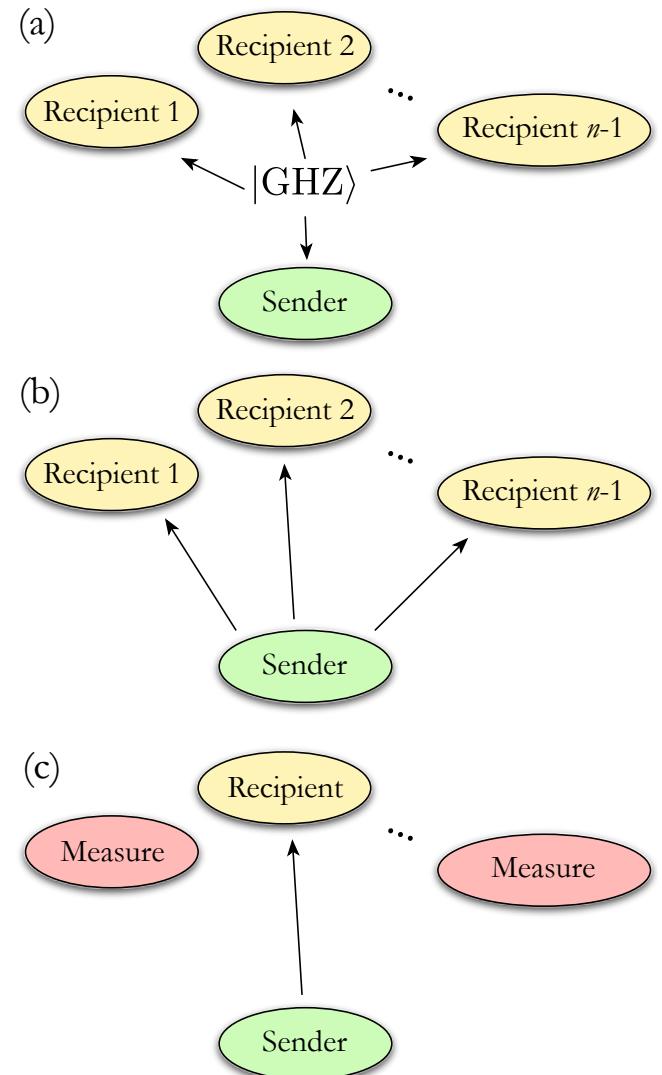


Figure 94: Open-destination quantum state teleportation. (a) GHZ state distribution from a central server between the sender and all potential recipients. (b) Sender performs quantum state teleportation, resulting in the state being teleported to all recipients in redundantly-encoded form. (c) All non-receivers measure out their qubits in the \hat{X} -basis, resulting in the teleported state arriving at the destination of just the chosen true receiver.

To implement this protocol, outlined in Fig. 94, rather than first distributing a Bell pair between sender and recipient, we distribute an n -party GHZ state between the sender and the $n - 1$ recipients,

$$|\psi_{\text{GHZ}}^{(n)}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}). \quad (27.13)$$

Next, note that performing an \hat{X} -basis measurement on one qubit in a GHZ state reduces it to an $n - 1$ -qubit GHZ state, up to a potential local \hat{Z} -correction,

$$\begin{aligned} \langle + |\psi_{\text{GHZ}}^{(n)}\rangle &= |\psi_{\text{GHZ}}^{(n-1)}\rangle, \\ \langle - |\psi_{\text{GHZ}}^{(n)}\rangle &= \hat{Z} |\psi_{\text{GHZ}}^{(n-1)}\rangle. \end{aligned} \quad (27.14)$$

Performing this contraction repeatedly ultimately reduces us all the way down to a single Bell pair, since,

$$|\psi_{\text{GHZ}}^{(2)}\rangle = |\Phi^+\rangle. \quad (27.15)$$

Thus, to teleport from sender to recipient, all non-recipients simply measure their qubits in the \hat{X} -basis, and publicly report their classical measurement outcomes for the purposes of performing local corrections. What is left is a Bell pair between sender and recipient. This Bell pair may then be employed in the conventional teleportation protocol to perform the teleportation.

The key observation now is that the dynamics of this entire system must be invariant under the time-ordering of the \hat{X} -basis measurements. Thus, whether they are performed at the beginning of the protocol (thereby directly reducing us to standard 2-party teleportation), or deferred until later, makes no difference to the final state obtained by the recipient. This is the basis for the ability of the protocol to broadcast the teleported state, without first specifying the intended recipient.

The full protocol is described in Alg. XXII. This protocol was successfully experimentally demonstrated photonically using 5 qubits (?), with high teleported state fidelities.

D. Quantum gate teleportation

Using quantum *state* teleportation as a primitive building block, quantum *gate* teleportation may be implemented (?). Here rather than teleporting a quantum state from one physical system to another, we teleport the action of a quantum gate onto a physical system (archetypically a maximally entangling 2-qubit gate, such as a CNOT gate).

The general outline of the derivation of the protocol for teleporting a CNOT gate onto a 2-qubit state is shown in Alg. XXIII.

Most notably, gate teleportation is useful when attempting to apply 2-qubit entangling operations using non-deterministic gates, in which case gate teleportation

function OpenDestTeleportation($|\phi\rangle_A$, $|\psi_{\text{GHZ}}^{(n)}\rangle$):

1. An n -qubit GHZ state is distributed between the sender and $n - 1$ potential recipients,

$$|\psi_{\text{GHZ}}^{(n)}\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}). \quad (27.16)$$

2. The sender completes their side of the protocol as per the ordinary 2-party quantum state teleportation protocol.
3. The sender's classical measurement outcome is broadcast to all potential recipients.
4. Wait until the potential recipients have decided between themselves (classically) who shall be the recipient.
5. All non-recipients measure their qubits in the \hat{X} -basis and publicly announce their measurement outcomes.
6. The recipient calculates the parity of the announced measurement outcomes, which determines whether they apply a local \hat{Z} -correction.
7. The recipient completes their side of the protocol as per the usual quantum state teleportation protocol, obtaining $|\phi\rangle$.
8. \square

Algorithm XXII: Open-destination quantum state teleportation of a single qubit from a single sender to $n - 1$ potential recipients.

allows the non-deterministic elements to be performed offline as a resource state preparation stage, overcoming the non-determinism during the gate application stage.

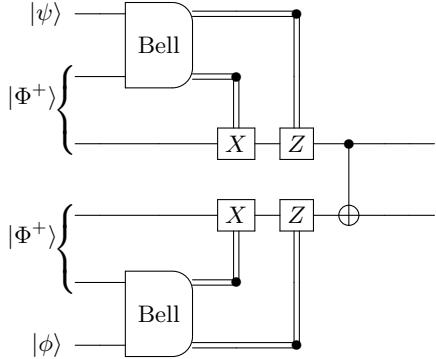
Specifically, when a CNOT gate acting directly upon two qubits fails, it corrupts those qubits, whereas if it fails during a state preparation stage, it can simply be reattempted until a success occurs, without corrupting the target qubits. A concatenated version of the gate teleportation protocol forms the basis for constructing near-deterministic entangling gates in linear optics, to be explained in detail in Sec. XLII.A.

Quantum gate teleportation effectively reduces the problem of implementing CNOT gates to:

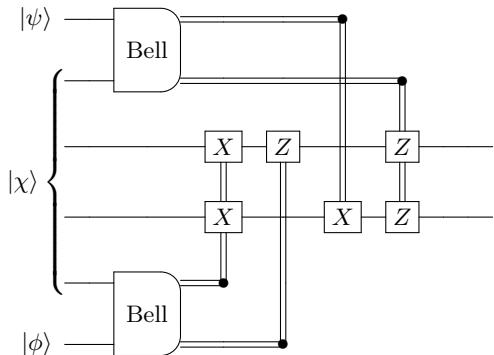
1. Offline preparation of highly-entangled 4-qubit resource states. This needn't be deterministic, since the resource state does not depend on the state to which the CNOT gate ought to be applied.
2. Two Bell measurements.
3. Some configuration of local Pauli operators, dependent upon the Bell measurement outcomes.

Importantly, like quantum state teleportation, there is no need for a quantum communications channel between

```
function GateTeleportation( $|\psi\rangle_A|\phi\rangle_B$ ):
    1. We wish to apply a CNOT gate to  $|\psi\rangle_A|\phi\rangle_B$ .
    2. Introduce two additional qubits,  $C$  and  $D$ .
    3. Teleport states  $|\psi\rangle_A \rightarrow |\psi\rangle_C$ ,  $|\psi\rangle_B \rightarrow |\psi\rangle_D$ .
    4. Apply  $\text{CNOT}|\psi\rangle_C|\phi\rangle_D$ .
```



5. The CNOT is a Clifford gate and can therefore be commuted to the front of the Pauli operators to yield a CNOT followed by some different configuration of Pauli operators.
6. The CNOT now acts jointly upon the Bell pairs that were acting as a resource for the state teleportation, independent of $|\psi\rangle_A|\phi\rangle_B$.
7. Group the CNOT gate and Bell pairs together, and treat them as a 4-qubit resource state preparation stage, which does not depend on $|\psi\rangle_A|\phi\rangle_B$.
8. Prepare the 4-qubit resource state, $|\chi\rangle = \text{CNOT}_{2,3}|\Psi^+\rangle_{1,2}|\Psi^+\rangle_{3,4}$, offline in advance.
9. If the CNOT is non-deterministic, employ REPEAT UNTIL SUCCESS to prepare $|\chi\rangle$.
10. The output state is $\text{CNOT}_{C,D}|\psi\rangle_C|\phi\rangle_D$.
11. \square



Algorithm XXIII: Teleporting a CNOT gate onto a 2-qubit state.

the two parties holding the qubits to which the gate is applied – classical communication is sufficient.

The gate teleportation idea is conceptually interesting as it converts the problem of ‘gate application’ to that of ‘state preparation’³⁶, by commuting all the entangling operations to the beginning of the protocol. Cluster state quantum computing (Sec. XL.B) is actually the extremity of this logic, whereby an entire quantum computation is transformed into a sequence of state and gate teleportations. One may interpret this to mean that teleportation is a universal resource for quantum computation (?).

The resource states required for gate teleportation are highly-entangled 4-qubit states, which are challenging to prepare, especially in the optical context. Thus, as with cluster states, if the preparation of these resource states were to be outsourced to a specialised provider, they could be in high demand.

Note that this technique works for the CNOT gate because it is a Clifford gate (i.e it commutes with the classically-controlled Pauli gates to yield a different combination of classically-controlled Pauli gates). Thus, this technique does not automatically apply to *any* 2-qubit gate.

E. Entanglement swapping

The obvious approach to sending a qubit from Alice to Bob is to send a qubit from Alice to Bob (duh!). However, over long distances this may accrue impractical error rates, particularly losses. The other alternative is to employ the quantum state teleportation protocol (Sec. XXVII.C) to teleport the state between the two parties. However, this requires that Alice and Bob first share an entangled Bell pair, which must itself be distributed across the same distances. Entanglement swapping (?) is the process of taking two Bell pairs, one held by each party, and swapping the entanglement between them such that the two parties share an entangled state. This procedure can be bootstrapped to progressively swap the entanglement over longer and longer distances, yielding *quantum repeater networks* (Sec. XXIX). The procedure for this protocol is shown in Alg. XXIV and Fig. 95

In a sense, entanglement swapping can be regarded as ‘indirect’ entanglement distribution, whereby entanglement is created between two distant parties who do not directly exchange any quantum information.

Alternately, note that the entanglement swapping is structurally almost identical to two instances of quantum

³⁶ The resource state is prepared from two Bell pairs and a single CNOT gate, which is locally equivalent to a 4-qubit GHZ state. In the absence of a direct source of Bell pairs, they can be prepared using separable single-qubit states and a CNOT gate. Thus, the full resource state may be prepared from separable single qubits via three CNOT gates.

```

function EntanglementSwapping( $|\Phi^+\rangle^{\otimes 2}$ ):
    1. Alice locally prepares the Bell pair,
         $|\Phi^+\rangle_{A_1, A_2}.$  (27.17)

    2. Bob locally prepares the Bell pair,
         $|\Phi^+\rangle_{B_1, B_2}.$  (27.18)

    3. The net initial state is,
         $|\psi\rangle_{\text{in}} = |\Phi^+\rangle_{A_1, A_2} |\Phi^+\rangle_{B_1, B_2}.$  (27.19)

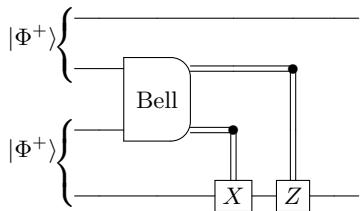
    4. Alice sends qubit  $A_1$  to third-party Eve.
    5. Bob sends qubit  $B_1$  to third-party Eve.
    6. Eve performs a Bell projection between  $A_1$  and  $B_1$ , yielding,
         $\langle \Phi^+ |_{A_1, B_1} |\psi\rangle_{\text{in}} = |\Phi^+\rangle_{A_2, B_2}.$  (27.20)

    7. In the case of the other Bell projection outcomes ( $\langle \Phi^- |_{A_1, B_1}$ ,  $\langle \Psi^+ |_{A_1, B_1}$  or  $\langle \Psi^- |_{A_1, B_1}$ ), local corrections (Pauli operators) are made by Alice and/or Bob, as dictated by classical communication from Eve,
         $\langle \Phi^+ |_{A_1, B_1} |\psi\rangle_{\text{in}} = |\Phi^+\rangle_{A_2, B_2},$ 
         $\langle \Phi^- |_{A_1, B_1} |\psi\rangle_{\text{in}} = \hat{Z}_{B_2} |\Phi^+\rangle_{A_2, B_2},$ 
         $\langle \Psi^+ |_{A_1, B_1} |\psi\rangle_{\text{in}} = \hat{X}_{B_2} |\Phi^+\rangle_{A_2, B_2},$ 
         $\langle \Psi^- |_{A_1, B_1} |\psi\rangle_{\text{in}} = \hat{X}_{B_2} \hat{Z}_{B_2} |\Phi^+\rangle_{A_2, B_2}.$  (27.21)

    8. Alice and Bob now possess a joint Bell pair between qubits  $A_2$  and  $B_2$ ,
         $|\psi\rangle_{\text{out}} = |\Phi^+\rangle_{A_2, B_2}.$  (27.22)

    9.  $\square$ 

```



Algorithm XXIV: Entanglement swapping protocol between two parties. Two Bell pairs held locally by two users, $|\Phi^+\rangle_{A_1, A_2} |\Phi^+\rangle_{B_1, B_2}$, are converted to a single Bell pair shared between the users, $|\Phi^+\rangle_{A_2, B_2}$.

state teleportation side-by-side. This is not a coincidence, and entanglement swapping can indeed be thought of as Bell pair state teleportation.

Now if instead of Alice and Bob we have a long chain of these operations in series, then the entanglement can be swapped across the entire length of the chain, enabling

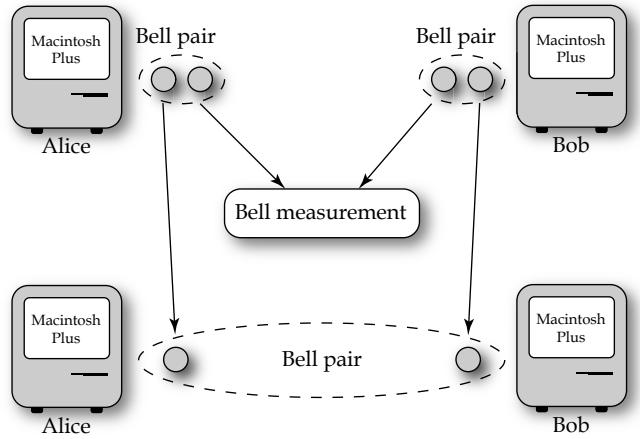


Figure 95: Entanglement swapping between two nodes. Each node initially holds a Bell pair (dotted ellipses) comprising two qubits (grey circles). One qubit from each pair is sent to the repeater between them, which measures them in the Bell basis. After local unitary corrections, the two nodes share an entangled pair.

the preparation of end-to-end entangled pairs, which can be employed for state teleportation.

The advantage to this approach is that the range of each repeater can be much smaller than the entire length of the channel, easing constraints imposed by errors, notably loss. Furthermore, the entanglement swapping needn't be actually performed in any chronologically linear sequence. The operations could be arbitrarily ordered, since the measurements are independent and commute. Thus, if some segments are detected as failing (e.g. qubits are lost), just those segments can be performed again without requiring the entire protocol to start from scratch, unlike the naïve direct communication technique. This DIVIDE AND CONQUER approach can drastically improve performance of the network in terms of channel capacity, improving the exponential dependence of loss on distance.

The protocol is conceptually very similar to teleportation, where instead of teleporting a qubit state, we are teleporting entanglement. Because of this similarity, it inherits similar error propagation characteristics as for teleportation discussed previously. That is, errors acting on the qubits upon which the Bell measurements are performed are effectively teleported onto the remaining qubits. Then, entanglement purification can be implemented as a higher-level layer on top of the repeaters, enabling high-fidelity entanglement distribution.

Each Bell measurement can be implemented non-deterministically using a PBS, mitigating the need for interferometric stability, as before, but therefore introducing non-determinism into the protocol.

F. Quantum cryptography

One of the most widely demonstrated class of quantum protocols is the cryptographic ones. Most importantly, these protocols allow, at least in principle, provably secure communication between two parties, immune to any attack. Because these protocols are so important and thoroughly researched, we dedicate Part. VII entirely to this topic.

G. Superdense coding

Superdense coding is a hybrid quantum/classical communications protocol for increasing classical bit-rates between two parties, who share entanglement as a resource.

Suppose Alice wishes to send classical information to Bob over a quantum channel. The HSW Theorem (??) tells us that Alice can send information to Bob at a maximum rate of one bit per qubit. However, if Alice and Bob share Bell pairs, superdense coding allows information to be transmitted at a maximum rate of two bits per qubit.

Let Alice and Bob begin with the shared Bell state,

$$|B_{00}\rangle = \frac{1}{\sqrt{2}} (|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B), \quad (27.23)$$

where the first qubit, A , belongs to Alice and the second qubit, B , belongs to Bob. This entangled pair is provided to them by a third-party entanglement server. The protocol exploits the fact that all four Bell states are locally-equivalent, and can be transformed into one another using operations performed only by Alice. Specifically, the four Bell states can be prepared from $|B_{00}\rangle$ via the local operations,

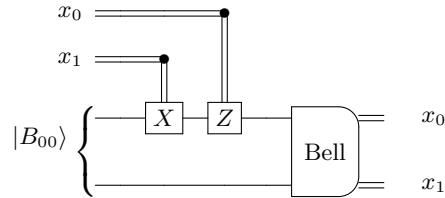
$$\begin{aligned} |B_{00}\rangle &= (\hat{\mathbb{1}} \otimes \hat{\mathbb{1}})|B_{00}\rangle \\ &= \frac{1}{\sqrt{2}} (|0\rangle|0\rangle + |1\rangle|1\rangle), \\ |B_{01}\rangle &= (\hat{Z} \otimes \hat{\mathbb{1}})|B_{00}\rangle \\ &= \frac{1}{\sqrt{2}} (|0\rangle|0\rangle - |1\rangle|1\rangle), \\ |B_{10}\rangle &= (\hat{X} \otimes \hat{\mathbb{1}})|B_{00}\rangle \\ &= \frac{1}{\sqrt{2}} (|0\rangle|1\rangle + |1\rangle|0\rangle), \\ |B_{11}\rangle &= (\hat{Z}\hat{X} \otimes \hat{\mathbb{1}})|B_{00}\rangle \\ &= \frac{1}{\sqrt{2}} (|0\rangle|1\rangle - |1\rangle|0\rangle). \end{aligned} \quad (27.24)$$

Suppose Alice wishes to send Bob the two-bit string $x \in \{0,1\}^2$. She applies local operations on her qubit to transform the shared Bell state into the Bell state $|B_x\rangle$. There are four such states, therefore this encodes two classical bits of information. She then sends her qubit to Bob, who already holds the other half of the entangled pair.

Now by measuring in the Bell basis Bob can determine which two-bit string Alice encoded. The algorithm is described in Alg. XXV.

function SuperdenseCoding($|B_{00}\rangle$, x):

1. Alice and Bob share the Bell pair $|B_{00}\rangle$.
2. Alice encodes the two-bit string $x \in \{0,1\}^2$ into a choice of the four possible Bell pairs.
3. Alice prepares the respective Bell pair using operations local to only her half of the shared state ($\hat{\mathbb{1}}$, \hat{X} , \hat{Z} or $\hat{Z}\hat{X}$).
4. Alice sends her qubit to Bob.
5. Bob measures in the Bell basis, with four possible measurement outcomes.
6. The measurement outcome corresponds to the bit-string x .
7. \square



Algorithm XXV: Superdense coding protocol for communicating two classical bits via transmission of a single qubit. The protocol requires the two parties share a Bell pair as a resource, provided by a third-party.

Note that the protocol in a sense ‘cheats’, since it assumes a resource of Bell pairs between Alice and Bob, which doesn’t come for free. However, in an environment where both parties have access to the same entanglement server or repeater network (Sec. XXIX), in addition to their own direct line of quantum communication, they can utilise this protocol to double classical communication rates from one bit per qubit to two.

However, this doubling in communication rate requires using quantum infrastructure, which, at least for the foreseeable future, will come at a greater cost than our present-day commodified classical hardware. It may therefore be the case that the technological effort of implementing this protocol outweighs the gain, or that for the same effort other classical bandwidth-increasing technologies could be employed.

Alternately, in a future quantum world where such technologies are cheap off-the-shelf commodities, as with our current classical ones, why not double our classical network bandwidths if we can?

H. Quantum metrology

The goal of quantum metrology is to estimate an unknown phase with the greatest degree of precision. This finds many applications, perhaps most notably the recent gravity wave measurement protocols (?). The shot-noise limit (SNL) represents the maximum achievable precision using classical states, whereas the Heisenberg limit (HL) is the best that can be achieved using quantum resources. The goal of quantum metrology is to beat the SNL, ideally saturating the HL.

Achieving the SNL is easily done using a Mach-Zehnder interferometer (Sec. XXII.B) fed with coherent states (Sec. XXIII.A), which are not true quantum states. Referring to Fig. 71, if a coherent state is inputted into one arm of the interferometer, with no phase-shift ($\tau = 0$) all the coherent amplitude would exit the corresponding output port. If on the other hand there were a π phase-shift, all the amplitude would exit the other output port. For intermediate τ there will be varying degrees of coherent amplitude distributed between the two outputs. Thus, the relative amplitude exiting the two output ports acts as a signature for the internal phase-shift τ .

Improving upon this, HL metrology can be achieved using NOON states (Sec. XXIII.C) (?). An alternate recent proposal (known as the MORDOR protocol, after the authors), employs only single-photon states (Sec. XXIII.B) and passive linear optics, which, although not saturating the HL, significantly beats the SNL (??). This was recently experimentally demonstrated by (?). Squeezed states have also been shown to beat the SNL.

NOON states in particular are difficult to prepare, as they cannot be deterministically prepared using linear optics, and no current source natively prepares them directly. Thus, outsourcing these state preparation stages could be of great value to end-users of metrology, were there a specialised server dedicated to this task.

I. Quantum state & process tomography

In Sec. XII.D we introduced QST and QPT, as procedures by which to experimentally reconstruct unknown density matrices or process matrices respectively. It is conceivable that these tomographic procedures might want to be performed over a quantum network in a distributed fashion.

Consider the case where a node joins an existing ad hoc network. Before thinking about routing its packets through the network, it must understand the network's relevant cost metrics. Suppose that metric is one that is calculated directly from a channel's process matrix. Then, to characterise the channels in the network connecting the node to its new nearest neighbours, it could apply distributed QPT, whereby the new node is responsible for preparing the complete basis of input states

required for QPT, which are transmitted to the chosen neighbour across the respective channel, after which the recipient performs all the necessary measurements in the required bases. Purely classical communication is obviously required to communicate measurement settings and outcomes.

In this simple example scenario it is immediately clear that QPT of new links in a network is perfectly suited to distributed implementation. In fact, having a node attempt to characterise a channel from start to finish could be entirely unrealistic if the channel ran over long distances – the owner of the node would never be able to reach the other end of the channel in time for the photons' arrival! This necessitates a cooperative protocol.

J. Quantum clock synchronisation

— by Tim Byrnes.

Clock synchronisation is a fundamental task with widespread applications, ranging from navigation, telecommunications, and financial transactions, to the internet as a whole, and many scientific applications. Of these, the global positioning system (GPS) has become a day-to-day necessity for much of humanity, having been increasingly incorporated into smartphones and other commodity devices.

The GPS system famously relies upon very precise clock synchronisation to perform its task through a process of quadrangulation from a constellation of several satellites. Due to the high speed of light, we require highly synchronised clocks, accurate to the nanosecond level. This allows positioning to be performed to the level of meters, a level of precision required for many routine applications. GPS satellites have atomic clocks that are stable to one part in 10^{13} , so that active correction can maintain this level of accuracy.

The great success of the GPS system has created a further demand for increasingly precise navigation. For example, autonomous vehicles would immediately benefit from a more precise navigation system.

In principle, technology for more stable clocks already exists, with atomic clocks exceeding stabilities of those on satellites being routinely produced, and optical atomic clocks now reaching stabilities of one part in 10^{18} (?). An outstanding question is then how to synchronise these clocks given their remarkable stabilities.

Previous works have examined the problem of clock synchronisation in space. In the proposal of Jozsa *et al.*, many copies of shared entanglement in a singlet state is first distributed and stored on the clock states of an atomic clock (?). The measurement is then performed by one party, which collapses the states simultaneously across all parties, and the time evolution of the states begins.

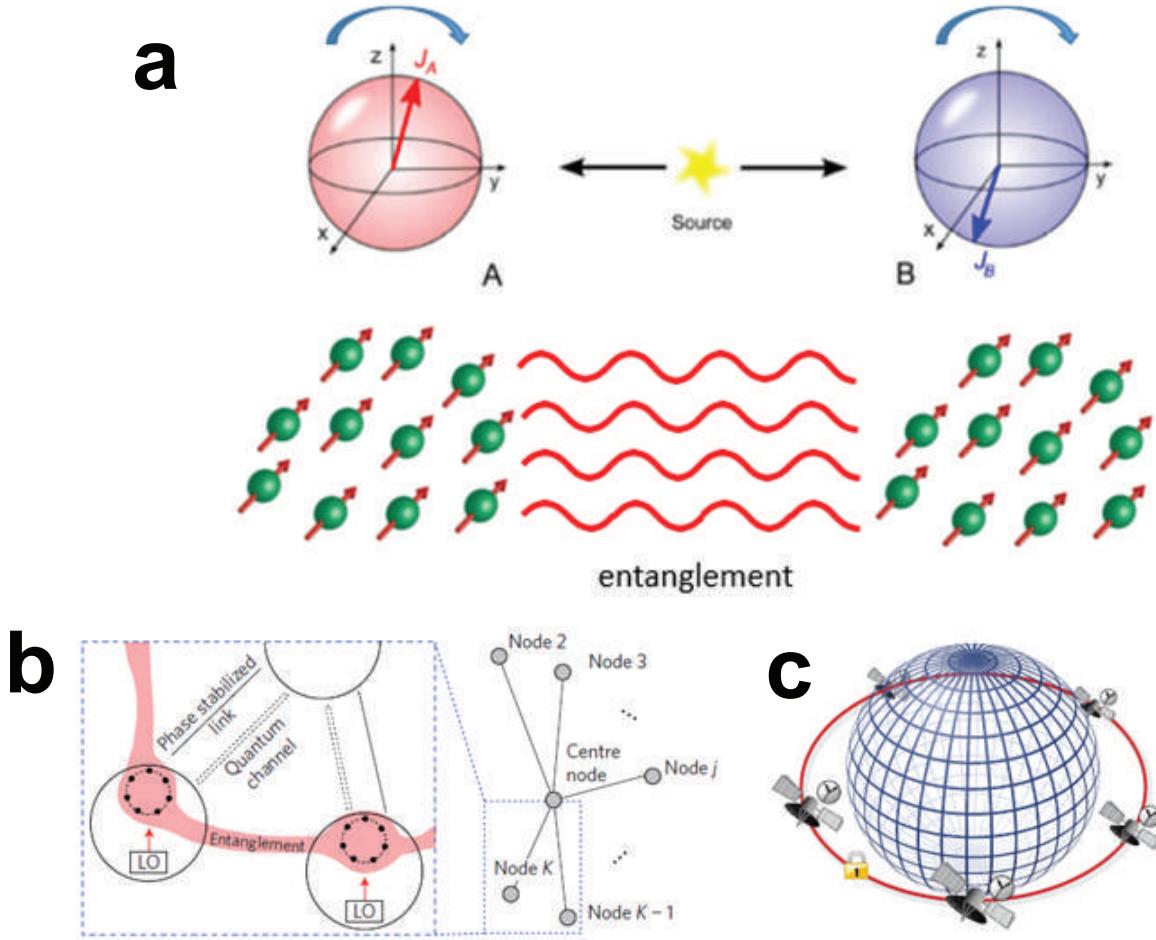


Figure 96: Better figure, this is shit Quantum clock synchronisation schemes. (top) The proposal by Jozsa & Dowling *et al.* where a singlet state is shared and measured in an ensemble of qubits (?). (bottom) The proposal by Lukin & Ye *et al.* where a GHZ state is distributed between satellites to measure the frequency drift at the Heisenberg limit (?).

Classical information is exchanged between them, which reveals the time elapsed since the measurement, which can be used to synchronise the clocks. While the original protocol only allowed clock synchronisation between two parties, similar ideas were used to extend this to the multiparty context (??).

In a more recent proposal, a shared GHZ state is prepared across all nodes in the quantum network, which allows for quantum metrologically enhanced detection of the clock signal drift at the Heisenberg limit (?). The use of shared resources acts to improve the overall precision, allowing for an optimal scheme for the qubit resources that are employed. Several other proposals have also been made, which are quantum versions of Eddington's slow clock transport protocol where the qubit keeps time of the transmission (??).

Experimentally, there have been several demonstrations of the protocol, albeit at relatively short distances. Continuous time-bin entangled photons were used as the entanglement resource to obtain a time-correlation be-

tween a distance of 3km (?), and another technique based on Hong-Ou-Mandel interferometry was performed across a 4km fibre link (?). Several other demonstrations based on nuclear magnetic resonance (NMR) (??) have also been performed.

There are however several outstanding problems with the quantum clock synchronisation scheme as presented above. In the scheme of Jozsa *et al.*, if one starts in a perfect singlet state, the scheme works as intended, but if one instead starts with the state,

$$\frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B - e^{i\delta}|1\rangle_A|0\rangle_B), \quad (27.25)$$

(a Bell pair augmented by a local phase) one obtains an offset to the synchronisations between the two parties. In practice, such a phase could arise from decoherence induced noise, or differences in the basis conventions chosen by the two parties. Thus, in practice entanglement purification would be required to produce a singlet state with $\delta = 0$ prior to executing the protocol. However, it

was argued that to perform the entanglement purification quantum circuit correctly, the timing of the quantum gates would need to be controlled, which requires synchronised clocks (?) – this renders synchronisation impossible. It has previously been shown that such a phase cannot be eliminated using asynchronous entanglement purification (?), and hence the protocol remains incomplete in the general case where imperfect singlet pairs are shared.

A quantum protocol for clock synchronisation based on shared entanglement is given in Alg. XXVI.

K. Quantum-enabled telescropy

— by Zixin Huang.

For the direct imaging of an object, diffraction limits the resolution of the image. When we consider two neighbouring points on the object separated by a small angle, the minimum angular separation resolvable is,

$$\theta_{\min} = 1.22 \frac{\lambda}{D}, \quad (27.32)$$

known as the Rayleigh criterion, D being the diameter of the aperture.

Current optical interferometers have limited baseline lengths, and thus limited resolution. In principle one can build a telescope array with a synthetic aperture of arbitrary D , however, phase-locking the entire system is extremely difficult over long distances. A quantum information protocol has been developed to side-step this problem.

The light arriving from the distant object is a thermal state, but the average photon number per mode is much less than 1, therefore higher order terms are negligible. The state that reaches the telescope is therefore approximated by,

$$|\psi_{\text{image}}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + e^{i\phi}|1\rangle_A|0\rangle_B), \quad (27.33)$$

where ϕ is the relative phase-shift between the two telescopes, which depends on the difference in distance of propagation. If ϕ can be measured accurately, this can give a precise estimate on the location of the object,

$$\phi = \frac{b \sin(\theta)}{\lambda}, \quad (27.34)$$

where λ is wavelength.

Often the light that arrives will be formed by a mixture of photons from different sources that emit incoherently, and different locations give rise to different phase-shifts ϕ , resulting in a density matrix of the form,

$$\hat{\rho} = \frac{1}{2} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & \mathcal{V}^* & 0 \\ 0 & \mathcal{V} & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (27.35)$$

function ClockSync($|\Psi^-\rangle$):

1. Distribute a Bell state between Alice and Bob,

$$|\psi_0\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B) \\ = \frac{1}{\sqrt{2}}(|+\rangle_A|-\rangle_B - |-\rangle_A|+\rangle_B). \quad (27.26)$$

2. The joint system is freely-evolving under the Hamiltonian,

$$\hat{H} = \hbar\omega(\hat{Z}_A + \hat{Z}_B). \quad (27.27)$$

3. Over time t_{free} this evolves to,

$$|\psi_1\rangle = e^{-i\frac{\hat{H}}{\hbar}t_{\text{free}}} \frac{1}{\sqrt{2}}(|+\rangle_A|-\rangle_B - |-\rangle_A|+\rangle_B) \\ = e^{-i\omega t_{\text{free}}} \frac{1}{\sqrt{2}}(|+\rangle_A|-\rangle_B - |-\rangle_A|+\rangle_B), \quad (27.28)$$

yielding only an irrelevant global phase.

4. Alice measures her qubit in the \hat{X} basis ($|\pm\rangle\langle\pm|$), and classically announces the time at which she performed the measurement, t_A , and her measurement outcome, $m_A = \pm$.

5. The system evolves for time t_{ev} ,

$$|\psi_+\rangle = e^{-i\hat{Z}t_{\text{ev}}} |-\rangle_B \\ \propto |0\rangle_B - e^{-i\omega t_{\text{ev}}} |1\rangle_B, \\ |\psi_-\rangle = e^{-i\hat{Z}t_{\text{ev}}} |+\rangle_B \\ \propto |0\rangle_B + e^{-i\omega t_{\text{ev}}} |1\rangle_B. \quad (27.29)$$

6. Bob measures his qubit in the \hat{X} basis, with measurement probabilities,

$$m_+ \propto \sin^2(\omega t_{\text{ev}}), \\ m_- \propto \cos^2(\omega t_{\text{ev}}), \quad (27.30)$$

and infers t_{ev} .

7. Bob sets his clock to,

$$t_B = t_A + t_{\text{ev}}. \quad (27.31)$$

8. \square

Algorithm XXVI: Algorithm for performing quantum clock synchronisation between two parties using shared Bell pairs and classical communication. [Check this for errors!](#)

where \mathcal{V} is the visibility, reflecting decoherence.

If we interfere the two modes at a 50:50 beamsplitter, the photon will exit port 1 with probability,

$$p_{\text{coin}} = \frac{1}{2}(1 + \text{Re}[\mathcal{V}e^{-i\delta}]), \quad (27.36)$$

from which \mathcal{V} can be determined by taking measurements while sweeping through δ .

The problem with implementing the measurement is the difficulty of transporting the single photon state over long distances without incurring loss or additional phase-shifts.

Instead of sending a valuable quantum state directly over a noisy quantum channel, one can distribute a Bell pair between the two telescopes, then we teleport the original quantum state from one telescope to the other. The entangled state is known, the preparation can be repeated, and one can use an entanglement distillation protocol to eliminate the phase noise.

Now, we can use the entangled pair directly to measure the visibility, as in Fig. 97. We post-select on the measurement results, considering the events where a single photon is observed at A and B simultaneously.

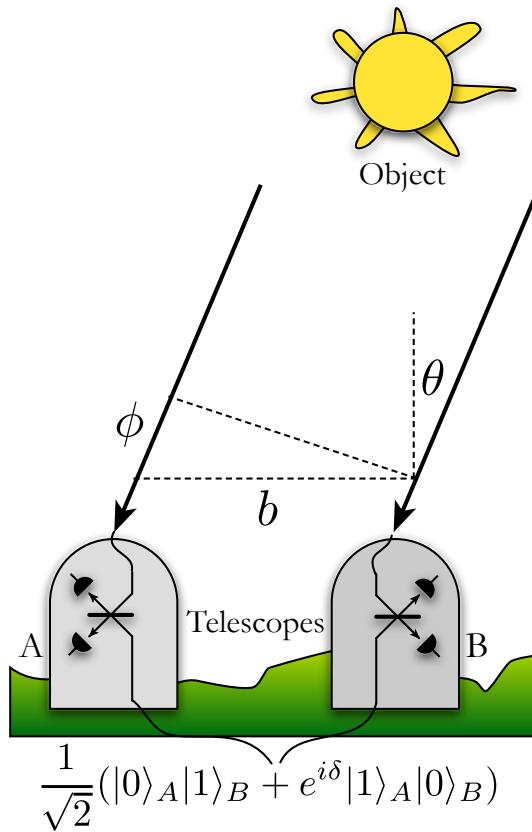


Figure 97: Architecture for quantum-enabled telescopes using two widely separated telescopes, which have shared Bell pairs. The basic idea is that the Bell pair mediates teleportation of one telescope's photon to the other, at which point an interferometric technique measures their phase-difference, thereby determining ϕ . It is assumed the baseline separation is large, $b \gg 1$.

The variable delay line is now applied to the entangled state when the photon is sent to A , producing the entangled state,

$$|\psi_{\text{shared}}\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B + e^{i\delta}|1\rangle_A|0\rangle_B), \quad (27.37)$$

where δ is determined by a controllable delay, allowing completion of the protocol to determine ϕ .

Thinking futuristicly, in a future large-scale quantum internet, whereby Bell pairs are a readily available resource across the globe, quantum-enabled telescopes needn't be limited to pairs of telescopes, but could expand to become large-scale telescope arrays comprising numerous telescopes, all sharing pairwise entanglement, distributed via the quantum internet (see Fig. 98). Using a 2D grid would enable θ to be measured along different axes, and the increased number of detectors would increase signal strength.

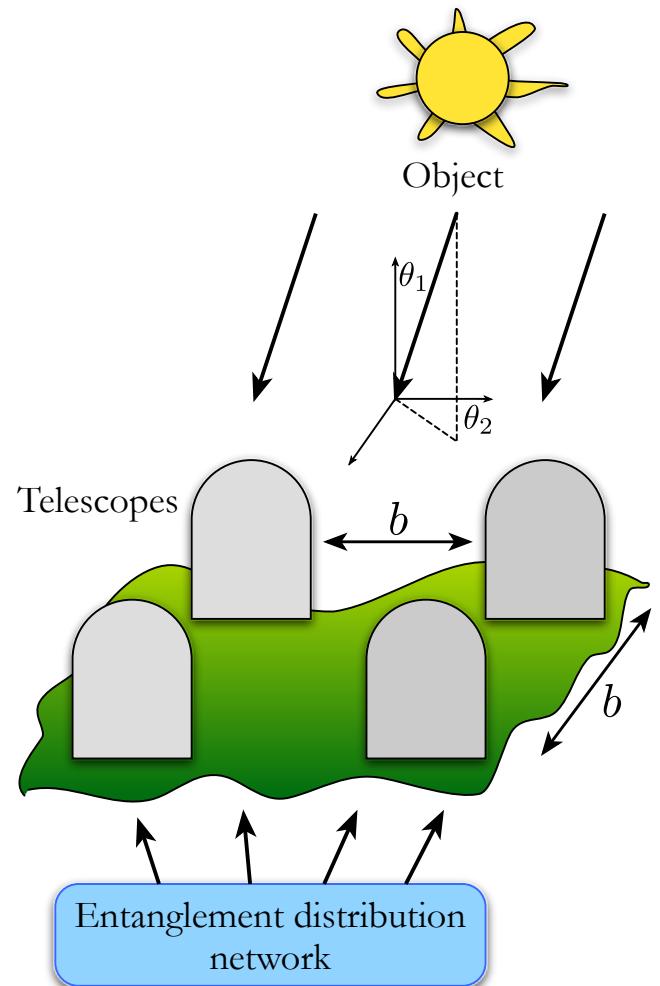
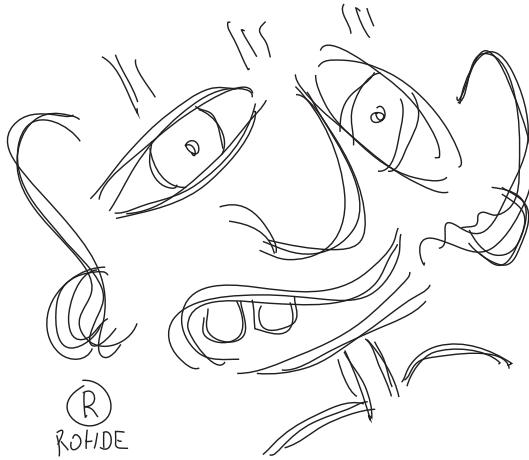


Figure 98: An array of quantum-enabled telescopes, paired with one another via Bell pairs, distributed over a quantum network, where $b \gg 1$. With more telescopes in the array, signal strength can be increased, and if the array has a 2D grid topology, rather than a linear one, the angles of incident light fields can be measured along multiple axes.

— *Etiam capillus unus habet umbram.*



Part VI

Entanglement distribution

XXVIII. ENTANGLEMENT – THE ULTIMATE QUANTUM RESOURCE

As we have seen, the diversity of quantum states that may be communicated, and protocols implemented over the quantum internet is extremely diverse, encompassing many different types of encodings and communications protocols.

Given this plethora of protocols and encodings, discussed in detail in Part. V, one might ask whether there is a single primitive resource that might be applicable to all, or at least most of these quantum protocols, thereby reducing the technological requirements of the nodes and quantum channels forming the network mediating them – if networks were able to specialise in a very limited number of tasks, we might reasonably expect them to be better optimised and exhibit better performance than a ‘Jack of all trades, master of none’ network!

It turns out that there is one particularly useful quantum resource, that finds applicability in many of these protocols – *distributed entanglement*, which comes in many flavours and varieties, some of which we discuss now.

A. Bell states

Foremost, Bell pairs (Sec. XXIII.G) – the simplest entangled states – are an utterly indispensable resource for countless quantum protocols. In brief, Bell pairs find applicability in, amongst many others, the following key protocols:

- Cluster states (Sec. XL.B): a Bell pair is also a

2-qubit cluster state, a supply of which can be employed in fusion strategies to prepare larger cluster states, enabling universal, distributed MBQC.

- Quantum state teleportation (Sec. XXVII.C): a shared Bell pair between Alice and Bob forms the elementary quantum resource upon which the state teleportation protocol is constructed.
- QKD (Sec. XXXVII.A): the E91 QKD protocol is built upon a reliable stream of distributed Bell pairs, enabling private communication with perfect information theoretic security.
- Modularised quantum computation (Sec. XLIII.D): using Bell pairs, entanglement swapping (Sec. XXVII.E) can be employed to fuse neighbouring, but potentially distant modules together using operations local to each module.
- Superdense coding (Sec. XXVII.G): a shared Bell pair enables the communication of two classical bits of information via transmission of a single qubit, thereby doubling classical channel capacity.
- Quantum-enabled telescropy (Sec. XXVII.K): a shared Bell pair between two telescopes allows a photon received at one telescope to be teleported to the other, at which point interferometric techniques yield extremely sensitive phase information.

We see that Bell pairs form a ubiquitous resource, covering many of the most significant quantum protocols in quantum computation, distributed quantum computation, quantum state teleportation, and quantum cryptography.

B. GHZ states

Beyond Bell pairs, multi-qubit GHZ states (Sec. XXIII.E) (the direct generalisation of Bell pairs to n qubits) are useful in a variety of settings.

For the purposes of quantum anonymous broadcasting (Sec. XXXVII.D), multi-party GHZ entanglement is the primitive resource upon which the cryptographic protocol is constructed. As with Bell pairs, GHZ states are a known state and infinitely reproducible. They can also be purified. Thus, GHZ entanglement distribution is another useful primitive, which future quantum hubs might specialise in preparing and distributing.

Additionally, quantum gate teleportation (Sec. XXVII.D) of a maximally-entangling 2-qubit gate (e.g a CNOT or CZ gate) is mediated via a shared 4-qubit GHZ state. In a distributed environment the sharing of such a state between two parties (2 qubits per party) enables implementation of a truly distributed 2-qubit entangling gate.

C. Cluster states

Finally, cluster states (Sec. [XL.B](#)) are a primitive resource for measurement-based quantum computation. Owing to their handy ability to fuse together to one another, forming larger clusters, the preparation and distribution of relatively small cluster states lends itself well to distributed implementation by specialised providers. Providers could distribute small cluster states, which are subsequently fused together using simple 2-qubit entangling operations to form desired topologies, held either locally or distributed in the cloud.

D. Why specialise in entanglement distribution?

These observations warrant special treatment of entanglement distribution as a fundamental building block in the quantum era. One might envisage a quantum internet in which a central server(s), who specialises in only entangled state preparation and distribution, serves the sole role of pumping out Bell pairs or other entangled states across the quantum internet to whomever requests them, who subsequently use them for protocols such as those mentioned above. This could be in the form of a server transmitting over fibre networks, across free-space, or via a satellite in orbit, transmitting at an intercontinental level.

What's the advantage of this approach to quantum networking? Why specialise in entanglement distribution, rather than implementing more capable networks with the ability to perform arbitrary operations? There are numerous:

1. Dedicated servers can specialise in this one particular task, as can be the transmission infrastructure.
2. The entanglement servers are entirely passive, not involved interactively with clients.
3. The server needn't concern itself with the nitty-gritty of the protocols implemented by the end-user. It acts purely as a provider of a single resource, remaining uninvolved in their subsequent applications.
4. Because servers are providing a single standardised product, they can be commodified, enabling mass production of the hardware devices and the associated economy of scale. For example, mass production of simple ground-based Bell state relays, or the construction of a comprehensive globe-enveloping constellation of satellites, would inevitably improve economies of scale.
5. Unlike generic quantum states, Bell pairs, GHZ states and cluster states are known states that are infinitely reproducible, without having to worry about no-cloning limitations.

6. Photonic Bell pairs are easily prepared via type-II SPDC at very high repetition rates ($\sim 100\text{MHz}-1\text{GHz}$), enabling rapid state preparation.
7. Small entangled states like Bell pairs are relatively 'cheap' to prepare, and can be readily manufactured using widely accessible, present-day technology that has already been well-demonstrated on Earth and in space.
8. QoS is a lesser issue in most scenarios. We can employ a SEND-AND-FORGET protocol for the distribution of entanglement (much like classical UDP) – since every state is identical, we needn't be concerned about missing ones. Instead, we can simply wait for the next one (a REPEAT-UNTIL-SUCCESS strategy), knowing it will be exactly the same. We also call this the SHOTGUN approach – keep firing away until we hit something, and if we lose a few, who cares?
9. Rather than transmitting quantum states between distant parties directly, if we instead use state teleportation (Sec. [XXVII.C](#)) mediated by Bell states, the state to be transmitted will not be corrupted if the communications channel fails (e.g via loss). Instead we can wait for the next successfully transmitted Bell pair until we are ready to teleport the state, which then proceeds without directly utilising the quantum communications channel, accumulating its associated costs, or risking losing the state altogether should link failure occur. Only classical communication is required to complete the protocol, which can be regarded as error-free for all intents and purposes.
10. Entanglement purification may be employed by parties to improve the cost metrics associated with their shared entanglement, thereby partially overcoming the limitations imposed by the quantum communication channels.
11. If no direct link exists between server and clients, bootstrapped entanglement swapping can be employed to concatenate servers to create longer-distance 'virtual' links. This is the basis for *quantum repeater networks*, to be discussed next in Sec. [XXIX](#).

E. Why not distributed entangling measurements?

In addition to entanglement distribution, entangling measurements, e.g Bell state projections (Sec. [XXIV.D](#)), may be used as a primitive for many protocols. This is effectively entangled state distribution in reverse, whereby two clients transmit states to a host, who performs a joint

entangling measurement upon them. For example, in the modularised model for cluster state quantum computing, two adjacent but distant modules might transmit optical qubits to a satellite, which projects them into the Bell basis, thereby creating a link between the respective modules via entanglement swapping. This isn't as powerful as entangled state distribution, since it cannot be used for, for example, E91 QKD, but nonetheless remains a powerful primitive for many protocols.

So which ought our quantum hubs specialise in, entanglement distribution, entangling measurements, or both? For most practical purposes the former is far more powerful and robust. Let us take the example of fusing two remote cluster states together to form a larger, distributed virtual cluster. Imagine that their fusion operations are optically mediated by a satellite overhead. The options for satellite-mediated state fusion are (see Fig. 99):

- Downlink mode: the satellite uses the downlink to distribute an entangled Bell pair between the two nodes. Each node performs a Bell projection between their half of the Bell pair and their respective qubit from their local cluster state, thereby swapping the entanglement and creating a link.
- Uplink mode: each node takes an optical qubit from their cluster state (or entangles their cluster state qubit with an optical qubit), which is uplinked to the satellite. The satellite performs a Bell projection between the two received optical qubits, thereby implementing entanglement swapping between the two nodes, creating a link.

Mathematically, these two processes are almost identical in their operation, differing only in direction. However, the former has the key advantage that it requires no time-synchronisation operations on the server side, whereas the latter does, and satellite-based hardware is orders of magnitude more expensive than Earth-based hardware.

Both scenarios involve Bell projections. These entangling measurements require active synchronisation to ensure that the measured qubits arrive at the entangling measurement device (typically a PBS) simultaneously, so as to achieve high HOM-visibility, requiring synchronisation on the order of the photons' coherence length. This can be achieved either using a brute-force REPEAT-UNTIL-SUCCESS mode of operation (post-selecting on events where both qubits arrive within a required temporal window), or storing one qubit in quantum memory until the other arrives. However, post-selection is expensive, requiring a massive overhead in the number of trials, and quantum memory is technologically challenging to implement, more so in space.

In the former case, the time ordering of the Bell projections performed locally on the ground nodes is irrelevant. Although within each ground station the two qubits being projected must be synchronised, requiring quantum memories within ground stations.

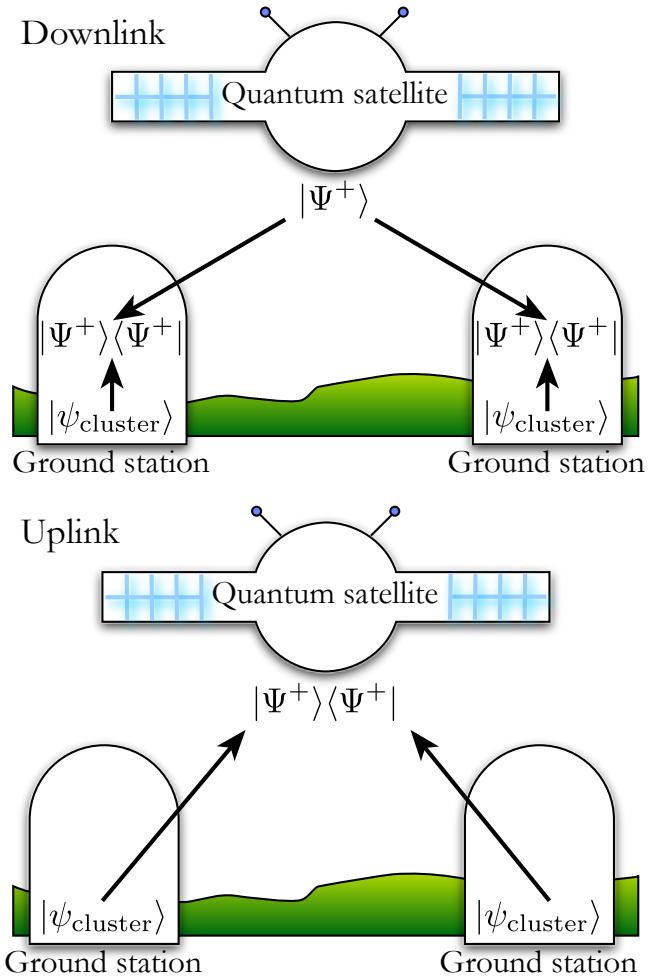


Figure 99: Satellite-mediated cluster state fusion operations for creating a link between two cluster states held by distant ground nodes. (top) Via entanglement distribution over a downlink channel. (bottom) Via distributed Bell projection over an uplink channel. When performing a distributed Bell measurement it is essential that the optical qubits arrive synchronously at the entangling measurement device, denoted $|\Psi^+\rangle\langle\Psi^+|$ (e.g a PBS), which is technologically challenging to implement on satellite given the unpredictable nature of the atmospheric quantum channel, necessitating on-board quantum memories to synchronise the qubits. This is likely to make downlinks cheaper, faster and more efficient than uplinks.

On the other hand, in the latter case it is essential that both optical qubits arrive at the satellite's entangling measurement device simultaneously, which is extremely difficult to enforce when our quantum channels are tracking moving targets in low-Earth orbit and traversing a turbulent atmospheric channel in between. An on-satellite quantum memory would be extremely costly!

This yields several key advantages in favour of entanglement distribution as opposed to server-side joint entangling measurements:

1. The challenging prospect of quantum memory may operate on Earth, far less onerous and expensive

- than incorporating this technology into a satellite in low-Earth orbit.
2. Because the server is not storing any qubits in quantum memory, it does not suffer downtime associated with the periods between receiving the first photon and waiting for the second – it can continue to spit out Bell pairs at maximum capacity.
 3. The satellite remains passive, implementing only the simplest of possible operations, reducing mass-production costs.
 4. The satellite does not require any interaction with its clients (classical or quantum).

5. Because Bell pairs are known, infinitely reproducible states, the server can operate in a UDP-like mode and it is not problematic if any given pair was lost. In the reverse direction, loss of a qubit could compromise the entire peripheral state associated with it in the ground station.
6. Entanglement purification can be employed to enhance the effective quality of the transmission channel.

We therefore anticipate that distributed entangling operations are likely to be mediated via entanglement distribution rather than distributed entangling measurements in the future quantum internet.

These observations lead us to naturally conclude that a quantum network specialised to this one particular task – entanglement distribution – would already be immensely useful, and on its own enable many key applications.

— *Credo in unum deum.*

XXIX. QUANTUM REPEATER NETWORKS

— by William Munro.

IN the previous section (Sec. XXVIII) we concluded that quantum networks specialising purely in entanglement distribution (Bell pairs in the simplest case), would already be extremely capable in enabling many distributed quantum protocols. This motivates the development of protocols for entanglement distribution over noisy, long-distance quantum networks.

Any useful future quantum internet is going to require the communication of quantum information over arbitrarily long distances. While intercity communication might be implemented via point-to-point connections, intracity and intercontinental communication will require extremely long distance links, well beyond the attenuation length of the optical fibres connecting them or the line-of-sight of satellites in orbit.

Quantum repeaters (???) are devices that allow high-quality entanglement to be shared between distant nodes,

when no direct line of communication is available from a server to its two clients. This is achieved by dividing long-distance links into a finite number of segments interspersed with repeaters (see Fig. 100).

For example, a satellite in low Earth orbit (Sec. LXXVII) may be outside simultaneous line-of-sight to two distinct ground stations, owing simply to the curvature of the Earth. But this can be overcome by relaying a channel through several satellites in line-of-sight of one another. This is achieved using a bootstrapped entanglement swapping and purification protocol (Secs. XXVII.E & XXVII.B). Most commonly, this entanglement is in the form of Bell pairs, which, as discussed previously in Sec. XXVIII, form a ubiquitous resource for many essential quantum protocols. The actual physical encoding of the entangled states may vary, but is most commonly and archetypically in the form of polarisation-encoded single-photons or CV states.

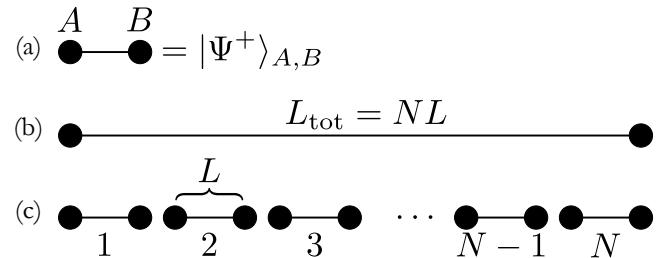


Figure 100: (a) Schematic representation of an entangled Bell pair $|\Psi\rangle_{A,B}$ shared between remote parties Alice and Bob. The two solid dots represent physical qubits, while the edge represents entanglement. (b) The link may be over a long distance L_{tot} . (c) Due to channel losses the link may be broken into N smaller segments of length L . The links for each of the smaller segments can be independently generated and combined to form the longer distance link.

The links are now over much shorter distances and so can be generated with far higher probability. Then by stitching these together using entanglement swapping (Sec. XXVII.E), we can generate our required long-range entanglement link.

Beginning from this simple principle, the field of quantum repeater networks has grown enormously, leading to several generations of repeater designs, of ever increasing power and sophistication, and ever more challenging technological demands.

A. First-generation repeaters

The above description is very hand-wavy, and of course things are a little more complicated in practise. We now examine these ideas in a little more detail, starting with a simple linear chain of repeater stations.

In a quantum repeater network, there are three main operations required:

1. Entanglement distribution (Sec. XXIX.A.1): to create entangled links between adjacent repeater nodes.
2. Entanglement purification (Sec. XXIX.A.2): to improve the quality of entanglement between nodes.
3. Entanglement swapping (Sec. XXIX.A.3): to join adjacent entangled links together to form longer distance links.

The basic operation of a repeater, as shown in Fig. 101, works as follows:

We begin our preparation of a long-range entangled link by creating multiple entangled pairs between adjacent repeater nodes (the number will depend both on the quality of the pairs we initially generate and also the target quality we want our final pair to have). Once we have enough pairs established between two repeater nodes, we perform entanglement purification, which converts multiple entangled links (pairs) into a fewer number with higher quality.

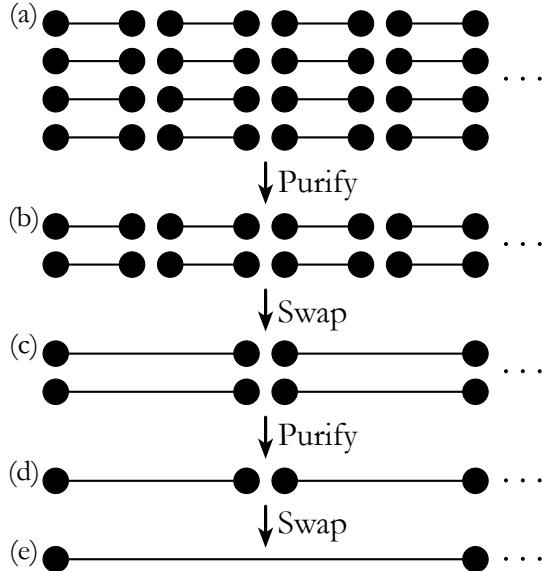


Figure 101: Basic operation of a (first-generation) quantum repeater network. (a) Preparation of multiple entangled links between adjacent repeater nodes. (b) They are then purified to create higher fidelity links. (c) Entanglement swapping between adjacent pairs creates links of twice the original length. (d) These new links are purified to create higher fidelity ones. (e) Entanglement swapping creates a link four times the original size. This process continues as necessary to reach target distance and purity.

These purification steps, shown in Fig. 101(a-b), are performed on the links between all adjacent repeaters, increasing the quality of the links between those adjacent repeater stations to the required degree. Entanglement swapping, as shown in Fig. 101(c), then creates links twice as long. The resulting entanglement links can then be used iteratively for further rounds of purification and

swapping until one generates a high quality link between the desired points in the network.

1. Entanglement distribution

Probably the most important operation for any quantum repeater setup is entanglement distribution, the process of creating entanglement between two remote parties (Alice and Bob) connected by a quantum channel (generally an optical fibre or free-space link). This can be implemented in a number of ways (????????), but can be broadly categorised into three basic schemas:

- Photon emission from quantum memories in the repeater nodes, followed by which-path erasure.
- Absorption of entangled photons by quantum memories.
- Photon emission at one node and absorption at another.

By far, the emission based schemes are the most common, which we will concentrate on here. Such schemes operate by using an entangling operation – *which-path erasure* – to entangle two quantum memories via photons to which they were coupled. Effectively the process teleports the action of an entangling gate (Sec. XXVII.D) acting on the photons onto the quantum memories to which they were entangled.

We now describe such a which-path entangling operation in the context of 2-level quantum memories coupled to polarisation-encoded photons. A closely related scheme for preparing cluster states on λ -configuration systems for the purposes of quantum computation is discussed in Sec. XLII.F.

Ideally one wants to initially generate a maximally entangled state of the form (?),

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|g\rangle\langle H| + |e\rangle\langle V|), \quad (29.1)$$

within the repeater node, where $|g\rangle$ and $|e\rangle$ are the two states (ground and excited) of the quantum memory, and $|H\rangle$ and $|V\rangle$ are the polarisation states of a single photon. The photons from the two repeater nodes (Fig. 102) are then transmitted to a beamsplitter (or PBS in this example), after which the state of the system is,

$$\begin{aligned} |\Psi\rangle = & \frac{1}{2}|g\rangle|g\rangle|H\rangle|H\rangle + \frac{1}{2}|e\rangle|e\rangle|V\rangle|V\rangle \\ & + \frac{1}{2}|g\rangle|e\rangle|HV\rangle|0\rangle + \frac{1}{2}|e\rangle|g\rangle|0\rangle|HV\rangle. \end{aligned} \quad (29.2)$$

One immediately notices that the $|g\rangle|e\rangle$ and $|e\rangle|g\rangle$ contributions are associated with two photons in one of the PBS exit modes, the other being in the vacuum state. However, the $|g\rangle|g\rangle$ and $|e\rangle|e\rangle$ terms have one photon in each of

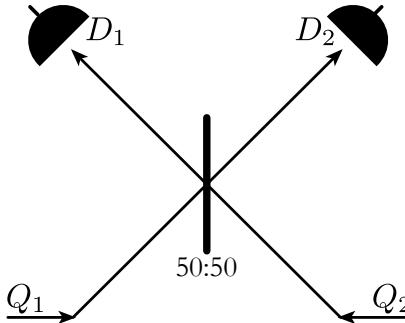


Figure 102: Entanglement distribution scheme based on quantum emitters and which-path erasure. Each node emits a photon entangled with the quantum memories present within that nodes. The photons from the adjacent repeater nodes then interfere on a beamsplitter (or polarising beamsplitter) which erases information about which path the photon took. The photons are then measured in an appropriate basis to project the quantum memories within the nodes onto an entangled state.

the output modes. They are of opposite polarisation, but measuring those photons in the diagonal/anti-diagonal (\hat{X}) basis erases this ‘which-path’ information yielding an equal superposition of the two alternative histories – an entangled Bell state of the form,

$$|\Psi_{\pm}\rangle = \frac{1}{\sqrt{2}}(|g\rangle|g\rangle \pm |e\rangle|e\rangle), \quad (29.3)$$

where the sign is given by the parity of the two photo-detection outcomes in the \hat{X} basis. This entangled state is stored in the quantum memories between nodes.

The scheme based on photon absorption by the quantum memories is effectively the time reversal of the emission-based scheme. Instead of using the beamsplitter to entangle the photons emitted from each memory, a source of entangled photon(s) is employed. Of course, the emission and absorption schemes can be used together in a hybrid architecture.

In any entanglement distribution scheme for quantum networks, the repeater nodes are spatially separated and one must consider channel losses, which are the dominant error source. Channel loss in this situation implies that we do not register a coincidence event between D_1 and D_2 , which heralds the entanglement. Thus our entanglement distribution success probability is reduced. In fact, the heralded probability of success can be expressed as,

$$p_{ED} = \frac{1}{2}e^{-L/L_0}p_{det}^2, \quad (29.4)$$

where L is the distance between the two repeater nodes with L_0 being the attenuation length of the channel, while p_{det} is the detector efficiency. Here we have ignored the source and coupling efficiencies. It is immediately obvious from this expression that the further the repeater nodes are apart, the lower the probability of success, on an exponentially decaying trajectory. The attenuation length of

typical telecom optical fibre is approximately 22.5km and so the average time to generate a distributed entangled pair is,

$$\begin{aligned} T_{av} &\sim \frac{L}{c \cdot p_{ED}} \\ &= \frac{2Le^{L/L_0}}{c \cdot p_{det}^2} \end{aligned} \quad (29.5)$$

where c is the speed of light in the channel. This grows exponentially against node separation and so places important constraints on the lifetime of the quantum memories. If we consider pure dephasing effects on our matter qubits, the state of our system can be represented by,

$$\hat{\rho}(F) = F|\Psi^+\rangle\langle\Psi^+| + (1-F)|\Psi^-\rangle\langle\Psi^-|, \quad (29.6)$$

where F is the fidelity of our entangled state given by,

$$F = \frac{1 + e^{-t/\tau_D}}{2}, \quad (29.7)$$

with t being the duration over which the entangled state is held in memory, while τ_D is the coherence time of the memory. If one only requires a single Bell pair and no further operation are performed, then $t = c/L$. However in a more general setting where multiple pairs are required, the time will be T_{av} on average, which is inversely proportional to the probability of generating the entangled state. The quality of the prepared remote entangled state may therefore not be sufficient for the tasks it is required for due to these finite memory lifetimes or operational gate errors. One needs to be able to purify these entangled resources.

2. Entanglement purification

The finite coherence-time of quantum memories and operational errors caused by quantum gates means some mechanism will be required to improve the fidelity of the distributed entangled state, especially if the spatial separation is large. This is generally achieved by entanglement purification (?????????) which as its name implies purifies the entanglement to a higher value. The purification operation uses either an error detection code (probabilistic but heralded operations) (???) or deterministic error correction codes (???). While the error correction codes purify in a deterministic way, they place tough constraints on both the required initial fidelity of entangled states and also the quality of the quantum gates implementing the purification (?). Given this, we will focus on the simplest error detection code which requires only a pair of shared entangled quantum memories (as shown in Fig. 103). This scheme is equivalent to the entanglement purification protocol described in Sec. XXVII.B, although the graphical notation is somewhat different.

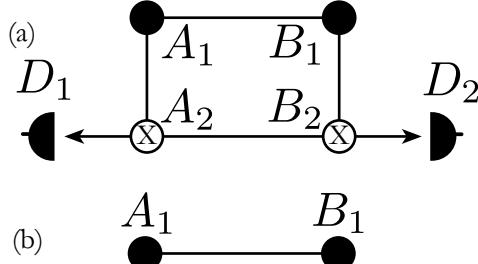


Figure 103: Entanglement purification: (a) The simplest purification scheme involving two pairs of shared remote entangled quantum memories ($A_1 - B_1$ and $A_2 - B_2$). The purification operation begins with Alice performing a CNOT operation between memories A_1 and B_1 . Similarly Bob performs a CNOT operation between his memories. Alice and Bob then measure qubits A_2 and B_2 in the computational (0, 1) basis and share their results. They discard the resulting state if between them they measured odd parity (0, 1 or 1, 0). They keep the state if they measured an even parity between them (0, 0 or 1, 1) which should have higher fidelity. (b) Two qubits are removed, leaving a residual 2-qubit state between A_1 and B_1 with improved fidelity.

In this simplest purification protocol, Alice and Bob share two pairs of entangled states of the form given by Eq. (29.6). These states are a mixture of only two Bell states. We begin our purification protocol by using local operations to transform $\hat{\rho}$ to,

$$\hat{\rho}(F) = F|\Psi^+\rangle\langle\Psi^+| + (1-F)|\Phi^+\rangle\langle\Phi^+|, \quad (29.8)$$

As shown in Fig. 103 we then apply a CNOT gate between Alice's two memories and Bob's two memories following by measuring A_2, B_2 in the computational basis. Upon measurement of even parity our resulting state $\hat{\rho}(F')$ has the form, but with new fidelity,

$$F' = \frac{F^2}{F^2 + (1-F)^2}. \quad (29.9)$$

It is immediately obvious that our resulting state $\hat{\rho}(F')$ is more entangled than $\hat{\rho}(F)$ when $F > 1/2$ (see Fig. 104). In fact the degree of entanglement as measured by the concurrence increases from,

$$C = 2F - 1, \quad (29.10)$$

to,

$$\begin{aligned} C' &= 2F' - 1 \\ &= \frac{2F^2}{F^2 + (1-F)^2} - 1. \end{aligned} \quad (29.11)$$

It is important to mention that the entanglement purification doesn't allow one to distribute a *perfect* Bell state. Rather it *asymptotically* approaches perfection (under ideal conditions) with repetition of the protocol.

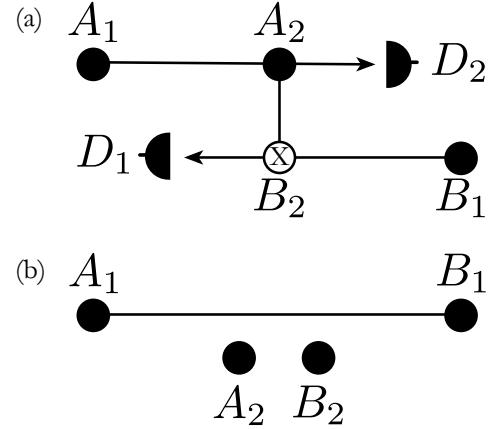


Figure 104: Plot of the increased fidelity and success probability for entanglement purification for a mixture of two Bell states with initial fidelity F . The dashed lines show how multiple pairs with an initial fidelity $F = 0.7$ can be purified iteratively to a final fidelity above 0.95.

The probability of obtaining the even parity outcome is,

$$p_{\text{even}} = \frac{F^2 + (1-F)^2}{2}. \quad (29.12)$$

Alternatively for the odd parity measurement results, which occur with probability,

$$p_{\text{odd}} = F(1-F), \quad (29.13)$$

the resulting state is an equal mixture of $|\Psi^+\rangle$ and $|\Phi^+\rangle$ and is not entangled at all. In this case we must start again from scratch with the entanglement distribution.

So far we have discussed one round of entanglement purification but the protocol naturally works in a recursive way where two copies of a state with the same fidelity are used for the next purification round. Using this bootstrapped approach one can in principle generate a near unit fidelity entangled pair from a finite fidelity pair (provided initial input fidelity $F > 1/2$).

There are two common variants of these purification protocols: the Deutsch and Dür variants:

- *Deutsch protocol* (?): This is an efficient purification protocol utilising Bell diagonal states that reaches a high fidelity in a few purification rounds. It is assumed that both entangled pairs have the same form. The purification protocol is the same as the one described above in Fig. 103, but begins with Alice (Bob) applying $\pi/2$ ($-\pi/2$) rotations about the X -axis on their qubits before the usual CNOT gates and measurements are performed. Two copies of the successfully purified pair can then be used in a recursive approach to purify either further. This in turns means multiple copies of the originally distributed states are required. We must have enough entangled pairs available to perform the multiple rounds of

purification that are required, which grows exponentially with the number of purification rounds.

- *Dür protocol (?)*: This uses the same core purification elements as shown in Fig. 103 but relaxes the traditional constraint that both Bell pairs must have the same fidelity. Instead we begin with two pairs of the same fidelity F , and perform the traditional purification. If successful we perform the next round of purification using the improved fidelity pair from the previous round and a fresh fidelity F pair. In effect this new auxiliary pair is used to boost the fidelity of the original pair higher. This can continue until we reach a limiting fidelity dependent on the original F . This limiting fidelity may be above the desired resultant fidelity, at which point we can terminate the purification protocol. A significant difference between the Deutsch and Dür protocols is that the number of memories in the Dür situation is linear in the number of nesting levels.

It is critical in repeater protocols to also discuss how fast these purification protocols can be performed. Even with ideal gates one has to wait for the parity information to be shared between the repeater nodes. For nodes separated by a distance L , the communication time for a single trial is L/c . However, remembering that purification is probabilistic but heralded in nature our waiting time could be many multiples of L/c . This will have a dramatic effect on performance, especially if performed at many different stages in the network with increasing distances between nodes.

3. Entanglement swapping

The entanglement distribution and purification scheme discussed previously allow one in principle to create high fidelity entangled states between adjacent repeater nodes. The next task is to extend the range of our entangled states, and this occurs via simple entanglement swapping (????). This was described previously in Sec. XVII.E, although the notation is modified.

Consider the situation where we have an entangled Bell pairs between nodes A_1 and A_2 and also between B_2 and B_1 . The entanglement swapping operation involves a Bell state measurement between the qubits A_2 and B_2 as shown in Fig. 105. After the Bell measurement we have the resultant state,

$$\hat{\rho}_{A_1, A_2}(F) \otimes \hat{\rho}_{B_2, B_1}(F) \rightarrow \hat{\rho}_{A_1, B_1}(F') \quad (29.14)$$

with,

$$F' = F^2 + (1 - F)^2, \quad (29.15)$$

where a local correction operation is performed on either A_1 or B_1 depending on the measurement outcome. It is

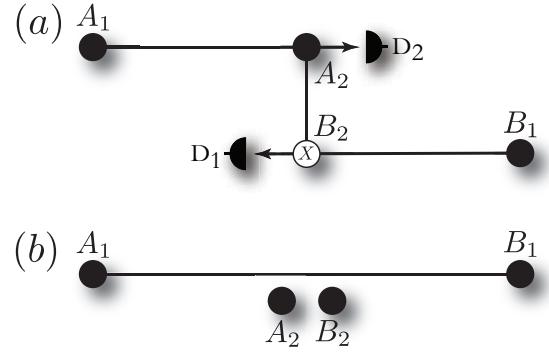


Figure 105: Entanglement swapping: (a) An entangled state is shared between Alice and a repeater station ($A_1 - A_2$), and also between the repeater station and Bob ($B_2 - B_1$). The entanglement operation begins by performing a Bell state measurement between A_2 and B_2 using a CNOT gate, and measurements at D_1 , D_2 . The measurements indicate which Bell state we have projected our state A_1 , B_1 onto. (b) The resultant entangled state between A_1 , B_1 with the qubits A_2 and B_2 disentangled from it.

clear that the longer range entangled state $\hat{\rho}_{F'}$ is less entangled than the states $\hat{\rho}_F$ used to generate it. In fact, to first order our fidelity drops from F to F^2 . This in turn means that we can not simply purify adjacent repeater pairs and swap them all to create the long range pairs. If we had n links, our final fidelity from all the swapping would scale as F^n . For high fidelity end-to-end entangled links we need to follow the approach outlined in Fig. 101. Finally, depending on how the Bell measurement is implemented, this process could be probabilistic (but heralded) or deterministic in nature. We assign the success probability as p_{ES} .

4. Performance

We now have all the operations required for a repeater to create long-range entanglement. The natural question to ask is how well it performs.

There are several important points to initially consider here. The majority of the repeater operations are probabilistic in nature (entanglement distribution and purification fundamentally, and entanglement swapping dependent upon implementation). While these probabilistic operations may be heralded, classical signalling must be performed between involved nodes to inform them of successes or failures. For entanglement distribution this time is just that associated with the signalling between adjacent nodes. However, purification and swapping are likely to require such signalling over the entire length of the network. This has a dramatic effect on the performance of the repeater network. The normalised rate for generating Bell pairs over a total distance L_{tot} is given

by,

$$R(n, k, L_{\text{tot}}) = \frac{1}{T_{n,k,L_{\text{tot}}} M_{n,k}} \quad (29.16)$$

where $T_{n,k,L_{\text{tot}}}$ is the time to generate a Bell pair over the total distance using an n -nested repeater configuration with k rounds of purification per nesting level. The

distance between repeater nodes is given by,

$$L = \frac{L_{\text{tot}}}{2^n}, \quad (29.17)$$

meaning there are $2^n - 1$ intermediate repeater nodes with Alice and Bob at the endpoints. In Eq. (29.16) we discount our rate by $M_{n,k}$, the total number of quantum memories used. The justification for this is that this provides a fairer comparison when different purification approaches are used. The Deutsch protocol for instance achieves its target fidelity much faster (fewer rounds) than the Dür protocol, but consumes far more resources in doing so.

Now it's straightforward, albeit tedious, to show that $T_{n,k,L_{\text{tot}}}$ is given by (?),

$$\begin{aligned} T_{n,k,L_{\text{tot}}} \sim & \frac{3^n}{2^{n-1} p_{\text{ED}}} \prod_{i=0}^{n-1} \left(\frac{3}{2}\right)^k \frac{1}{P_{\text{ES}}(n-i)} \prod_{j=0}^{k-1} \frac{1}{p_{\text{P}}(k-j, n-i)} \\ & + \sum_{m=1}^n \left(\frac{3^{n-m}}{2^{n-1}}\right) \prod_{i=0}^{n-m} \left(\frac{3}{2}\right)^k \frac{1}{P_{\text{ES}}(n-i)} \prod_{j=0}^{k-1} \frac{1}{p_{\text{P}}(k-j, n-i)} \\ & + \sum_{m=1}^n \sum_{q=0}^{k-1} \left(\frac{3^{n-m+q}}{2^{n-2m+q}}\right) \prod_{r=0}^q \frac{1}{p_{\text{P}}(k-r, m)} \prod_{i=0}^{n-m-1} \left(\frac{3}{2}\right)^k \frac{1}{P_{\text{ES}}(n-i)} \prod_{j=0}^{k-1} \frac{1}{p_{\text{P}}(k-j, n-i)} \end{aligned}$$

where p_{ED} is the probability of successfully distributing entanglement between adjacent repeater nodes, while $p_{\text{P}}(j, i)$ [$P_{\text{ES}}(i)$] represents the purification [entanglement swapping] probability at the i th nesting level with j rounds of purification. The factors of $3/2$ present in all entanglement distribution, purification and swapping operations is a multiplicative factor associated with the extra time required for the two pairs to be available for the various quantum operations (?).

It can be easily seen from this formula that,

$$T_{n,k,L_{\text{tot}}} \gg \frac{2L_{\text{tot}}}{c}, \quad (29.18)$$

especially if probabilistic gates are included. Next the resources scale polynomially with,

$$\begin{aligned} M_{n,k} & \sim 2^{(k+1)n} \\ & = \left(\frac{L_{\text{tot}}}{L}\right)^{k+1}, \end{aligned} \quad (29.19)$$

for the Deutsch protocol, which in turn implies it is efficient. However for long distances L_{tot} , our normalised rate $R(n, k, L_{\text{tot}}) \ll 1\text{Hz}$, especially when probabilistic CNOT gates and Bell state measurements are employed (?).

B. Second-generation repeaters & error correction

The previous approach for entanglement distribution over long distances based on first-generation quantum repeaters has its performance heavily constrained by both the probabilistic nature of the various quantum operations and the associated classical communication time. We know that the classical communication in entanglement distribution is only between the adjacent nodes, whereas for the purification and swapping operations it can be very long-range, potentially over the entire network length. This is the fundamental reason why the time to create a pair is of order $O(L_{\text{tot}}/c)$ or longer. This will not change significantly even if we have deterministic CNOT gates and Bell measurements as the entanglement purification protocols will remain probabilistic in nature (even though the swapping operations will be deterministic). We thus need to replace our usual entanglement purification protocols with a similar operation that is deterministic in nature (?).

The typical entanglement purification protocols are a form of quantum error detection code (?) (see Sec. XVI.A.6 for further discussion on quantum error detection and correction). Such codes herald whether an error has occurred or not, and in the situation considered above, detection of errors means one must discard the

entangled pairs associated with the purification protocol. No errors means the purification protocol has worked.

Error correction codes which operate in a deterministic fashion can also detect errors and can be used in this fashion (??). More critically, quantum error correction codes have the potential to correct some errors that have occurred, mitigating the need to completely discard states affected by errors. For normal error correction protocols used in quantum computations, we encode our physical qubits into logical qubits using the code, and then use syndrome measurements to determine where an error has potentially occurred.

Quantum communication however is different in this case as we must assume we have generated a number of imperfect Bell pairs between the repeater nodes before we utilise the error correction schemes. The error correction protocol in this case operates by using the error correction encoding circuit on Alice's qubits and the decoding circuit on Bob's (?) as illustrated in Fig. 106 for the 5-qubit code (??).

It's important to state here that error correction-based purification is deterministic in nature (there is however a significant cost that must be paid – the fidelity of the originally generated entanglement between adjacent nodes must be quite high) (??). There are no measurement events that need to be discarded. Instead the measurement results only inform us of which particular imperfect Bell states we have and the correction operation required to return to the desired state. In effect the measurement is updating the Pauli reference frame (?). This does not need to be executed immediately, and may be deferred until later. In turn this means once the measurements have been performed, we can immediately use the purified Bell state without having to wait for the classical signalling (at some stage the correction operation needs to be executed but this can be once the long distance entanglement has been generated).

Mitigating having to wait for the measurement results to be sent and received in both the quantum error correction-based purification and entanglement swapping protocols has a profound effect on the rate of generating long-range entangled pairs. We still need to perform long-range classical messaging (potentially between end nodes), and thus it's immediately obvious that the preparation time can scale solely as,

$$T = \frac{2L_{\text{tot}}}{c}, \quad (29.20)$$

which was the lower bound on the first-generation schemes (?).

Naïvely this seems to imply that the generation rate between end nodes cannot be faster than this. However, one can in fact do far better! This is shown in Fig. 107 (protocol described in caption). The key issue is that the generation rate depends on how long the adjacent nodes need to store part of an entangled state (????).

Fundamentally we know that the time to attempt to generate a single entangled Bell pair between two nodes is scaling as L/c (where L is the distance between those two nodes). With channel losses we need to make,

$$m = \frac{\log_{10}(\varepsilon)}{\log_{10}(1 - p_{\text{ED}})} - \log_{10}\left(\frac{\varepsilon}{p_{\text{ED}}}\right), \quad (29.21)$$

attempts to generate a single Bell pair with error probability ε . We can make these attempts simultaneously and not affect the generation time. Now by using a butterfly repeater design, as illustrated in Fig. 107, one immediately notices that the qubits with the repeater nodes are only used for duration $\sim 2L/c$. After this time those qubits have been freed up and are available to generate new entangled links. This means in turn that the time to generate the long-range entangled pair will scale as $T \sim O(2L/c)$ (independent of the overall distance L_{tot}) (????). The exact resources used depends heavily on the error correcting code, but we know they in principle scale as $M \sim O(\text{polylog}(L_{\text{tot}}))$ (?).

This is quite a dramatic decrease in both T and M compared to the first-generation. In fact one could expect the normalised rates to be on the order of kHz (?). However this is a significant cost in terms of the quality of the original Bell pairs that must be prepared. In the first-generation schemes a fidelity just over 50% was sufficient. However with the second-generation schemes using normal error correcting codes, it is likely this initial fidelity will have to be over 90% (??).

C. Third-generation repeaters

The use of error correcting codes significantly improves the performance of second-generation quantum repeaters compared to first-generation ones. The second-generation schemes are now limited by the communication time between adjacent repeater nodes to herald whether entanglement distribution was successful or not (??). The communication (both quantum and classical) is ultimately limited by the speed of light (either in fibre or over free-space). The natural question is whether we can improve performance even further.

The only remaining avenue at our disposal is to move from probabilistic to deterministic entanglement distribution. Remembering that we have losses in the channel, the only way to achieve deterministic entanglement distribution will be by transmitting encoded error-correctable states between repeaters. This means we must turn to loss-based error correction codes (??????).

1. Loss-tolerant codes

There are quite a number of error codes that can correct for loss events, but here for illustration we consider

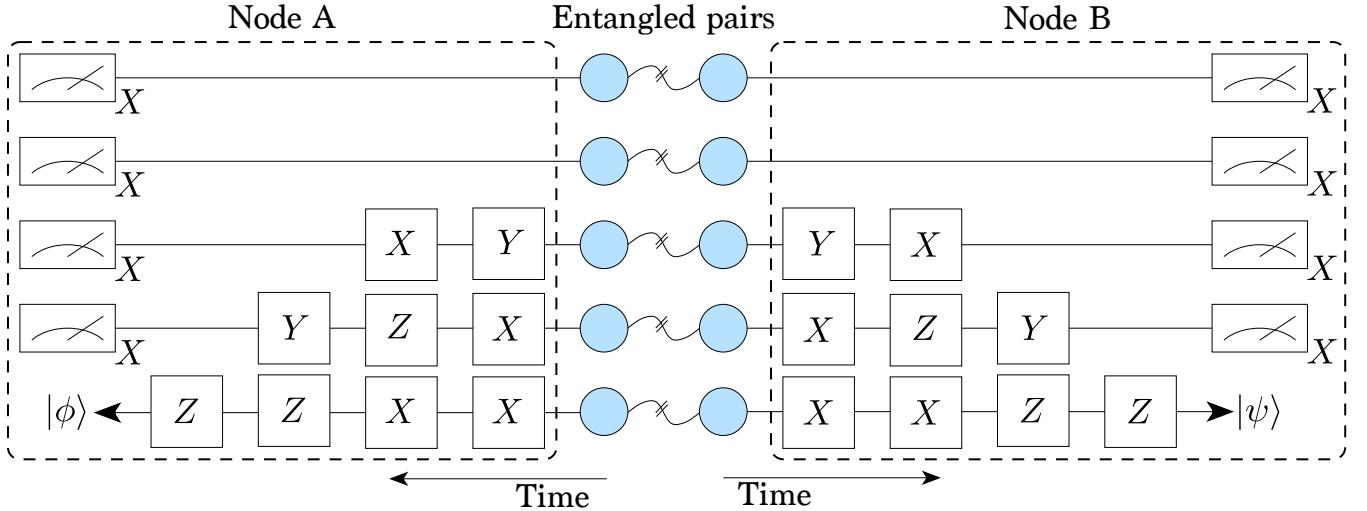


Figure 106: Purification circuit based on quantum error correction. The specific example shown is for the [[5,1,3]] code (??). We assume that entanglement distribution has allowed Alice and Bob to create 5 copies of their imperfect Bell pairs. The error correction circuit is executed independently between the two nodes. While we show the situation when the measurements at both sides are done directly on four pairs of entangled qubits (leaving us with one unencoded Bell pair), one can also use ancilla qubits to measure the appropriate syndromes. As soon as the measurements are complete both nodes' qubits are available for continued use as the error correction is deterministic and there are no failure events that need to be heralded. In this case the classical message between nodes just carries Alice's measurement results, allowing either node to interpret which Bell state was generated and for one of them to apply the bit-flip or phase-flip correction operation if needed to recover the desired Bell state. In many cases this correction is classically tracked in the Pauli frame, which keeps a record of whether \hat{X} and/or \hat{Z} corrections need to be performed at some stage (??). Note that it's not necessary to measure out all but one of the qubits involved in the entangled links. Instead the logical qubit can be maintained by the use of ancilla qubits within that node with the syndrome being measured with the help of the ancilla qubits. Entanglement swapping could then be performed on the logical qubits enabling a much more error resilient system.

parity codes in their simplest form (??). Other well-known approaches are based on cluster states (?).

Consider a four photon state of the form,

$$\begin{aligned} |\Psi\rangle = & \alpha|0\rangle_1|0\rangle_2 + |1\rangle_1|1\rangle_2 \otimes (|0\rangle_3|0\rangle_4 + |1\rangle_3|1\rangle_4) \\ & + \beta(|0\rangle_1|1\rangle_2 + |1\rangle_1|0\rangle_2) \otimes (|0\rangle_3|1\rangle_4 + |1\rangle_3|0\rangle_4), \end{aligned} \quad (29.22)$$

where $|0\rangle$ and $|1\rangle$ represent orthogonal degrees of freedom (e.g polarisation). This state can be rewritten in the form,

$$|\Psi\rangle = \alpha|\Phi_{12}^+\rangle|\Phi_{34}^+\rangle + \beta|\Psi_{12}^+\rangle|\Psi_{34}^+\rangle, \quad (29.23)$$

and thus the state has been encoded into terms of a tensor product of two redundantly encoded Bell states. Now photon loss will remove one of these photons. As an example, let us consider what happens when photon 4 is lost. The resultant state can be represented by the density matrix,

$$\hat{\rho} = |\zeta^+\rangle\langle\zeta^+| + |\zeta^-\rangle\langle\zeta^-|, \quad (29.24)$$

where,

$$\begin{aligned} |\zeta^+\rangle &= \alpha|\Phi_{12}^+\rangle|0\rangle_3 + \beta|\Psi_{12}^+\rangle|1\rangle_3, \\ |\zeta^-\rangle &= \alpha|\Phi_{12}^+\rangle|1\rangle_3 + \beta|\Psi_{12}^+\rangle|0\rangle_3. \end{aligned} \quad (29.25)$$

We immediately notice that $|\zeta^-\rangle = \hat{X}_3|\zeta^+\rangle$ and so by measuring the third photon in the \hat{X} basis our state reduces to the pure state $\alpha|\Phi_{12}^+\rangle \pm \beta|\Psi_{12}^+\rangle$, where the \pm sign is given by the \hat{X} measurement outcome. This is then correctable using local operations. After the loss event of photon 4 and the measurement of the third photon our state thus becomes,

$$\alpha|\Phi_{12}^+\rangle + \beta|\Psi_{12}^+\rangle, \quad (29.26)$$

which has exactly the same information in it as $|\Psi\rangle$ but without the redundant encoding.

It is now straightforward to re-encode back to our original state, Eq. (29.23). We considered photon loss only on the fourth qubit. However the same principle applies for any lost photon. Unfortunately we can only tolerate the loss of a single photon using this encoding, so the loss rate must be small.

The above example illustrates how the smallest optical loss code work. The general code with $n - 1$ redundancy can be written as (??),

$$|\Psi\rangle = \alpha|\Phi_e\rangle_1 \dots |\Phi_e\rangle_n + \beta|\Psi_o\rangle_1 \dots |\Psi_o\rangle_n \quad (29.27)$$

where $|\Phi_{e,o}\rangle$ are the even and odd parity m photon states

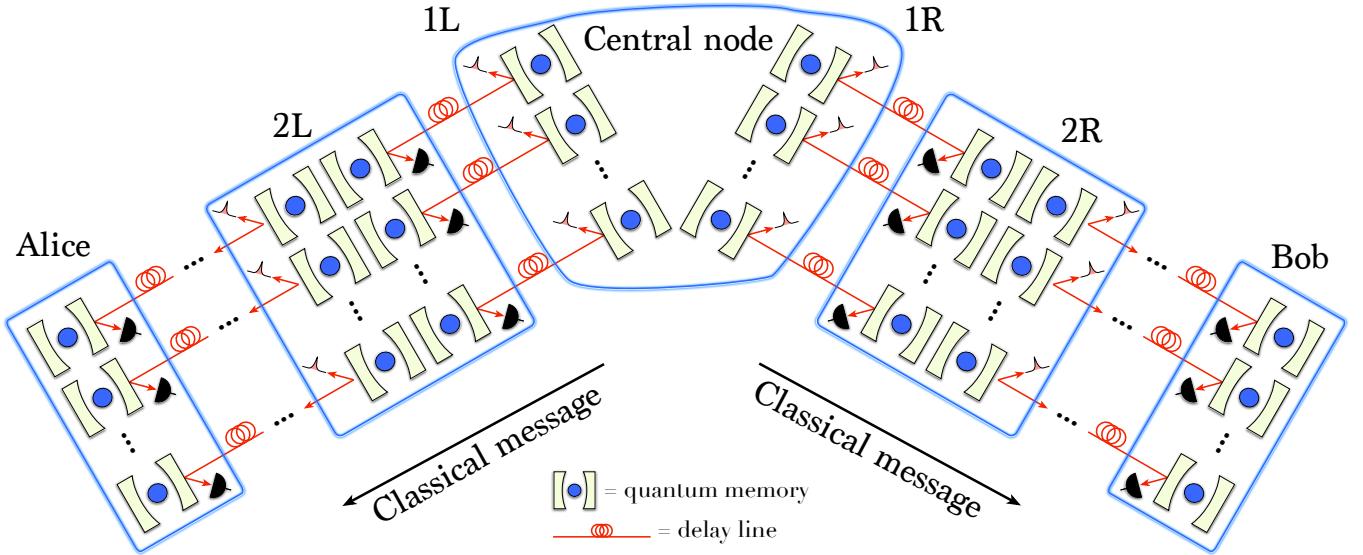


Figure 107: A butterfly design quantum repeater network protocol that reduces the requirements on all the quantum memory times to only that associated with the signalling time between adjacent repeater nodes (?). Enough pairs must be generated between the node to ensure that we can use them in the error correction code in a single round trip time between adjacent nodes. The scheme relies on multiple entangled pairs being generated temporally, starting from the mid point of the network. The protocol begins with the central node creating links to both the left and right nearest neighbour nodes in sufficient number to allow an error correction code to be implemented. Once they are created the error correction circuits are applied to the links left and right of this central node (effectively creating encoded logical links). Entanglement swapping at the middle node is then applied between these logical links, creating a logical link between the left and right adjacent nodes. The left and right nodes can then do the same to their next adjacent repeater nodes, error correcting as they go, until the desired end-to-end entangled link is achieved.

given by,

$$|\Phi_{e,o}\rangle = \frac{1}{\sqrt{2}}(|+\rangle_1 \dots |+\rangle_m \pm |-\rangle_1 \dots |-\rangle_m), \quad (29.28)$$

with $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. This redundancy-based parity code is composed of n logical qubits each containing m photons. For this code to correct loss errors we have two constraints,

- At least one logical qubit must arrive without photon loss.
- Every logical qubit must have at least one photon arrive successfully.

If these constraints are met, the loss events during transmission between adjacent repeater nodes can be corrected. Of course, such codes can not correct more than fifty percent errors and so the distance between repeater nodes is limited. Remembering that the probability of a photon being successfully transmitted through a channel of length L with attenuation length L_0 is given by $p = e^{-L/L_0}$, the maximum distance between repeater nodes is $L/L_0 \sim 0.69$ (which corresponds to approximately 17km in present-day commercial telecom fibre). This is much shorter than what we would typically consider for the first and second-generation schemes.

2. Operation

Let us now describe the operation of the third-generation repeater scheme depicted in Fig. 108 in detail (??).

It begins at the left hand node by Alice encoding her message into a redundant parity code created on a series of matter qubits using local quantum gates within that repeater node.

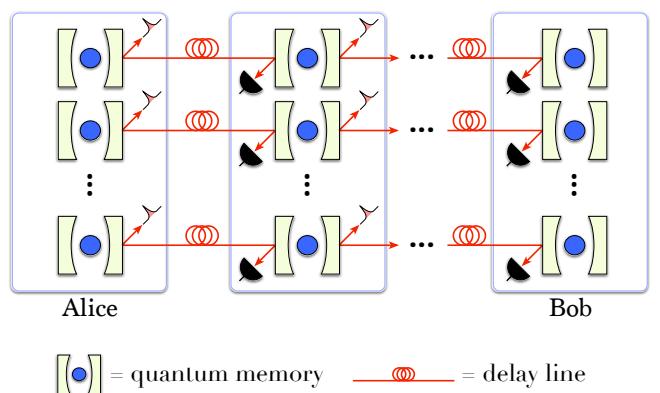


Figure 108: Transmission of a quantum signal using loss-based error correction codes in a quantum network.

The quantum state is then transferred/teleported via

photons which are transmitted through a lossy channel to the adjacent repeater node. Here, two specific operations occur: first the information encoded on each photon is transferred to a matter qubit within that repeater node and then that photon is measured. The photon measurement is critical as it heralds which photons have been lost and allows us to measure the remaining qubit in that block in the \hat{X} basis, which removes the damaged parity blocks from our encoded state, leaving our information intact. We can now add the full redundancy back into our encoded state in the matter qubits. The fully encoded state can then be transferred to photons and transmitted to the next repeater node where the same procedure occurs again. This continues until our state reaches the last repeater node where Bob is.

There is one immediate observation that can be made from this scheme. The matter qubits (quantum memories) within the local nodes are only used to encode and error correct the redundancy code as well as transmitting those quantum states as photons. Entanglement is not stored within the nodes while the photons are being sent to the adjacent repeater nodes. This in turn means the resources within that repeater node can be used immediately again

(once the photons have been transmitted), and so the rate of communication is now limited by the time to perform the local operations within a node, rather than the round trip time between adjacent nodes.

The focus so far has been only on loss-based errors but this code is fault-tolerant to general errors as well (?). Furthermore, this redundancy code was only an illustrative example that photon loss in the channel can be corrected. Many other codes can be used in a similar fashion (???). Finally the scheme we have presented in Fig. 108 transmits a quantum signal from Alice and Bob. It can however be adapted to use the butterfly design from Fig. 107 to create remote entanglement between Alice and Bob while maintaining the performance advantages our direct transmission scheme gave.

D. Resource scalings across repeater generations

As can seen the various quantum repeater generations take quite different approaches as to how they distribute entanglement between Alice and Bob over a long distance (?). It is useful thus to summarise in Tab. V the performance of the various repeater approaches and their requirements.

Repeater generation	T_{av}	Resources consumed	L_{max}	Local gate precision
First-generation	$O(L_{\text{tot}}/c)$	$O(\text{poly}(L_{\text{tot}}))$	arbitrary	arbitrary
Second-generation	$O(2L/c)$	$O(\text{polylog}(L_{\text{tot}}))$	arbitrary	high
Third-generation	$O(t_{\text{local}})$	$O(\text{polylog}(L_{\text{tot}}))$	$L/L_0 < 0.69$	fault tolerant levels

Table V: Quantum repeater approaches and their expected performance scalings. T_{av} corresponds to the time between which the protocol can be attempted (the time to generate a single Bell state is at least L_{tot}/c) (?). The generation rate is $R \sim 1/T_{\text{av}}$. Further given are the resources (quantum memories) required as well as the precision for the local gate operations within repeater nodes. L_{max} is the maximum spacing between repeater nodes. L_{tot} is the total distance between Alice and Bob while L is the distance between adjacent repeater nodes. t_{local} is the time required to perform the local operations within the repeater node, while L_0 is the attenuation length of the channel/fibre.

The table clearly shows that the average time to generate the Bell pair between the end nodes of the repeater network decreases significantly as we move to higher generation quantum repeaters. In the first-generation our generation rate is $O(c/L_{\text{tot}})$, which increases to $O(c/L)$ for the second-generation schemes, and finally to $O(1/t_{\text{local}})$ for the third-generation ones. The difference here could be more than nine orders of magnitude. Next the number of quantum memories required decreases from $O(\text{poly}(L_{\text{tot}}))$ for the first-generation approach to $O(\text{polylog}(L_{\text{tot}}))$ for the higher ones. The higher generation schemes however come at quite a cost, with the requirement for fully fault-tolerant (or near fault-tolerant) quantum gates within nodes. In fact, it's likely that Alice and Bob will have

multiple potential routes between themselves.

E. The transition to quantum networks

The previous quantum repeater networks we discussed have been simple point-to-point linear networks. While there may have been a number of ways to establish end-to-end entangled links between Alice and Bob, they knew they were connected via a simple, direct, linear chain.

Of course this is highly unrealistic. Alice and Bob are likely to be members of a complex quantum network, that supports multiple users simultaneously and offers multiple routes from a given source to a given destina-

tion (see Sec. X). This leads to a number of interesting considerations going forward.

- For large-scale networks the users may not know its exact network topology or even the best route between them. In fact there could be multiple paths between Alice and Bob. Probing the entire network to establish the best route would be slow and costly (in practice). Still, every node should have a unique identifier (quantum IP address), uniquely indicating its location and resource availability in the network.
- Most complex networks dynamically change in time as resources become congested or nodes break. This in turn means using a butterfly approach to create Alice and Bob's links is problematic as one does not know the middle point between them to start the entanglement creation process. If one has to determine the route in advance and restrict access to those parts of the network required to establish the entire links, congestion will quickly follow. The generation rate will be very slow.
- Finally, it's unlikely that repeater nodes will be equally spatially separated (making the first-generation repeater schemes extremely hard to use in this situation).

The above issues lead us to a network model where Alice and Bob suspect there is a route between them but do not know the exact route, which is likely to be dynamic, i.e. the availability of routes and their relative costs are liable to change. In such a case, if Alice wants to send a message to Bob, she uses her knowledge of Bob's rough location (from the quantum IP address) and her knowledge of the nodes close to her to send a message to a repeater node, who will have more knowledge of Bob's part of the network. This node can then forward the message to further nodes (who know even more about Bob's location) until it finally reaches Bob. The quantum IP address is essential here as that identifier indicates to the repeater node who to forward to next. In principle as the message (or entanglement) is being established node-by-node, those repeater nodes who have already been used are free to work for other users.

There is another interesting aspect of our general complex quantum networks. There are likely to be many paths between Alice and Bob which could be attempted in superposition fashion. This will not only increase the capacity between Alice and Bob but also its robustness.

F. Repeater synchronisation

— by Peter P. Rohde.

When employing non-deterministic entanglement sources in a quantum repeater network there is no guar-

antee that pumping the source will actually yield an output entangled pair. Indeed this is extremely unlikely for sources such as SPDC. For this reason quantum memories will be required when performing entanglement swapping, so as to temporally synchronise the unpredictable arrival times of qubits.

However, future technologies may enable push-button entanglement sources, in which case there is no ambiguity in the preparation times of pairs. In this instance quantum memories may be avoided entirely. Instead we can trigger all the sources at exactly the right times so as to ensure that at every joint measurement device in the repeater network the photons arrive simultaneously.

Consider a simple quantum repeater network, comprising a linear chain of alternating Bell pair sources and entanglement swappers, as shown in Fig. 109.

Let t_i be the triggering time of the i th source, and $\tau_i^\leftarrow (\tau_i^\rightarrow)$ be the channel propagation times from the i th source to the entanglement swapper immediately to its left (right) in the chain. Imposing the race-time condition that two photons arriving at a measurement device are simultaneous,

$$t_i + \tau_i^\rightarrow = t_{i+1} + \tau_{i+1}^\leftarrow, \quad (29.29)$$

where we set $t_1 = 0$ as an arbitrary reference. This yields the linear system of equations,

$$\hat{T} \cdot \vec{t} = \vec{\delta}, \quad (29.30)$$

where,

$$\hat{T} = \begin{pmatrix} -1 & 0 & 0 & 0 & \dots \\ 1 & -1 & 0 & 0 & \\ 0 & 1 & -1 & 0 & \\ 0 & 0 & 1 & -1 & \\ \vdots & & & & \ddots \end{pmatrix}, \quad (29.31)$$

$$\vec{t} = \begin{pmatrix} t_2 \\ t_3 \\ t_4 \\ t_5 \\ \vdots \end{pmatrix}, \quad (29.32)$$

$$\vec{\delta} = \begin{pmatrix} \tau_2^\leftarrow - \tau_1^\rightarrow \\ \tau_3^\leftarrow - \tau_2^\rightarrow \\ \tau_4^\leftarrow - \tau_3^\rightarrow \\ \tau_5^\leftarrow - \tau_4^\rightarrow \\ \vdots \end{pmatrix}. \quad (29.33)$$

Solving,

$$\vec{t} = \hat{T}^{-1} \cdot \vec{\delta}, \quad (29.34)$$

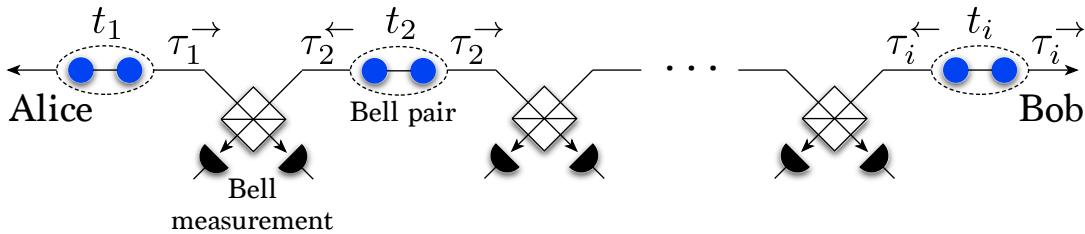


Figure 109: A linear quantum repeater network comprising Bell pair sources, and Bell measurements for entanglement swapping (there is no purification stage in this example). Upon success this yields a single Bell pair between the leftmost and rightmost photonic qubits. t_i is the triggering time of the i th source, and τ_i^{\leftarrow} (τ_i^{\rightarrow}) are the channel propagation times between the i th source and the entanglement swapper to its immediate left (right). With push-button sources, if the t_i are chosen appropriately, as per Eq. (29.34), we can satisfy the race-time condition of simultaneous arrival times of photons at Bell measurements, mitigating the need for quantum memories to synchronise them.

where,

$$\hat{T}^{-1} = \begin{pmatrix} -1 & 0 & 0 & 0 & \dots \\ -1 & -1 & 0 & 0 & \\ -1 & -1 & -1 & 0 & \\ -1 & -1 & -1 & -1 & \\ \vdots & & & & \ddots \end{pmatrix}, \quad (29.35)$$

yields the required triggering times for all sources (relative to $t_1 = 0$) to ensure synchronisation of photon arrival times at all entanglement swappers.

The total execution time of the protocol (i.e time taken to successfully distribute an entangled pair) is given by,

$$t_{\text{exec}} = \max_i(t_i + \max(\tau_i^{\rightarrow}, \tau_i^{\leftarrow})) - \min_i(t_i), \quad (29.36)$$

just the difference between the latest photon arrival time and the earliest source triggering time.

— *Acta non verba.*

XXX. THE IRRELEVANCE OF LATENCY

ENTANGLEMENT distribution can be executed in a highly varying manner of ways – from transmitting optical qubits through space via satellites, to across land surfaces via optical fibre, to dumping solid-state qubits into cargo containers and shipping them via land or sea freight. These bring with them associated transmission latencies. The former two distribute entanglement at the speed of light with latencies on the order of microseconds, whereas the latter induces enormous latencies on the order of days or weeks.

At first glance it may appear that this renders the Sneakernet™ approach to entanglement distribution useless. Who wants to wait several weeks to communicate their qubits?

If these transmission methods were being utilised for direct transmission of quantum data, this would certainly be a major concern. However, we are not employing them to communicate unknown quantum data packets directly. Rather we are using them to distribute many instances of

completely identical Bell states. This changes the impact of latency entirely. That is to say, we treat known entangled states as a *resource* rather than as an actual unit of data, and provided we can store it (i.e we have a good quantum memory), whether it arrives sooner or later is not terribly important. More important is that we have a ‘buffer’ of entangled states at hand to draw upon when needed.

If our goal is to transmit a quantum state between two parties, the obvious approach is to send the qubits directly over the quantum channel. Alternately, they could initially share Bell pairs then employ quantum state teleportation to teleport the state between parties. In this case all that matters is that they hold a shared Bell pair in time for execution of the teleportation protocol. It could have been distributed between them at any point in the past, held in a quantum memory until needed. The latency is now determined entirely by the latency of the *classical* channel, which communicates the associated local corrections required to complete the teleportation protocol. In most classical networks, communication rates are on the order of the speed of light, with very little latency.

We see that the latency associated with entanglement distribution does not affect the latency of quantum state transmission when implemented via teleportation. The quantum network could continually be sharing entangled pairs between parties in a UDP-like mode, who hold them in quantum memory. They ensure that Bell pairs are being distributed at a sufficient rate that parties have a buffer of entangled pairs sufficient to accommodate demand for future teleportations. This irrelevance of quantum latency is a uniquely quantum phenomena, not applicable to any classical protocols³⁷.

Teleportation-based quantum communication is additionally favourable in that shared Bell pairs can be purified

³⁷ One minor exception might be to treat randomness as a resource for randomised classical computation, i.e for application in **BPP** algorithms. In that restricted instance the latency of our source of random bit-strings is also irrelevant since randomness is invariant under temporal displacement and can be buffered for future use.

before being utilised, allowing errors accrued during quantum communication to be minimised, something not so straightforward (or impossible) when transmitting data qubits directly.

— *Ceterum censeo carthaginem delendam esse.*

XXXI. THE QUANTUM SNEAKERNET™

— by Simon Devitt.

To do. Simon Devitt.

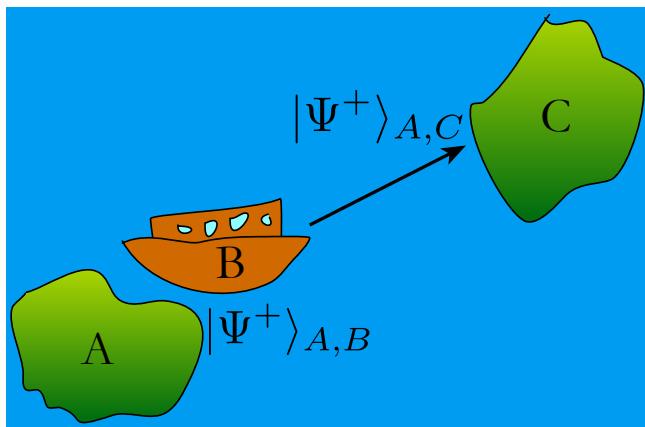
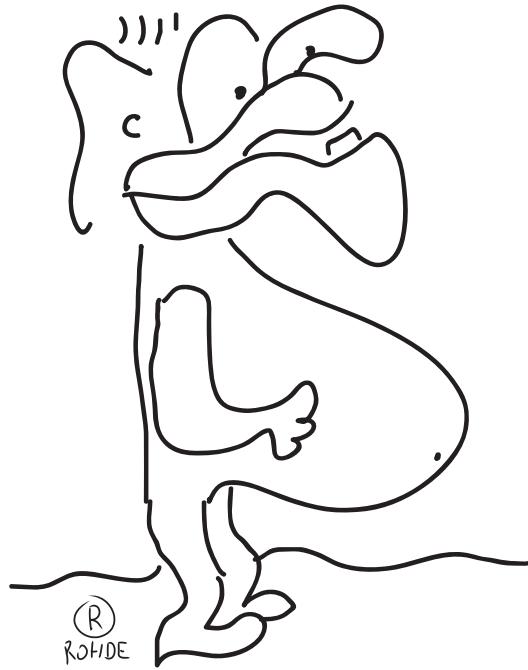


Figure 110: In the quantum Sneakernet™ entanglement distribution protocol, rather than using conventional flying qubits, we take highly error-corrected stationary qubits and physically transport them over long distances as freight. The error correction must be sufficient to maintain coherence over the time-scale of the journey, which could be anything from hours (when flying) to weeks (by sea^a). The mode of transport only transports one half of an error-corrected Bell pair, transferring the initial entanglement between source and vessel, $|\Psi^+\rangle_{A,B}$, to be between source and destination, $|\Psi^+\rangle_{A,C}$. Because all Bell pairs are identical and infinitely reproducible, latency presents no problem. For example, when using a Bell pair to teleport a qubit over long distances, when the Bell pair arrives is unimportant, as long as it's available at the time of teleportation. Thus, our Bell pair cargo carriers can simply operate in the background, transporting Bell pairs with as much bandwidth as possible, which may then be buffered by the recipient until required, latency being unimportant provided the coherence lifetime of the error correcting code is long enough.

^a Or indefinitely if the destination is Australia and you're fleeing war crimes.

— *Veritas vos liberabit.*



Part VII

Quantum cryptography

“I would rather have questions that can’t be answered than answers that can’t be questioned.” — Richard Feynman.

UNDoubtedly, quantum technologies will be most impactful (and disruptive!) in the area of information security, something of fundamental importance to us all on a daily basis, vital to the entire world economy. Quantum technologies will be important both in terms of breaking and maintaining security, with the former mandating interest in the latter.

In Sec. XLIV we discussed encrypted outsourced quantum computation as an important concept in future cloud quantum computing. In this section we will step back from full-fledged distributed quantum computation, instead focussing on more elementary protocols for simple secure communication or protocols.

Today, the ability to communicate secretly with others is completely taken for granted in all but a few nations and resides in every smartphone and desktop PC. Furthermore, the encryption technologies available to the average consumer are extremely strong, the same as those used by large organisations, including world governments.

XXXII. WHAT IS SECURITY?

“The only true wisdom is knowing you know nothing.” — Socrates.

“Ignorance more frequently begets confidence than does knowledge.” — Charles Darwin.

BEFORE describing any specific cryptographic protocols, let us define what is meant by ‘security’ in a cryptographic context. We differentiate between *information theoretic security*, as opposed to *computational security*:

- Information-theoretic security: the laws of quantum information bound the amount of information that can be extracted from a system, irrespective of measurement or computational operations. Thus, such security can be regarded as attack-independent, making no assumptions about our adversary’s capabilities.
- Computational security: is based on the assumption that an adversary’s computational resources are insufficient to perform cryptanalysis or brute-force cracking.

Clearly the former makes a far stronger statement about the security of a protocol than the latter.

Classical public- and private-key encryption protocols are typically based upon the assumption of computational security (e.g the computational complexity of performing integer factorisation in the case of RSA public-key encryption, or solving a complex satisfiability problem in the case of private-key encryption), whereas quantum encryption protocols are typically information theoretically secure (e.g the one-time pad using QKD).

XXXIII. CLASSICAL CRYPTOGRAPHY

WE begin with an introduction into *classical* cryptography, so as to understand its limitations, which logically leads us into how quantum mechanics can assist in overcoming them. We only scratch the surface of this extremely well-researched field, reviewing some of the most important and widely used protocols. For a deeper understanding of classical cryptography we refer the interested reader to the excellent and comprehensive (?).

A. Private-key cryptography

Private- (or symmetric-) key cryptography is perhaps the most basic (and useful) cryptographic primitive, enabling encryption of a channel between two parties who share a secret-key – a random bit-string of length determined by the encryption algorithm. The same secret-key is employed for both encryption and decryption operations (hence ‘symmetric’), making it of utmost importance that it be retained secret.

Private-key cryptography has a long history, in fact going back to ancient times, enabling the secret sharing of diplomatic messages between emperors and empires, e.g the so-called *Caesar cipher*, a simple substitution cipher based on shifting the letters of the alphabet. However it was a niche technology that very few utilised, since it had to be implemented by hand without computers or automation.

Today there are countless freely available private-key cryptographic protocols available online, and some have been standardised by standards institutes. Currently, the Advanced Encryption Standard (AES) is a standard endorsed by the US government, replacing the earlier standardised Data Encryption Standard (DES) whose mere 56-bit key-length is today considered insecure in light of present-day computing power. AES is a block cipher, meaning that it divides data into small blocks of 128 bits, each of which are encrypted independently, and operates with key lengths of up to 256 bits (referred to as AES256), making it very robust against (even quantum) brute-force attacks (Sec. XXXIV). The length of the plaintext and ciphertext is the same, meaning there is no bandwidth overhead when communicating encrypted data across a network.

B. One-time pad cipher

There is one and only one *provably* secure (in the sense of information-theoretic security as opposed to computational security) encryption protocol – the *one-time pad*. This protocol requires Alice and Bob to share a random secret-key as long as the message (plaintext) being communicated between them. The two bit-strings undergo bit-wise XOR operations to form the ciphertext. Mathematically,

$$c = s \oplus k, \quad (33.1)$$

where \oplus is the bitwise XOR operation (equivalently addition modulo 2), and c , s and k are the ciphertext, plaintext and key bit-strings respectively, all of which are of the same length,

$$|c| = |s| = |k|. \quad (33.2)$$

The security of this protocol is easy to see intuitively – with an appropriate choice of key, *any* plaintext of the same length could be inferred from *any* ciphertext. This means that there is no possibility of performing any kind of frequency analysis, as the ciphertext string has maximum entropy (inherited from the maximum entropy of the random key, and assuming a strong cryptographic random bit generator) and thus no correlations. Since every possible valid plaintext can be recovered using an appropriate key, a cracking algorithm is unable to find a unique plaintext matching the ciphertext, since all are equally valid decryptions.

Importantly, the secrecy of the one-time pad strictly requires that a key never be reused. A fresh key must be generated for each message sent, otherwise trivial frequency analysis techniques can be employed to compromise security. If the same key k is used to encode two messages s_1 and s_2 , yielding ciphertexts,

$$\begin{aligned} c_1 &= s_1 \oplus k, \\ c_2 &= s_2 \oplus k, \end{aligned} \quad (33.3)$$

then we trivially obtain,

$$\begin{aligned} c_1 \oplus c_2 &= (s_1 \oplus k) \oplus (s_2 \oplus k) \\ &= (s_1 \oplus s_2) \oplus (k \oplus k) \\ &= s_1 \oplus s_2, \end{aligned} \quad (33.4)$$

which is independent of the key. Now a frequency analysis on the bitwise XOR of two plaintexts can be applied, without requiring any knowledge of the key whatsoever.

Needless to say, the requirement for keys of the same length as the plaintexts, which cannot be reused, raises the obvious criticism that now secret-key-sharing is as difficult as sharing a secret message in the first place. This reduces the problem of perfect secrecy of arbitrary messages to the secrecy of shared randomness.

Although during the Cold War Soviet diplomats would literally carry briefcases between countries full of paper with random data for use in a one-time pad, it is clearly not suitable for everyday applications!

C. Public-key cryptography

While private-key cryptography solves the problem of end-to-end cryptography, it has one main downfall – how does one share a private-key between two parties? After all, if we had the ability to secretly share keys between ourselves, wouldn't we just use that same method to directly communicate, bypassing the unnecessary cryptographic protocol?

Public- (or asymmetric-) key cryptography addresses this issue by replacing the private-key with two keys (known as a key-pair), one used solely for *encryption*, the other solely for *decryption*. Importantly, these two keys are non-trivially related and cannot be efficiently computed from one another. To send a message to a friend I can send him my encryption (public) key that he is only able to use for preparing an encrypted message for me. No security is required when sharing the public-key since an eavesdropper can't use it for decryption. Finally, I am able to decrypt the message using my decryption (private) key, which I kept completely to myself and never shared with anyone.

RSA (?) was the first published public-key cryptographic protocol, and forms the backbone for most encryption used on the internet today. It achieves its security based on the (strongly held, but unproven) belief

that factorising large integers into constituent primes is a computationally hard problem – a so-called ‘trapdoor function’. The algorithm is built upon number theory using modular arithmetic. Alg. XXVII described the RSA key-generation, encryption, and decryption protocols.

```

function RSAGenerateKey():
  1. p = randomPrime()
  2. q = randomPrime()
  3. if (p = q) goto 1
  4. n = pq
  5.  $\lambda = \text{lcm}(p - 1, q - 1)$ 
  6. e = coprime( $\lambda$ ) s.t. e <  $\lambda$ 
  7. d =  $e^{-1} \pmod{\lambda}$ 
  8. publicKey = {n, e}
  9. privateKey = {n, d}
  10. keyPair = {publicKey, privateKey}
  11. return(keyPair)
  12.  $\square$ 

function RSAEncrypt(plaintext, publicKey):
  1. ciphertext = plaintexte (mod n)
  2. return(ciphertext)
  3.  $\square$ 

function RSADecrypt(ciphertext, privateKey):
  1. plaintext = ciphertextd (mod n)
  2. return(plaintext)
  3.  $\square$ 

```

Algorithm XXVII: Number-theoretic algorithms based on modular arithmetic for RSA key generation, encryption and decryption.

Since RSA, numerous other public-key cryptosystems have been developed, based on different choices of trapdoor function. Most notably, elliptic-curve cryptography has gained much attention. However, RSA remains the most widely used and well-studied public-key cipher.

To mitigate the need for constant one-on-one exchange of public-keys, many key servers exist around the globe, which maintain databases of people's public-keys. These servers are in a position of trust, vouching for the identities associated with their stored public-keys.

D. Key exchange protocols

A downside of RSA is that ciphertexts are in general much longer than plaintexts, unlike private-key protocols where the ciphertext is always the same length as the plaintext. For this reason it is typically not used to directly encrypt long messages, since the memory and bandwidth overheads would be undesirable. Instead, RSA is typically employed in conjunction with private-key cryptography in

a key exchange protocol. Here, the public-key system communicates a private *session key* between parties, which is subsequently employed in a private-key protocol, without incurring the time and memory overhead that RSA does. In Sec. [XXXIV.B.1](#) we point out that quantum computers are not believed to be able to efficiently crack private-key protocols, but instead effectively reduce their key length by a factor of $1/2$, which is easily counteracted with longer keys to restore security.

Numerous key exchange protocols have been formulated, the best known being the Diffie-Hellman protocol (?), which has been widely adopted on the internet.

E. Digital signatures

Rather than cryptographically ensuring the secrecy of messages, a user may wish to prove their identity when sending a message, such that the recipient can be certain it originated from who it says it does, and accurately conveys what they said. This is achieved using *digital signatures*.

Digital signatures can be easily implemented using the RSA protocol, by reversing the roles of the public and private-keys. Now the public-key can only be used for decrypting a message, and the private-key can only be used for encrypting it. As before, it is computationally hard to infer one from the other. The protocol for sending and verifying a digitally signed message is shown in Alg. [XXVIII](#).

The key point from the security perspective is that the private-key cannot be efficiently inferred from the public-key. So although everyone has access to Alice's public-key, no one is able to counterfeit messages since they cannot create encrypted signatures without access to her private-key – signatures can be easily verified but not created.

Because this protocol is implemented using ordinary RSA, albeit with reversed roles for the key-pair, it shares the same security strengths and vulnerabilities as RSA public-key cryptography (Sec. [XXXIV](#)).

Like RSA cryptography, key servers exist, maintaining databases of people's public-keys and their associated identities.

Because RSA-encrypted messages are long, digital signature protocols typically do not sign the full document directly. Instead they create a message digest of the document using a cryptographic hash function, which is signed using RSA. These hash functions have the property that they cannot be forged or manipulated, providing an accurate summary of a document, but with extremely low memory overhead (256 bits is typical). Hash functions are discussed next in Sec. [XXXIII.F](#).

```

function DigitalSignature(message, keyPair):
    1. Alice prepares a short digest of her
       message using a cryptographic hash
       function, such as SHA256 (Sec. XXXIII.F),
       digest = SHA256(message)

    2. Alice encrypts the digest using her
       private-key. This forms the ‘digital
       signature’,
       signature = RSAEncrypt(digest, privateKey)

    3. Alice transmits the digital signature and
       original message to Bob,
       signedMessage = {message, signature}

    4. Bob hashes the received message,
       hash = SHA256(message)

    5. Bob uses Alice’s public-key to decrypt her
       digital signature,
       decryptedHash =
           RSADecrypt(signature, publicKey)

    6. Bob compares his calculated hash with
       Alice’s decrypted hash for consistency.
       If the two hashes are identical, Bob
       concludes the message was authentic,
       if(decryptedHash = hash)
           return(pass)
       else
           return(fail)

    7. □

```

Algorithm XXVIII: Protocol for digitally signing a message using public-key cryptography (RSA) and a cryptographic hash function (SHA256).

F. Hashing

Hash functions are functions that map a long bit-string of arbitrary length to a short, fixed-size bit-string with quasi-random behaviour,

$$f_{\text{hash}} : \{0, 1\}^n \rightarrow \{0, 1\}^m, \quad (33.5)$$

for an n -bit input and m -bit output hash, where n is variable and m is fixed. They are an example of ‘one-way functions’ that are computationally easy to compute in the forward direction, but extremely hard to invert. That is, given a hash, it is computationally unviable to find input strings that map to that value.

Hash functions have broad applicability throughout computer science, but here we are most interested in *cryptographic hash functions* for use in cryptography, which impose strong conditions on the difficulty of inversion and their quasi-random characteristics. Most notably, the desired characteristics of a cryptographic hash function include:

- The distribution of hashes ought not exhibit any biases, following a uniform distribution with quasi-random behaviour.

- Changing a single bit in the input string ought to flip approximately half the bits of the hash on average.
- It's computationally efficient to calculate a hash from an input.
- It's computationally complex to find an input that hashes to a given value (that is, they are one-way or trapdoor functions).
- The hashes of two very similar inputs ought to yield hashes that are very different.
- Two different inputs are extremely unlikely to hash to the same value.

The standard cryptographic hash function with mainstream adoption is the 256-bit Secure Hashing Algorithm (SHA256), which generates 256-bit hashes. The algorithm is extremely efficient to implement digitally, and exhibits $O(n)$ runtime for input string length n .

Cryptographically, hash functions are useful for creating message digests, which act as a highly condensed checksum of a document that can be utilised in a digital signature (Sec. XXXIII.E).

Note that because the function in general maps longer strings to shorter ones, there are necessarily *collisions* – multiple inputs for a given output. However, for strong cryptographic hash functions their behaviour is sufficiently random that two distinct messages will almost certainly yield completely different hashes (even if the messages are very similar), making it all but impossible for someone to make the claim that Alice said something she did not. This property is extremely important for the security of digital signatures.

XXXIV. ATTACKS ON CLASSICAL CRYPTOGRAPHY

“While I thought that I was learning how to live, I have been learning how to die.” — Leonardo da Vinci.

HAVING introduced the main classes of classical cryptographic protocols, we now turn our attention to their weaknesses and vulnerabilities, both against adversaries with classical or quantum computational resources.

A. Classical attacks

All known classical attacks against any respected classical cryptosystem involve tremendous computational resources. After all, were this not the case the cryptosystem would be considered weak and would never have become widely adopted in the first place!

1. Brute-force

The most obvious approach to cracking a cryptosystem is to systematically try out all possible keys until we find one that correctly decodes the encrypted message. This is also the most naïve approach, and one which is computationally intractable for real-world key lengths. Specifically, for a key length of k bits ($k = 256$ for AES256), there are 2^k possible keys to try, and on average we will wait for 2^{k-1} trials before choosing the right one. Clearly an average waiting time of 2^{255} is not plausible!

2. Cryptanalysis

Far better than waiting the age of the universe for the right key to turn up, is *cryptanalysis*. Here we study patterns between input and output strings from a cipher utilising a particular key. There are many variations on this, but include techniques such as (?):

- Known plaintext attacks (KPA): Through alternate means of espionage, the attacker is able to possess *both* a ciphertext and its associated plaintext. Knowing both the input and output to the encryption algorithm may then reveal information about the key relating them. This technique was important to Alan Turing's successful cracking of the German Enigma encryption protocol during World War II.
- Chosen plaintext attack (CPA): The same as a KPA except that the adversary has the ability to choose what the known plaintext is, a more challenging prospect to orchestrate.
- Linear cryptanalysis: A technique for representing ciphers as linear systems, to which KPA are applied.
- Differential cryptanalysis: We analyse how changes in input bits propagate through the cipher to modulate output bits. Typically this type of technique operates as a CPA.

3. Integer factorisation

In the case of RSA encryption, whose security derives from the believed computational hardness of factorising large integers, the most efficient known classical algorithm for integer factorisation is the general number field sieve (GNFS), with time-complexity,

$$O(\exp(O(1)(\log n)^{\frac{1}{3}}(\log \log n)^{\frac{2}{3}})), \quad (34.1)$$

which scales poorly for large n , keeping in mind that present-day implementations of RSA accommodate key lengths of up to 4,096 bits, as for example is implemented by the widely-used Pretty Good Privacy (PGP) package.

B. Quantum attacks

Having established that classical attacks against strong classical cryptosystems are quite limited by their implausible computational requirements, what if our adversary now has quantum computational resources? Does this change the game?

1. Brute-force

A brute-force attack by a quantum computer does not offer us the exponential improvement attacker Eve might hope for. However, we can gain a quadratic improvement by cleverly exploiting Grover's search algorithm (Sec. XLI).

To do this, we treat the brute-force cracking algorithm as a satisfiability problem, similar to how Grover's is employed to enhance **NP**-complete problems. Specifically, our oracle implements the code's decryption operation, taking as input a qubit string representing the key. After decoding the message with the key, the oracle runs an appropriate test on the decrypted message to determine whether it is a legitimate decoded message. For example, it could run an English language test – a message decoded incorrectly with the wrong key will appear very random and almost certainly won't pass such a test. The oracle tags an element passing this test, which the Grover algorithm searches for, yielding the associated key.

Note that when performing a brute-force attack against a private encryption key, a quadratic speedup effectively halves the key length in terms of algorithmic runtime, since $O(\sqrt{2^k}) = O(2^{k/2})$. Thus, in the quantum era private-key lengths will need to be doubled to maintain an equivalent level of security against brute-force attacks.

This same technique of treating encryption as an oracle within a quantum search algorithm can be utilised to invert hash functions. However, in this case there will necessarily be multiple solutions owing to collisions.

2. Cryptanalysis

In the case of private-key cryptosystems such as AES, no quantum-enhanced cryptanalytic techniques have been described, which offer an exponential enhancement. Thus, modulo doubling key-lengths to counter a Grover attack, these cryptosystems are not regarded as being compromised by quantum computing.

3. Integer factorisation

In the case of RSA public-key cryptography the attack is more direct – with access to a scalable quantum computer, Shor's algorithm can be employed to efficiently

factorise large integers, allowing private-keys to be retrieved from public-keys. Unlike the brute-force attacks, which yielded only a quadratic enhancement, Shor's algorithm is exponentially faster than the classical GNFS, requiring runtime of only,

$$O((\log n)^2 (\log \log n) (\log \log \log n)). \quad (34.2)$$

Compare this with the classical case given in Eq. (34.1).

XXXV. BITCOIN & THE BLOCKCHAIN

ONE of the most exciting new cryptographic applications that has emerged in recent years is the Blockchain, a secure distributed ledger for recording the execution of contracts and transactions. This has enabled cryptocurrencies, most notably Bitcoin, to emerge as a secure digital alternative to conventional fiat currencies.

More recent developments, such as the Ethereum project, develop the distributed ledger further to allow executable code to be committed to the Blockchain, opening the prospects for self-enforcement and -execution of completely arbitrary 'smart contracts', a potential game-changer for the operation of financial and derivative markets.

In the Blockchain protocol, the validity of contracts and transactions is recognised collectively by participants using an encrypted digital ledger. The ledger records the complete history of all Blockchain transactions, which are digitally signed (Sec. XXXIII.F) by network participants using elliptic-curve public-key cryptography. A democratic process ensures that, provided a single user doesn't monopolise the network, recorded transactions are legitimate, recognised collectively and democratically. This is secured by network participants digitally signing off on transactions as they take place.

The Bitcoin protocol builds on top of the Blockchain to create a secure digital cryptocurrency. This requires the introduction of another sub-protocol, *mining*, where units of currency ('coins') are created. The protocol cryptographically ensures that there is an upper-bound on the number of coins that can exist, thereby preventing forgery and an inflationary blowout in the money supply.

The mining process is based upon the computational hardness of inverting (double) SHA256 hashing. A legitimate Bitcoin is defined by a string with a hash satisfying a thoughtfully chosen constraint, specifically one which hashes to a value within some range,

$$\epsilon_{\text{lower}} \leq \text{SHA256}(\text{SHA256}(x_{\text{coin}})) \leq \epsilon_{\text{upper}}. \quad (35.1)$$

This is slightly weaker than inverting hash functions, but is nonetheless a task that can only be approached via brute-force hashing in the forward direction. This associates computational complexity with the mining process, and hence computational integrity of the money supply, whilst

upper bounding the number of unique coins that can exist. This technique is known as ‘proof-of-work’, for associating something of value with proof that certain amounts of computation were invested into achieving it³⁸. This idea was originally borrowed from the Hashcash protocol, where proof-of-work is employed to associate work (and hence monetary value) with sending emails so as to eliminate automated spamming bots.

The two key algorithms for Bitcoin and the Blockchain are therefore hashing and public-key digital signatures. Both of these are subject to enhanced quantum attacks.

Inverse hashing does not have any known quantum algorithm with exponential improvement, however using a Grover search one can achieve a quadratic speedup, using the same idea as for enhancing **NP**-complete problems by treating the hash function as a search oracle. This however does not pose a fundamental security concern as it will speed up the Bitcoin mining process, but does not circumvent the upper-bound on the number of coins that may be in existence. Already classical mining has pushed the Bitcoin money supply close to its asymptotic maximum and there is limited room for additional mining³⁹.

Elliptic-curve public-key cryptography, like RSA, has a known efficient quantum attack via Shor’s algorithm. In the context of implementing digital signatures this implies that an adversary could fraudulently sign off on illegitimate transactions, thereby committing falsified contracts to the Blockchain.

A detailed investigation into the vulnerability of the Blockchain to quantum attacks was performed by (?). However, it is near impossible to predict the future rate of growth in quantum computer technology and hence over what kind of timescale the Blockchain will be compromised. But it is certain that a full compromise is inevitable at some point in the future when scalable, universal quantum computing becomes a reality.

To address this security threat, quantum-resistant hashing and public-key cryptographic protocols will need to be developed. In the former case this can easily be achieved by increasing hash lengths so as to offset the quadratic enhancement offered by Grover’s algorithm. In the latter

case this will require post-quantum public-key cryptosystems (to be discussed in the next section, Sec. XXXVI).

Evidently, the lifespan of existing Blockchain technologies is limited and in the quantum future post-quantum Blockchain algorithms will be required to ensure the survival of cryptocurrencies.

XXXVI. THE END OF CLASSICAL CRYPTOGRAPHY?

THE vulnerability of RSA to attacks by quantum computers raises the question whether this spells the end of classical cryptography and compromises the security of much of the present-day internet.

Thankfully, there are two saving graces. First of all, much research is being carried out into *post-quantum classical cryptography*. That is, public-key cryptosystems based upon trapdoor functions that reside outside of **BQP** and are therefore not efficiently attacked by quantum computers. One such line of research is to construct cryptosystems based upon **NP**-complete problems, such as the McEliece protocol (?). Recall from Fig. 1 that **NP**-complete is strongly believed to reside completely outside of **BQP**. However, while many computer scientists might be comfortable with such a level of security, it is nonetheless based on the unproven conjecture that **NP**-complete and **BQP** do not intersect, i.e $\text{NP} \not\subseteq \text{BQP}$. What would be much more satisfying would be protocols demonstrating information-theoretic security rather than computational security. Here, quantum mechanics can help us – *quantum cryptography*.

XXXVII. QUANTUM CRYPTOGRAPHY

AS quantum physics can compromise some important aspects of classical cryptography, can it perhaps be similarly exploited to make new cryptosystems that are immune even to quantum adversaries? Thankfully the answer is yes... at least some of the time.

A. Quantum key distribution

Aside from quantum computing, a central use for quantum technologies is in cryptography (?). The demand for secure cryptography is now extremely important in the context of electronic commerce and general security of information transmission in the internet age. Electronic currencies such as Bitcoin depend on cryptographic protocols in order to secure the value of assets, assign ownership certificates, and secure the currency against fraud. However, such protocols are based upon the computational complexity of certain mathematical problems (i.e computational security), and are not fundamentally secure in the presence of limitless computational resources, or quantum computers. Therefore, using quantum mechanical

³⁸ In future implementations of Blockchain protocols, the proof-of-work required for a given protocol can be arbitrarily manipulated to accommodate for technological advances in computational power, for example via the adoption of quantum computing. The amount of work required to satisfy the constraint grows as we narrow the range $\epsilon_{\text{upper}} - \epsilon_{\text{lower}}$, providing us with much leverage to manipulate the complexity of the proof-of-work, and hence the rate of growth in the money supply, equivalently the rate of inflation.

³⁹ Bitcoin mining has gained so much traction and become so competitive that desktop PCs have become uneconomical for mining. Instead miners are resorting to utilising specialised hardware in the form of CUDA cores, FPGAs and ASICs (or by secretly using the company supercomputer while the boss isn’t looking).

protocols based on physical principles (i.e information-theoretic security) rather than computational limitations, are favourable for future-proofing ourselves.

Quantum key distribution (QKD) protocols facilitate shared, secret randomness, where any intercept-resend (or man-in-the-middle) attack may be detected and rejected, guaranteed by the laws of quantum physics (specifically the Heisenberg uncertainty principle and no-cloning theorem). This shared, secret randomness may subsequently be employed in a one-time pad cipher, presenting us with true information-theoretic security.

The central notion to QKD protocols, in their numerous manifestations, is that measurement of quantum states invokes a wave-function collapse. When measuring a state in a basis for which that state is not an eigenstate, this necessarily changes the state. QKD relies on this simple result from quantum mechanics to reveal any eavesdropper performing an intercept-resend attack via the changes to transmitted quantum states that this would induce.

QKD is a relatively mature technology with already several commercial systems being available off-the-shelf⁴⁰ and initial space-based implementations have been successfully demonstrated (?).

It's easy to see the utility of quantum networks in enabling commodity deployment of QKD – users desire to communicate photons across long-range ad hoc networks, with low loss and dephasing. A global quantum internet would allow quantum cryptography to truly supersede classical cryptography, bypassing the vulnerabilities faced by classical cryptography in the era of quantum computing.

1. BB84 protocol

The first described QKD scheme was the *BB84* (?) protocol, which exploits the fact that states encoded in the \hat{Z} -basis but measured in the \hat{X} -basis (and vice versa) collapse randomly, yielding completely random measurement outcomes, whereas states measured in the same basis in which they were encoded always correctly communicate a single bit of information.

Implemented photonically, BB84 requires only the transmission of a sequence of single photons, polarisation-encoded with random data.

The BB84 protocol is described in detail in Alg. XXIX in the context of polarisation-encoded photons, which is the most natural (but not only) setting for this protocol. An example evolution of the protocol is illustrated in Fig. 111.

To understand the secrecy of the protocol as described in Alg. XXIX, suppose an eavesdropper, Eve, were to

function BB84():

1. Alice chooses a random bit, 0 or 1.
2. Alice randomly chooses a basis, \hat{X} or \hat{Z} .
3. Depending on the choice of basis, she encodes her bit into the polarisation of a single photon as:

$$\begin{aligned} |0\rangle_Z &\equiv |H\rangle, \\ |1\rangle_Z &\equiv |V\rangle, \end{aligned} \quad (37.1)$$

or,

$$\begin{aligned} |0\rangle_X &\equiv \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle), \\ |1\rangle_X &\equiv \frac{1}{\sqrt{2}}(|H\rangle - |V\rangle). \end{aligned} \quad (37.2)$$

4. Encoding into the randomly chosen basis, she transmits the randomly chosen bit to Bob.
5. She does not announce the choice of bit or basis.
6. Bob measures the bit in a randomly chosen basis, \hat{X} or \hat{Z} .
7. The above is repeated many times.
8. Upon receipt of all qubits, Alice (publicly) announces the basis used for encoding each bit sent.
9. Qubits where Bob measured in the opposite basis to which Alice encoded are discarded, as they will be decorrelated from Alice.
10. The remaining measurement outcomes are guaranteed to yield identical bits between Alice and Bob.
11. Remaining is roughly half as many bits as were sent, which are random, but guaranteed to be identical between Alice and Bob.
12. Alice and Bob sacrifice some of their bits by publicly communicating them to check for consistency. This rules out intercept-resend attacks.
13. Privacy amplification may be used to distill the partially compromised key into a shorter but more secret one.
14. \square

Algorithm XXIX: BB84 QKD protocol using polarisation-encoded photons. Upon completion of the protocol, Alice and Bob share a random bit-string for use in a one-time pad cipher, yielding perfect information-theoretic security.

perform an intercept-resend attack on the channel between Alice and Bob. At that stage in the protocol Alice had not yet announced her choice of encoding bases, and Eve will not know the bases in which to measure states without randomly collapsing them onto values inconsis-

⁴⁰ Examples of companies selling off-the-shelf QKD hardware include MagiQ and ID Quantique.

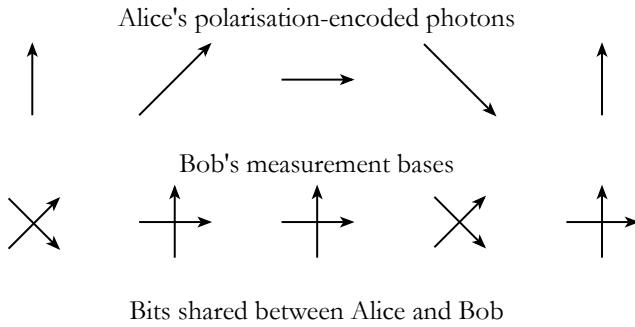


Figure 111: Example execution of the BB84 protocol for securely sharing random bit-strings between Alice and Bob as per Alg. XXIX. At the conclusion of the protocol, some bits are discarded (red ‘X’), with those remaining guaranteed to be secret between the two parties.

tent with Alice’s encoding. Thus, by sacrificing some of their shared bits, via openly communicating them to one another for comparison, such an attack will be detected with asymptotically high probability. Now Alice and Bob have great confidence that they have a shared, secret, random bit-string, which may subsequently be employed in a one-time pad with perfect secrecy.

The BB84 protocol has no measurement timing, mode-matching or interferometric stability requirements, making it a very robust protocol, readily achievable with present-day photonics technology. The scheme has been adapted to physical architectures beyond just polarisation-encoded photons, such as CV encodings (see Sec. XXXVII.A.4).

2. Privacy amplification

When Alice and Bob sacrifice and compare a randomly chosen subset of their key bits to detect eavesdroppers, they also need to accept the inescapable fact that their qubits propagated through imperfect channels and were subject to noise en route. This has the same effect as an eavesdropper – it corrupts some of the bits – and it’s impossible to distinguish which took place, a noisy channel (which is ok) or an eavesdropper (which is not).

Because the channel was necessarily noisy, Alice and Bob *must* tolerate some number of corrupted bits. But if the corruption came from Eve rather than the noisy channel they would effectively be tolerating her knowing some of the key. We don’t want her to know *any* of the key!

Privacy amplification is a mathematical technique based on hashing algorithms for taking a shared key with a number of unknown compromised bits and distilling it to a shorter key of which Eve has almost zero knowledge.

Specifically, if we know that Eve has compromised t of our n shared random bits, privacy amplification allows

us to distill a new key from the compromised one of approximately length $n - t$ over which Eve knows almost nothing.

This is an information-theoretic security result, not a computational security one, thereby rescuing the perfect security of the BB84 QKD protocol.

3. E91 protocol

E91 is slightly different to BB84. Here Alice and Bob share an entangled Bell pair provided by a central authority. Then both Alice and Bob measure their qubits in random bases. As with BB84, after measuring all qubits, they compare their choices of random bases. When they coincide, they have a shared, random bit. When they don’t, they discard their result. From here the remainder of the protocol is the same as for BB84. The protocol is summarised in Alg. XXX.

Fix this up. Discuss using Bell violation to prove security.

```
function E91(|Φ+):
```

1. A central server shares a Bell pair between Alice and Bob,

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B). \quad (37.3)$$

2. Alice randomly measures her qubit in either the \hat{X} or \hat{Z} basis.
3. Bob randomly measures his qubit in either the \hat{X} or \hat{Z} basis.
4. Alice and Bob share what their measurement bases were (classically and unencrypted).
5. When Alice and Bob’s bases were consistent they store the measurement outcomes as a shared random bit.
6. Alice and Bob sacrifice some of their bits by publicly communicating them to check for consistency. This rules out intercept-resend attacks.
7. Privacy amplification may be used to distill the partially compromised key into a shorter but more secret one.
8. □

Algorithm XXX: E91 QKD protocol using polarisation-encoded photonic Bell pairs. Upon completion of the protocol, Alice and Bob share a random bit-string for use in a one-time pad cipher.

Like BB84, E91 has no mode-matching or interferometric stability requirements, and Alice and Bob both only require single-photon detection. Unlike BB84, however, E91 requires a central authority that is able to prepare entanglement on-demand as a resource.

An advantage of E91 over BB84 is that it does not require a direct quantum communications link between Alice and Bob. The protocol could be mediated from above by a Bell pair-producing satellite within line-of-sight of both Alice and Bob.

4. Continuous-variable protocols

— by Zixin Huang.

Like quantum computing, QKD protocols may be adapted to the CV domain also. Alg. XXXI describes a simple such scheme based on encoding in phase-space, where the basis states are coherent states of different amplitudes and phases. The goal is the same as BB84 – to securely share a random bit-string for use in a one-time-pad.

Note that this protocol provides information-theoretic security, as per BB84, despite the fact that coherent states are non-orthogonal, forming an over-complete basis in phase-space.

Conceptually, the operation of the CV QKD protocol is virtually identical to photonic BB84, differing only in that now the different choices of encodings correspond to phase-space transformations. Like BB84, if Eve were to perform an intercept-resend attack she would probabilistically re-encode in the wrong quadrature, thereby revealing herself to Alice and Bob, who could then terminate and start over.

```
function CV_QKD():
    1. Alice chooses two Gaussian-distributed
       random numbers with mean zero,
        $x_A = \mathcal{N}(0, V_{\text{mod}})$ ,
        $p_A = \mathcal{N}(0, V_{\text{mod}})$ , (37.4)
       where  $V_{\text{mod}}$  is the modulation variance.
    2. Alice prepares the coherent state,
        $|\alpha\rangle = |x_A + ip_A\rangle$ . (37.5)
    3. Alice transmits  $|\alpha\rangle$  to Bob.
    4. Bob randomly measures either  $\hat{x}$  or  $\hat{p}$  using
       homodyne detection.
    5. Alice and Bob use classical communication
       to determine for which transmissions
       their preparation and measurement were
       consistent.
    6. The remainder of the protocol proceeds as
       per BB84.
    7.  $\square$ 
```

Algorithm XXXI: CV QKD protocol using coherent states, encoded in the quadrature basis.

5. Security

Importantly, unlike classical cryptographic protocols, QKD makes no assumptions about the computational complexity of inverting encoding algorithms or trapdoor functions. The protocols are information-theoretically secure, and therefore no physically realisable computer, even a quantum computer, can compromise them. Thus, usual cryptanalytic techniques, like linear and differential cryptanalysis (?), or the ability to factor large numbers, that are employed to attack other encryption protocols, do not compromise QKD.

However, this is not to say that QKD is actually perfectly secure in real-life. Recent history has demonstrated that this is certainly not the case, with many attacks against various quantum cryptographic protocols being described and successfully demonstrated. The reason for this schism between theory and experiment is that no experiment ever *perfectly* mimics the theoretical proposal it is trying to implement. Laboratory components might be imprecise in an unfortunate way, opening up avenues for attack, or they might perform unwanted additional actions that leak information to Eve. The prospects for such so-called ‘side-channel attacks’ must be carefully considered and satisfactorily addressed.

The best known attack against photonically implemented BB84 is the ‘photon-number splitting attack’. This attack targets implementations where Alice’s photon source does not produce perfect single-photon states, but may have some amplitude of higher photon-number. Weak coherent states or SPDC states exhibit this property. The attack is very simple. Eve simply performs a man-in-the-middle attack, but not of an intercept-resend variety. Rather than intercepting the entire channel, she inserts a low reflectivity beamsplitter and measures only the reflected mode, the other following its desired trajectory to Bob. Now there is a chance that Eve can extract just one of the multiple photons in the signal, such that Bob still receives a photon. Eve holds the split-off signal in memory until the classical communication of encoding bases, at which point she measures all her split signals in the correct basis, thereby recovering the associated secret-key bit.

This trivial attack vector clearly demonstrates the importance of well-considered engineering decisions when physically implementing QKD. No piece of hardware is ever 100% to specification!

6. Public-key cryptography

The BB84 protocol is used exclusively for private-key cryptography. For many applications (notably digital signatures and easy key exchange with unidentified parties), public-key cryptosystems would be highly desirable.

Are there any viable public-key quantum protocols

that could fill the vacancy of the soon-to-be-compromised RSA? Unfortunately the answer is ‘not yet’. As appealing as it would be, and despite many highly intelligent people putting their minds to it, to-date no one has presented a viable public-key quantum cryptosystem.

This is problematic since when quantum computing becomes a reality it will immediately compromise the classical public-key cryptosystems we all rely on on a daily basis, and it would be highly desirable for a quantum replacement to be available to fill its shoes.

B. Quantum Enigma machines

While the BB84 and other related QKD protocols are perfectly secure, they suffer the major drawback that because they are based upon the one-time pad, the number of successfully communicated key-bits must equal the message length and cannot be reused (not even once). This means that for real-time or high-bandwidth applications, the quantum communications channel must have similarly high bandwidth to be applicable.

What would be more useful would be a quantum equivalent of private-key cryptography, whereby a short key can be used to lock a much longer message.

This led to the proposal for *quantum Enigma machines*⁴¹ (?), based on the phenomenon of quantum data locking (?). The protocol is shown in Fig. 112.

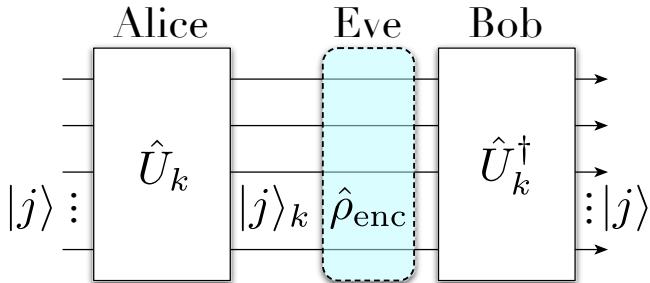


Figure 112: Schematic of the quantum Enigma machine protocol for quantum private-key cryptography. $|j\rangle$ is the message, k is the key, and \hat{U}_k is the encryption operation associated with key k , chosen randomly from the Haar measure. The encrypted message is $|j\rangle_k$. An eavesdropper intercepting the channel without knowledge of the key observes the state $\hat{\rho}_{\text{enc}}$.

Quantum data locking is a uniquely quantum phenomenon, and presents one of the most striking differences between classical and quantum information theory (?). Quantum data locking occurs when the accessible information about a classical message encoded into a quantum

state decreases by a much larger amount when “locked” with a much smaller key.

Here Alice and Bob share a short secret-key of length K , and wish to communicate a message M of length n -bits. The secret-key is assumed to be unconditionally secure, and can be established via conventional QKD. First off, they agree upon a set of K Haar-random unitaries (but it is sufficient to consider 2-designs) , $\{\hat{U}_k\}$, associating one with each of the possible K keys. The set of $\{\hat{U}_k\}$ unitaries need not be a secret.

Alice then encodes her message into the state $|j\rangle$. Her encryption operation is to apply \hat{U}_k to this state according to Alice and Bob’s shared secret-key.

$$|j\rangle_k = \hat{U}_k |j\rangle, \quad (37.6)$$

which Bob is easily able to decrypt using the inverse unitary,

$$\hat{U}_k^\dagger |j\rangle_k = |j\rangle. \quad (37.7)$$

To characterise the security of the scheme, we first note that in the absence of knowing the key or message, and assuming all j and k are equally likely, Eve observes the mixed state,

$$\hat{\rho}_{\text{enc}} = \frac{1}{2^n} \sum_{j=1}^{2^n} |j\rangle_A \langle j| \otimes \frac{1}{K} \sum_{k=1}^K U_k |\psi_j\rangle_E \langle \psi_j| U_k^\dagger. \quad (37.8)$$

The security can now be quantified in terms of the accessible information ⁴² between this state and the plaintext state, which can be upper-bounded as (?),

$$I_c \leq n + \frac{1}{2^n} \max_{|\phi\rangle} \sum_{j,k} |\langle \phi | j \rangle_k|^2 \log |\langle \phi | j \rangle_k|^2. \quad (37.9)$$

It was shown that this quantity can be made arbitrarily small with key-size scaling as,

$$m = O(\epsilon \log n), \quad (37.10)$$

for an ϵ that can be made arbitrarily small, given $\log K \sim \log(1/\epsilon)$, therefore the key size is exponentially smaller than the length of the message. Thus, we have an encryption scheme that requires very frugal requirements in key-size versus message length. It is further showed that such a scheme can be made robust against noise and loss.

Note that the security of the scheme is information-theoretic, and does not make any assumptions about computational security (except we require the assumption that Eve has finite-time quantum memory (?) in order to preserve composable security required for QKD (?)).

⁴¹ The Enigma machine was the classical cryptosystem employed by the Germans during World War 2 for military and intelligence communication.

⁴² The accessible information is the maximum of the mutual information between Alice’s input states and measurements performed on the encoded state by Eve.

Thus, this represents a strong form of quantum private-key cryptography, requiring only very short keys compared to the message length.

Since then, quantum data locking has found its application for Boson Sampling (?). However, the scheme is very challenging to implement on a large scale over long distances. Consider an optical implementation, where the message is photonically encoded. Now the scheme represents a complex, multi-mode, generalised Mach-Zehnder interferometer, meaning that the channel between Alice and Bob, who might be far apart, must be interferometrically stabilised (Sec. XXII) on the order of the photons' wavelength (hundreds of nanometers for optical frequencies), which is extremely challenging over long distances.

C. Hybrid quantum/classical cryptography

As discussed, the RSA public-key cryptosystem is vulnerable to an efficient quantum attack, whereas private-key schemes like AES are not (believed to be). Thus, combining QKD schemes with private-key classical schemes does not compromise security in the quantum era.

Why would we combine quantum and classical encryption techniques when QKD is already provably secure, whereas the classical schemes are not?

In the near future, as QKD schemes begin their roll-out in space and on Earth, random bits from the QKD implementation will be very expensive and exhibit low bandwidth. Suppose we wanted to securely videoconference across the globe. For just a single user this would require megabits per second of shared random bits, which will quickly saturate the capacity of overhead quantum satellites.

Instead, let us use the QKD system to securely share just a 256-bit private session key between two users. This is subsequently employed for AES256 encryption that operates entirely over the classical network, which we regard as extremely cheap and high-bandwidth. Importantly, unlike one-time pad implementations, this session key may be reused. Now we have a hybrid system which is not quantum-compromised, but which overcomes the cost and bandwidth issues associated with emerging QKD networks.

While such a hybrid scheme is not information-theoretically secure (AES is not proven to be quantum-safe), the computational security assumptions are far stronger than for say RSA, since there are no known efficient quantum attacks against strong private-key schemes.

D. Quantum anonymous broadcasting

The previously described protocols all focussed on preserving the secrecy of messages. Alternately, it may not be the message that is sensitive, but rather the identity

of the person who says it. *Anonymous broadcasting* is a protocol for achieving this.

Consider the following scenario. A group of users share a classical broadcast channel that anyone is able to transmit to, and everybody is able to listen to unencrypted. But it is of importance that the identity of whoever broadcasts to the channel must be kept secret from all users. (?) described a scheme for achieving this quantum mechanically using shared GHZ states – *quantum anonymous broadcasting* (QAB).

Let there be a (trusted⁴³) server that distributes GHZ states (of arbitrary numbers of qubits) to a group of users, one qubit per user. This can be prepared as described in, for example, Sec. XXIII.E. Now if every user measures in the $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ basis the joint *parity* (i.e whether an even or odd number of +'s were measured) is guaranteed to be even. For example, all users might measure $|+\rangle\langle +|$, or exactly 2, but never exactly 1 or 3.

On the other hand, if a \hat{Z} gate were applied to any one qubit, this would flip the parity outcome. Note that a GHZ transforms according to,

$$\hat{Z}_i \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n}) \rightarrow \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} - |1\rangle^{\otimes n}) \quad \forall i, \quad (37.11)$$

for any qubit i . This invariance in the location of the \hat{Z} gate is the basis for the anonymity of the protocol. If a user wishes to broadcast '0' he does nothing, whereas if he wishes to broadcast '1' he applies a \hat{Z} gate to his local qubit.

Finally, all users measure their qubits in the \pm -basis and publicly (without encryption) broadcast their measurement outcomes. All users now see all other users' measurement outcomes and are able to calculate the collective parity of the measurements. Now if the parity is even, the speaker must have said '0', whereas if it is odd he must have said '1'. The protocol is shown in Fig. 113 and described in detail in Alg. XXXII.

Note that the scheme can be slightly simplified by rather than the speaker applying the \hat{Z} to his qubit, upon announcing his measurement outcome he instead simply lies about his outcome and flips it. This follows simply because a \hat{Z} gate prior to a \pm measurement bit-flips the classical measurement outcome, $\hat{Z}|\pm\rangle = |\mp\rangle$.

There are no constraints on time-ordering of the measurements, nor, much like BB84, are there any interferometric stability requirements (not including the GHZ preparation stage), making this protocol very experimentally practical and robust over long distances.

⁴³ Note that if the server is not trusted, he could easily conspire to reveal people's identities by distributing $|+\rangle^{\otimes n}$ states instead of GHZ states.

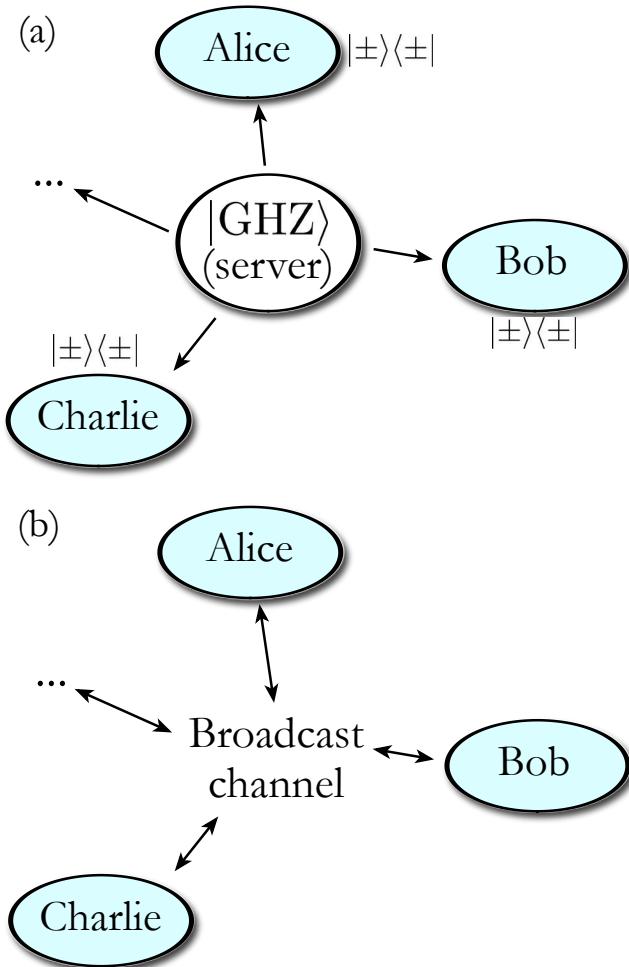


Figure 113: Protocol for quantum anonymous broadcasting. (a) A central trusted server prepares GHZ states and distributes them amongst a group of users, one qubit per user. All users measure in the \pm -basis. (b) All users classically broadcast their measurement outcomes yielding shared random parity. During broadcast, the broadcaster lies about his measurement outcome to flip the joint parity if he wishes to transmit ‘1’, or tells the truth to transmit ‘0’. The joint parity encodes the message of the anonymous user, which all listeners are able to recover. Importantly, only one user may broadcast at a time, otherwise the recovered message will be given by the XOR of all the simultaneously broadcast messages.

Because of the time invariance in the measurements, distribution and measurement of the GHZ states can be performed well in advance of the actual message broadcast. This allows us to treat ‘shared parity’ as a fundamental resource (Sec. XXVIII) for the QAB cryptoprotocol.

Since the parity-sharing can be isolated from the broadcasting stage it is unimportant if the GHZ source is non-deterministic or the channels for distributing it lossy. We can instead simply repeat GHZ distribution over and over at high repetition rate, post-selecting upon measurement outcomes where all users signal that they successfully received and measured their photons.

```

function QuantumAnonymousBroadcasting(message, speaker):
    1.  $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$ 
    2.  $|\psi\rangle \rightarrow (\hat{Z}_{\text{speaker}})^{\text{message}}|\psi\rangle$ 
    3. for(i in users) {
        4. outcomei = measureInXBase(|ψi⟩)
    5. }
    6. parity =  $\sum_i \text{outcome}_i \text{ (mod 2)}$ 
    7. return(parity)
    8. □

```

Algorithm XXXII: Protocol for quantum anonymous broadcasting. The GHZ state is distributed in advance, one qubit per user. The measurement outcomes are classically broadcast without encryption. The final parity of the classical measurements reflects the message bit without identifying the speaker.

For these reasons, this scheme lends itself readily to photonic implementation, provided a reliable GHZ preparation circuit. The scheme has since been ported to operate on distributed toric codes to facilitate error correction of the distributed GHZ states (?).

E. Quantum voting

The security of voting systems, and the anonymity of their voters are pressing issues in the modern free-world, and there have been countless high-profile instances of voting systems being compromised nefariously.

Based on similar ideas to quantum anonymous broadcasting is quantum voting, whereby a group of parties can anonymously vote such that no party, including the tallyman, is able to learn any individual voter’s vote, but at the conclusion all are able to see the collective outcome of the vote.

There are a multitude of different models for voting, and a number of quantum implementations for them have been described. The two most well-known are:

- Binary voting: whereby each party votes ‘yes’ or ‘no’.
- Anonymous surveys: whereby each party votes a number and we wish to determine the sum of all the votes.

Fig. 114 and Alg. XXXIII describe a quantum implementation for anonymous surveys, using the protocol described by (?).

Conceptually, the scheme is similar to quantum anonymous broadcasting in that it hides votes in phases within an entangled state, which are not accessible to individual parties, but are rather a global property of the state. The scheme relies on the preparation and distribution of a particular entangled state as a resource. Unfortunately

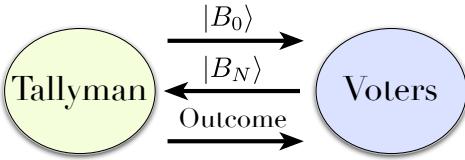


Figure 114: Protocol for quantum voting via anonymous surveying. The tallyman prepares the entangled state $|B_0\rangle$, half of which is shared with the voters. The voters each perform local phase-shift operations on their half of the state, as per Alg. XXXIII. The phase-transformed state is then returned to the tallyman, who is able to extract the phase and hence the cumulative vote.

this particular state is not one which is known how to be trivially prepared optically, and would therefore lend itself well to the outsourcing of preparation and distribution to a capable host via the quantum internet.

F. Quantum secret sharing

To do

function QuantumVoting($|B_0\rangle$):

1. Prepare ballot state for N voters,

$$|B_0\rangle = \frac{1}{\sqrt{N+1}} \sum_{n=0}^N |N-n\rangle_T |n\rangle_V, \quad (37.12)$$

in the photon-number basis, where T and V represent the tallyman and the voters.

2. Each voter j applies their vote operator,

$$\hat{v}_j = e^{i\pi \hat{n}_V \frac{\nu_j}{N+1}}, \quad (37.13)$$

where \hat{n}_V is the photon-number operator, and ν_j is the vote cast by j .

3. Following the m th vote,

$$|B_m\rangle = \frac{1}{\sqrt{N+1}} \sum_{n=0}^N e^{in\Delta_m} |N-n\rangle_T |n\rangle_V, \quad (37.14)$$

where,

$$\Delta_m = \frac{2\pi}{N+1} \sum_{j=0}^m \nu_j, \quad (37.15)$$

is a scaled sum of the votes.

4. The state observed by the T is,

$$\text{tr}_V(|B_m\rangle\langle B_m|) = \frac{1}{N+1} \sum_{n=0}^N |n\rangle_T \langle n|_T, \quad (37.16)$$

which contains no voting information, similarly for V .

5. After all N voters have voted, the voter state V is transferred to the tallyman T .

6. Define the tally operator as,

$$\hat{T} = \sum_{n=0}^N n |T_n\rangle\langle T_n|, \quad (37.17)$$

where,

$$|T_n\rangle = \frac{1}{\sqrt{N+1}} \sum_{j=0}^N e^{inj \frac{2\pi}{N+1}} |N-j\rangle |j\rangle. \quad (37.18)$$

7. The tallyman finds the expectation value of the tally operator,

$$r = \langle B_N | \hat{T} | B_N \rangle = \sum_{j=0}^N \nu_j, \quad (37.19)$$

yielding the sum of all the votes.

8. return(r)

9. \square

Algorithm XXXIII: Protocol for performing secure quantum anonymous surveying.

XXXVIII. ATTACKS ON QUANTUM CRYPTOGRAPHY

— by Zixin Huang.

FOR a QKD system, information-theoretic security is achieved only when security against collective, coherent attacks is proven. We mustn't make assumptions about the limitations of our adversaries. For a more comprehensive review on quantum cryptography, see Ref. (?).

Hacking attacks in this context exploit weaknesses in the physical implementation, rather than weakness of the theory – so-called ‘side-channel attacks’. Some examples of weaknesses which allow zero-error attacks include:

- **Losses:** Genuinely lossy channels or components are indistinguishable from ideal ones where some of the signal has been tapped off by Eve. Therefore, we must always assume the worst, that whatever is lost from our system is in the hands of Eve.
- **Imperfect components:** Our physical implementation might just not be operating strictly according to the theory.
- **Correlations:** Mutual information between signals in our system may leak information from the secure system to the environment, where Eve might be waiting patiently.

When considering the security of noisy channels, one must assume that all noise is due to manipulation by an eavesdropper – the worst-case scenario. An attempted attack is considered successful if it can be proven that the eavesdropper can gain a non-negligible (i.e not exponentially small) amount of mutual information with the final secret-key established between Alice and Bob, without alerting them. Some of the best-known attacks follow:

A. Hacking discrete-variable protocols

1. Beamsplitter & photon-number-splitting attacks

The security proofs for many discrete variable protocols assumes that Alice's signals consist of single photons. However, true single-photon sources are not yet widely available, and QKD systems often make use of strongly-attenuated laser pulses, and there is some probability of the source emitting multiple photons. This fact can be exploited by Eve, who employs the photon number splitting (PNS) attack (?): Eve can perform a non-demolition measurement to determine the number of photons in the signal, she steals the excess photons and sends the rest to Bob. She stores these in her quantum memory and wait until the classical communication between Alice and Bob, hence finding out Alice's preparation basis.

A beamsplitting (BS) attack translates the fact that any signal lost over a channel is acquired by Eve. Here, Eve

induces losses in the communications channel by putting a beamsplitter outside Alice's device, then forwards the remaining photons to Bob. The BS attack does not modify the optical mode that Bob receives: it's therefore always possible for lossy channels, and does not introduce any errors.

A method used to counter the PNS attack is the decoy-state method (??). In the decoy-state protocol, Alice randomly replaces some of her signal states with multi-photon pulses from a decoy source. Eve cannot distinguish between decoy pulses from the encoding signals, and can only act identically on both. In the post-process stage, Alice public announces which states were the decoy pulses. The trusted parties can then characterise the action of the channel on the multiple-photon pulses and detect the presence of a PNS attack.

2. Trojan horse and flash-back attacks

Another family of hacking that can be used against discrete-variable QKD is the Trojan horse attacks. These attack involve Eve probing the settings of Alice and Bob by sending light into their devices and collecting the reflected signal. The first of this kind of attack actually came for free for the eavesdropper (?): it was discovered that some photon-counters emit light when a photon is detected (?). If the emitted light carries correlated information about which detector was triggered, it must be prevented from leaking outside the secure space and becoming accessible to Eve.

In general, Eve probes into the optical channel that Alice and Bob use to communicate. She send her own states into Alice's system, which will reflect off the same apparatus Alice uses to encode her signal. Eve's states can be imprinted upon some information about the encoding used by Alice, when Eve measures them. She can then use the result of this measurement, combined with some operation on Alice's signals to make a best estimate of the quantum state that Alice sent to Bob, thus giving her some non-negligible mutual information with the key (?).

3. Detector attacks

The faked-state attack is based on the weak-laser implementation of BB84. Here, Eve manipulates Bob's detectors to force him to measure in the same basis. It exploits the fact that the detectors may have a dead-time, and the eavesdropper can trigger the detector whenever she chooses. It follows that Bob's detection outcomes are controlled by the eavesdropper.

Eve can also go beyond detector blinding. She can send in a powerful laser pulse to optically damage components in the QKD system and permanently change its char-

acteristics (?). If the new characteristics then assist the eavesdropper in an attack without Alice or Bob being notified, the security of the QKD system would be severely compromised.

A more detailed discussion on attacks on physical implementation can be found in (?).

B. Attacks on continuous-variable protocols

1. Attacks on the local oscillator

In order to measure Alice's signal states, Bob needs to carry out quadrature measurements, which are defined with respect to Alice's local oscillator. Since it is difficult to maintain coherence between Alice's and Bob's local oscillators, often the Alice sends the local oscillators through the channel together with the signal states. This leaves open some side-channels which Eve can exploit. For example, she can reduce the error which the trusted parties would expect (e.g. of an intercept and resend attack), if she replaces the signal and local oscillators with squeezed states. Other attacks of the local oscillators include Eve exploiting the wavelength dependence of the beam splitters, shape of the local oscillators pulse, phase noises and the non-linearity photo-detectors.

2. Saturation attacks on the detector

The security proofs of CV-QKD assume a linear relationship between the quadratures of the state and the measurements, but in reality, homodyne detection have a finite range of linearity. The detectors can also be saturated. By causing Bob's measurement results to overlap with the saturation region, Eve can artificially reduce the trusted parties' error estimation.

Proposed countermeasures include using a Gaussian post-selection filter to ensure that the measurement results used for key generation fall within the linear region of the detector, and to use random attenuation of Bob's signal to monitor whether the measurements are linearly related to the inputs (?)

3. Trojan horse attacks

Continuous variable protocols are also vulnerable to trojan horse attacks. By sending Trojan states into Alice's encoder, Eve can gain information on how Alice's states have been modulated. Active monitoring of incoming light is suggested as a countermeasure.

C. Quantum digital signatures

— by Zixin Huang.

As discussed previously, a classical digital signature has three main tasks. It ensures that the message

- was created by the claimed sender.
- has not been altered.
- is non-repudiable: the sender cannot deny having sent this message.

A digital signature is a vital tool for a huge range of modern applications, ranging from software distribution, financial transaction, emails, cryptocurrencies and voting etc (?).

A quantum digital signature (QDS) scheme involves one sender and potentially many receivers. It consists of three stages, each consists of a corresponding algorithm.

1. Key generation: a private key is obtained by the sender, and public keys are delivered to the receivers.
2. Signature: the sender chooses a message M and uses the private key to generate a signature, $\sigma = \sigma(M)$. She sends the pair $(M, \sigma(M))$ to the desired receiver(s)
3. Verification: a receiver receives the message and signature pair $(M, \sigma(M))$ and the public key, he/she checks whether to accept the message as having originated from the claimed sender.

After the key generation phase, it is important that the actions of the involved parties are determined by this point. That is, they decide to accept or reject the signature without further classical or quantum communication.

Ensuring all the parties receive the same and correct quantum public key is a non-trivial task, given that quantum states need to be distributed. Instead of using the cryptographic primitives from classical digital signatures, when dealing with QDS schemes, one often aims to ensure that the QDS scheme has the following properties:

1. Unforgeability: a dishonest party cannot send a message pretending to be someone else.
2. Transferability: if a receiver receives a signature, they should be confident that any other receiver should also accept the signature.
3. Non-repudiation: the sender cannot deny having sent the message.

There are various protocols that fall under the category of QDS. The protocols we describe in this section deals with the task of signing a classical message using tools from quantum information theory.

1. The (classical) Lamport one-time signature

QDS schemes are inspired Lamport's one-time signatures (?). Here we rely on a classical one-way function f . Given x , it is easy to evaluate $f(x)$, but if $f(x)$ is given, it is hard to invert $f(x)$ to find x . The algorithm is as follows

1. Key generation: Alice chooses two random inputs x_0 and x_1 , and evaluates $f(x_0)$, $f(x_1)$. She publicly broadcasts the pair $(0, f(x_0))$, $(1, f(x_1))$ whilst keeping x_0, x_1 secret.
2. Signing: Alice sends the message b along with her stored corresponding secret key x_b .
3. Verification: the receivers evaluate $f(x_b)$ and check if this agrees with the public key in order to choose to accept or reject.

The security of such a scheme comes from the fact that $f(x)$ is classically difficult to invert, and therefore an adversary who has access to only the public key cannot find the secret key to forge a signature. However, any scheme that is based on computational-complexity arguments cannot offer information-theoretic security, and schemes based purely on the physics of quantum mechanics are much more desirable.

2. The Gottesman-Chuang QDS

In 2001, Gottesman and Chuang (?) proposed the first QDS protocol. The idea is to use the non-orthogonality of state to realize a "quantum one-way function", where the difficulty to invert the function is based on the laws of physics rather than computational complexity.

If we have a quantum state $|f(x)\rangle$ based on the classical description $f(x)$, no one should be able to characterize $|f(x)\rangle$ with certainty, unless they already know $f(x)$. The function f does not need to be difficult to invert, since the computational complexity in the classical protocol is replaced by non-orthogonality.

1. A function f is chosen and is made public, it takes the input x and generates $f(x)$ that describes the quantum state.
2. Key generation: for the private key, Alice chooses a pair of bit-string $\{x_0^i, x_1^i\}$ where $1 \leq i \leq L$. The bit-string x_0 (x_1) will be used to sign the 0's (1's) in the message. L is determined by the security level required.
3. For the public key, Alice generates multiple copies of $\{|x_0^i\rangle, |x_1^i\rangle\}$. She distributes to each receiver the corresponding states. To check that the public keys are the same, each of the receivers interacts by performing SWAP tests. A SWAP test provides an affirmative answer without disturbing the states.

4. Signing: to sign a message, Alice chooses the message value b , and sends (b, x_b^i) to the receiver. This can be done classically.

5. Verification: the receiver uses x_b^i and generates the corresponding quantum state $|f(x_b^i)\rangle$ and checks whether this is consistent with the stored public key.

Whilst the Gottesman-Chuang QDS protocol is intuitive, it is also highly impractical due to the experimental requirements. Firstly, the protocol requires for long-storage quantum memories, since the signature and the verification can be separated by long periods of time – this requirement is one of the major bottle-necks for building a scalable quantum computer. Secondly, multiple copies of the public key are needed, and each receiving party need to perform the SWAP test, the latter involves more quantum communication and ancillary qubits

3. A practical QDS protocol

Since the development of the Gottesman-Chuang QDS protocol, the field has undergone major developments and lifted all the restricting requirements of the protocol.

The storage of the quantum state until verification renders the scheme impractical. This requirement can be removed by replacing the quantum public key with classical verification keys, which is now different for each receiver (?). After removing the quantum memory requirement, the other experimentally challenging task is ensuring that the same quantum public key was sent to different receivers. This step can be bypassed by adding an extra step, replacing the SWAP test with a "symmetrization" (?). This ensures that even if the states distributed by Alice are not identical, the classical verification keys shared by the receivers will be "symmetric" (?).

We explain the symmetrization protocol now: Alice generates two sets of BB84 states and sends them to Bob and Charlie respectively. Bob (Charlie) either measures it or forwards it to Charlie (Bob), and he similarly measures the states he receives from Charlie (Bob). Depending on the outcome, Bob knows for certainty which state Alice did *not* send. For example, if the measurement outcome is $|0\rangle$, he knows for sure Alice did not send $|1\rangle$. Bob stores the sequence of eliminated state, and whether it was received directly from Alice or Charlie. This classical information forms Bob's and Charlie's eliminated signatures, and will be the verification keys.

Now we are equipped to describe a memory-less QDS protocol that can be realized with the same technology as QKD. The protocol works for three parties, but can be generalised. In this protocol, Alice signs the message, Bob first receives the message and needs to authenticate it, and Charlie receives the forwarded message from Bob, and verifies that the initial source was indeed Alice.

```

function QDS():
    1. Key generation: Alice performs QKD (up
       to the point where the raw keys are
       obtained) with Bob and Charlie separately,
       twice. Now, Alice has four bit-strings:
        $A_0^B, A_1^B, A_0^C, A_1^C$ , Bob has two:  $K_0^B, K_1^B$ , and
       Charlie has  $K_0^C, K_1^C$ . The private key is
       the concatenation of the two corresponding
       strings,  $(A_0^B || A_0^C, A_1^B || A_1^C)$ .
    2. Error rates of the quantum channel are
       estimated at this stage.
    3. Bob and Charlie perform a symmetrization
       by exchanging secretly half of their
       strings via another QKD link. The new
       strings for Bob are  $S_0^B, S_1^B$  and Charlie
        $S_0^C, S_1^C$ . They are composed of half of the
       string initially sent to Bob and half of
       that to Charlie, but Alice does not know
       from which parts of Bob's and Charlie's
       strings they came from.
    4. The verification keys for Bob and Charlie
       are  $(S_0^B, S_1^B)$  and  $(S_0^C, S_1^C)$ .
    5. Signing: to sign a message  $M$ , Alice sends
        $(M, A_M^B || A_M^C)$  to Bob.
    6. Verification: Bob checks the mismatch rate
       between the signature received  $A_M^B || A_M^C$ 
       and his stored verification key  $S_M^B$ . If
       the error is compatible with the channel
       noise, he accepts. Charlie receives a
       message with Alice's signature, but from
       Bob. He performs a similar check and may
       choose to accept the message as having
       originated from Alice.
    7.  $\square$ 

```

Algorithm XXXIV: A quantum digital signature (QDS) protocol compatible with a QKD network.

XXXIX. QUANTUM CRYPTO-ASSETS

A SIDE from the value of outsourcing actual computations is the value of users' data itself. In the classical world we can generically refer to high-value data (especially in the context of digital tokens representing units of crypto-currencies) as *crypto-assets*. Similarly, and especially given the nature of data likely to be subject to quantum treatment, we anticipate future *quantum crypto-assets* – high-value quantum states. Although the ways in which such assets are handled will be highly application-dependent, we comment on several specific use-cases that will foreseeably arise from observing recent developments in classical crypto-assets.

A. Secure quantum data

Suppose Alice wishes to store quantum data offsite, for example in a repository for safekeeping, or within remote

or decentralised data structures (such as a Blockchain). Since her data is not held locally, there is the concern that an unauthorised third-party might simply steal it. How can she ensure its integrity, without maintaining any quantum data (assuming she can locally maintain classical data)?

This is a trivial problem to solve. Essentially we can think of this as a trivialised special case of blind quantum computing (Sec. ??), where the outsourced computation is just the identity operation. Employing the same ideas as the blind cluster state quantum computing protocol (Sec. ??), Alice simply takes each qubit in her quantum data structure, and with equal probability ($p = 1/4$) applies one of the four Pauli operators to it ($\hat{1}, \hat{X}, \hat{Y}$ or \hat{Z}). She of course keeps track of which ones were applied, which requires 2 classical bits per qubit.

From her perspective (or anyone she chooses to share the 2 classical bits with), she is always able to perfectly recover the hidden qubit, simply by applying the same Pauli operators a second time to invert them. However, from the perspective of a third-party without access to the 2 classical bits, they observe a perfect depolarising channel,

$$\mathcal{E}_0^{\text{depolarising}}(\hat{\rho}) = \frac{\hat{1}}{4}, \quad (39.1)$$

yielding the completely mixed state, independent of the input state, from which no state information can be inferred. Thus, this approach confers *perfect* information-theoretic security.

B. Quantum atomic swaps

In a crypto-market we inherently wish to engage in the free exchange of different types of crypto-assets. In the context of Blockchain-based asset ledgers, we may wish to directly exchange tokens residing on entirely distinct Blockchains. For example, we may wish to trade a Bitcoin for an Ether (Ethereum coin). Enabling this kind of exchange is vital for creating a fully functional crypto-market of arbitrarily interconvertible assets.

It's essential that such exchanges be performed with integrity, such that in a potentially anonymous transaction one party cannot run off with everything. The obvious solution here is to employ a trusted third-party to mediate the transaction, as is done in many real-world high-value exchanges. But this is undesirable for several reasons:

- Trust: Both parties must have complete confidence in the integrity of the mediating third-party. This necessarily introduces risk, which manifests itself as an indirect transaction cost.
- Monetary cost: A third-party is most likely to charge for the service they provide. Even if this margin is

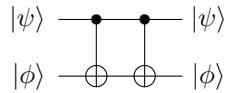
slim, in high-volume markets this becomes a consideration.

- Latency: Mediation introduces an additional layer of communication, with associated latency. Even in today's high-frequency markets, minute latency improvements yield major competitive advantage.
- Resource overheads: Mediation imposes greater resource requirements, most notably communication or computational ones.
- Regulatory: 'Credible' mediators typically comply with regulatory and taxation agencies. Traders of an anarcho-capitalistic mindset would rather avoid such nonsense altogether, and establish a truly globalised free-market. — *Lux libertas. Lux tua nos ducat.*

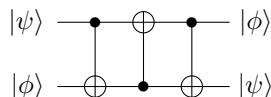
This motivates the question, can we securely implement *direct* exchanges, in the absence of any trusted mediating authority or regulatory oversight?

In classical Blockchain technology, *atomic swaps* can be employed for this purpose. Such algorithms allow the direct exchange of crypto-assets, cryptographically enforced to guarantee one of two outcomes: a successful mutual exchange, or no exchange at all. There is cryptographically no possibility for a partial exchange to occur, in which one party ends up with both assets.

With quantum crypto-assets we can easily construct such *quantum atomic swaps* by exploiting some well-known identities relating CNOT and SWAP gates. The first identity is that two consecutive CNOT gates cancel one another to yield an identity operation,

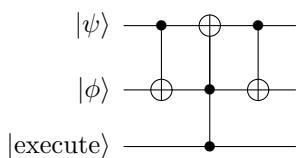


Second, a sequence of three alternating CNOT gates in series yields a SWAP operation,



Important is that in neither of the above two identities does partial implementation (i.e. not executing one of the CNOTs) manifest itself as a one-way transaction, the key security consideration in the construction of a cryptographic atomic swap.

Note that these two decompositions for the identity and SWAP gates differ only via the presence of the central CNOT gate within the latter decomposition. By replacing it with a doubly-controlled CNOT gate, the additional control qubit effectively specifies whether or not a regular CNOT gate is applied between the first two qubits,



This ancillary 'execute' qubit (restricted to $|0\rangle$ or $|1\rangle$, i.e. effectively a classical bit), acts as a toggle between the two modes of operation, without changing the underlying physical circuit implementation – it's now software-controlled, rather than hardware-controlled.

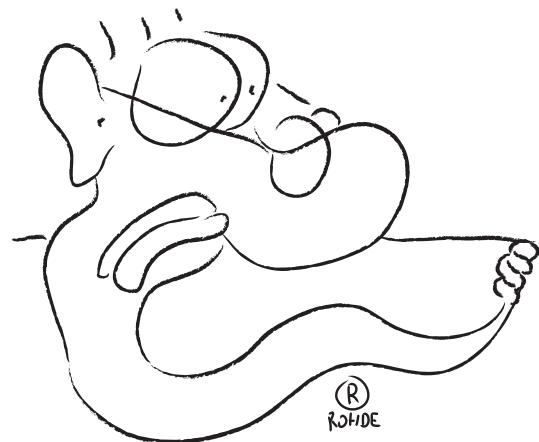
C. Quantum smart contracts

An atomic swap implementation with an 'execute' control signal is particularly useful in an environment involving *smart contracts* – self-executing generalisations of conditional contracts, such as credit default swaps (CDSs) – where the execution of an exchange depends upon an algorithmically-determined outcome. In this instance, the 'execute' qubit will be held and controlled by the smart contract algorithm. The algorithm making this choice can essentially be arbitrary, enabling extremely sophisticated contractual arrangements and exotic derivative instruments to be implemented in a self-enforced manner, without reliance upon third parties.

In the most general case in the quantum era, not only will the crypto-assets under exchange be quantum in nature, but the smart contract algorithms controlling their execution could be too (in principle any **BQP** algorithm). This paves the way towards generic *quantum smart contracts*, the direct quantum extension of existing classical smart contract techniques.

The implications of complimenting smart contracts with quantum algorithms (even if only trading classical crypto-assets) are potentially immense. Should quantum-enhanced algorithms facilitate improved pricing models for example, crypto-markets could benefit from improved market efficiency, more accurate price signals from futures markets (i.e. reduced risk spreads), enhanced allocation of capital, and greater investor confidence, with higher market volume and liquidity.

— *A mari usque ad mare.*



Part VIII

Quantum computing

SINCE quantum computing is perhaps the most exciting of the emerging quantum technologies, which we treat as the foremost application for the quantum internet, we now introduce quantum computing, covering models and physical implementations for realising it, and some of its well-known algorithmic applications.

XL. MODELS FOR QUANTUM COMPUTATION

WE begin by reviewing the models for quantum computation that we will refer to throughout this work. There are various approaches to implementing and representing quantum computations. We now briefly introduce the ones most relevant to our discussions on networked quantum computation.

A. Circuit model

The *circuit model* is the conventional and most intuitive approach for expressing quantum algorithms, decomposing them into chronological sequences of elementary operations, comprising state preparation, single- and multi-qubit gates, measurement, and classical feedforward. We recommend referring to the introductory sections of (?) for a far more comprehensive introduction to quantum circuits than is presented here. This model will be naturally intuitive to those familiar with classical circuit diagrams, albeit with some important differences, such as time-ordering.

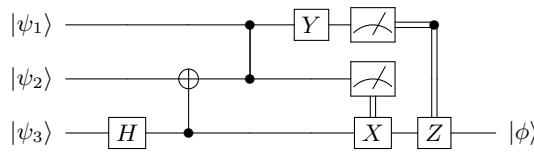


Figure 115: Simple example of a quantum circuit on 3 qubits, comprising several single- and 2-qubit quantum gates and measurements. Rows represent qubits, and time flows from left-to-right.

Fig. 115 illustrates a simple 3-qubit quantum circuit comprising all of these elements. The interpretation of this diagram is as follows:

- Horizontal lines represent individual qubits.
- Time flows from left to right (feedback is not allowed in the typical formalism for this representation).
- The three input qubits are labelled on the far-left as $|\psi_1\rangle$, $|\psi_2\rangle$ and $|\psi_3\rangle$.

- Single-qubit gates are denoted as boxes containing the name of the associated unitary operation. Here, the examples are the Hadamard (\hat{H}), Pauli bit-flip (\hat{X}), Pauli bit-phase-flip (\hat{Y}), and Pauli phase-flip (\hat{Z}) gates,

$$\begin{aligned}\hat{H} &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \\ \hat{X} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \hat{Y} &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ \hat{Z} &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned}\quad (40.1)$$

- 2-qubit gates are denoted by vertical lines between the respective qubits.
- The maximally-entangling 2-qubit controlled-NOT (CNOT) gate is denoted via a control (\bullet) and a target (\oplus),

$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (40.2)$$

This is the quantum equivalent of the classical XOR gate, flipping the target (\hat{X}) if the control is on.

- All quantum gates have the same number of input as output qubits. This is a necessary condition for the unitarity of quantum gates ($\hat{U}^\dagger \hat{U} = \hat{\mathbb{1}}$).
- The maximally-entangling 2-qubit controlled-phase (CZ) gate is denoted by two targets (\bullet) (the gate operates symmetrically on its two qubits),

$$\text{CZ} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad (40.3)$$

applying a phase-gate (\hat{Z}) to the target if the control is on.

- The ‘meter’ symbol represents a classical measurement in the Pauli \hat{Z} -basis (the computational or logical basis).
- Double lines represent classical feedforward of measurement outcomes, controlling a subsequent gate.

The circuit in Fig. 115 can be interpreted mathematically as implementing the following operation,

$$\begin{aligned} |\phi\rangle &= \hat{Z}_3^{m_1} \cdot \hat{X}_3^{m_2} \cdot \hat{M}_2 \cdot \hat{M}_1 \cdot \hat{Y}_1 \\ &\quad \cdot \text{CZ}_{1,2} \cdot \text{CNOT}_{3,2} \cdot \hat{H}_3 \cdot |\psi_1\rangle \otimes |\psi_2\rangle \otimes |\psi_3\rangle, \end{aligned} \quad (40.4)$$

where m_1 and m_2 are the binary measurement outcomes of the two single-qubit \hat{Z} -basis measurements, \hat{M}_1 and \hat{M}_2 .

Using the circuit model, arbitrary quantum computations can be elegantly and intuitively represented. To enable *universal* quantum computation within this model, a *universal gate set* must be available at our disposal. Most commonly, this is chosen to be the maximally-entangling 2-qubit CZ or CNOT operation, in addition to arbitrary single-qubit gates. Any quantum (i.e **BQP**) algorithm may be efficiently decomposed into a polynomial-depth circuit comprising elements from this universal gate set. Note that the universal gate set is not unique, and there are many distinct sets. However, this set must contain at least one entangling operation acting on two or more qubits (such as a CZ or CNOT gate), and at least one non-Clifford gate⁴⁴.

B. Cluster states

The *cluster state* model for quantum computation (???) (also referred to as the *one-way, measurement-based, or graph state* models for quantum computation) is an extremely powerful paradigm that warrants treatment of its own, owing to its significant distinction from the more familiar circuit model, and its applicability to distributed models for quantum computation, to be discussed in Sec. XLIII.B.

In the cluster state model, we begin by preparing a particular, highly-entangled state, called a *cluster state* or *graph state*. The state is associated with a graph G , comprising vertices, V , and edges, E ,

$$G = (V, E), \quad (40.5)$$

of some topology, although rectangular lattice graphs are usually considered as they are sufficient for universal quantum computation⁴⁵. That is, they act as a ‘substrate’ for implementing arbitrary quantum computations.

In the graph, vertices represent qubits initialised into the,

$$|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad (40.6)$$

state, and edges represent the application of maximally entangling CZ gates between vertices,

$$|\psi\rangle_{\text{cluster}} = \prod_{e \in E} \hat{\text{CZ}}_e \cdot \bigotimes_{v \in V} |+\rangle_v. \quad (40.7)$$

⁴⁴ The Clifford group is that which commutes with the CNOT gate, such as the Pauli group.

⁴⁵ Note that the graph upon which a cluster state resides is not to be confused with the network graph. Rather it is just a convenient graphical representation for a class of multi-qubit states.

Alternately, but equivalently, cluster states may be defined in the stabiliser formalism. Specifically, a cluster state is defined to be the joint +1 eigenstate of all the stabilisers,

$$\hat{S}_v = \hat{X}_v \prod_{i \in n_v} \hat{Z}_i, \quad (40.8)$$

where there is one stabiliser \hat{S}_v per vertex v , and n_v is the set of vertices neighbouring v . The cluster state therefore satisfies,

$$\hat{S}_v |\psi\rangle_{\text{cluster}} = |\psi\rangle_{\text{cluster}} \forall v, \quad (40.9)$$

and the full set of stabilisers, \hat{S}_v , over all vertices v is sufficient to fully characterise the cluster state, $|\psi\rangle_{\text{cluster}}$, for a given graph topology.

An example of a rectangular lattice cluster state is presented in Fig. 116. Cluster states are easily encoded optically using photonic polarisation encoding (Sec. XIII.A), and therefore readily lend themselves to optical networking.

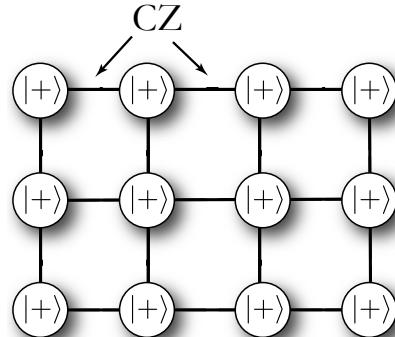


Figure 116: Example of a 4×3 rectangular lattice cluster state. Each vertex in the graph represents a qubit initialised into $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. Edges represent the application of CZ gates between qubits (CZ gates commute, so the order is unimportant). Of sufficient dimension, states of this topology enable universal measurement-based quantum computation, whereby computation proceeds purely via single-qubit measurements, and all entangling operations have been commuted to the state preparation stage. Because CZ gates commute, the preparation of cluster states is time-independent, and easily implemented in a distributed or parallelised manner. The time-ordering of the single-qubit measurements is dependent on the structure of the graph and the algorithm.

Having prepared this state, the computation is implemented purely via a well-orchestrated routine of single-qubit measurements. The order and basis in which they are performed (which depends on previous measurement outcomes in general – i.e we require fast-feedforward) then stipulates the computation. In the context of distributed computation (Sec. XLIII.B), this requires classical communication between nodes.

Mapping a circuit model computation to a cluster state topology can be most naively performed by taking a circuit acting on n_{qubits} qubits with depth n_{depth} , preparing

an $n_{\text{qubits}} \times n_{\text{depth}}$ rectangular lattice cluster, and ‘etching’ the circuit directly into the cluster state substrate. To perform this mapping we choose a universal gate set comprising CZ and single-qubit gates, retaining vertical edges where CZ gates ought to be present, eliminating the remaining vertical edges. Now we have a substrate that looks topologically very much like its equivalent circuit construction, and the computation proceeds chronologically in the same manner. The only conceptual distinction is that in the circuit model gates are directly applied chronologically to the set of qubits, whereas in the cluster state gate teleportation (Sec. XXVII.D) is effectively implemented upon each measurement, with the action of gates accumulating as these teleportations are successively applied. A simple example of this notion is shown in Fig. 117.

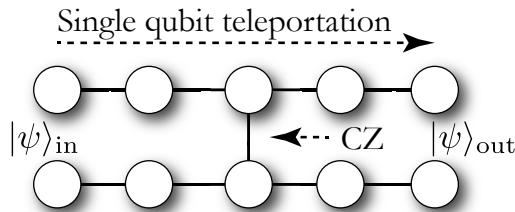


Figure 117: Simple example of a cluster state that performs a computation comprising single-qubit operations and a CZ gate between two logical qubits. Let the two horizontal chains represent our two logical qubits. After inputting our input state from the left, we progressively measure out the cluster state qubits chronologically from left-to-right. Upon each single-qubit measurement, the choice of measurement basis teleports the action of a single-qubit gate. These accumulate sequentially. When we reach the point of measuring the two cluster state qubits joined with the vertical edge, the logical qubits accumulate the action of a CZ gate between them, since this is identically what that vertical edge physically corresponds to. Reaching the final two qubits, one from the upper rail and one from the lower, we obtain our two output logical qubits.

The distinctive feature of this model is that all the entangling CZ gates are performed at the very beginning of the protocol, during the state preparation stage. The algorithm itself is purely measurement-based, requiring only single-qubit measurements (no entangling measurements).

An alternate interpretation of the cluster state model is that it is a complicated network of state and gate teleportation protocols (Sec. XXVII.C). Specifically, a CZ gate with a $|+\rangle$ state as a resource, followed by measurement of one of the two qubits acts as a single-qubit teleporter, as shown in Fig. 118⁴⁶. Thus, with a substrate state of CZ gates applied between $|+\rangle$ states, the single-qubit measurements progressively teleport the input state through

the graph topology, at each stage accumulating the action of more gates, which are related to the choices of the previous single-qubit measurement bases, and the graph topology.

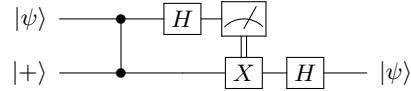


Figure 118: The single-qubit teleporter, based upon a CZ gate, a single-qubit measurement, and classical feedforward.

The cluster state formalism has proven very useful, enabling the development of models for linear optics quantum computing (Sec. XLII.A), orders of magnitude more efficient than the originally proposed protocol. It has been found that bonding strategies – i.e the order in which smaller clusters are fused into larger ones when using non-deterministic gates – plays a major role in resource overhead, and much work has been performed on efficient preparation strategies for various topologies (??????????????).

These cluster states are highly valuable, given their computational power, and the ability to communicate them from Alice, who is able to prepare them, to Bob, who lacks the technology, would be a boon for Bob.

It would be most practical, economical, and resource efficient to have a single, well-equipped server with the ability to prepare such states, who does so on behalf of everyone else, and communicates the fresh cluster states to them over the quantum internet (for a price, perhaps).

Importantly, the preparation of cluster states is readily parallelised. All the entangling CZ operations commute, the order in which they are applied is irrelevant, and a rectangular lattice cluster is completely uniform. Thus, the graphs representing smaller cluster states may be easily ‘fused’ together to form larger cluster states using, for example, CZ gates. Several other types of entangling gates can also be employed, such as polarising beamsplitters – so-called *fusion gates* (?). This allows the preparation of cluster states to be performed in a ‘patchwork quilt’-like manner – a number of nodes each prepare small lattice clusters, they are all put side-by-side, and stitched together using CZ gates. This type of distributed state preparation is a perfect application for in-parallel distributed quantum processing (Sec. XLIII.B).

Consider the scenario whereby Alice requests a large cluster state from Bob, but, while she was unable to prepare the cluster state herself, she has the technological ability to perform the measurement-based computation on the state (i.e simple single-qubit measurements). This would effectively bypass the need for secure quantum computation (Sec. XLIV) on Bob’s hardware altogether, enabling computation with *perfect* secrecy, since no foreign parties would be involved in the computation stage,

⁴⁶ This is an alternative, but equivalent implementation for quantum state teleportation to that presented in Sec. XXVII.C.

and no secret data is communicated – only the *substrate* for the computation is communicated, which could be used for any purpose whatsoever. By commuting all the technologically challenging aspects of a quantum computation to the state preparation stage, we can effectively mitigate the need for blind quantum computing entirely, since the ‘hard work’ has been done in advance by the host, and Alice gets to fulfil the computation on her own, completely bypassing poor old Bob, who was just dying to read Alice’s secret love letters before processing them into Hallmark cards.

There are several cluster state identities we will utilise later, summarised in Fig. 119.

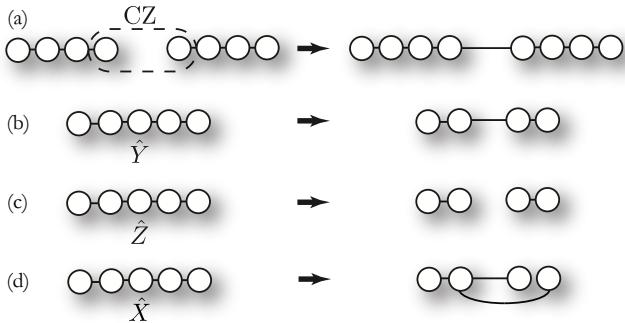


Figure 119: Several cluster state identities, demonstrated in the case of linear clusters. (a) a CZ gate between two qubits creates an edge between them in the graph. (b) Measurement of a qubit in the Pauli \hat{Y} basis removes that qubit from the graph, whilst creating new edges between the neighbouring qubits. (c) Measurement of a qubit in the Pauli \hat{Z} basis removes that qubit and any neighbouring edges. (d) Measurement of a qubit in the Pauli \hat{X} basis removes that qubit, leaving one of its neighbours as a ‘dangling node’.

When using non-deterministic gates (i.e ones that probabilistically fail) to prepare cluster states, there are approaches to nonetheless preparing ideal cluster states. There have been two main approaches that have become particularly well known.

The first is to use the ideas of *micro-clusters* and cluster state recycling to incrementally build up larger clusters, progressing as a random walk, which is biased in the direction of state growth. This approach is discussed in more detail in Secs. XLII.B & XLIII.D.

The second approach is to borrow techniques from percolation theory to simply tolerate defects in a cluster state lattice by working around them (?). Specifically, if the defect probability (i.e probability of a missing vertex or edge) is below some *percolation threshold*, $p_{\text{defect}} \leq \epsilon_{\text{threshold}}$, in the asymptotic limit we are guaranteed that routes exist through the lattice, enabling the required flow of information. This allows defective graphs to be employed for quantum computation.

C. Adiabatic quantum computation

— by Zixin Huang.

Adiabatic quantum computation (AQC) began as an approach for solving optimisation problems, but now has evolved into a universal alternative to the circuit model. AQC is based on an idea that is distinct from the circuit model (?); in the circuit model, the computation is encoded into a series of gates, whereas in AQC the computation starts from an initial Hamiltonian whose ground state is easy to prepare, and evolves to a final state that encodes the solution to the computational problem.

The adiabatic theorem forms the backbone of AQC, which states that for a system initially prepared in an eigenstate of a time-dependent Hamiltonian $\hat{H}(t)$, the is dictated by the Schrödinger equation,

$$i \frac{\partial |\psi(t)\rangle}{\partial t} = \hat{H}(t)|\psi(t)\rangle, \quad (40.10)$$

which will keep approximately to the instantaneous ground state, if $\hat{H}(t)$ varies sufficiently slowly.

In the circuit model, the cost of the computation is equated with the number of gates. In AQC, the cost of adiabatic algorithms is defined to be the dimensionless quantity (?),

$$\text{cost} = t_f \cdot \max_s \|\hat{H}(s)\|, \quad (40.11)$$

where t_f is the algorithmic runtime, and $\|\cdot\|$ denotes the operator norm, the largest singular value of the operator.

The worst-case runtime of an adiabatic algorithm scales as,

$$t_f = O\left(\frac{1}{\Delta^{(3)}}\right), \quad (40.12)$$

where $\Delta^{(3)}$ is the minimum eigenvalue gap between the ground state and the first excited state of the Hamiltonian (?). This is the reason why AQC has generated much interest – it has a rich connection to the well-studied field of condensed matter physics.

The analysis for AQC involves bounding the eigenvalue gap of a complicated many-body Hamiltonian, which is a notoriously difficult problem. Nevertheless, a number of examples are known exhibiting a gap between the classical and quantum cases.

Some of the well-known examples of AQC algorithms include:

- Adiabatic Grover algorithm: as its name suggests, this is the adiabatic version of Grover’s search algorithm. The initial Hamiltonian is,

$$\hat{H}_0 = \hat{\mathbb{1}} - |\phi\rangle\langle\phi|, \quad (40.13)$$

where $|\phi\rangle$ is the uniform superposition state $|+\rangle^{\otimes n}$. The final Hamiltonian is given by,

$$\hat{H}_1 = \hat{\mathbb{1}} - |m\rangle\langle m|, \quad (40.14)$$

where $|m\rangle$ is the marked item we are searching for (?).

- Adiabatic Deutsch-Jozsa algorithm: given a binary function,

$$f : \{0, 1\}^n \rightarrow \{0, 1\}, \quad (40.15)$$

which is either constant or balanced, the Deutsch-Jozsa algorithm can determine which type of function it is, exhibiting quantum speedup. An adiabatic implementation (?) is derived to match the speedup obtained in the original circuit model implementation.

- Adiabatic glued-trees problem: we consider two binary trees of depth n . Each tree has $2^{n+1} - 1$ vertices, and the two trees are randomly glued together, as shown in Fig. 120. A leaf is chosen from the left, connected to a random leaf on the right, which is in turn glued to another leaf on the left and so on. Every leaf on is connected to two on the other side, creating a random cycle that alternates between the two trees. The computation problem is to start from the left root and find a path to the right root in the smallest possible number of steps, given that one can only see adjacent vertices. Classical algorithms require at least a sub-exponential number of queries, but there exists a polynomial-time quantum algorithm based on quantum walks (?).
- Adiabatic PAGERANK algorithm: the PAGERANK vector is a crucial tool in data mining and information retrieval, employed heavily by the Google search engine. The goal of the algorithm is to rank the importance, impact or influence of some entity in a network (websites on the internet in the case of a Google search). The algorithm achieves this by representing the network as a flow network, whereby flow from one vertex to another represents a vote by the source vertex for the importance of the destination vertex. By finding a steady-state flow to each edge in the network, the net flow into vertices represents their cumulative importance, as determined collectively by participants in the network. However, the magnitude of votes cast by vertices is weighted by their own ranking. Therefore, the algorithm aims to filter out bogus gaming of the system via the creation of dummy nodes that cast votes en masse to rig the election. The PAGERANK approach to ranking websites has been by far the most successful and robust such algorithm, and was perhaps the key development behind the global success of Google. The best-known classical algorithms for solving the PAGERANK problem are via: representing the flow network as a matrix equation, from which the solution is obtained via solving an eigenvalue equation; representing the flow network

as a random walk, whereby walkers randomly follow paths influenced by the flows, which in the long-time limit yields walks approximating the solution. The current best-known quantum PAGERANK algorithm outperforms all currently known classical algorithms with polynomial or even exponential speedup (?), but better future classical algorithms have not been ruled out and are under active investigation.

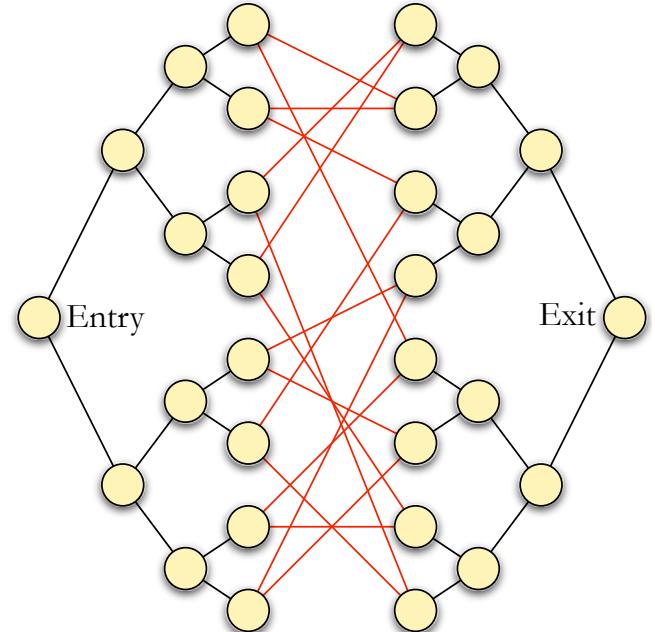


Figure 120: A glued-trees graph of depth $n = 4$. The ‘gluing’ edges, chosen randomly such that each leaf vertex connects to exactly two of them, are shown in red.

The computational power of the circuit model and the adiabatic model for quantum computing are equivalent up to a polynomial overhead. In the circuit model, a set of gates is said to be universal if any unitary operation may be approximated to arbitrary accuracy by a quantum circuit involving only those gates. The analog of such a set of gates in AQC is to efficiently map any circuit to a Hamiltonian.

If we have in the circuit model an n -particle pure state $|\psi\rangle$, acted upon by unitary \hat{U} with circuit depth L , a time-dependent Hamiltonian $\hat{H}(t)$ is universal if:

- The ground state of $\hat{H}(t_f)$ is equal to $\hat{U}|\psi\rangle$ with non-zero probability.
- The number of particles in $\hat{H}(t_f)$ is polynomial in n and L , and t_f is also polynomial in n and L .

It has been proven that AQC can efficiently simulate the entire circuit model (?), which implies that it is itself universal for efficient quantum computation.

D. Restricted models for quantum computation

In the near future we are unlikely to have devices with the full power and versatility of universal quantum computers. Instead, we will gradually evolve towards that challenging goal via many incremental, intermediate steps. Those steps will take us along a path of restricted quantum computers, that solve specific problems of relatively small size, and are not fault-tolerant. Probably we can expect them to contain on the order of hundreds of qubits in the coming few years, as of the time of writing this.

We dedicate Sec. LXXXIX to discussing these so-called *Noisy Intermediate-Scale Quantum* (NISQ) devices (?), speculating on what that might exactly entail for our quantum journey in the near future.

One thing is for certain – full-fledged quantum computing will remain an extremely ambitious goal for some time, but we will learn a lot from traversing the path towards it, and hopefully uncover new restricted quantum applications along the way.

E. Fault-tolerance

In Sec. XVII we discussed QoS in the context of quantum networks, where we wish to protect the quantum information being communicated via packets of quantum data. In particular, QEC allows us to detect and correct errors introduced into quantum data during transmission across noisy channels.

Much more broadly, in the context of an entire quantum computation we will want to achieve the same goal, except that our techniques will need to extend far beyond defending individual quantum states against errors during transmission, but defending an entire computation and all the information residing within it at every stage throughout its execution.

This is achieved by extending techniques from QEC to achieve *fault-tolerant quantum computing*. The primary difficulty here is that a quantum computation is not a passive operation, but involves the successive application of a potentially enormous number of quantum gates, each of which subject to its own error processes, all of which must be mitigated for the computation to succeed.

Because a quantum computation is not a passive operation but highly active, fault-tolerance protocols are also active and it does not suffice to simply perform an encoding at the beginning and error correction at the end. Instead, error correction procedures must be applied repeatedly throughout execution, at each stage projecting the encoded computation onto an error-free state.

The concept of Fault-tolerance in computation is not a new idea, it was first developed in relation to classical computing (???). However, in recent years the precise manufacturing of digital circuitry has made large scale

error correction and fault-tolerant circuits largely unnecessary.

The basic principle of Fault-tolerance is that the circuits used for gate operations and error correction procedures should not cause errors to cascade. Quantum gates not only covert errors, i.e. a Hadamard operation can convert an X -error into a Z -error and visa versa, but multi-qubit gates can also *copy* errors. Hence if quantum circuits are not designed carefully, a correctable number of *physical* errors could occur which are consequently copied so many times that they overwhelm the error-correction capabilities of the encoding scheme.

Fault-tolerance, in the context of error correction, is a function of how circuits and protocols are implemented, not a function of the underlying physical hardware. It is assumed that all single qubit gates can introduce single qubit errors at some probability, p , and it is assumed that all two-qubit gates will *copy* pre-existing errors that exist at the input and also has the possibility of introducing a two-qubit correlated error on the two-qubits, with probability, p .

In some cases there are examples of higher order gates being defined as primitives, for example the three qubit Toffoli gate. However, it should be noted that in almost all cases, the physical implementation of these multi-qubit gates occur through an implicit decomposition into single and two-qubit gates. This is due to the fact that the vast majority of physical systems, the highest order coupling term in a system Hamiltonian is weight two. Higher weight coupling terms, which would be required to enable native multi-qubit gates (i.e. weight three terms in the Hamiltonian would be needed to natively enact a Toffoli gate), simply do not arise in natural and easily controllable quantum systems.

To determine how errors are copied by gate operations, and error operator E is conjugated through the gate operation to create a new error operator, $E' = G^\dagger EG$, for some gate unitary, G . A single qubit example is the transformation of X -errors to Z -errors and visa versa through a Hadamard gate, due to the identity $\hat{X} = \hat{H}\hat{Z}\hat{H}$ and $\hat{Z} = \hat{H}\hat{X}\hat{H}$.

A two-qubit example is more involved as we need to check all combinations of error mappings on both qubits involved in the gate. If $\hat{G} = \text{CNOT}$, we can examine how X - and Z -errors change via G ,

$$\begin{aligned} \text{CNOT}(I \otimes X)\text{CNOT} &= I \otimes X \\ \text{CNOT}(X \otimes I)\text{CNOT} &= X \otimes X \\ \text{CNOT}(I \otimes Z)\text{CNOT} &= Z \otimes Z \\ \text{CNOT}(Z \otimes I)\text{CNOT} &= Z \otimes I \end{aligned} \tag{40.16}$$

where the notation, $A \otimes B$, are error operators, $\{A, B\} \in \{I, X, Y, Z\}$, on qubits one and two of the gate and, $\hat{G} = \hat{G}^\dagger = \text{CNOT}$.

So, for a controlled-not operation, X -errors are copied from control qubit to target, and Z -errors are copied from

target to control. pre-existing X -errors on the target qubit or Z -errors on the control qubit are unchanged through the gate.

The fact that quantum circuits can cause errors to be copied implies that if circuits are designed badly, errors can cascade during error correction protocols even when only one or two *physical* errors actually took place. Considering that error correction codes have a finite correcting power, i.e. the Steane code will deterministically correct an arbitrary *single* qubit error, but if more than a single error occurs between correction cycles, logical errors are likely to be induced.

Fault-tolerance is a discrete feature of a quantum circuit construction, either a construction is fault-tolerant or it is not. However, what is defined to be fault-tolerant can be a function of what type of error-correction code is used. For example, for a single error correcting code, fault-tolerance is defined as:

1. A single error will cause **at most** one error in the output for each logical qubit block.

However, if the quantum code employed is able to correct multiple errors, then the definition of fault-tolerance can be relaxed, i.e. if the code can correct three errors then circuits may be designed such that a single failure results in at most two errors in the output (which is then correctable). In general, for a code correcting $t = \lfloor (d-1)/2 \rfloor$ errors, fault-tolerance requires that $\leq t$ errors during an operation does not result in $> t$ errors in the output for each logical qubit.

F. The threshold theorem

The threshold theorem is a consequence of fault-tolerant circuit design and the ability to perform dynamical error correction. Rather than present a detailed derivation of the theorem for a variety of noise models, we will instead take a very simple case where we utilize a quantum code that can only correct for a single error, using a model that assumes uncorrelated, errors on individual qubits. For more rigorous derivations of the theorem see (????).

Consider a quantum computer where each physical qubit experiences either an X and/or Z error independently with probability p , per gate operation. Furthermore, it is assumed that each logical gate operation and error correction circuit is designed fault-tolerantly and that a cycle of error correction is performed after each elementary *logical* gate operation. If an error occurs during a logical gate operation, then the fault-tolerant constructions ensure this error will only propagate to at most one error in each block, after which a cycle of error correction will remove the error.

Hence if the failure probability of un-encoded qubits per time step is p , then a single level of error correction will ensure that the logical step fails only when two (or

more) errors occur. Hence the failure rate of each logical operation, to leading order, is now $p_L^1 = cp^2$, where p_L^1 is the failure rate (per logical gate operation) of a 1st level logical qubit and c is the upper bound for the number of possible 2-error combinations which can occur at a physical level within the circuit consisting of the correction cycle + gate operation + correction cycle (?).

We now repeat the process, re-encoding the computer such that a level-2 logical qubit is formed, using the same $[[n, k, d]]$ quantum code, from n , level-1 encoded qubits. It is assumed that all error correcting procedures and gate operations at the 2nd level are self-similar to the level-1 operations (i.e. the circuit structures for the level-2 encoding are identical to the level-1 encoding). Therefore, if the level-1 failure rate per logical time step is p_L^1 , then by the same argument, the failure rate of a 2-level operation is given by, $p_L^2 = c(p_L^1)^2 = c^3 p^4$. This iterative procedure is then repeated (referred to as concatenation) up to the k th level, such that the logical failure rate, per time step, of a k -level encoded qubit is given by,

$$p_L^k = \frac{(cp)^{2^k}}{c}. \quad (40.17)$$

Eq. 40.17 implies that for a finite *physical* error rate, p , per qubit, per time step, the failure rate of the k th-level encoded qubit can be made arbitrarily small by simply increasing k , dependent on $cp < 1$. This inequality defines the threshold. The physical error rate experienced by each qubit per time step must be $p_{th} < 1/c$ to ensure that multiple levels of error correction reduce the failure rate of logical components.

Hence, provided sufficient resources are available, an arbitrarily large quantum circuit can be successfully implemented, to arbitrary accuracy, once the physical error rate is below threshold. Initial estimates at the threshold, which gave $p_{th} \approx 10^{-4}$ (????) did not sufficiently model physical systems in an accurate way. Recent results (?????) have been estimated for more realistic quantum processor architectures, showing significant differences in threshold when architectural considerations are taken into account. The most promising thresholds that have been calculated for expected, circuit level noise, are based on surface codes (????), with thresholds slightly less than 1%. This has now become the target for experimental groups as a large number of scalable systems architectures utilise the surface code as the underlying correction model (??????).

— *Aqua vitae.*

XLI. QUANTUM ALGORITHMS

THE ultimate goal of quantum computing is to implement algorithms with a quantum speedup compared to classical algorithms. The degree of speedup achieved varies between algorithms, and it is important to note

that not every classical algorithm exhibits any speedup when implemented quantum mechanically.

To provide context for the excitement of quantum computing and motivate interest in their development, we now summarise some of the key quantum algorithms that have been described exhibiting quantum speedup.

A. Deutsch-Jozsa

The first quantum algorithm demonstrating a provable improvement over the best classical algorithm was the Deutsch-Jozsa algorithm (?). Unfortunately the algorithm solves a very contrived problem, designed for the purposes of demonstrating post-classicality rather than solving a problem of actual practical interest. Nonetheless, the algorithm is straightforward to explain and understand, making it a useful starting point in understanding quantum algorithms and the computational enhancement they may offer.

The algorithm relies on a ‘black box’, referred to as an *oracle*, which takes an input bit-string and outputs a single bit, evaluating the function $f(x)$ for the n -bit input bit-string x . In this contrived problem $f(x)$ is guaranteed to be either *uniform* or *balanced*. In the former case, the output to the oracle is always $f(x) = 0$ or always $f(x) = 1$, but it doesn’t matter which, they simply must always be the same. In the latter case, the output is $f(x) = 0$ for exactly half the inputs x , and $f(x) = 1$ for the other half of x , but the ordering of which inputs generate which outputs may be arbitrary. The goal of the algorithm is to determine whether $f(x)$ is uniform or balanced using the least number of queries to the oracle.

While it’s clear that the dimensionality of the input state space is exponentially large, 2^n , it is fairly obvious that a trivial **BPP** algorithm exists for solving this problem with confidence exponentially asymptoting to unity against the number of oracle queries. We simply evaluate the oracle for randomly chosen inputs. If we measure any occurrences of measurement outcomes that are not all 0 or all 1 we know with certainty that the function must have been balanced. If on the other hand we measure all 0s or all 1s for more than half the input state space x , we know with certainty the function was uniform.

However, if the function were balanced, there is the possibility that it might conspire against us to fool us into thinking the function was uniform until we evaluate half plus one of the input states, requiring $O(2^n)$ oracle queries, although this will occur with exponentially low probability against the number of queries. Thus, the algorithm can be approximated with exponential asymptotic certainty in **BPP**. But considering the *worst* case rather than the *average* case, we may have to perform an exponential number of evaluations, $O(2^n)$, to know the answer with absolute certainty.

The Deutsch-Jozsa algorithm solves this rather spe-

cialised problem in the worst case using only a single quantum evaluation of the oracle.

The algorithm implementing the Deutsch-Jozsa protocol and its circuit diagram are shown in Alg. XXXV. The engine room of the algorithm is in the Hadamard transform, $\hat{H}^{\otimes n}$, which prepares an equal superposition of all 2^n possible input bit-strings x , which are then evaluated in superposition by the oracle. To ensure unitarity, the oracle is defined to implement the transformation⁴⁷,

$$\hat{U}_f|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle. \quad (41.1)$$

That is, it flips bit y if $f(x) = 1$ (equivalently addition modulo 2 or an XOR operation). An inverse Hadamard transform subsequently yields a measurement outcome with one of two possibilities:

- The 0 and 1 terms outputted from the oracle interfere perfectly constructively, if the function was uniform.
- They interfere perfectly destructively, if the function was balanced.

Then, with a single-shot measurement of the inverse Hadamard transformed output from the oracle we establish whether $f(x)$ was balanced or uniform with certainty. This exhibits an exponential worst case speedup compared to a randomised classical sampling algorithm (which is classically optimal).

B. Quantum search

The problem of finding specific entries in unstructured data spaces is a ubiquitous one. This class of *search algorithms* have amongst the broadest applicability of any class of algorithms. Computer scientists have invested excruciating man-hours⁴⁸ into organising and structuring data so as to minimise the resource overhead (in time and/or space) associated with extracting desired components. However, the methodology for achieving this, and the favourability of associated resource overheads, is highly dependent on the structure of the underlying data, or whether there even is any. To this end, numerous data structures and algorithms have been developed, accommodating for every mutation and variation of the posed problem imaginable. Often, there is a tradeoff between the overheads induced in time and memory, as

⁴⁷ Note that the seemingly more obvious choice of $\hat{U}_f|x\rangle = |f(x)\rangle$ is not unitary. This trick of introducing an additional ancillary state to enable unitary construction of arbitrary functions is a common one in quantum algorithm design, as will be discussed further in Sec. XLI.C.

⁴⁸ Presently, most computer science research institutions are equal opportunity employers.

```

function DeutschJozsa(f,n):
    1. Prepare the  $n + 1$ -bit state,
        
$$|\psi\rangle_0 = |0\rangle^{\otimes n}|1\rangle. \quad (41.2)$$

    2. Apply the  $n + 1$ -bit Hadamard transform,
        
$$|\psi\rangle_1 = \hat{H}^{\otimes(n+1)}|\psi\rangle_0$$

        
$$= \frac{1}{\sqrt{2^{n+1}}} \sum_{x=0}^{2^n-1} |x\rangle(|0\rangle - |1\rangle), \quad (41.3)$$

        where  $x$  enumerates all  $n$ -bit binary bit-strings.
    3. Apply the unitary oracle, implementing the transformation,
        
$$\hat{U}_f|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle, \quad (41.4)$$

        where  $\oplus$  denotes addition modulo 2, yielding,
        
$$|\psi\rangle_2 = \hat{U}_f|\psi\rangle_1. \quad (41.5)$$

    4. Apply another Hadamard transform,
        
$$|\psi\rangle_3 = \hat{H}^{\otimes n}|\psi\rangle_2. \quad (41.6)$$

    5. The full evolution is thus given by,
        
$$|\psi\rangle_{\text{out}} = (\hat{H}^{\otimes n} \otimes \hat{\mathbb{1}}) \cdot \hat{U}_f \cdot \hat{H}^{\otimes(n+1)}|0\rangle^{\otimes n}|1\rangle. \quad (41.7)$$

    6. Measure the first  $n$  qubits to determine the probability of measurement outcome  $|0\rangle^{\otimes n}$ .
    7. This probability is given by,
        
$$P_0 = \left| \frac{1}{2^n} \sum_{x=0}^{2^n-1} (-1)^{f(x)} \right|^2. \quad (41.8)$$

    8. Depending on whether  $f(x)$  was uniform or balanced, the alternating sign terms in this sum interfere constructively or destructively, yielding  $P_0 = 1$  or  $P_0 = 0$  respectively.
    9. Thus, a single measurement outcome suffices to determine whether  $f(x)$  was balanced or uniform.
    10.  $\square$ 

```

Algorithm XXXV: Deutsch-Jozsa algorithm for evaluating whether the function $f(x)$ is balanced or uniform, exhibiting exponential worst case speedup compared to the best classical **BPP** algorithm.

well as in pre-processing and data structure maintenance requirements.

For example, *hash tables* enable theoretical $O(1)$ lookup times on data with a *key-value pair* data structure. In a key-value pair each data entry (value) is tagged with a unique identifier (key) used for lookup. The value can observe any structure whatsoever, whereas the key is designed so as to minimise search times. When storing telephone numbers one might represent entries as key-value pairs, where the keys are people's names, and the values their respective telephone numbers. An efficient algorithm for mapping keys to physical memory addresses would imply efficient lookup of telephone numbers by name.

In the absence of a key-value representation one might simply store data in sorted form. However, this requires pre-sorting the entire data space, which may become costly for large data sets, and requires continual rearrangement whenever the data space is modified, making it computationally costly for mutable datasets.

For the end user, who wishes to find data elements, the worst-case data space is one with no order or underlying structure. Suppose we want to find whether a number exists in the telephone directory, but we don't know its associated name. In this instance, it can easily be seen that the best one can hope for, in terms of algorithmic runtime, is to simply look through the data space brute-force until we find what we are looking for. It is clear that with an unstructured space of N elements, this brute-force search algorithm requires on average $O(N)$ queries to find the desired entry. We call this the *unstructured search problem*.

The brute-force classical algorithm, despite already being technically 'efficient' (i.e $O(N)$ linear runtime), could nonetheless become unwieldy for very large datasets. Google doesn't want to exhaustively scan their entire collection of data-centres each time they want to lookup a database element. The quantum search algorithm, first presented by Grover (?), provides a solution to this problem using only $O(\sqrt{N})$ runtime (oracle queries), a quadratic enhancement. Whilst this falls far short of the exponential quantum enhancement one might have hoped for, which has also shown to be optimal (?), it is nonetheless still extremely helpful for many purposes, given the broad applications for this algorithm.

We will formulate the quantum search algorithm as an oracular algorithm, where the oracle takes as input an n -bit string, and outputs 1 if the input matches the entry we are looking for, otherwise 0. This formulation of the problem makes the algorithm naturally suited to solving satisfiability problems (many of which are **NP**-complete and of great practical interest).

The Grover quantum search algorithm is shown explicitly in Alg. **XXXVI**.

```
function Grover(f,n):
```

1. Using a Hadamard transform, prepare the n -qubit equal superposition of all 2^n logical basis states,

$$\begin{aligned} |\varphi\rangle &= \hat{H}^{\otimes n}|0\rangle^{\otimes n} \\ &= \frac{1}{\sqrt{2^n}} \sum_{x=0}^{2^n-1} |x\rangle, \end{aligned} \quad (41.9)$$

where x denotes a bit-string of length n .

2. The oracle is defined as a unitary black-box, which tags a target element T using a phase-flip,

$$\begin{aligned} \hat{U}_T|x\rangle &= (-1)^{f(x)}|x\rangle \\ &= (\hat{1} - 2|T\rangle\langle T|)|x\rangle, \end{aligned} \quad (41.10)$$

where $f(x) = \{0, 1\}$ is the black-box function determining whether input x is the target element T ($f(x) = 1$) or not ($f(x) = 0$).

3. The Grover diffusion operator is defined to implement,

$$\hat{U}_s = \hat{1} - 2|T\rangle\langle T|. \quad (41.11)$$

4. repeat $O(N)$ times:

5. $|\varphi_{i+1}\rangle = \hat{U}_s \cdot \hat{U}_T|\varphi_i\rangle$.

6. \square

7. What about measurement stage

Algorithm XXXVI: Grover's algorithm for performing a quantum search over an oracle, exhibiting a quadratic quantum enhancement.

C. Oracles

From Sec. XLI.B we know that the quantum search algorithm requires a black box oracle that evaluates a classical function or database lookup query as a subroutine. How do we construct these oracles?

If we consider the case of solving a satisfiability problem, to find a satisfying input such that the function evaluates to $f(x) = 1$, then it would naively appear that the required quantum operation needs to implement the transformation,

$$\hat{U}_f|x\rangle = |f(x)\rangle. \quad (41.12)$$

However, it's easy to see that in general such a transformation is non-unitary, and therefore cannot be implemented quantum mechanically.

To overcome this obstacle and enable unitary implementation, we introduce additional ancillary bits to $f(x)$, and first write it as a reversible classical circuit, which then always lends itself to a unitary quantum construction. Specifically, if instead of attempting to construct

Eq. (41.12) we introduce some extra qubits and attempt to implement the transformation,

$$\hat{U}_f|x\rangle|0\rangle = |x\rangle|f(x)\rangle, \quad (41.13)$$

then it can easily be seen that this preserves orthonormality and may be implemented unitarily for any function $f(x)$, even if $f(x)$ is not an invertible function.

To provide the simplest possible example, consider the classical XOR gate, given by,

$$y_1 = x_1 \oplus x_2. \quad (41.14)$$

This operation is obviously not reversible (and hence not unitary), since for a given output y_1 , the inputs x_1 and x_2 are not unique. However, if we trivially modify the transformation to be,

$$\begin{aligned} y_1 &= x_1 \oplus x_2, \\ y_2 &= x_2, \end{aligned} \quad (41.15)$$

then a quick back-of-the-envelope calculation verifies that the transformation (simply a CNOT gate) is now reversible, unitary, and output y_1 encodes the desired function evaluation.

Having re-expressed our desired oracle function as a reversible classical circuit, we now simply make direct substitutions of all the reversible classical gates with their logically equivalent quantum counterparts, yielding a quantum implementation of the function that is unitary and preserves quantum coherence. The progression for this example is shown in Fig. 121, and the general procedure for oracle construction is shown in Alg. XXXVII.

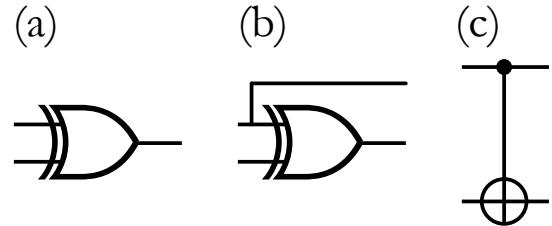


Figure 121: Simple example of the progression of turning a classical function into a quantum oracle. (a) A simple classical circuit, the XOR gate, which is not invertible, implementing 2-bit modulo 2 addition. (b) By introducing an ancillary bit we are able to make a reversible implementation of the same circuit. (c) With a reversible circuit construction, we can directly substitute the reversible gates for their unitary equivalents, yielding a quantum oracle implementation of the starting function.

D. Quantum Fourier transform

The quantum Fourier transform (QFT) is not an algorithm per se, but rather a unitary operator that finds widespread use in other quantum algorithms – a quantum

```

function BuildOracle(f):
    1. Let f be an arbitrary,
       efficiently-computable function in BPP.
    2. Express f as a classical logic circuit.
    3. Make ancillary bits available.
    4. Rewrite the classical circuit as a
       reversible classical circuit, exploiting
       the introduced ancillary bits as
       necessary.
    5. Make direct substitutions of all
       reversible classical gates with their
       quantum counterparts that implement the
       equivalent logical operations in qubit
       space.
    6. Return  $\hat{U}_f$ .
    7.  $\square$ 

```

Algorithm XXXVII: Outline of the general procedure for constructing quantum oracles from classical logic descriptions.

subroutine of sorts. The QFT operator is so ubiquitous as a component in other quantum algorithms, that it warrants special treatment.

The QFT matrix simply contains coefficients taken from the discrete Fourier transform (DFT). Specifically, the $N \times N$ QFT matrix has elements,

$$\hat{\text{QFT}}_{j,k} = \frac{1}{\sqrt{N}} \omega^{(j-1)(k-1)}, \quad (41.16)$$

where,

$$\omega = e^{\frac{2\pi i}{N}}, \quad (41.17)$$

is a complex root of unity.

In matrix form the N -dimensional QFT is therefore,

$$\hat{\text{QFT}}_N = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \dots & \omega^{(N-1)(N-1)} \end{pmatrix}, \quad (41.18)$$

which is symmetric, $\hat{\text{QFT}} = \hat{\text{QFT}}^\top$. Note that all matrix elements are phases with magnitude $1/\sqrt{N}$,

$$|\hat{\text{QFT}}_{i,j}| = \frac{1}{\sqrt{N}} \quad \forall i, j. \quad (41.19)$$

Equivalently, in terms of basis state transformations this implements the map,

$$\hat{\text{QFT}}|j\rangle \rightarrow \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \omega^{jk} |k\rangle, \quad (41.20)$$

on the N basis states.

Unlike the closely-related Hadamard transform, $\hat{H}^{\otimes n}$, which shares many properties⁴⁹ with the QFT, the QFT is a highly entangling operation. The QFT of any dimension has an efficient circuit implementation, making it an important subroutine in many algorithms. Specifically, the n -qubit QFT (i.e a transformation on 2^n amplitudes) requires only $O(n^2)$ elementary gates (Hadamards and CZs). If an approximation of the QFT is sufficient for one's purposes, this can be further reduced to only $O(n \log n)$.

Note that while the circuit only requires $O(n^2)$ gates, it implements a Fourier transform on 2^n amplitudes, giving this subroutine an exponential quantum improvement over classical DFTs.

The circuit construction for the QFT is shown in Alg. XXXVIII.

E. Phase-estimation

The goal of the phase-estimation algorithm (?) is to calculate the eigenvalue of a given unitary operator for a given eigenvector of that unitary. Since the eigenvalues of unitary operators are always phases of amplitude 1, we can write,

$$\hat{U}|\psi\rangle = e^{2\pi i \theta} |\psi\rangle, \quad (41.25)$$

where θ parameterises the phase of the eigenvalue and $|\psi\rangle$ is an eigenvector of \hat{U} .

The algorithm for this is baked from several ingredients. To implement phase-estimation with n bits of precision, we require:

- An n -qubit Hadamard transform, $\hat{H}^{\otimes n}$.
- n controlled- \hat{U} s⁵⁰.
- An n -qubit inverse quantum Fourier transform, $\hat{\text{QFT}}_n^\dagger$ (Sec. XLI.D).

Alg. XXXIX shows the circuit implementation of the phase-estimation algorithm. The output is a binary representation of $2^n \theta$, where the precision is determined entirely by the number of qubits in the output register, which can be arbitrarily large in principle.

The phase-estimation algorithm is not extremely useful as a standalone product, but acts as a necessary subroutine in many important algorithms, such as Shor's algorithm (Sec. XLI.G) and topological data analysis (Sec. XLI.I).

⁴⁹ Like the QFT, all matrix elements of the Hadamard transform have magnitude $1/\sqrt{N}$, differing only in their phases, which are all simple \pm .

⁵⁰ Any unitary has a coherently controlled-unitary equivalent, of the form $\hat{U}_{\text{controlled}} = |0\rangle\langle 0| \otimes \hat{1} + |1\rangle\langle 1| \otimes \hat{U}$. These always have an efficient circuit construction, given \hat{U} .

function QFT($|x\rangle$):

1. Apply the recursively-defined circuit below to the n -qubit input state $|x\rangle$, where,

$$\hat{R}_\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix}, \quad (41.21)$$

is a generalised phase-flip gate.

2. The amplitudes in the n -qubit output state $|y\rangle$ are given by the quantum Fourier transform of the input amplitudes, implementing the transformation,

$$QFT|j\rangle \rightarrow \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \omega^{jk} |k\rangle, \quad (41.22)$$

on the logical basis states, or in matrix form,

$$\frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \omega & \omega^2 & \dots & \omega^{N-1} \\ 1 & \omega^2 & \omega^4 & \dots & \omega^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \omega^{N-1} & \omega^{2(N-1)} & \dots & \omega^{(N-1)(N-1)} \end{pmatrix}, \quad (41.23)$$

where there are,

$$N = 2^n, \quad (41.24)$$

logical basis states.

3. The quantum circuit requires $O(n^2)$ gates.
4. Return $|y\rangle$.
5. \square

Algorithm XXXVIII: The quantum Fourier transform (QFT) algorithm. The circuit is defined recursively, defining the n -qubit QFT in terms of an $n - 1$ -qubit QFT.

F. Quantum simulation

The field of quantum computation was originally inspired by Feynman's observation that quantum systems cannot be efficiently classically simulated, and therefore maybe computers based on quantum principles could handle this problem. Indeed they can, as was shown by (?).

It requires little imagination to recognise that the applications for quantum simulation are enormous, given the multitude of quantum systems under active investigation

function PhaseEstimation(\hat{U} , $|\psi\rangle$, n):

1. For unitary \hat{U} with eigenvector $|\psi\rangle$ and respective eigenvalue $e^{2\pi i\theta}$, apply the circuit below, on the input state $|0\rangle^{\otimes n}|\psi\rangle$.

2. The register of qubits at the output to the QFT encodes $2^n\theta$ in binary representation, where n determines the number of bits of precision in the output.

3. \square

Algorithm XXXIX: Quantum phase-estimation algorithm.

by researchers across countless fields.

Consider a quantum system comprised of a global Hamiltonian, which may be decomposed into smaller local interaction terms,

$$\hat{H} = \sum_{i=1}^N \hat{H}_i, \quad (41.26)$$

where each \hat{H}_i acts on a subspace of dimension m_i within the larger system. The evolution of the entire system is given by the unitary operator,

$$\hat{U} = e^{i\hat{H}t}. \quad (41.27)$$

We wish to simulate this evolution.

From the Baker-Campbell-Hausdorff lemma we can approximate this as,

$$\hat{U} \approx \left(e^{i\hat{H}_1 \frac{t}{n}} \dots e^{i\hat{H}_N \frac{t}{n}} \right)^n + O\left(\frac{t^2}{n}\right). \quad (41.28)$$

This representation effectively decomposes the global evolution into nN discretised stages of,

$$\hat{U}_j = e^{i\hat{H}_j \frac{t}{n}}, \quad (41.29)$$

each of which operates on an m_i -dimensional subspace, and may therefore be directly efficiently implemented as a unitary gate within the circuit model on a quantum computer.

Clearly the error terms vanish in the limit of $n \rightarrow \infty$, whereby the simulation becomes exact. However, this requires an infinite number of gates via infinitesimal discretisation. We would rather approximate the solution

using a finite number of gates. It follows from Eq. (41.28) that for simulation accuracy δ , the number of discrete steps scales as,

$$n = O\left(\frac{t^2}{\delta}\right), \quad (41.30)$$

which scales efficiently with the duration of time being simulated and the accuracy of the simulation.

This approach is efficient and applies to any Hamiltonian which may be decomposed into local terms as per Eq. (41.31). However many other quantum simulation algorithms have since been described for simulating different types of quantum systems with different Hamiltonian structures (??). The algorithm is summarised in Alg. XL.

```
function HamiltonianSimulation( $\hat{H}$ ,  $t$ ,  $\epsilon$ ):
```

1. Hamiltonian to be simulated is of the form,

$$\hat{H} = \sum_{i=1}^N \hat{H}_i, \quad (41.31)$$

where \hat{H}_i are local Hamiltonians operating on low-dimensional spaces, and \hat{H} is the global Hamiltonian for the entire system.

2. For each \hat{H}_i construct a quantum circuit implementing the unitary,

$$\hat{U}_j = e^{i\hat{H}_j \frac{t}{n}}. \quad (41.32)$$

3. The degree of discretisation, n , is chosen according to,

$$n = O\left(\frac{t^2}{\delta}\right), \quad (41.33)$$

where t is evolution time and δ is the accuracy.

4. Apply the circuits for simulating the local Hamiltonians according to the sequence,

$$\hat{U} = \left(\prod_{j=1}^N \hat{U}_j \right)^n. \quad (41.34)$$

5. The algorithm has runtime,

$$O\left(\frac{Nt^2}{\delta}\right). \quad (41.35)$$

6. \square

Algorithm XL: Quantum Hamiltonian simulation algorithm.

G. Integer factorisation

By far the most influential quantum algorithm, and one of the first, is Shor's integer factorisation algorithm (?). The problem is simply to find $x, y \in \mathbb{Z}^+$ given $z = xy$.

While this algorithm is known to reside in **BQP** (since it has an efficient quantum algorithm), it is strongly believed not to be **BQP**-complete. Similarly, while it is known to reside in **NP** (since it can be efficiently classically verified using simple multiplication), it is strongly believed not to be **NP**-complete, thereby placing it in the 'limbo zone' of **NP**-intermediate complexity.

This problem is of immense interest to the field of cryptography, since finding private RSA keys (Sec. XXXIII.C) computationally can be reduced to this problem. An adversary with access to an efficient factoring algorithm could completely compromise RSA cryptography. For this reason, Shor's algorithm is responsible for the large investments made into quantum computing by nation states, and their military and intelligence agencies!

Shor's algorithm works by first reducing integer factorisation to another problem, *period finding*. For the function,

$$f(x) = a^x \bmod N, \quad (41.36)$$

find its period $r \in \mathbb{Z}^+$, the smallest integer such that,

$$f(x) = f(x+r) \bmod N. \quad (41.37)$$

With the ability to solve the period finding problem, an efficient classical algorithm exists for transforming this solution to a solution for factoring.

The algorithm derives its power from a quantum Fourier transform subroutine (Sec. XLI.D), and has runtime,

$$O((\log N)^2 (\log \log N) (\log \log \log N)), \quad (41.38)$$

making it quantum-efficient. By contrast, the best-known classical algorithm, the *general number field sieve*, requires time,

$$O\left((e^{1.9(\log N)^{\frac{1}{3}} (\log \log N)^{\frac{2}{3}}})\right). \quad (41.39)$$

The algorithm is described in Alg. XLI.

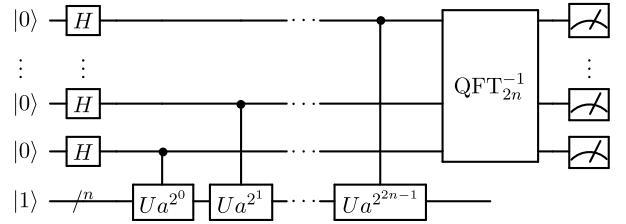
H. Quantum machine learning

With present-day classical computing power becoming ever more powerful, the desire to employ machine learning algorithms has become immense. Machine learning techniques have become indispensable in many fields. Quantum machine learning has emerged as an exciting application for quantum computing power, to enhance the power of machine learning algorithms. We devote the entirety of Sec. LXXIX to this topic.

```

function Shor(a,N):
  1. For function,
     $f(x) = a^x \bmod N,$  (41.40)
    we wish to find the smallest integer  $r$ 
    such that,
     $f(x) = f(x + r) \bmod N.$  (41.41)
  2. Run the circuit below.
  3. Efficient classical processing of the
    measured output registers reveals  $r$  with
    high probability.
  4. Classically verify that,
     $f(x) = f(x + r) \bmod N.$  (41.42)
  5. If it isn't then repeat the circuit to
    obtain more samples.
  6.  $\square$ 

```



Algorithm XLI: Shor's quantum algorithm for integer factorisation.

I. Topological data analysis

The internet currently comprises extraordinary amounts of data, from which useful information must be extracted if this vast amount of data is to be utilised effectively. For example, firms like Google and Facebook must extract meaningful information from their databases of user behaviour in order to make appropriate advertising suggestions. This task is extremely valuable – a small improvement in a recommendation engine, for example, could be worth many millions of dollars in revenue.

Performing analyses like these is extremely computationally challenging when dealing with such enormous datasets as Facebook's user database or Google's website database, so-called *big data analysis*.

One avenue for the analysis of large, complex datasets is via homology theory, which yields *topological data analysis* (TDA). In particular, the so-called *Betti numbers* characterise the nature of interconnectedness within a dataset. Specifically, the k th Betti number, β_k , is the number of k -dimensional holes and voids in a dataset. For example, β_0 , β_1 and β_2 represents the number of connected components, 1-dimensional holes, and 2-dimensional voids in a dataset respectively.

Calculating Betti numbers exactly is known to be

PSPACE-complete⁵¹, and the best-known classical approximation algorithm has exponential runtime,

$$O\left(2^n \log\left(\frac{1}{\delta}\right)\right), \quad (41.43)$$

for accuracy δ on n data-points.

Recently, improved quantum algorithms for approximating the Betti numbers have been presented (???), with polynomial runtime of only,

$$O\left(\frac{n^5}{\delta}\right). \quad (41.44)$$

An elementary photonic experimental demonstration of this algorithm has been performed using a small dataset (?).

Taking a dataset with well-defined distances between data-points⁵², we begin by applying a distance cutoff ϵ to define connections between data-points. We define k -simplices within the dataset, which are fully connected subsets of $k+1$ data-points, with $k(k+1)/2$ undirected edges between them. The full set of simplices defines the dataset's *simplicial complex* for a given distance cutoff ϵ . Construction of the so-called Vietoris-Rips simplicial complex is described in Fig. 122.

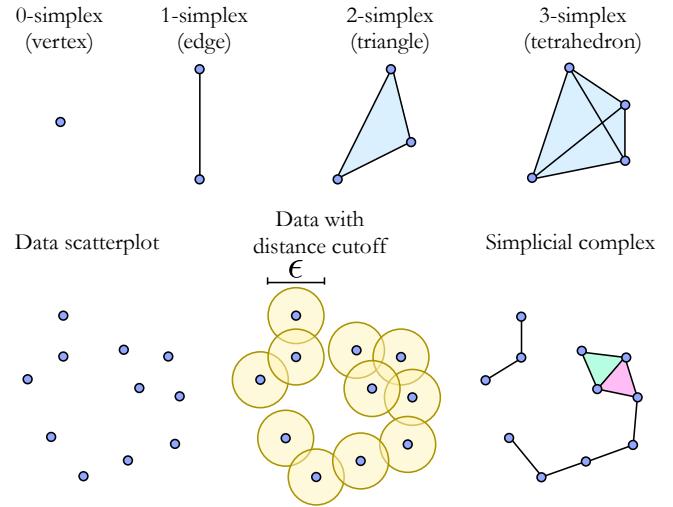


Figure 122: (top) k -simplices are constructed as fully connected subsets of the data of different dimensions k . (bottom) A distance cutoff is applied to create edges within a scatterplot of raw data, from which the simplicial complex is constructed. The shown example of a simplicial complex contains two 2-simplices (coloured triangles), and 6 1-simplices (edges). The final simplicial complex is highly dependent on the choice of cutoff ϵ . In general, as ϵ is increased the complex will contain more simplices of higher dimension, since vertices will have more immediate neighbours.

⁵¹ **PSPACE** is the complexity class of problems requiring polynomial memory with unbounded runtime, a class that is not classically efficient.

⁵² 'Distance' could be any arbitrary metric of any dimension.

The first step in the algorithm is to construct the simplicial complex superposition state,

$$|\psi\rangle_k^\epsilon = \frac{1}{\sqrt{|S_k^\epsilon|}} \sum_{s_k \in S_k^\epsilon} |s_k\rangle, \quad (41.45)$$

where s_k denotes a k -simplex from the simplicial complex S_k^ϵ . This superposition can be constructed by employing a Grover search using a set-membership oracle function,

$$f_\epsilon(s_k) = \begin{cases} 1, & \text{if } s_k \in S_k^\epsilon \\ 0, & \text{otherwise} \end{cases}, \quad (41.46)$$

yielding quadratically enhanced efficiency in the simplicial complex state preparation.

From the superposition state, the uniform mixture of simplices state,

$$\hat{\rho}_k^\epsilon = \frac{1}{|S_k^\epsilon|} \sum_{s_k \in S_k^\epsilon} |s_k\rangle\langle s_k|, \quad (41.47)$$

is easily prepared with the addition of CNOT gates and ancillary qubits⁵³.

The quantum TDA algorithm then takes the simplicial complex mixed state and estimates the full set of Betti numbers by employing a phase-estimation algorithm (Sec. XLIE), which induces an exponential algorithmic runtime improvement. The full TDA algorithm is summarised in Alg. XLII.

Performing the TDA across a range of ϵ yields a *barcode* representation for the dataset's topology. Topological features which persist over large ranges of ϵ can then be regarded as robust features of the dataset, whereas ones which only persist over a small range of ϵ can be regarded as localised, non-persistent features, which might correspond to noise for example, and be filtered out prior to further analysis. The barcode representation thereby gives us an extremely robust characterisation of the topology of the data in a scale-dependent way.

As an example for how this type of technique might be applied, consider Facebook's user database. The distance metric might relate to how similar users' interests are, or how closely related they are via their friendship networks. Then examining the barcode representation of the data by scanning over ϵ would provide insight into the persistence and robustness of these relationships at different scales within the network. At different scales we could investigate topological relationships and features at the level of individuals, family or friendship networks, communities, common interest groups, between cities, across demographic characteristics, or between nations, for example.

⁵³ Using parallel CNOT gates one can transform an arbitrary superposition state into a redundantly encoded equivalent, $\hat{U}_{\text{CNOTs}} \sum_i \lambda_i |\psi_i\rangle |0\rangle \rightarrow \sum_i \lambda_i |\psi_i\rangle |\psi_i\rangle$, following which tracing out the redundant copy takes us to its uniform mixture, $\hat{\rho} = \sum_i |\lambda_i|^2 |\psi_i\rangle\langle\psi_i|$.

function TDA(dataPoints):

1. Implement a Grover search on the set of data-points with set-membership oracle function,

$$f_\epsilon(s_k) = \begin{cases} 1, & \text{if } s_k \in S_k^\epsilon \\ 0, & \text{otherwise} \end{cases}, \quad (41.48)$$

which prepares the simplicial complex superposition state,

$$|\psi\rangle_k^\epsilon = \frac{1}{\sqrt{|S_k^\epsilon|}} \sum_{s_k \in S_k^\epsilon} |s_k\rangle, \quad (41.49)$$

2. Using ancillary qubits, CNOT gates and a trace-out operation, prepare the simplicial complex uniform mixed state,

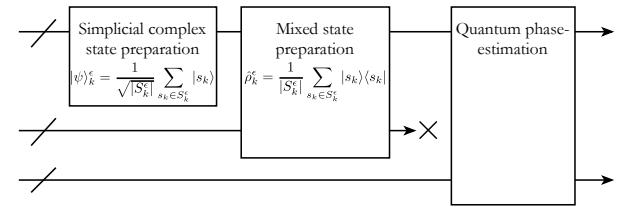
$$\hat{\rho}_k^\epsilon = \frac{1}{|S_k^\epsilon|} \sum_{s_k \in S_k^\epsilon} |s_k\rangle\langle s_k|, \quad (41.50)$$

3. Perform quantum phase-estimation on the simplicial complex mixed state.
4. Classical post-processing reveals the Betti numbers.
5. Algorithm has runtime,

$$O\left(\frac{n^5}{\delta}\right), \quad (41.51)$$

for accuracy δ .

6. \square



Algorithm XLII: Quantum topological data analysis algorithm for calculating Betti numbers.

J. Sampling problems

The algorithms described previously are examples of *decision problems*, whereby the computation answers a question, providing a well-defined output for a well-defined input. Another entirely different class of problems are the so-called *sampling problems*, whereby the goal is to accurately reproduce samples taken from some probability distribution. By their nature, these algorithms are statistical and generally their output cannot be associated with the answer to a decision problem. Nonetheless, despite being an entirely different category of problems, they reside in distinct sampling complexity classes, some of which are classically efficient to simulate, others not.

The simplest example of a classical sampling problem

is the propagation of balls through a Galton board, as shown in Fig. 123. This is a computationally easy problem, whose simulation requires only calculating a binomial distribution, which is classically straightforward.

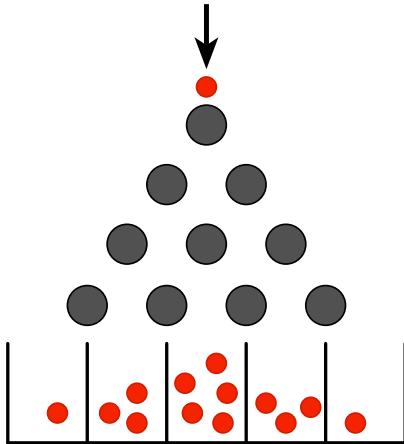


Figure 123: The Galton board yields a very simple classical sampling problem. Balls (red) are input at the top and allowed to fall freely through the pyramid network of pegs (grey), at which balls bounce to the left or right with 50% probability. At the output the balls are collected into buckets, which populate according to a binomial distribution. The computational problem is to produce statistically accurate samples from such a device, which classically is efficiently implemented by sampling the binomial distribution.

An equivalent quantum sampling problem would be to replace the pegs in the Galton board with beam splitters, and the balls with photons. Now we have an analogous problem, but defined in terms of photonic wave-function amplitudes rather than classical transition probabilities. When generalised to the multi-photon context, this problem turns into the so-called BOSONSAMPLING problem, which is discussed in detail in Sec. XLII.D.1, a problem believed to not have an efficient classical simulation algorithm (?).

In computationally hard sampling problems, e.g. quantum ones, the goal is generally not to reconstruct the complete probability distribution. This is generally impossible in the quantum context, since the number of basis states we are sampling from is exponentially large, and therefore cannot be efficiently fully reconstructed. Thus the goal of *sampling* is distinct from the goal of *reconstruction*. In the case of sampling problems, we are satisfied with incomplete data, provided that we only have to take a polynomial number of samples for the sake of computational efficiency.

However, this limitation to having incomplete data makes the problem of verification a conceptually challenging one. We are unable to simply compare our statistical results with that of a reliable reference, since our massively incomplete samples are almost certainly going to be entirely different ones, providing no benchmark for comparison. Therefore verification by comparison is ef-

fectively ruled out, requiring more elaborate verification techniques, something which has become a highly active area of research on its own. Addressing the verification problem is of course an important one – when we reach the milestone of achieving ‘quantum supremacy’, we’d sure like to be able to convincingly prove to the world that we in fact did!

Countless other quantum sampling problems have been described. Most notably, the IQP (instantaneous quantum protocol) sampling problem is a very simple prescription for a quantum algorithm that has been shown to likely be classically hard despite having very shallow circuit depth and all-commuting gates (?).

Although not as obviously algorithmically useful as decision problems, sampling problems have received much interest owing to their generally very simple construction. For example, BOSONSAMPLING requires only single-photon inputs, evolved via a passive linear optics beam-splitter network, and measured using photo-detection. IQP sampling requires only single-qubit Hadamard gates and generalised multi-qubit controlled-phase gates⁵⁴, thereby sampling from the logical state,

$$|\psi_{\text{out}}\rangle = \hat{H}^{\otimes n} \cdot \hat{U}_{\text{CZs}} \cdot \hat{H}^{\otimes n} |0\rangle^{\otimes n}. \quad (41.52)$$

Because CZ gates commute, they can be performed in parallel, giving the IQP circuit construction very low circuit depth. The IQP protocol is shown in Alg. XLIII.

These reduced resource requirements makes both analysis and physical construction of some sampling problems far simpler than a universal quantum computer, yet nonetheless they implement computationally hard problems. For these reasons, many researchers regard non-universal sampling problems as being likely candidates for the first demonstration of quantum supremacy.

K. Shallow quantum circuits

— by Zixin Huang.

As discussed in Sec. LXXXIX, we have good reason to believe that quantum computers have super-classical capabilities. Recently, an unconditional proof of a quantum speed-up has been shown for a particular class of circuits.

It is incredibly difficult to compare quantum versus classical polynomial time computation, and currently we have only noisy gates in the lab, further complicating the goal of achieving quantum supremacy. Given n qubits, circuit depth d , and error rate ϵ , intuitively, we demand the condition,

$$nd \ll \frac{1}{\epsilon}, \quad (41.59)$$

⁵⁴ Generalised CZ gates are simply gates diagonal in the logical basis, where the diagonal elements are arbitrary phases, $\hat{U}_{\text{CZ}} = \text{diag}(e^{i\theta_1}, \dots, e^{i\theta_n})$.

```

function IQP_sampling():
    1. Prepare the  $n$ -qubit state,
         $|\psi_{\text{in}}\rangle = |0\rangle^{\otimes n}.$  (41.53)

    2. Apply the  $n$ -qubit Hadamard transform,
         $\hat{H}^{\otimes n}.$  (41.54)

    3. Apply some choice of generalised CZ gates,
         $\hat{U}_{\text{CZs}} = \text{diag}(e^{i\theta_1}, \dots, e^{i\theta_n}).$  (41.55)

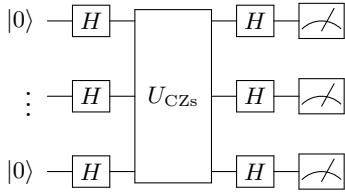
    4. Apply another  $n$ -qubit Hadamard transform,
         $\hat{H}^{\otimes n}.$  (41.56)

    5. The output state is,
         $|\psi_{\text{out}}\rangle = \hat{H}^{\otimes n} \cdot \hat{U}_{\text{CZs}} \cdot \hat{H}^{\otimes n} |0\rangle^{\otimes n}.$  (41.57)

    6. Measure all qubits in the computational basis, yielding bit-string  $\vec{x}$ , which occurs with probability,
         $P_{\vec{x}} = |\langle \vec{x} | \hat{H}^{\otimes n} \cdot \text{diag}(e^{i\theta_1}, \dots, e^{i\theta_n}) \cdot \hat{H}^{\otimes n} |0\rangle^{\otimes n}|^2.$  (41.58)

    7. Repeat protocol  $O(\text{poly}(n))$  times.
    8.  $\square$ 

```



Algorithm XLIII: The IQP sampling problem, which is believed to be a classically hard problem.

to hold, in order for a quantum algorithm to run successfully. If d is large, then noise dominates and our circuit is classically simulable. On the other hand, if we have large n and small d (constant depth circuits), we may observe a quantum speed up in the near future. Given these factors, we are motivated to examine ‘shallow circuits’. Now we will see that there is a distinction between constant depth circuits run using classical versus quantum algorithm.

The result is for the following hidden linear function (HLF) problem (?). Given an $n \times n$ symmetric binary matrix A , we specify a quadratic form,

$$q(\vec{x}) = \vec{x}^T A \vec{x} \pmod{4}. \quad (41.60)$$

The goal is to find a binary vector $\vec{z} \in \{0, 1\}^n$ such that,

$$q(\vec{x}) = 2\vec{z}^T \vec{x}, \quad (41.61)$$

for all \vec{x} in the binary null-space of A .

The quantum algorithm that solves the 2D HLF problem is similar to the Bernstein-Vazirani problem (?). The algorithm, shown in Alg. XLIV, has similar structure to the IQP sampling problem (Alg. XLIII).

```

function ShallowQuantumCircuits():
    1. Prepare the  $n$ -qubit state,
         $|\psi_{\text{in}}\rangle = |0\rangle^{\otimes n}.$  (41.62)

    2. Apply the  $n$ -qubit Hadamard transform,
         $\hat{H}^{\otimes n}.$  (41.63)

    3. Apply the oracle,
         $\hat{U}_q | \vec{x} \rangle = i^{q(\vec{x})} | \vec{x} \rangle,$  (41.64)

        for bit-string  $\vec{x} \in \{0, 1\}^n$ , composed of a product of CZ gates and phase-shifts,
         $\hat{S}_i = \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix}.$  (41.65)

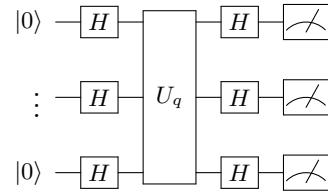
    4. Apply another  $n$ -qubit Hadamard transform,
         $\hat{H}^{\otimes n}.$  (41.66)

    5. The output state is,
         $|\psi_{\text{out}}\rangle = \hat{H}^{\otimes n} \cdot \hat{U}_q \cdot \hat{H}^{\otimes n} |0\rangle^{\otimes n}$ 
         $= \frac{1}{2^n} \sum_{\vec{x}, \vec{z} \in \{0, 1\}^n} i^{q(\vec{x})} (-1)^{\vec{z}^T \vec{x}} | \vec{z} \rangle.$  (41.67)

    6. Measure all qubits in the computational basis, yielding measurement outcomes  $\vec{z} \in \{0, 1\}^n$ , with probabilities,
         $P(\vec{z}) = |\langle \vec{z} | \hat{H}^{\otimes n} \cdot \hat{U}_q \cdot \hat{H}^{\otimes n} |0\rangle^{\otimes n}|^2.$  (41.68)

    7. Repeat protocol  $O(\text{poly}(n))$  times.
    8.  $\square$ 

```



Algorithm XLIV: Quantum computing using shallow circuits, where circuit depth scales as $O(\log d)$.

Here a quantum circuit \hat{U}_q acts on the computational basis states $\vec{x} \in \{0, 1\}^n$ via a generalised controlled-phase operation,

$$\hat{U}_q | \vec{x} \rangle = i^{q(\vec{x})} | \vec{x} \rangle, \quad (41.69)$$

for some oracle function $q(\vec{x})$. The solution is obtained by measuring the state $|\Psi_q\rangle$ in the computational basis.

Here,

$$\begin{aligned} |\Psi_q\rangle &= \hat{H}^{\otimes n} \cdot \hat{U}_q \cdot \hat{H}^{\otimes n} |0\rangle^{\otimes n} \\ &= \frac{1}{2^n} \sum_{\vec{x}, \vec{z} \in \{0,1\}^n} i^{q(\vec{x})} (-1)^{\vec{z}^T \vec{x}} |\vec{z}\rangle. \end{aligned} \quad (41.70)$$

The circuit \hat{U}_q can be decomposed into a product of CZ gates and phase-shift gates, which all commute, thereby allowing the shallow circuit-depth.

If we lay out the circuit in 2D (hence the 2D HLF problem), it only requires classically-controlled Clifford gates between nearest neighbour qubits on a 2D grid. An example is shown in Fig. 124. An instance of the HLF problem can be described by a graph $G(A)$ with n vertices such that the off-diagonal part of A is the adjacency matrix of $G(A)$. Here $A_{i,j} = 0$ unless i, j are nearest neighbour vertices of the grid or $i = j$. We place qubits at each vertex and every edge corresponds to an input bit. The HLF problem is solved by applying nearest-neighbour CZ gates and S_i phase gates at the correct site.

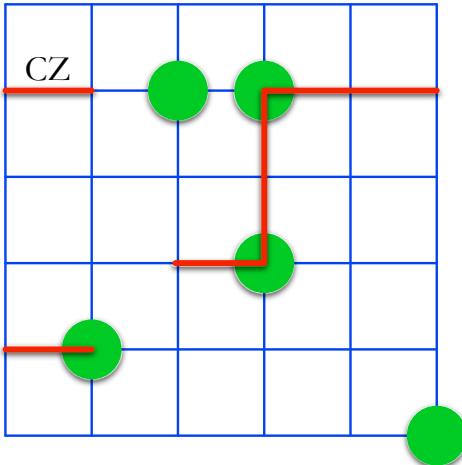


Figure 124: Implementation of the HLF problem on a 2D grid. The qubits are represented as green circles (some omitted for clarity), and CZ gates are drawn in red. To implement the circuit \hat{U}_Q , the qubits at the vertex of the graph $G(A)$ embedded into a 2D grid. Then \hat{U}_q can be decomposed into a product of nearest-neighbour CZ gates and S .

Now, comparing the quantum algorithm to a classical one, it has been proven that any probabilistic classical circuit with bounded input that solves the 2D HLF problem with high probability, $\epsilon \leq 1/8$, must have depth increasing logarithmically with input size (?). The speed-up here is provided by quantum non-locality.

The fact that a problem can be solved with constant depth quantum circuits, but not classical ones shows that there is a clear exponential computational separation between them.

— *Sic semper tyrannis.*

XLI. PHYSICAL ARCHITECTURES FOR QUANTUM COMPUTING

THE models for quantum computation introduced in the previous section are abstractions of algorithms in terms of elementary operations. But elementary operations must ultimately be physically realised. There are countless physical architectures for realising quantum computations, far too many to describe here, each with their own advantages and disadvantages, and it is far from clear which physical architecture(s) will ultimately win the quantum race.

Here we will summarise some of the physical architectures most applicable to networking. Since we reasonably anticipate that future quantum networking will be optically mediated, we focus on pure-optical and hybrid-optical architectures, on the basis that these will naturally lend themselves to optical interfacing.

A. Universal linear optics

With single-photon encoding of qubits in the quantum network, the obvious architecture to implement quantum computation is linear optics quantum computing (LOQC) (?) (KLM), since the states being processed by the computer are of the same form as the states traversing the network. See (??) for excellent introductions to this what has become a very broad and exciting field.

LOQC allows universal quantum computing to be implemented using single-photon polarisation or dual-rail encoding, with only linear optics interactions, i.e. beamsplitter/phase-shifter networks (?), with the addition of quantum memory, and fast-feedforward, whereby some photons are measured, and the remaining part of the optical circuit is dynamically reconfigured based on the measurement outcomes. The former is readily available technology today, and elementary demonstrations have been performed (??), but the latter two have proven to be somewhat more challenging.

Originally it was believed that universal optical quantum computation, specifically the implementation of 2-qubit entangling gates (such as CNOT or CZ gates), would require extremely (and unrealistically) strong optical nonlinearities that implement a non-linear sign-shift (NS) gate,

$$NS : \alpha|0\rangle + \beta|1\rangle + \gamma|2\rangle \rightarrow \alpha|0\rangle + \beta|1\rangle - \gamma|2\rangle, \quad (42.1)$$

in the photon-number basis, up to normalisation (which is determined by the post-selection success probability). That is, it applies a π phase-shift to only the $|2\rangle$ component of a photon-number superposition. The breakthrough result by KLM demonstrated that this is in fact not the case at all. Instead, the NS gate can be implemented nondeterministically using post-selected linear optics. Two such NS gates allow the construction of a single CZ gate.

The construction of the KLM NS and CZ gates are shown in Figs. 125 & 126. Equivalently, a CNOT gate may be trivially constructed via conjugation by Hadamard gates, based on the identity $\hat{H}\hat{Z}\hat{H} = \hat{X}$.

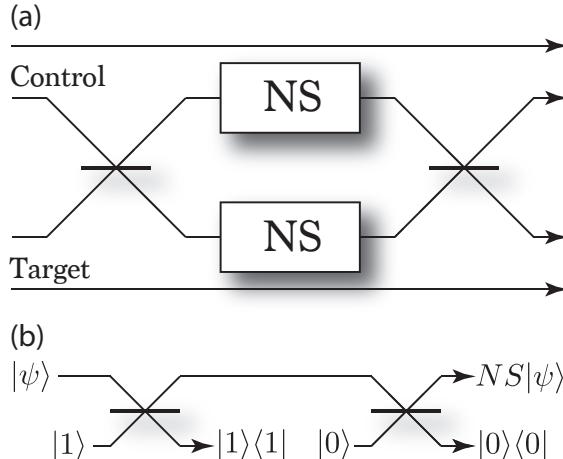


Figure 125: (a) A KLM CZ gate, employing dual-rail encoding, constructed from two non-linear sign-shift (NS) gates, which apply a π phase-shift to only $|2\rangle$ terms in the photon-number basis. (b) Construction of the non-deterministic linear optics NS gate. Two ancillary states – one $|1\rangle$ and one $|0\rangle$ – are employed, and two photo-detectors post-select upon detecting $|1\rangle\langle 1|$ and $|0\rangle\langle 0|$ respectively. The beamsplitter reflectivities in (a) are 50:50, and in (b) chosen such that the amplitudes obey Eq. (42.1).

Clearly this non-determinism is of immediate concern, since concatenating multiple gates would have exponentially decreasing success probability, making the protocol inefficient – if the probability of a single gate succeeding is p , and we require that a circuit comprising n of them all succeed, the success probability is clearly p^n .

The first key observation then is that gate teleportation can be used to shift this non-determinism to a resource state preparation stage, as described in detail in Sec. XXVII.D. However, this is not the end of the story, since gate teleportation requires Bell state projections, which are themselves non-deterministic using purely linear optics (either using PBSs or CNOT gates).

— *Ignotum per ignotius.*

The final insight provided by KLM is that by concatenating these non-deterministic CNOT gates, we can inductively build up higher-level CNOT gates with ever increasing success probabilities, asymptoting to unity with high- (but polynomial-) depth concatenation. By combining these key insights, KLM were able to show that near-deterministic CNOT gates can be constructed using an efficient (polynomial) resource overhead, thereby enabling efficient universal quantum computation⁵⁵. A

⁵⁵ Note that all single-qubit gates are trivially and deterministically

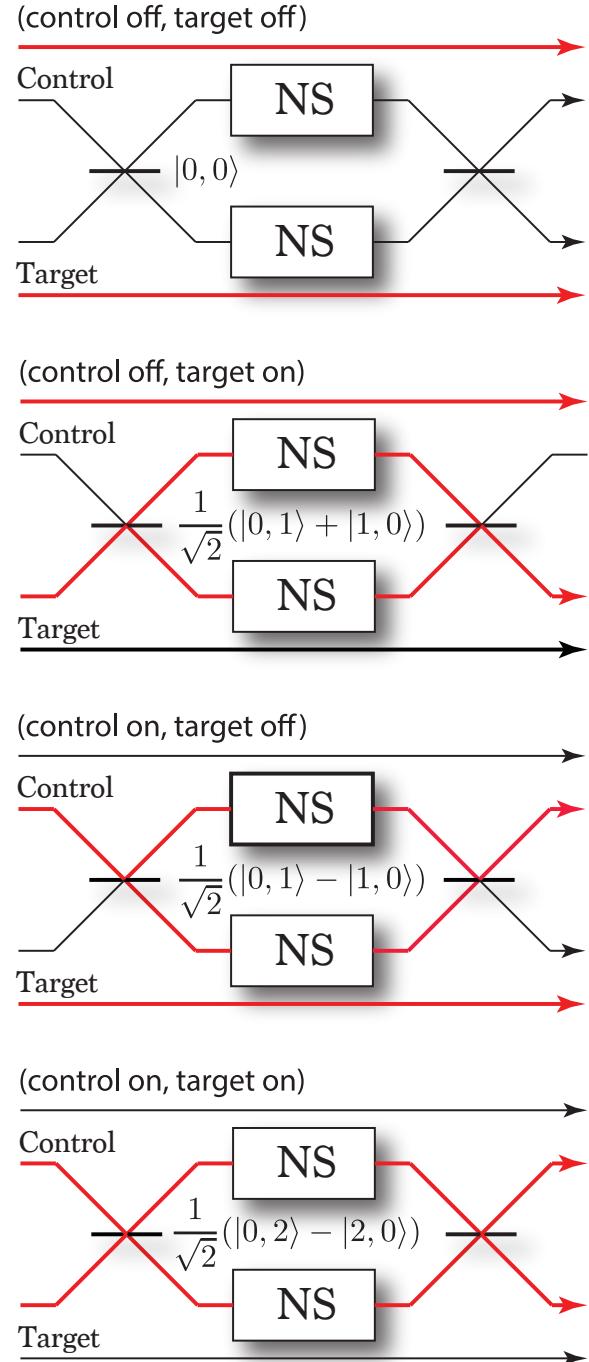


Figure 126: Evolution of the four logical basis states through the KLM CZ gate. The NS gates do nothing in the first three cases, since they are operating only on vacuum and single-photon terms, which are left unchanged by the NS gate. In the last case, where both control and target are on, HOM interference results in photon bunching after the first beamsplitter, thereby creating two-photon terms. These terms inherit the π phase-shift from the NS gate transformation, after which the final beamsplitter reverses the HOM photon bunching, yielding the same logical basis state with an acquired π phase-shift.

sketch of the general KLM formalism is shown in Fig. 127.

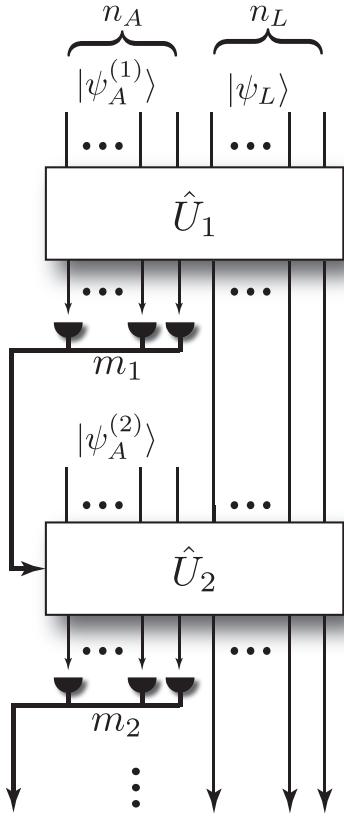


Figure 127: KLM architecture for universal LOQC. n_L optical modes are associated with logical qubits in the state $|\psi_L\rangle$, with the remaining n_A modes acting as ancillary states, $|\psi_A\rangle$. A round of passive linear optics is applied, \hat{U}_1 . Then the ancillary modes are measured, yielding some set of measurement outcomes m_1 . These are classically processed to determine what the next round of passive linear optics, \hat{U}_2 , ought to be. This repeats some polynomial number of times, from which an arbitrary quantum computation can be implemented. The BOSONSAMPLING and quantum walk models are equivalent to taking just the first stage of this protocol: one round of input state, passive linear optics, and measurement.

Evolution via linear optics implements transformations of the form of Eq. (25.1), and may be implemented using the experimental architectures described in Sec. XXV.A and Fig. 86.

The measurements are implemented simply by number-resolved photo-detectors, implementing measurement projectors of the form $\hat{\Pi}_n = |n\rangle\langle n|$, for the measurement outcome of n photons (Sec. XXIV.A).

Since the original presentation of a universal LOQC gate set by KLM, numerous alternate implementations

implemented using wave-plates or beamsplitters, for polarisation or dual-rail encoding respectively. Thus, we need only concern ourselves with the challenges associated with implementing 2-qubit entangling gates.

have been presented and experimentally demonstrated, with various pros and cons (??????).

Significant progress is being made on reconfigurable, integrated LOQC devices (?), but switching times remain orders of magnitude slower than that required for fast-feedforward. The resource overhead associated with overcoming the non-determinism of entangling gates is substantial in the original KLM proposal. But despite being improved upon by cluster state approaches, to be discussed next (Sec. XLIL.B), resource scaling remains daunting. It therefore seems most likely that certain elements from LOQC might be combined into hybrid architectures, to be discussed in detail in Sec. XLII.F.

B. Cluster state linear optics

Although the original KLM scheme is universal, and ‘efficient’⁵⁶, resource usage can be reduced by orders of magnitude by combining concepts from LOQC with the cluster state formalism (Sec. XL.B) or related concepts (??????).

Specifically, instead of using our non-deterministic KLM CZ gates within the circuit model formalism, they could be employed for the preparation of cluster states, since after all a CZ gate directly creates an edge in a cluster state graph.

We now review approaches for cluster state-based LOQC using non-deterministic entangling gates. A further discussion of this topic continues in Sec. XLIII.D, where we introduce modularised quantum computing from a cluster state perspective also using non-deterministic gates. We recommend beginning this topic here, and then skipping ahead to Sec. XLIII.D for continued discussion if interested.

1. Fusion gates

As introduced in Sec. XL.B, a cluster state may be defined by the action of CZ gates upon a graph of qubits initialised into the $|+\rangle$ state. As we saw in the previous section, implementing these CZ gates is troublesome using linear optics, as it is non-deterministic and carries the burden of a large resource overhead. Nonetheless, it was shown early on (?) that by combining non-deterministic CZ gates with the cluster state formalism yields LOQC protocols far more efficient than the original KLM protocol for LOQC.

It was then noted (?) that CZ gates aren’t required at all for the preparation of optical cluster states. Instead, parity measurements (Sec. XXIV.D) operating in a rotated basis

⁵⁶ From a purely computer scientist’s definition of ‘efficient = polynomial’.

may be used to fuse smaller cluster states into larger ones, albeit acting destructively on two of the qubits, and also being non-deterministic, with a success probability of $1/2$. These gates have become known as *fusion gates*, of which there are two types:

- Type-I: destroy only a single photon, but require efficient number-resolved detection.
- Type-II: destroy two photons, but only require on/off detectors, since the gate succeeds upon coincidence events only and preserves photon-number.

Both types of gates have several highly favourable characteristics:

- Unlike the KLM CZ gate, only HOM stability is required (Sec. [XXII](#)). At no stage in the cluster state preparation procedure is any interferometric (i.e wavelength-scale) stability required.
- Gate failure is heralded by measurement of the wrong photon-number.
- Gate failure measures the respective qubits in the computational basis, thereby simply removing those qubits from the cluster state graph, whilst preserving the remainder of the state, which can be ‘recycled’ for reuse.

The explicit construction of the linear optics fusion gates is shown in Fig. [128](#).

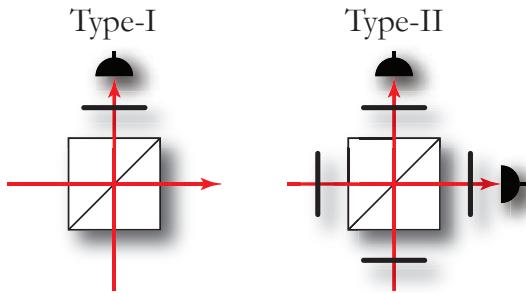


Figure 128: Linear optics cluster state fusion gates for polarisation-encoded photons. Both type-I and type-II gates employ a single polarising beamsplitter to mediate the entangling measurement. The black bars represent Hadamard gates in the polarisation basis (waveplates). The type-I gate only measures a single photon, the other freely exiting the gate, which forms a part of the final cluster state. The type-II gate consumes two photons. Because the type-I gate does not measure in coincidence, as per the type-II gate, it requires number-resolved photo-detection, whereas bucket detectors suffice for type-II.

2. Fusion strategies

If a large cluster state has n edges in its graph, single-shot state preparation will succeed with probability p^n

if individual gates succeed with probability p , implying that on average $1/p^n$ attempts will need to be made until success. Clearly this exponentiality doesn’t lend itself to efficient implementation.

Thankfully, cluster states needn’t be prepared in a single shot, since individual gate failures do not destroy the entire graph, but rather only cause localised damage to the graph in the vicinity of the gate.

Despite their non-determinism, numerous authors have examined approaches for efficiently preparing arbitrarily large cluster states using these destructive, non-deterministic gates ([???](#)). We refer to these schemes as ‘fusion strategies’ – simple algorithms for how to arrange qubits geometrically and the order in which to attempt bonding them.

These principles can be extended beyond LO to other schemes where entangling gates are inherently non-deterministic or sometimes fail in a heralded manner, e.g hybrid architectures (Sec. [XLII.F](#)), where a beam splitter mediates entanglement via which-path erasure, but only successfully projects onto an entangled state with probability $1/2$.

The key feature of all these fusion strategies is to employ ‘micro-clusters’ as a primitive resource, which enable multiple bonding attempts between them via redundant vertices. We will now outline several of these schemes.

a. *Linear clusters* We begin with discussion of linear clusters as these are a particularly useful primitive resource for more advanced strategies. We briefly sketch out the formalism introduced by (?) for linear state preparation, which is applicable to a number of different variants of entangling gates.

A key observation was that although numerous strategies yield efficient state preparation, exact efficiencies are highly dependent on the ordering of bonding operations – which clusters do we choose to bond together first?

Consider a non-deterministic KLM-type CZ gate⁵⁷, which upon failure destroys its two input photons by measuring them in the computational \hat{Z} -basis, and leaves the number of qubits unchanged upon success and bonds them together.

To analyse the operation of such a non-deterministic protocol, we begin by defining a vector \vec{n}_t at time t , which stores purely classical information. Specifically, the vector tells us how many clusters of every length we have stored in memory (except single-qubit clusters, which we assume ‘come for free’⁵⁸). For example, the second element of the

⁵⁷ In reality, no one would use KLM-type gates for preparing cluster states, owing to their complexity compared to fusion gates. Rather, we use this gate for illustrative purposes, since its operation upon success and failure are very simple for exposition.

⁵⁸ As in beer.

vector tells us how many 3-qubit linear clusters we have in our possession.

We then define a strategy, \mathcal{S} , for choosing clusters we have stored and bonding them together to form larger clusters. The strategy acts on our cluster vector and updates it accordingly,

$$\vec{n}_{t+1} = \mathcal{S}(\vec{n}_t). \quad (42.2)$$

That is, the length vector can be thought of as a series of ‘buckets’ containing clusters of different lengths, and the strategy simply probabilistically shuffles the contents of the buckets around each time it is applied. The process can be thought of as a random walk, guided by a probabilistic update rule. Fig. 129 outlines the protocol.

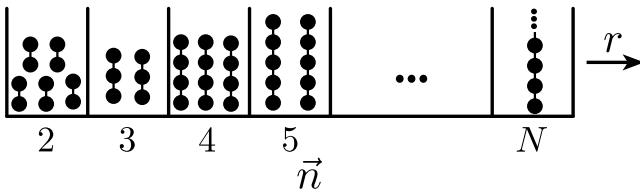


Figure 129: Protocol for preparing linear cluster states using non-deterministic entangling gates and quantum memory. Each ‘bucket’ holds a resource of micro-clusters of a given length, represented by the vector \vec{n} . Beginning with a resource of single qubits we repeatedly attempt bonding operations between micro-clusters in the buckets, according to a fusion strategy, \mathcal{S} . Proceeding as a biased random walk, the contents of the buckets shuffle around until ultimately (hopefully) clusters of the target length N are prepared and steadily flow out as output at rate r clusters per update operation. We assume a free supply of single qubits as a resource.

The strategy description, \mathcal{S} , is also responsible for taking care of updating the elements of \vec{n} according to an update rule, which dictates how many photons are lost or gained upon success or failure of the non-deterministic gate. For example, the CZ gate we have employed here for our toy model destroys two qubits upon failure, but upon success creates a cluster of length given by the sum of the lengths of the clusters acted upon by the gate.

We are then interested in the *rate*, r , at which large clusters are output from the protocol. This is simply extracted by defining a single parameter which counts the number of clusters in \vec{n} above some predetermined target length, and normalises it by the total time taken to reach that point. The rate parameter converges asymptotically for long runtimes. Formally, the rate of preparation is given by,

$$r = \lim_{t \rightarrow \infty} \frac{N_t}{t}, \quad (42.3)$$

where N_t is the total number of clusters of length greater than the target length at time t . The preparation rate is bounded by $0 \leq r \leq 1$. If the rate r converges to a positive, finite value in the limit of large t , this implies

state preparation proceeds in linear time and is therefore efficient.

This completes the theoretical analysis for different strategies and gate types, allowing us to explore different approaches tailored to different physical systems and their varying gate implementations.

One of the key outcomes was that a BALANCED strategy is optimal in terms of preparation rate. This is simply a strategy which preferentially always bonds clusters of equal length, beginning with the largest ones available. Asymmetric strategies, which bond clusters of differing lengths were found to be far less efficient.

Example simulated state preparation rate results are shown in Fig. 130 for various types of entangling gates and gate success probabilities.

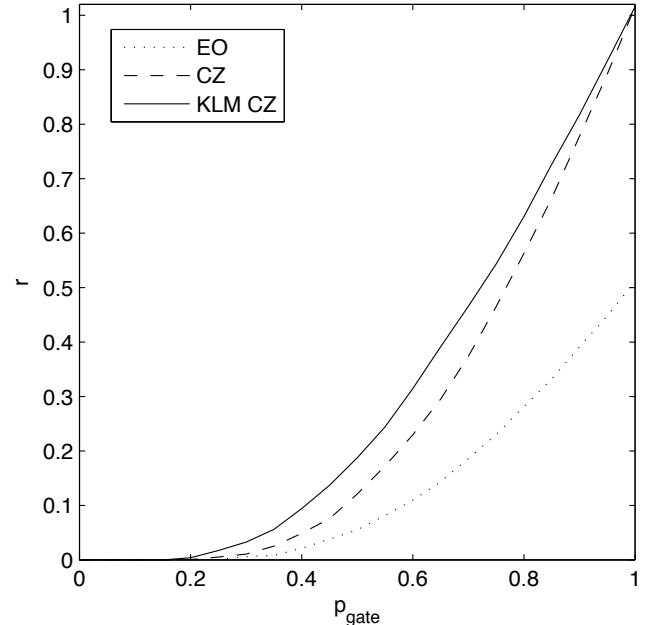


Figure 130: Linear micro-cluster preparation rates for three different types of entangling gates as described in (?), against gate success probability, p_{gate} . Here a BALANCED fusion strategy is employed, whereby we only attempt to join clusters of equal length, always prioritising the largest ones available. This strategy was empirically found to perform better than any asymmetric strategies.

b. Lattice clusters As we learnt from Sec. XL.B, linear cluster states are not universal for quantum computation. What is required is lattices, where the rows and columns respectively map to logical qubits and time in the circuit model. There are numerous approaches one could employ to assemble such lattice clusters using non-deterministic gates, however the easiest to treat for illustrative purposes is to take a resource of linear clusters, prepared as described earlier, and weld them together according to some algorithm, enabling more complex two-dimensional topologies.

The central strategy is similar as for linear clusters – we construct recyclable micro-clusters, which enable multiple bonding attempts, since gate failures only cause localised damage. The key difference now is that these redundant vertices must emanate in multiple directions so as to allow the more complex 2D topology.

Fig. 131 illustrates several topologies for micro-clusters, beginning with the linear micro-cluster that we employed previously for preparing 1D clusters, and two variations of micro-clusters that can be employed for 2D state preparation.

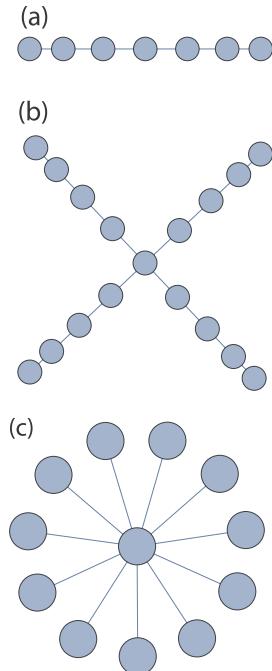


Figure 131: (a) Linear, (b) plus (+), and (c) star micro-cluster states.

The +-cluster simply comprises four linear clusters emanating in the four directions, welded together at a central vertex. It's self-evident how this is subsequently applied to 2D state preparation – we lay out the +-clusters in a grid, and attempt nearest neighbour bonding in each direction for every neighbouring pair of micro-clusters.

The star-cluster similarly allows multiple bonding attempts in each direction. But now the dangling bonds are not uniquely associated with a particular direction, and may therefore be utilised when bonding to a neighbouring micro-cluster in any direction. This implies a modest efficiency improvement, since leftover vertices in any given direction needn't be wasted upon a successful bond in that direction. However, these micro-clusters are not as efficient to prepare as the +-clusters, since they do not straightforwardly arise from two fused linear clusters, which are highly efficient to prepare. Rather they must be prepared via a sequence of repeated successful bonding operations to the central node, where a single gate failure destroys the entire state.

In addition to an efficiency improvement in terms of the number of required physical qubits, minimising the number of redundant qubits that must be removed via \hat{Y} measurements upon completion of the bonding strategy has another key benefit – error accumulation (?). Whenever a cluster state qubit is measured, the action of any error process that acted on that qubit will be teleported to its neighbour(s). For example, if we measure the first qubit in a linear cluster, which was previously acted upon by a depolarising channel, the depolarisation process will be teleported to the neighbouring second qubit in the cluster. Thus, with high levels of redundancy, although this increases our chances of successfully joining two micro-clusters, it similarly increases the accumulation of errors. There is therefore a direct tradeoff between two undesirable error mechanisms – gate failure, and logical errors. This tradeoff must be carefully managed in a real-world implementation.

Having made this observation about error accumulation, is there a topology that is optimal? Yes there is – the so-called snowflake cluster, shown in Fig. 132. Here we take the +-cluster topology and replace the linear clusters emanating in each direction with binary tree graphs of some depth, d . This variation of micro-clusters has been studied in great detail both in optical and non-optical contexts (?).

The endpoints (leaves) of each tree now provide the bonding opportunities for joining two neighbouring micro-clusters. There are $O(2^d)$ such opportunities. The bonding attempts proceed as expected, always exploiting the trees' outermost leaves.

Now the key feature is that when two sub-trees are successfully bonded via their leaves we do not need to measure out *all* the leftover redundant vertices to reduce the graph to the desired residual topology. Instead we can ‘prune’ away entire sub-trees by performing \hat{Z} measurements at the base of their trunks. All vertices above the trunk will thereby be detached from the graph and needn't all be individually measured. Correspondingly, any error processes that had acted on the pruned vertices will not be teleported onto the main cluster, only the measurements acting on the trunks will contribute.

Formally, for a linear subgraph of length n , there will be $O(n)$ leftover redundant qubits on average, which must *all* be measured out using \hat{Y} measurements. Thus, the residual state will have accumulated the action of $O(n)$ independent error processes. On the other hand, for a snowflake subgraph, the trees' depth scales as $d = O(\log n)$, and therefore at most $O(\log n)$ qubits must be measured to prune away unwanted branches. Fig. 133 presents an example of how the pruning process works.

Keeping in mind that for an error process with error rate p , the net probability of an error occurring for m independent channels is $1 - p^m$, thus reducing m from linear to logarithmic is highly favourable in terms of the accumulation of errors.

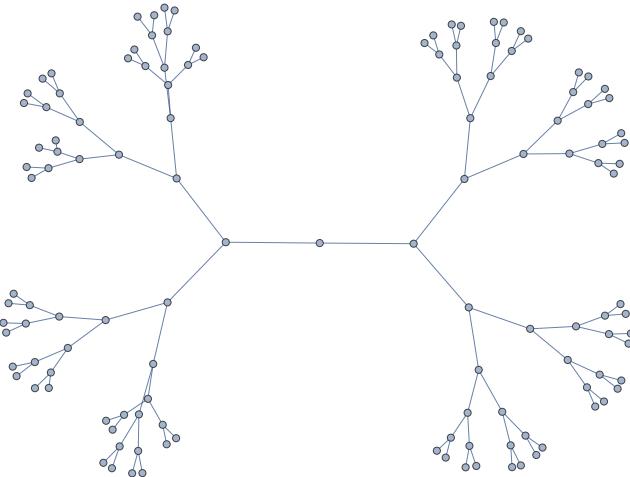


Figure 132: Snowflake micro-cluster comprising a binary tree structure, with depth $d = 7$. Multiple copies of this micro-cluster can be placed side-by-side and fused together via attempting to bond the most outward available leaf qubits from neighbouring clusters. The tree structure allows excess qubits to be ‘pruned’ via their trunks rather than leaves, bypassing the need to measure out every single leftover qubit, as is the case, for example, for $+$ -clusters. This reduces the number of required pruning measurements from linear to logarithmic, similarly reducing the accumulation of errors associated with pruned qubits.

c. *On-demand cluster state preparation* A beautiful feature of the cluster state model is that the entire cluster needn’t be prepared in its entirety for computation to proceed. Instead the state can be grown via the fusion of additional qubits on-demand as the computation proceeds. This arises simply because all the entanglement in the graph is nearest neighbour only, i.e very short-range. So long as a gate failure doesn’t lay its fingers on the leftmost column of qubits in the cluster we are in business. This means fewer quantum memories are required, which are very challenging optically. In non-optical, specifically matter qubit systems, this additionally means that physical qubits can be reused on-the-fly.

The computation therefore proceeds as alternating applications of:

1. Measure the leftmost column of physical qubits to evolve the computation by a single step.
2. Bond on a new column of qubits to the rightmost column.

The qubits in between the left and rightmost columns act as a buffer to give us some leeway when bonding operations fail. The architecture is shown in Fig. 134.

(?) performed an analysis of this approach in the optical context and found that high-depth MBQC can be efficiently implemented using non-deterministic entangling operations, with significantly reduced quantum memory requirements compared to full in-advance state preparation.

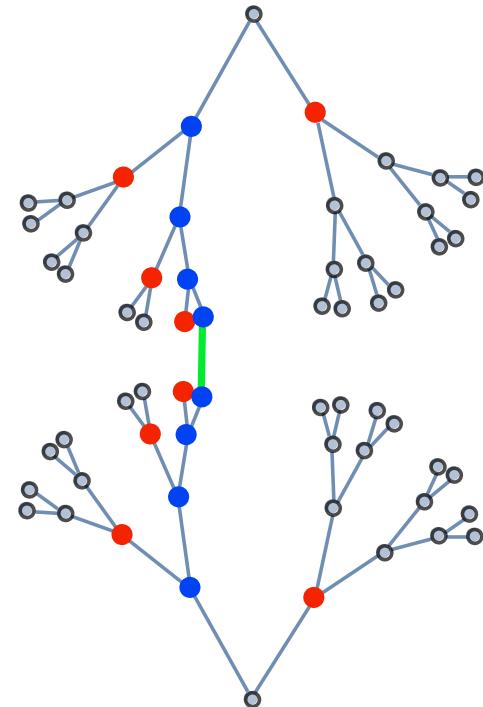


Figure 133: Pruning snowflake micro-clusters upon successful fusion of their outer leaves. Green indicates where the successful fusion operation took place. To remove all redundant nodes we measure the qubits marked in blue in the \hat{Y} -basis and the ones marked in red in the \hat{Z} basis. This will discard all the other qubits marked in grey, modulo the two root qubits at the far top and bottom, which are left with a direct link between them. The total number of measured qubits scales logarithmically with the number of leaves.

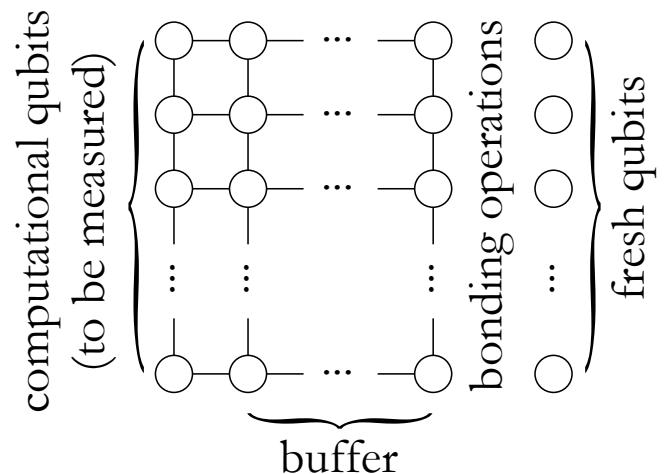


Figure 134: On-demand cluster state preparation. Every time a column of computational qubits are measured away from the lefthand side, thereby evolving the computation by a single step, we dynamically bond on a fresh column of qubits to the righthand side. The buffer in between provides the redundancy necessary when using non-deterministic gates.

In addition to technologically simplifying the architecture by reducing the number of required quantum memories, physical qubits are in existence within the computation for substantially reduced periods of time, since they are only prepared on-demand. This correspondingly reduces error rates.

C. Weak cross-Kerr non-linearities

In Sec. XLII.A we showed that with strong non-linearities at our disposal, scalable photonic quantum computing is possible. However, the interaction strengths of such materials available in the lab today are minuscule compared to the full π phase-shift required for NS and CNOT gate constructions. In LOQC this problem is circumvented using measurement-induced non-linearities, which while being sufficiently strong, are also necessarily non-deterministic, mandating substantial error correction resource overheads to accommodate gate failure.

More recently, as an alternative to using post-selection to simulate strong optical non-linearities, it was shown that by introducing strong coherent states, the strength of the non-linear interaction can be effectively amplified arbitrarily, allowing even very weak non-linearities to be employed for deterministic entangling gate operations (?), compensated for using strong coherent states. Thankfully, strong coherent states are an easy resource to come by nowadays!

In this hybrid linear/non-linear optical architecture, a coherent state is used as a ‘qubus’ (quantum bus), which is entangled with photonic qubits via weak non-linear interactions, thereby mediating long-range entangling operations. This can provide the sufficient entangling power needed to enable scalable, universal quantum computation.

Let us describe the operation of a parity measurement (i.e Bell analyser) device using weak non-linearities. This gate could subsequently be employed as a fusion gate (Sec. XLII.B.1) for building cluster states, and is therefore a resource for universal quantum computation. Simple extensions of this design idea easily extend to other non-trivial operations, such as CNOT and CZ gates or QND measurements.

The key ingredient here is the cross-Kerr interaction, which obeys the Hamiltonian,

$$\hat{H}_{\text{ck}} = \hbar\chi\hat{n}_a\hat{n}_b, \quad (42.4)$$

where \hat{n}_a and \hat{n}_b are the photon-number operators for modes a and b respectively, and χ is the interaction strength, which is typically very small in the lab. This Hamiltonian generates the unitary transformation,

$$\hat{U}_{\text{ck}} = e^{i\theta\hat{n}_a\hat{n}_b}. \quad (42.5)$$

The magnitude of the induced phase-shift is now proportional to photon-number and,

$$\theta = \chi t, \quad (42.6)$$

where t is the interaction time. For strongly entangling gates we need $\theta \approx \pi$.

Applying this operation between a coherent state and a photon-number state implements the two-mode transformation,

$$\hat{U}_{\text{ck}}|\alpha\rangle|n\rangle = |\alpha e^{i\theta n}\rangle|n\rangle. \quad (42.7)$$

Note that the phase-shift accumulated by the coherent state is proportional to the photon-number in the other mode, thereby entangling the two modes via their shared dependence on n , effectively a photon-number-controlled phase-shift operation.

Consider the parity measurement circuit shown in Fig. 135. Let us begin with two polarisation-encoded photonic qubits, which might be separated over long distances, and a coherent state in the shared qubus mode,

$$|\psi_{\text{in}}\rangle = (\alpha_{HH}|H\rangle|H\rangle + \alpha_{HV}|H\rangle|V\rangle + \alpha_{VH}|V\rangle|H\rangle + \alpha_{VV}|V\rangle|V\rangle)|\alpha\rangle. \quad (42.8)$$

Applying the cross-Kerr interactions leaves us in the state,

$$\begin{aligned} \hat{U}_{\text{gate}}|\psi_{\text{in}}\rangle &= [\alpha_{HH}|H\rangle|H\rangle + \alpha_{VV}|V\rangle|V\rangle]|\alpha\rangle \\ &\quad + \alpha_{HV}|H\rangle|V\rangle|\alpha e^{i\theta}\rangle \\ &\quad + \alpha_{VH}|V\rangle|H\rangle|\alpha e^{-i\theta}\rangle. \end{aligned} \quad (42.9)$$

Now the qubus state is in some superposition of $|\alpha\rangle$, $|\alpha e^{i\theta}\rangle$ and $|\alpha e^{-i\theta}\rangle$. The key observation is that these three coherent basis states become highly distinguishable for large $|\alpha|$ with non-zero θ . Specifically, a homodyne measurement on the qubus that projects onto $x = 0$, approximately leaves us in the state,

$$|\psi_{\text{out}}\rangle = \alpha_{HH}|H\rangle|H\rangle + \alpha_{VV}|V\rangle|V\rangle, \quad (42.10)$$

which corresponds to a maximally-entangling parity or Bell projection.

Clearly, since $\langle\alpha|\alpha e^{\pm i\theta}\rangle \neq 0$, there is some probability of error, associated with confusing the coherent basis states and hence their associated photonic qubit states. However, we asymptote towards perfect behaviour in the limit of large coherent qubus amplitudes, since,

$$\lim_{|\alpha|\rightarrow\infty} \langle\alpha|\alpha e^{\pm i\theta}\rangle = 0 \quad \forall \theta \neq 0. \quad (42.11)$$

This type of non-local gate lends itself very naturally to distributed, network-based implementation, where the quantum internet is employed to mediate the qubus, potentially over long distances. Note that unlike polarising beamsplitter-based fusion gates, this gate is non-destructive, and does not require measuring any photonic qubits, only the qubus.

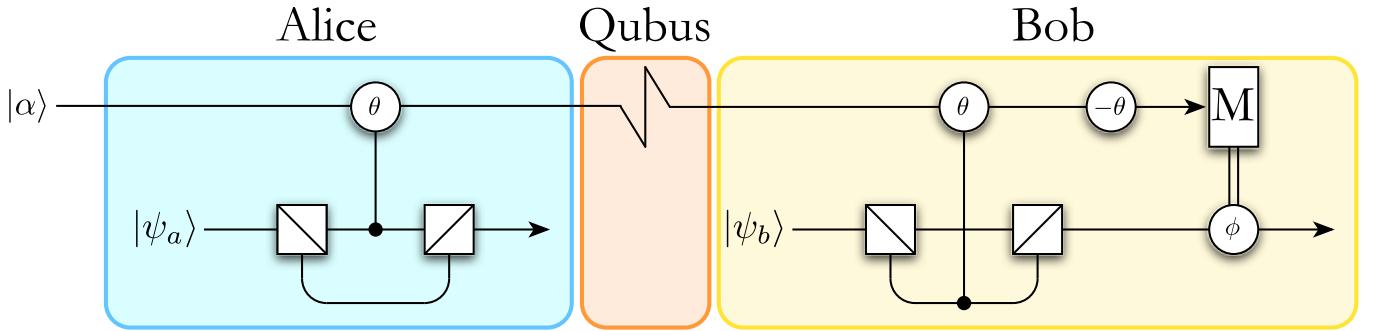


Figure 135: Construction of a two-mode, polarisation-encoded, photonic parity gate using an ancillary qubus coherent state $|\alpha\rangle$, mediating entangling photonic interactions between $|\psi_a\rangle$ and $|\psi_b\rangle$ via weak cross-Kerr non-linearities (denoted by the controlled- θ gates). The polarising beamsplitters are used to switch between polarisation and dual-rail encoding. The qubus is an optical channel, potentially across long distances over the quantum internet for performing distributed gates. The measurement M is a homodyne measurement, which feedforwards an X -quadrature measurement outcome to a local phase correction. This gate could be used for preparing distributed photonic cluster states, a resource for universal quantum computation.

The downside of this protocol, and other qubus-based protocols based on the same idea, is its sensitivity to loss. This is because the qubus is effectively in a cat state (Sec. XIII.E.2), whose sensitivity to decoherence increases rapidly with the coherent amplitude $|\alpha|$. This will effectively place hard limits on how remote Alice and Bob can be in a distributed setting, depending on the loss characteristics of the quantum channel shared between them. It also presents the engineer with a direct tradeoff between decoherence of the qubus (undesirable) and distinguishability of the qubus basis states (desirable), both of which increase with $|\alpha|$.

D. Passive linear optics

While the KLM protocol (and subsequent improvements, e.g. using cluster states) are universal for quantum computing, some of the key technological requirements are very challenging, and unlikely to be achieved in the short-term. However, simplified yet non-universal models for optical quantum computing can abandon some of the more challenging requirements, nonetheless implementing a restricted set of post-classical quantum computations. In particular, we consider protocols requiring only photon-number state preparation, passive linear optics evolution [as per Eq. (25.1)], and photo-detection.

Optically, the two main contenders for this are multi-photon quantum walks (??????) and BOSONSAMPLING (??), both closely related in that they require only passive linear optics and single-photon states, whilst mitigating the need for active switching, quantum memory and dynamic fast-feedforward. Since, evidence has been presented that similar passive linear optics protocols may implement computationally hard problems using states of light other than photon-number states (????).

These protocols involve nothing more than evolving

multiple single-photon states through beamsplitter networks and measuring the output photo-statistics. This is equivalent to just taking the first stage of the KLM protocol shown in Fig. 127.

Both quantum walks and BOSONSAMPLING have been subject to extensive experimental investigation in recent years (??????????).

Because these models are entirely passive, they can be made cloud-based very trivially: Alice prepares her permutation of single photons as the input state, sends it to Bob over the quantum network, who applies the passive operations before returning the state to Alice. In this case, no intermediate client/server interaction is required. Alternately, she could classically communicate a bit-string to Bob indicating the input photon-number configuration, in case she is unable to prepare it herself.

The BOSONSAMPLING and quantum walk models are based on single-photon encoding. However, passive linear optics could also be applied to other states of light. In particular, passive linear optics acting upon multi-mode coherent states implements the *classical* computation of matrix multiplication.

1. BOSONSAMPLING

BOSONSAMPLING is the problem of sampling the output photon-number statistics of a linear optics interferometer fed with single-photon inputs. While not universal for quantum computing (in fact no one has any idea what to use it for at all!), there is strong evidence that it is a classically hard problem (??).

The computational hardness of BOSONSAMPLING relates to the fact that the amplitudes in the output superpositions are proportional to matrix permanents, which are known to be $\#P$ -hard in general. This is believed to be a classically hard complexity class, even harder

than **NP**-complete in the complexity hierarchy, requiring exponential classical time to evaluate (see Fig. 1 for the believed complexity relationships). This yields computationally complex sampling problems.

a. *The BOSONSAMPLING model* For an m -mode interferometer, and input state,

$$|\psi\rangle_{\text{in}} = |T_1, \dots, T_m\rangle, \quad (42.12)$$

where there are T_i photons in the i th input mode, the output superposition takes the form,

$$|\psi\rangle_{\text{out}} = \sum_S \gamma_{S,T} |S_1, \dots, S_m\rangle, \quad (42.13)$$

where S sums over all possible photon-number configurations at the output, of which there are,

$$|S| = \binom{n+m-1}{n}, \quad (42.14)$$

where there are n photons in total in m modes. It is assumed that,

$$m = O(n^2), \quad (42.15)$$

which, for large m , puts us into the anti-bunched (i.e. binary photon-number) regime with high probability⁵⁹, rendering non-number-resolved photo-detectors sufficient for physical implementation. However, this ‘no-collision’ subspace remains exponentially large,

$$|S_{\text{no collision}}| = \binom{m}{n}. \quad (42.16)$$

The amplitudes $\gamma_{S,T}$ are given by,

$$\gamma_{S,T} = \frac{\text{Per}(U_{S,T})}{\sqrt{S_1! \dots S_m! T_1! \dots T_m!}}, \quad (42.17)$$

and the associated configuration probabilities by,

$$\begin{aligned} P_{S,T} &= |\gamma_{S,T}|^2 \\ &= \frac{|\text{Per}(U_{S,T})|^2}{S_1! \dots S_m! T_1! \dots T_m!} \end{aligned} \quad (42.18)$$

where $\text{Per}(\cdot)$ denotes the matrix permanent, and $U_{S,T}$ is a sub-matrix of U – the transfer matrix associated with the particular input-to-output sample configuration – obtained by taking S_i copies of the i th row, and T_j copies of the j th column of the linear optics unitary matrix U .

⁵⁹ That is, we are unlikely to observe more than a single photon in any given output mode, placing us into a binary photo-detection regime. This condition has become known as the ‘bosonic birthday paradox’ (?).

For the purposes of the original complexity proof, the unitary is chosen randomly from the Haar-measure⁶⁰, although it remains an open question as to what is the full class of unitaries that yield computationally hard problems.

b. *The relationship to matrix permanents* The observation that output probability amplitudes are related to matrix permanents as per Eq. (42.17) is the most important one, as this is ultimately responsible for the computational hardness of the BOSONSAMPLING problem, since permanents are in general $\#P$ -hard problems, a classically inefficient complexity class.

The permanent of a square matrix is defined as,

$$\text{Per}(A) = \sum_{\sigma \in S_n} \prod_{i=1}^n A_{i,\sigma_i}, \quad (42.19)$$

which sums over $n!$ terms (super-exponential), where S_n is the symmetric group, the group of permutations on n elements, of which there are $n!$. Note the similarity with the definition for the matrix determinant, defines identically, but with the addition of an alternating \pm -sign in the terms. Despite this similarity, determinants reside in **P**, with an efficient classical algorithm. For this reason, Fermionic sampling yields an easy computational problem, since Fermionic sampling differs only in replacing the permanent with the determinant. The best-known classical algorithm for evaluating permanents by (?) has exponential runtime,

$$O(2^n n^2). \quad (42.20)$$

To see how matrix permanents naturally arise in this setting, it is easiest to explain by example. In Fig. 136 we illustrate a simple interferometer, fed with two photons. We wish to calculate the output amplitude of measuring a photon in each of the modes 2 and 3, given photons input at modes 1 and 2. To evaluate this amplitude we simply need to add up the amplitudes of all possible paths yielding the desired outcome. In this simple example this sum-of-paths is given by,

$$\begin{aligned} \gamma_{\{2,3\}} &= \underbrace{U_{1,2} U_{2,3}}_{\text{don't swap}} + \underbrace{U_{1,3} U_{2,2}}_{\text{swap}} \\ &= \text{Per} \begin{pmatrix} U_{1,2} & U_{2,2} \\ U_{1,3} & U_{2,3} \end{pmatrix}, \end{aligned} \quad (42.21)$$

from which it is immediately clear that the amplitude is given by the sum of $2! = 2$ paths⁶¹, the permanent of

⁶⁰ The Haar-measure generalises the notion of a uniform distribution to higher-dimensional topologies than the real numbers, in this instance to the $\text{SU}(n)$ group.

⁶¹ The number of paths scales as $n!$ in general, which corresponds to the $n!$ order of the symmetric group, S_n , in the definition of the permanent from Eq. (42.19).

the 2×2 matrix obtained from taking the columns (rows) of \hat{U} where a photon is present at the respective input (output) mode.

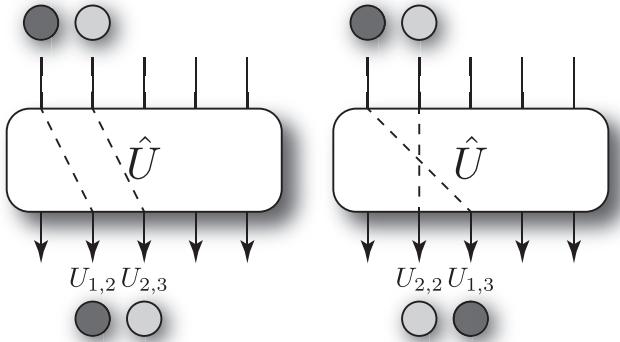


Figure 136: A linear optics interferometer \hat{U} , fed with 2 single-photon inputs, one in each of the first two modes. To calculate the output amplitude of one photon in each of the modes 2 and 3, we sum the amplitudes of all possible paths consistent with that output. In this example there are only two such paths – either both photons pass straight through, or they swap positions. This summation yields a 2×2 matrix permanent, given by Eq. (42.21).

In Fig. 137 we present to next most sophisticated example of an interferometer fed by 3 photons, for which the sum-of-paths has $3! = 6$ terms, given by,

$$\begin{aligned} \gamma_{\{1,2,3\}} &= U_{1,1}U_{2,2}U_{3,3} + U_{1,1}U_{3,2}U_{2,3} \\ &\quad + U_{2,1}U_{1,2}U_{3,3} + U_{2,1}U_{3,2}U_{1,3} \\ &\quad + U_{3,1}U_{1,2}U_{2,3} + U_{3,1}U_{2,2}U_{1,3} \\ &= \text{Per} \begin{pmatrix} U_{1,1} & U_{2,1} & U_{3,1} \\ U_{1,2} & U_{2,2} & U_{3,2} \\ U_{1,3} & U_{2,3} & U_{3,3} \end{pmatrix}, \end{aligned} \quad (42.22)$$

and it is now clear upon inspection that the amplitude is given by a 3×3 matrix permanent.

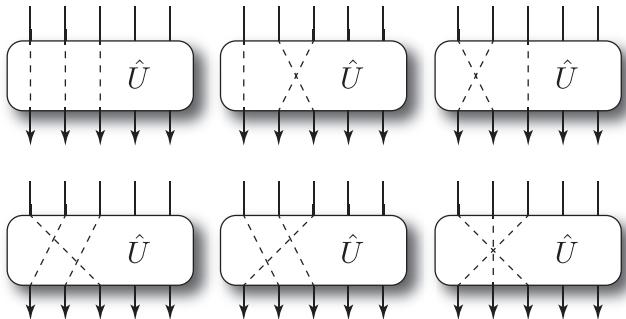


Figure 137: A linear optics interferometer \hat{U} , fed with 3 single-photon inputs, in modes 1, 2 and 3. To calculate the output amplitude of one photon in each of the modes 1, 2 and 3, we sum the amplitudes of all possible paths consistent with that output. This summation yields a 3×3 matrix permanent, given by Eq. (42.22).

c. Problem description The computational problem is simply to sample this probability distribution $P_{S,T}$, which the linear optics network can implement efficiently, but it is believed a classical computer cannot. The full model is shown in Fig. 138.

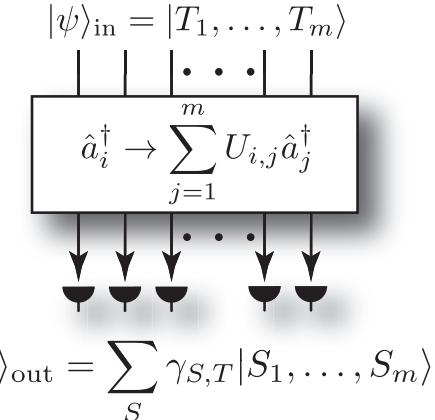


Figure 138: The BOSONSAMPLING model for non-universal linear optics quantum computing. S (output) and T (input) are represented in the photon-number basis. After application of the Haar-random linear optics unitary to the input multi-mode Fock state, the output superposition is sampled with coincidence photo-detection.

For comparison, the equivalent classical protocol using distinguishable photons that evolve independently through the network would be described by,

$$P_{S,T} = \frac{\text{Per}(|U_{S,T}|^2)}{S_1! \dots S_m! T_1! \dots T_m!}, \quad (42.23)$$

which yields a classically efficient sampling problem. Thus, for BOSONSAMPLING the permanents are of complex-valued matrices, whereas for the equivalent classical problem the permanents are of positive real-valued matrices.

Very importantly, note that BOSONSAMPLING does *not* let us efficiently *calculate* matrix permanents. Rather, it samples across a distribution of an exponential number of permanents. This is because, for an exponentially large sample space, with only a polynomial number of measurement trials, we are unlikely to gain more than binary accuracy about individual amplitudes, which is insufficient for determining any particular permanent. It appears that God knows how to efficiently solve matrix permanents, but conspires against us such that we remain ignorant of them. — *Deus magnus est.*

The size of a boson-sampler required to exhibit post-classicality is under active debate, as has undergone much historical revision (?). But some recent estimates suggest that as many as $n = 50$ photons in $m = 2,500$ modes might be a rough guide for such a threshold (?). Needless to say, this is already an extremely challenging technological goal, suggesting that although the BOSONSAMPLING problem is far simpler than universal LOQC, it is far from simple.

BOSONSAMPLING in the presence of various error models, such as loss, source non-determinism and mode-mismatch, has been extensively investigated (?????).

d. Multiplexed BOSONSAMPLING As discussed in Sec. [XXIII.B](#), SPDC is the most common present-day implementation of single-photon sources. However, despite their ready availability, they suffer from non-determinism, with single-photon heralding probability given by Eq. (23.3). To improve upon this, multiplexed sources can be employed (?), improving effective single-photon preparation probabilities asymptotically to unity, as given by Eq. (23.4).

However, rather than employing a multiplexed SPDC source in place of each of the required n single-photons, we can instead employ a larger multiplexer that routes $N \gg n$ sources to n modes, which is far more efficient than n independent multiplexed single-photon sources.

The model is shown in Fig. 139. We begin by operating N SPDC sources in parallel. Clearly if N is sufficiently large with respect to n , it becomes asymptotically certain that at least n photons will be heralded. When this occurs, the successfully prepared n photons – in whatever configuration they happen to occur – are routed to the first n modes of the BOSONSAMPLING interferometer \hat{U} by the multiplexer (which is classically controlled by the SPDC heralding outcomes), and the protocol proceeds as usual.

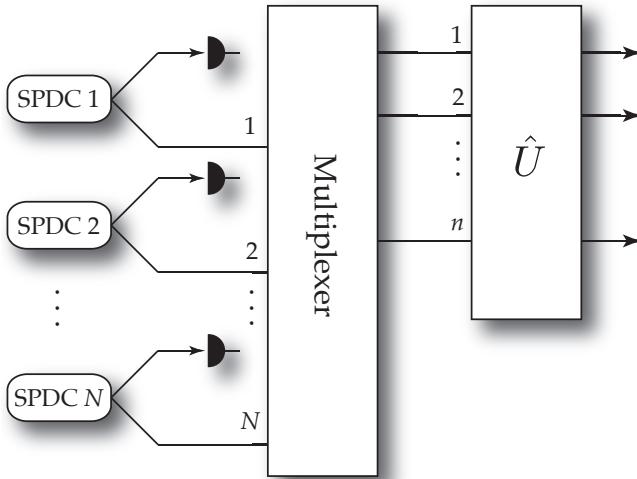


Figure 139: Model for multiplexed BOSONSAMPLING. We operate $N \gg n$ SPDC sources in parallel, which are multiplexed to the first n modes of the interferometer \hat{U} . With sufficiently large N it becomes asymptotically certain that at least n single-photons will be heralded, thereby successfully preparing the desired BOSONSAMPLING input state.

Specifically, the probability of at least n successful

single-photon heralding events occurring is,

$$P_{\geq n} = \sum_{i=n}^{\infty} \binom{N}{i} P_{\text{herald}}^i (1 - P_{\text{herald}})^{N-i}, \quad (42.24)$$

where,

$$P_{\text{herald}} = \chi^2(1 - \chi^2), \quad (42.25)$$

is the probability of a single SPDC source heralding the preparation of a single-photon. This quantity asymptotes to unity for $N \gg n$, as shown in Fig. 140.

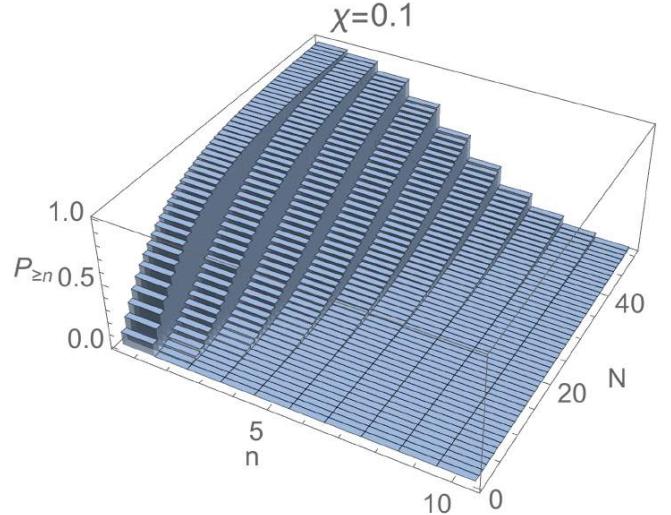


Figure 140: Probability of successfully preparing at least n photons for BOSONSAMPLING from a multiplexed source comprising a bank of N SPDC sources in parallel. For sufficiently large N , we prepare the desired n photons with probability asymptoting to unity.

However, although this procedure works in-principle, it comes at the expense of a large number of sources, N , and more challengingly, fast-feedforward. Keep in mind that if we were able to perform complex fast-feedforward, we might be able to do much more (and far more interesting things) than just BOSONSAMPLING in the first place (i.e universal LOQC)!

e. Scattershot BOSONSAMPLING A variation on SPDC-based BOSONSAMPLING, known as ‘scattershot’ BOSONSAMPLING, has been presented (?), which obviates the difficulty of fast multiplexing in the approach described previously. Here, rather than inputting an SPDC source into the first n of the m modes, we input a source into *every* mode, i.e m sources in total. We then accept all events with n heralding successes in total, irrespective of the configuration in which they occur. This has the effect of implementing n -photon BOSONSAMPLING with an additional layer of randomisation on the input modes (i.e a randomisation in the input configuration, T , which is ordinarily fixed). However, since the algorithm is already

randomised, this additional layer of randomisation does not undermine the complexity proofs, which hold as is. The scattershot model is shown in Fig. 141.

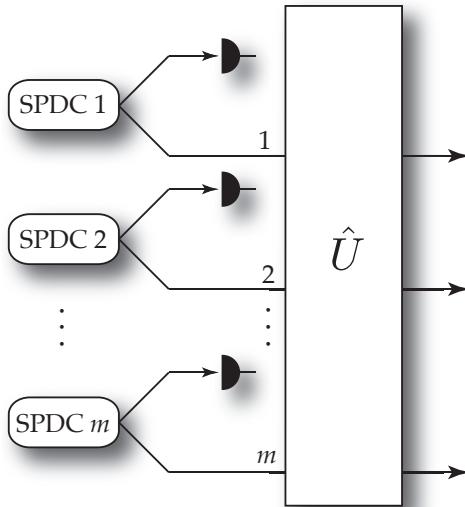


Figure 141: Model for ‘scattershot’ BOSONSAMPLING. An SPDC source is inputted into all m input modes. We post-select upon detecting a total of n photons in the heralding modes, irrespective of their configuration, yielding an n -photon instance of BOSONSAMPLING with randomised input configuration. Unlike multiplexed architectures, the scheme remains entirely passive, without requiring adaptive fast-feedforward.

By keeping all configurations of n photons, rather than just the $|T\rangle = |1\rangle^{\otimes n} |0\rangle^{\otimes(m-n)}$ case, we effectively boost the n -photon heralding probability from,

$$P_n = \chi^{2n} (1 - \chi^2)^n, \quad (42.26)$$

to,

$$P_n = \binom{n^2}{n} \chi^{2n} (1 - \chi^2)^{n^2}, \quad (42.27)$$

exhibiting a binomial enhancement in n -photon events, yielding a significant improvement in count-rates. For a given desired photon-number n , choosing the value for the squeezing parameter, χ , which maximises P_n , we obtain the optimised success probability,

$$P_n^{(\text{opt})} \approx \frac{1}{e\sqrt{2\pi(n-1)}}, \quad (42.28)$$

which exhibits only polynomial scaling against photon-number n , and is therefore scalable. This is shown in Fig. 142. This is in stark contrast to conventional BOSONSAMPLING, where the success probability decays exponentially with photon-number, and is therefore inefficient. Importantly, unlike the multiplexed approach, this efficiency improvement does not require any active elements, remaining in the true spirit of BOSONSAMPLING.

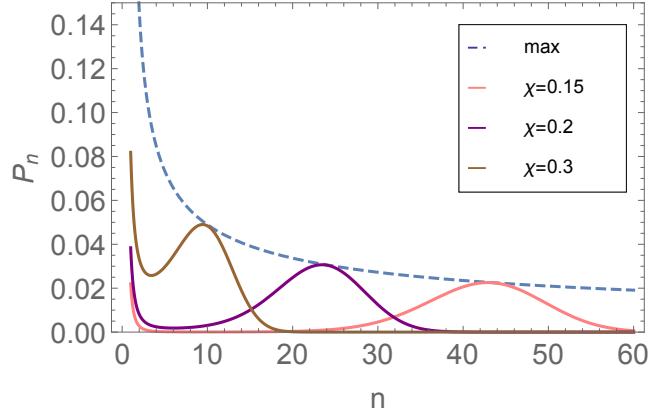


Figure 142: Probability of successfully implementing an instance of n -photon BOSONSAMPLING using the scattershot technique, whereby all input modes are fed with an SPDC source, and all n -photon heralding events are accepted, irrespective of their configuration.

2. Coherent states

A linear optics network acting on a tensor product of coherent state inputs implements simple matrix multiplication on the vector of coherent amplitudes. Specifically, Eq. (25.1) implies that for the input multi-mode coherent state,

$$|\psi\rangle_{\text{in}} = |\vec{\alpha}\rangle = |\alpha_1, \dots, \alpha_m\rangle, \quad (42.29)$$

where α_i is the coherent amplitude of the i th mode, the linear map now takes the form,

$$\beta_i = \sum_{j=1}^m U_{i,j} \alpha_j, \quad (42.30)$$

where the output state is the separable multi-mode coherent state,

$$|\psi\rangle_{\text{out}} = |\vec{\beta}\rangle = |\beta_1, \dots, \beta_m\rangle. \quad (42.31)$$

Equivalently, this could simply be expressed as the matrix equation,

$$\vec{\beta} = U \cdot \vec{\alpha}. \quad (42.32)$$

Of course this is not strictly a *quantum* computation, since:

- It can be efficiently classically computed using $O(m^2)$ operations⁶², thus residing in \mathbf{P} .
- Coherent states are considered classical states (i.e approximated by laser light) with strictly positive Wigner and P -functions.

⁶² Using the naïve element-wise approach, which can be further improved upon using more sophisticated contemporary algorithms.

- There is no entanglement between modes.
- The algorithm offers no quantum (exponential) speedup.

Despite offering no direct quantum advantage, we introduce this model for restricted computation, since it lends itself very elegantly to a form of homomorphic encryption, to be described in detail in Sec. [XLIV.D.2](#).

The applications for matrix multiplication needn't be stated, as it forms such a ubiquitous elementary primitive throughout linear algebra and in solving systems of differential equations, with applications too many to count.

3. Other linear optics sampling problems

Beyond photonic BOSONSAMPLING, much investigation has explored the computational hardness of other types of linear optics sampling problems, using states beyond just single photons.

a. Hard problems In addition to photonic BOSONSAMPLING, several authors have presented strong evidence that other classes of quantum states of light exist, which yield computationally complex sampling problems under the action of linear optics. Most notably, such evidence has been provided for the following:

- (?) considered two-mode squeezed vacuum (or SPDC) states, a type of Gaussian state with strictly positive Wigner function. This is the same as the scattershot model presented in Sec. [XLII.D.1](#).

$$|\psi\rangle_{\text{in}} = \sqrt{1 - \chi^2} \sum_{n=0}^{\infty} \chi^n |n, n\rangle. \quad (42.33)$$

- (?) considered photon-added coherent states and displaced single-photon states.

$$\begin{aligned} |\psi\rangle_{\text{in}} &\propto \hat{a}^\dagger |\alpha\rangle, \\ |\psi\rangle_{\text{in}} &\propto \hat{D}(\alpha) |1\rangle. \end{aligned} \quad (42.34)$$

- (?) considered photon-added or -subtracted squeezed vacuum states.

$$\begin{aligned} |\psi\rangle_{\text{in}} &\propto \hat{a}^\dagger \hat{S}(\chi) |0\rangle, \\ |\psi\rangle_{\text{in}} &\propto \hat{a} \hat{S}(\chi) |0\rangle. \end{aligned} \quad (42.35)$$

- (?) considered ‘cat’ states – superpositions of coherent states.

$$|\psi\rangle_{\text{in}} \propto |\alpha\rangle \pm |-\alpha\rangle. \quad (42.36)$$

Preparation of all of these classes of quantum states of light present their own technological challenges, some very daunting, and all much harder to prepare than single-photons. Thus, the ability to outsource their preparation would be a useful application for the quantum cloud.

b. Easy problems On the other hand, some classes of optical states are known to be efficiently classically simulable under linear optics evolution and photo-detection. This includes coherent states, thermal states, or any state with strictly positive P -function⁶³ (Sec. [XIII.E](#)) (??). Furthermore, Gaussian states evolved via linear optics and measured using Gaussian measurements have been shown to be computationally easy to simulate (??).

While such negative results might be somewhat depressing, it is extremely insightful to understand these regimes, in the interest of avoiding investing excruciating effort into trying to instead fruitlessly prove that they are hard.

Cite Saleh/Carlton paper on most optical states generating entanglement. Is it exponential entanglement? If so, is this a necessary but not sufficient condition for computational complexity.

The simplest example of an easy such problem is coherent state linear optics, as discussed in Sec. [XLII.D.2](#). Taking this notion further, recall from Sec. [XIII.E](#) that one of the phase-space representations for generic optical states is the P -function, which represents a density operator as a sum over coherent states,

$$\hat{\rho} = \iint P(\alpha) |\alpha\rangle \langle \alpha| d^2\alpha. \quad (42.37)$$

Here $P(\alpha)$ is a quasi-probability function. Importantly, iff the P -function is strictly non-negative, $P(\alpha) \geq 0 \forall \alpha$, the optical state may be trivially interpreted as a purely classical mixture of coherent states ($P(\alpha)$ would be a delta function for pure coherent states). If, on the other hand, the P -function exhibits negativity for any α , this interpretation breaks down and is indicative of the state exhibiting non-classical behaviour.

Alg. [XLV](#) describes an efficient classical algorithm for simulating the output photo-statistics of a linear optics sampler fed with strictly non-negative P -function input states (?).

4. Quantum walks

Photonic quantum walks (QWs) are the other main contender for implementing restricted quantum computation,

⁶³ A strictly positive P -function implies that the state can be considered a purely classical mixture of coherent states (each of which are classically efficient to simulate), according to some classical probability distribution.

```

function SimulatePositiveP( $\vec{P}$ ) :
  1. for(m ∈ modes) {
    2. Randomly choose a sample  $\alpha_m$  from
       probability distribution function  $P_m(\alpha)$ .
  3. }
  4. Evolve the set of input coherent state
     samples through the linear optics network,
     
$$\vec{\beta} = U \cdot \vec{\alpha}. \quad (42.38)$$

  5. for(m ∈ modes) {
    6. The probability of measuring  $n$ 
       photons in the  $m$ th mode is given by the
       distribution,
       
$$D_{m,n} = |\langle \beta_m | n \rangle|^2
                  = e^{-|\beta_m|^2} \frac{\beta_m^{2n}}{n!}. \quad (42.39)$$

    7. Choose  $n$  from this distribution.
  8. }
  9. return( $\vec{n}$ ).
 10.  $\square$ 

```

Algorithm XLV: Efficient classical algorithm for simulating any linear optics sampling problem whose input states have strictly non-negative P -functions, with output measured via photon-counting. \vec{P} is the vector of P -functions for all modes.

without requiring the full spectrum of challenging LOQC operations. The resource requirements are the same as for **BOSONSAMPLING**, the difference being that now instead of choosing a Haar-random unitary matrix for the interferometer, we choose one which encodes a graph. The photons are now referred to as ‘walkers’, and they evolve by following edges within the graph, ‘hopping’ between neighbouring vertices.

With only a single walker (photon), nothing computationally complex can occur in the system, since a single photon evolving under passive linear optics can be efficiently classically simulated⁶⁴. However, once multiple walkers are introduced we have a system with almost identical features to **BOSONSAMPLING**, differing only in the structure of the linear optics unitary.

There are two predominant varieties of quantum walks: discrete- (?) and continuous-time (?), which we will now introduce. Algorithms have been described for both the discrete- and continuous-time QW models.

⁶⁴ Note that the literature has described QW schemes, both discrete-time (?) and continuous-time (?), that are universal for quantum computation. However, such universal schemes require an exponential number of vertices in the underlying graph, which clearly does not lend itself to efficient optical representation.

a. *Continuous-time quantum walks* In the continuous-time QW model, a Hamiltonian, \hat{H}_{QW} , encoding the (Hermitian) adjacency matrix of the QW’s graph evolves the walker(s), generating a unitary evolution of the form,

$$\hat{U}_{\text{QW}}(t) = e^{-i\hat{H}_{\text{QW}}t}, \quad (42.40)$$

where $t \in \mathbb{R}_+$.

This model lends itself readily to optical wave-guide implementation, where evanescent coupling between neighbouring wave-guides is inherently a continuous-time process. Fig. 86(c) illustrates an example implementation of a linear optics, continuous-time quantum walk on a line in an integrated wave-guide device.

In the context of linear optics, the evolution is best described using the coupled oscillator Hamiltonian,

$$\hat{H}_{\text{QW}} = \sum_{i,j=1}^m c_{i,j} \hat{a}_i^\dagger \hat{a}_j, \quad (42.41)$$

where \hat{a}_i^\dagger (\hat{a}_i) is the photonic creation (annihilation) operator for the i th of the m modes, and the Hermitian matrix $c_{i,j}$ encodes the coupling strength between the i th and j th modes, which could correspond identically to the QW’s graph adjacency matrix.

b. *Discrete-time quantum walks* In the discrete-time QW model, each walker has access to an ancillary ‘coin’ Hilbert space, which is used to record the direction of the walker through the graph. At each discrete time-step the coin is used to update the position (vertex) of the walker, before applying a unitary ‘coin’ operator to the coin Hilbert space. The addition of the coin space is necessary to enable such quantum walks to reside on arbitrary graph topologies, whilst retaining unitarity in their evolution.

We will briefly summarise the discrete-time QW model, as it most readily lends itself to linear optics implementation, and illustrate the parallels with **BOSONSAMPLING**. First, let us consider the standard simple example scenario of a single quantum walker, on a linear graph topology, using a Hadamard coin,

$$\hat{C} = \hat{H} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (42.42)$$

which has been experimentally demonstrated using both bulk-optics (?) and time-bin encoding (??). The walker is defined by two Hilbert spaces: the position x , and the coin $c = \pm 1$, where $+1$ (-1) indicates that the walker is moving to the right (left). The basis states are then $|x, c\rangle$, and the state of the walker takes the form,

$$|\psi\rangle = \sum_{x,c} \lambda_{x,c} |x, c\rangle. \quad (42.43)$$

The evolution of the walk is given by the coin and step operators,

$$\begin{aligned}\hat{C}|x, \pm 1\rangle &\rightarrow \frac{1}{\sqrt{2}}(|x, +1\rangle \pm |x, -1\rangle), \\ \hat{S}|x, c\rangle &\rightarrow |x + c, c\rangle.\end{aligned}\quad (42.44)$$

The Hadamard coin operator could be replaced with any arbitrary SU(2) matrix. The total time-evolution of the walk is then given by,

$$|\psi(t)\rangle = (\hat{S}\hat{C})^t|\psi(0)\rangle,\quad (42.45)$$

where $t \in \mathbb{Z}_+$. Upon measurement, the probability of the walker being at position x is simply given by summing the probabilities over the coins at a given position,

$$P(x) = |\lambda_{x,-1}|^2 + |\lambda_{x,+1}|^2.\quad (42.46)$$

An example of this kind of quantum walk is shown in Fig. 143.

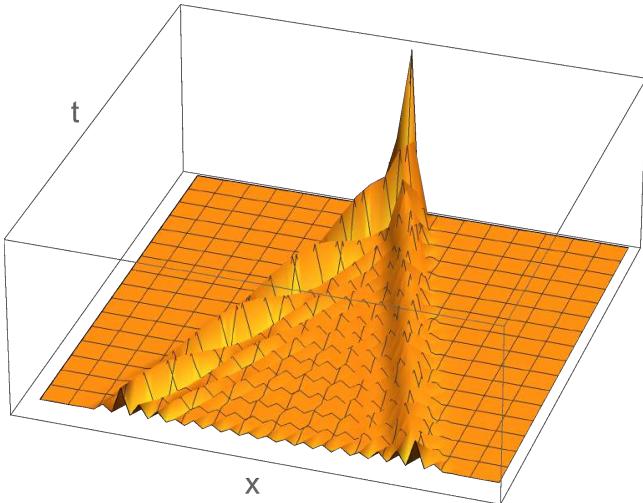


Figure 143: Evolution on a 1D quantum walk on a line with a Hadamard coin operator. The walker begins localised at the origin and then spreads out as a superposition over the position space (x) with time (t). A key feature of this distribution is that its variance grows quadratically with time, compared with linear growth for the equivalent classical random walk. This enhanced spreading forms the basis of the quantum walk search algorithm, with quadratic enhancement compared to a classical search (?).

The single-walker walk on a linear graph is not of computational interest as it can be efficiently classically simulated. However, the formalism is easily logically generalised to multiple walkers on arbitrary graph topologies. We will illustrate this using the formalism of (?), where *walker operators*, rather than walker basis states are evolved under time-evolution (i.e we operate in the Heisenberg picture rather than the Schrödinger picture). In an optical context, walker operators are identically photonic creation operators. The walker operators are of the form

$\hat{w}(x, c)^\dagger$, where x denotes the vertex number currently occupied by the walker, and c denotes the previous vertex occupied by the walker. The single-walker basis states are then of the form $\hat{w}(x, c)^\dagger|0\rangle$, where $|0\rangle$ is the vacuum state containing no excitations. Notice the parallels to the previous example of a linear walk, where the coin degree of freedom specifies the direction the walker is following, which effectively acts as memory of the previous position.

The coin and step operators now take the form,

$$\begin{aligned}\hat{C} : \hat{w}(x, c)^\dagger &\rightarrow \sum_{j \in n_x} A_{c,j}^{(x)} \hat{w}(x, j)^\dagger, \\ \hat{S} : \hat{w}(x, j)^\dagger &\rightarrow \hat{w}(j, x)^\dagger.\end{aligned}\quad (42.47)$$

Here n_x denotes the set of vertices neighbouring x . The coin operators $A^{(x)}$ are $SU(|n_x|)$ unitary matrices representing the weights of edges within this neighbourhood. The step operator, on the other hand, is simply a permutation. A simple example is shown in Fig. 144.

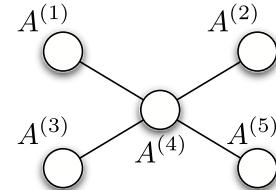


Figure 144: A simple example of a quantum walk on an irregular graph structure. Associated with each vertex x is an $SU(|n_x|)$ coin operator $A^{(x)}$, where $|n_x|$ is the number of neighbours to x .

The total time evolution is defined analogously to before,

$$\hat{U}_{QW}(t) = (\hat{S}\hat{C})^t,\quad (42.48)$$

where $t \in \mathbb{Z}_+$.

With this formalism, multiple walkers are easily accommodated for simply with the addition of extra walker operators. Specifically, the n -walker basis states are of the form,

$$|\vec{x}, \vec{c}\rangle \propto \prod_{i=1}^n \hat{w}(x_i, c_i)^\dagger|0\rangle,\quad (42.49)$$

where we have ignored the normalisation factor, which is a function of the number of walkers in each basis state. Any graph topology can be represented, subject to the constraint that all $A^{(x)}$ are unitary. This implies that every vertex must have as many incoming as outgoing edges, which could be either directed or undirected, subject to this constraint.

Now the probabilities of measuring the walkers in different position configurations will be related to matrix permanents, in a similar manner to **BOSONSAMPLING**. But now the permanents will be of matrices that are

functions of the set of $A^{(x)}$ matrices characterising the graph, rather than a Haar-random matrix.

It was shown by (?) that any such walk can be efficiently represented using a linear optics decomposition comprising at most $O(|V|^2)$ optical modes. Such a decomposition for $|V| = 3$ is shown in Fig. 145.

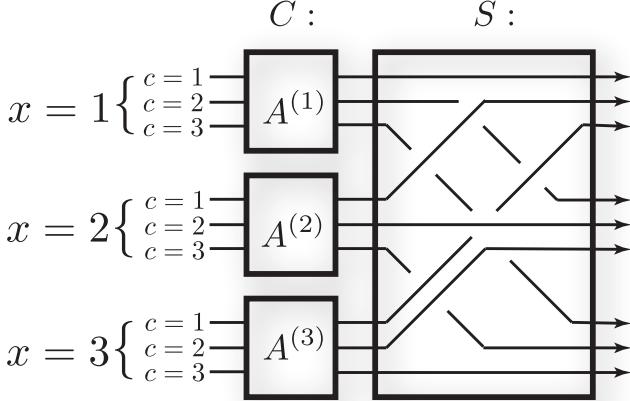


Figure 145: Linear optics decomposition for a single step of an arbitrary 3-vertex discrete-time quantum walk. The coin operators, $A^{(x)}$, may be arbitrary SU(3) matrices, whereas the step operator is simply a permutation of the optical modes. $O(|V|^2)$ optical modes are required, which scales efficiently.

Because of the graph structure of the walk, it lends itself naturally to distributed implementation. We might imagine that different subgraphs – or ‘widgets’ (??) – implement different computational primitives or subroutines. These widgets might be proprietary or expensive to implement, and are therefore best outsourced over a network.

E. Continuous-variables

— by Zixin Huang.

Until now we have focussed on optical systems where quantum information is encoded into discrete variables, such as photon-number or polarisation. However, quantum states of light can also be considered in terms of continuous-variables (CVs) in phase-space.

In this picture, using squeezed states as a resource, qubits can be closely approximated by vacuum states squeezed in orthogonal directions, where the closeness of the approximation is determined by the squeezing parameter⁶⁵. Squeezed state encoding of quantum information was introduced in Sec. XIII.E.3.

A universal gate set can be constructed, enabling universal quantum computation to be implemented using

such an encoding. All the necessary elements may be readily implemented using present-day quantum optics technology and numerous CV quantum protocols have been demonstrated in recent years (?). Most notably, very large-scale CV cluster states have been experimentally prepared in the laboratory (?).

1. Encoding quantum information using squeezed states

As discussed in Sec. XIII.E.3, position and momentum eigenstates are orthogonal and may therefore be employed to encode a single qubit. However, these eigenstates have infinite energy. That is, they are infinitely squeezed in phase-space. But position and momentum eigenstates can be closely approximated using finite, but strong squeezing, where there is a direct tradeoff between energy and the quality of the qubit approximation. The squeezed states in the two quadratures will now no longer be orthogonal, but will have a small overlap that asymptotes to zero as squeezing is increased. Thus, squeezed states can be used to approximate qubits using non-orthogonal basis states. Squeezed states may be prepared directly using a spontaneous parametric down-conversion (SPDC) process (Sec. XXIII.B) (??). CV encoding using squeezed states is illustrated in phase-space in Fig. 35.

Mathematically, there are two operators of interest, the single- and two-mode squeezing operators,

$$\begin{aligned}\hat{S}_{\text{single}}(\xi) &= \exp\left[\frac{1}{2}(\xi^* \hat{a}^2 - \xi \hat{a}^{\dagger 2})\right], \\ \hat{S}_{\text{two}}(\xi) &= \exp\left[\xi \hat{a}_1^\dagger \hat{a}_2^\dagger + \xi^* \hat{a}_1 \hat{a}_2\right],\end{aligned}\quad (42.50)$$

where $\xi \in \mathbb{C}$ is the squeezing parameter. Both of these may be realised physically using non-linear crystals with second order non-linearities, readily available in present-day labs.

2. Phase-shifters

A phase-shifter, implementing the unitary operation,

$$\hat{R}(\theta) = e^{i\theta \hat{n}}, \quad (42.51)$$

where \hat{n} is the photon-number operator, rotates a state in phase-space by angle θ about the origin, implementing the transformation between the position and momentum operators,

$$\begin{pmatrix} \hat{x}_\theta \\ \hat{p}_\theta \end{pmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \hat{x} \\ \hat{p} \end{pmatrix}. \quad (42.52)$$

It is evident upon inspection that this can be thought of as a single-qubit rotation in position/momentum space.

⁶⁵ In the limit of infinite squeezing the two orthogonally squeezed states becomes orthogonal, enabling the encoding of a genuine qubit.

3. Bell pairs

In the squeezed state basis one can prepare Bell pairs in an analogous manner to doing so using single-photon encoding. Namely, mixing two orthogonally squeezed states (squeezed and anti-squeezed) on a 50:50 beamsplitter generates Bell-type entanglement. This is shown in Fig. 146.

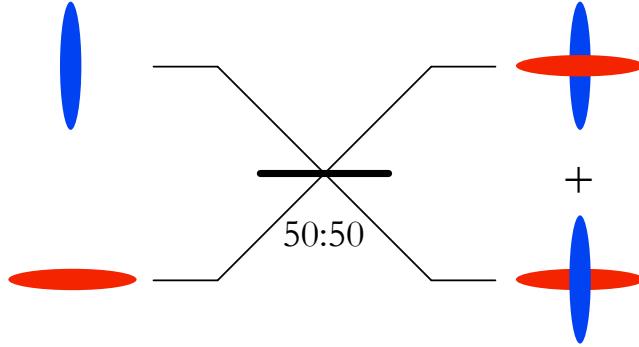


Figure 146: Preparation of an (approximate) CV Bell pair using a 50:50 beamsplitter to entangle two states, squeezed in orthogonal directions in phase-space.

4. Measurement

Measurement of CV states may be performed using homodyne detection (Sec. XXIV.C), which performs a projection along an arbitrary axis in phase-space, allowing \hat{x} and \hat{p} , or any linear combination of the two, to be directly sampled, i.e referring to Eq. (42.52), we can access the observables \hat{x}_θ and \hat{p}_θ for any θ .

Using a 50:50 beamsplitter to implement the reverse of Bell pair preparation, one can construct a Bell analyser for performing projections in the Bell basis, gifting us a highly-cherished entangling operation, useful for all manner of protocols (Part. V). The measurement is deterministic, and requires only homodyne detections.

5. Logical operations

In the position/momentum picture, the logical generalisations of the single-qubit Pauli \hat{X} and \hat{Z} gates may be thought of as displacements (Sec. XXV.B) in the real and imaginary directions in phase-space (?),

$$\begin{aligned}\hat{X}(s) &\equiv \hat{D}(s), s \in \mathbb{R}, \\ \hat{Z}(t) &\equiv \hat{D}(it), t \in \mathbb{R},\end{aligned}\quad (42.53)$$

where $\hat{D}(\alpha)$ is the phase-space displacement operator from Eq. (13.21). These have the logical action,

$$\begin{aligned}\hat{X}(s)|x\rangle &= |x + s\rangle, \\ \hat{Z}(t)|p\rangle &= |p + t\rangle.\end{aligned}\quad (42.54)$$

The logical generalisation of the CZ gate is,

$$\hat{U}_{\text{CZ}} = e^{\frac{i}{2}\hat{x}_1\hat{x}_2}, \quad (42.55)$$

which transforms two-mode quadrature eigenstates as,

$$\hat{U}_{\text{CZ}}|s\rangle_1|t\rangle_2 = e^{\frac{i}{2}s_1t_2}|s\rangle_1|t\rangle_2. \quad (42.56)$$

Intuitively, it is evident upon inspection of the form of the generalised CZ gate that it leaves logical basis state amplitudes unchanged, but adds state-dependent phases to them. Qualitatively, this is exactly what a regular CZ gate does in the space of two qubits, except that the basis is discrete rather than continuous.

A full set of circuit model CV gates, universal for CV quantum computation is summarised in Tab. VI (?).

6. Cluster states

The CV model for quantum computation can be shown to be universal by operating within the cluster state model. As discussed in Sec. XL.B, the basic primitive from which the universality of cluster states arises is the single-qubit teleporter (Fig. 118), which enables arbitrary quantum information to be teleported through the substrate state, accumulating the action of single- and 2-qubit quantum gates in the process, thereby building up a large, arbitrary quantum computation as measurements proceed. Thus, to demonstrate the universality of the CV model we must first demonstrate an analogous circuit for single-mode CV state teleportation.

An optical circuit for a single-mode teleporter is shown in Fig. 147. Evidently, it is structurally almost identical to the standard single-qubit teleporter, just swapping out some operations for their direct CV equivalents.

This circuit has the desired property that, with classical feedforward and local corrections, we can teleport an arbitrary CV qubit state from one mode to another using the CV generalisation of the CZ gate as a resource, and accumulate single-mode unitaries in the process.

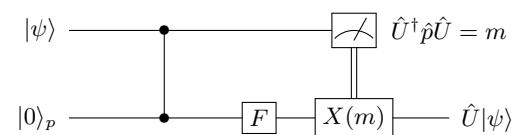


Figure 147: The single-mode teleporter using CV state encoding, structurally almost identical to an ordinary single-qubit teleporter, just substituting the operations with their respective CV generalisations. The circuit teleports the input state $|\psi\rangle$ from the first mode to the second mode, accumulating the action of the single-mode gate \hat{U} applied to the first mode prior to measurement.

Having demonstrated an equivalent single-mode teleporter, we have all we need to construct arbitrary

Qubit model gate	CV equivalent	Implementation
Pauli X	$\hat{X}(s) = \exp[-is\hat{p}]$	Displacement
Pauli Z	$\hat{Z}(t) = \exp[it\hat{x}]$	Displacement
Phase gate	$\hat{P}(\eta) = \exp[i\eta\hat{x}^2]$	Single-mode squeezer & quadrature rotation
Hadamard	$\hat{F} = \exp[i\frac{\pi}{8}(\hat{p}^2 + \hat{x}^2)]$	Phase-shift
CZ	$\hat{U}_{\text{CZ}} = \exp[\frac{i}{2}\hat{x}_1\hat{x}_2]$	Two beam splitters & two squeezers
CNOT	$\hat{U}_{\text{CNOT}} = \exp[-2i\hat{x}_1\hat{p}_2]$	CZ & Hadamards
Non-linear phase gate	$\hat{U}_{\text{NL}} = \exp[it\hat{x}^n], n \geq 3$	Probabilistic measurements

Table VI: Logical generalisations of a universal gate set to the CV model for quantum computing using squeezed state encoding.

measurement-based quantum computations using the generalised CZ gates as the primitive entangling operation for preparation of the substrate graph state.

The addition of any non-Gaussian projective measurement allows universal quantum computation using CV cluster states. A class of such gates is the non-linear phase gate,

$$\hat{U}_{\text{NL}} = \exp(it\hat{x}^n), \quad (42.57)$$

where $n \geq 3$ (the cubic phase gate).

In order to perform universal quantum computation, one needs to implement Hamiltonians of arbitrary degree. In the Heisenberg picture, any Gaussian operation is at most quadratic in the Hamiltonian, and the commutators are also at most quadratic. If one can implement the cubic phase gate, the commutators will now be of degree 3 or 4, and by induction one can construct a Hamiltonian of any degree (?).

7. Fault-tolerance

As with any architecture for quantum computation, we must give consideration to the inevitable presence of noise corrupting our computation. To overcome this problem, and enable scalable quantum computation, we must demonstrate the capacity for the architecture to be made fault-tolerant.

It was shown by (?) that the CV cluster state model can be made fault-tolerant using quantum error correcting codes, sufficient to enable arbitrary scalability.

Having demonstrated fault-tolerance, it can be concluded that CV optical quantum computation is viable, and scalable, with efficient error correction resource overheads.

F. Hybrid light-matter architectures

It is unlikely that future, large-scale quantum computers will be purely optical. Some other technologies have a more favourable outlook in terms of scalability. Nonetheless, when it comes to networking quantum computers, optics is the natural approach, motivating investigation into hybrid architectures, where qubits are represented

using some non-optical system, but entangling operations (EOs) between them are mediated by optical states and linear optics (?).

The natural example is matter qubits which couple to single-photon states, whereupon which-path erasure between coupled optical modes teleports entanglement onto the physical matter qubits. Similarly, measurement of the matter qubits may be performed by stimulating the emission of photons from them. This idea has been applied to λ -configuration atomic qubits (?), shown in Fig. 148, and atomic ensemble qubits (?) (Sec. XX.A.3). The protocol is described in Alg. XLVI.

In principle, this technique could be applied to any physical system comprising natural or engineered λ -configured energy levels, which couple to accessible optical modes. This light-matter coupling may require carefully constructed optical cavities, as is the case for single atoms. Alternately, atomic ensemble qubits inherently undergo collective enhancement in their light-matter coupling, mitigating the need for optical cavities.

Optically-mediated atomic ensemble architectures are particularly attractive, as discussed in Sec. XX.A.3, owing to their long coherence lifetimes, room temperature operation, strong light-matter coupling, and robustness against qubit loss.

A novel ‘double heralding’ technique, introduced in (?), allows photon loss to be overcome during which-path erasure. Similarly, quantum states of light can be coupled to two-level quantum systems using Hamiltonians of the form shown in Eq. (20.2). The preparation of long-distance entanglement between atomic systems has been demonstrated (?).

The attractive feature of this type of approach is that the actual entanglement is generated using all-optical operations, despite the underlying logical qubits being stationary and potentially physically separated a long distance apart, mitigating the need for direct matter-matter interactions, and enabling distributed computation. Optical interfacing is discussed in Sec. XX.A. This allows the EOs to be performed remotely in the cloud, without physically moving the stationary qubits. Such hybrid systems present an interesting platform for cloud quantum computing – despite the qubits being stationary, we are able to outsource the interactions between them to distant servers or even satellites.

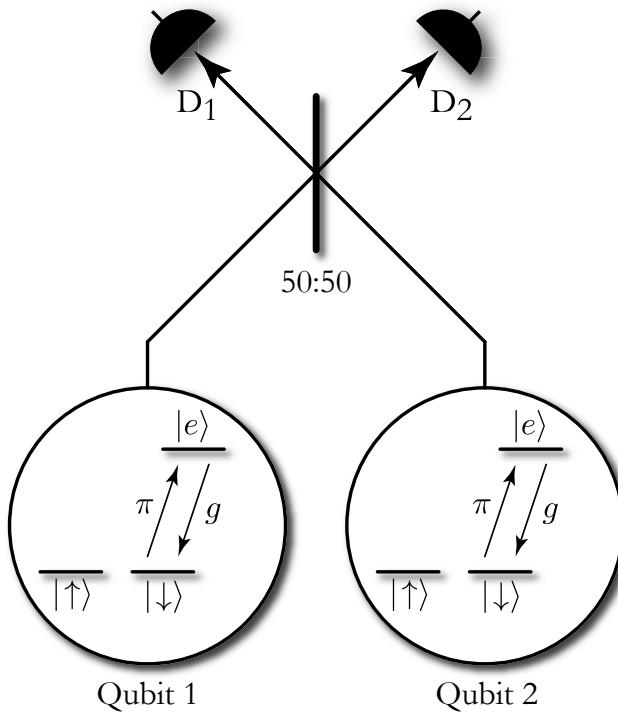


Figure 148: Two atomic systems in λ -configurations, each coupled with an optical mode. An EO between them is mediated via linear optics which-path erasure. Each system contains two degenerate ground states, which jointly encode a qubit ($|0\rangle \equiv |\uparrow\rangle$, $|1\rangle \equiv |\downarrow\rangle$), and an additional excited state ($|e\rangle$), which only couples to the $|\downarrow\rangle$ state. A π -pulse excites the electron from the $|\downarrow\rangle$ to the $|e\rangle$ state, after which emission of a photon is associated with a coherent relaxation back to $|\downarrow\rangle$. If the two optical modes are interfered on a 50:50 beamsplitter, and a single photon is detected between the two photo-detectors, D_1 and D_2 , the two emission processes become indistinguishable, and which-path erasure entangles the two qubits by projecting them onto a maximally entangled Bell pair. More complicated networks based on this EO allow the preparation of cluster states, enabling universal quantum computation. In a quantum networking context, the matter qubits could be held by a client, and the optical interferometry implementing the computation outsourced to the cloud, i.e. the PBS, D_1 and D_2 are implemented in the cloud. This would also facilitate the preparation of shared entangled states, where different clients possess parts of an entangled state, potentially physically separated over long distances.

Importantly, the beamsplitter mediating the which-path erasure EO is based upon HOM interference, and therefore does not require interferometric stability, making the outsourcing process relatively robust and suitable for long-range operation.

This protocol can be regarded as a variation on the entanglement swapping protocol (Sec. XXVII.E), whereby entanglement between matter qubits and optical modes is swapped onto entanglement between the distinct matter qubits.

Alternately, if there is no direct line of quantum communication between two qubits, an EO can be performed by

directly employing the same idea in reverse. We imagine that a third-party, such as a satellite, acts as a server for entangled Bell pairs. Two parties receive one qubit each from the pair. Then they perform an EO between their halves of the Bell pair and their local qubits. With appropriate local corrections, mediated by only cheap classical communication, this teleports the action of an EO onto the two qubits, creating a link between them.

Expanding upon this idea, we can envisage distributed models for quantum computation, where the qubits needn't even be of the same physical medium. We could, for example, entangle quantum dot qubits, atomic qubits, and atomic ensemble qubits with one another by coupling them to optical modes and performing which-path erasure between them. This enables distributed quantum computation between hosts possessing quantum infrastructure comprising different physical mediums (provided the photons emitted by those systems may be made indistinguishable, such that HOM interference is possible).

G. Atomic ensembles

To do

H. Ion traps

To do. Talk about optical interfacing.

I. Superconducting circuits

— by Chandrashekar Radhakrishnan.

Qubits may be engineered by considering the lowest two energy levels of a quantum system. Based on the spacing between their energy levels, such quantum systems are classified into two categories:

- Quantum harmonic oscillator type: have equal spacing between energy levels, given by,

$$E_n = \hbar\omega \left(n + \frac{1}{2} \right), \quad (42.63)$$

where $n \in \mathbb{Z}^+$ denotes the discrete energy level, ω is optical frequency, and E_0 is the lowest-lying ground state energy.

- Atomic type: have unequal spacing between energy levels, given by,

$$E_n = -\frac{E_0}{n^2}. \quad (42.64)$$

We illustrate these two cases in Fig. 149, with their corresponding energy level diagrams.

```
function WhichPathErasure():
```

1. Alice and Bob each prepare an equal superposition of the two logical basis states,

$$|\psi\rangle_{\text{in}} = \frac{1}{2}(|\uparrow\rangle_{A_1} + |\downarrow\rangle_{A_1})|0\rangle_{A_2} \cdot (|\uparrow\rangle_{B_1} + |\downarrow\rangle_{B_1})|0\rangle_{B_2}, \quad (42.58)$$

where A_1/B_1 denote the matter qubits, and A_2/B_2 denote their coupled optical modes.

2. Apply a π -pulse to each qubit, inducing a $|\downarrow\rangle \rightarrow |e\rangle$ transition,

$$|\psi\rangle_{\pi} = \hat{U}_{\pi}|\psi\rangle_{\text{in}} = \frac{1}{2}(|\uparrow\rangle_{A_1} + |e\rangle_{A_1})|0\rangle_{A_2} \cdot (|\uparrow\rangle_{B_1} + |e\rangle_{B_1})|0\rangle_{B_2}. \quad (42.59)$$

3. Wait for a coherent relaxation, inducing the transition $|e\rangle \rightarrow |\downarrow\rangle \hat{a}^{\dagger}$, which emits a single photon,

$$|\psi\rangle_{\text{relax}} = \hat{U}_{\text{relax}}|\psi\rangle_{\pi} = \frac{1}{2}(|\uparrow\rangle_{A_1} + |\downarrow\rangle_{A_1} \hat{a}_{A_2}^{\dagger})|0\rangle_{A_2} \cdot (|\uparrow\rangle_{B_1} + |\downarrow\rangle_{B_1} \hat{a}_{B_2}^{\dagger})|0\rangle_{B_2}. \quad (42.60)$$

4. Apply a 50:50 beamsplitter between the two optical modes,

$$|\psi\rangle_{\text{BS}} = \hat{U}_{\text{BS}}|\psi\rangle_{\text{relax}} = \frac{1}{2}(|\uparrow\rangle_{A_1} + |\downarrow\rangle_{A_1} [\hat{a}_{A_2}^{\dagger} + \hat{a}_{B_2}^{\dagger}]) \cdot (|\uparrow\rangle_{B_1} + |\downarrow\rangle_{B_1} [\hat{a}_{A_2}^{\dagger} - \hat{a}_{B_2}^{\dagger}]) \cdot |0\rangle_{A_2}|0\rangle_{B_2}. \quad (42.61)$$

5. Conditional upon detecting exactly one photon between the output optical modes, we obtain,

$$|\psi\rangle_{\text{out}}^{1,0} = \langle 1, 0 |_{A_2, B_2} |\psi\rangle_{\text{BS}} = \frac{1}{2}(|\uparrow, \downarrow\rangle_{A_1, B_1} + |\downarrow, \uparrow\rangle_{A_1, B_1}), \\ |\psi\rangle_{\text{out}}^{0,1} = \langle 0, 1 |_{A_2, B_2} |\psi\rangle_{\text{BS}} = \frac{1}{2}(|\uparrow, \downarrow\rangle_{A_1, B_1} - |\downarrow, \uparrow\rangle_{A_1, B_1}), \quad (42.62)$$

which is a Bell pair between the matter qubits.

6. \square

Algorithm XLVI: Using which-path erasure to entangle two λ -configuration matter qubits via post-selected linear optics. Note that the two matter qubits could in principle be arbitrarily physically separated. Only the emitted photons need be brought together locally for the implementation of a beam-splitter operation. This lends such entanglement generation protocols to distributed implementation.

To construct a qubit we should be able to use external fields to control and selectively drive transitions between only two energy levels in the system. Such a procedure is easy to achieve in atomic systems, but it is not possible to address only two levels in a quantum oscillator due to the

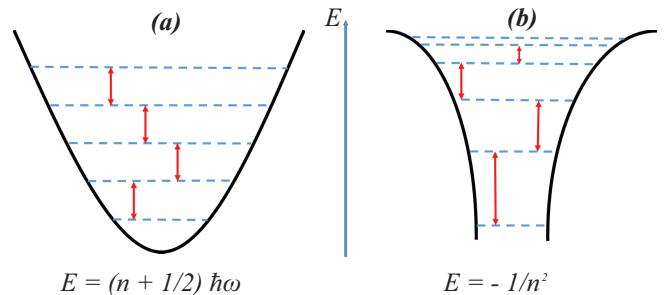


Figure 149: The energy levels of: (a) a quantum oscillator; and, (b) an atomic system. The quantum oscillator exhibits equidistant separation between energy levels, whereas for the atomic system the energy levels are non-uniform.

harmonicity (equal energy spacing) between energy levels. On the other hand, it's hard to work with individual natural atoms, mainly because of their size, which makes their individual isolation and control very challenging. To overcome this problem, we need to develop new quantum devices with anharmonic energy spectra. Such devices are referred to as *artificial atoms*, due to their similarity to natural atoms in the anharmonicity of their energy level spectrum.

One of the most widely used types of artificial atom are superconducting qubits (?????), a class of non-linear quantum circuits. An LC oscillator composed of an inductor L and capacitance C is a typical example of a linear quantum circuit with equal spacing. By introducing a Josephson junction into the linear quantum circuit we can make it non-linear, with anharmonic energy spectrum.

A Josephson junction (?) comprises two bulk superconducting materials, separated by a thin layer of insulating material. In the superconducting phase the superconductors contain Cooper-pairs, composed of paired electrons. These Cooper-pairs move from one superconducting layer to another through the insulating layer via quantum tunnelling. The quantum mechanical nature of Josephson junctions is determined by two important energy scales:

- Josephson coupling energy, E_J .
- Coulomb energy, E_C .

The ratio between these two energy scales determines the energy spectrum of the superconducting qubit. This yields three distinct types of qubits:

- Voltage-driven charge qubits (??).
- Flux-driven flux qubits (??).
- Current-driven phase qubits (?).

These circuits and energy level diagrams for these are illustrated in Fig. 150.

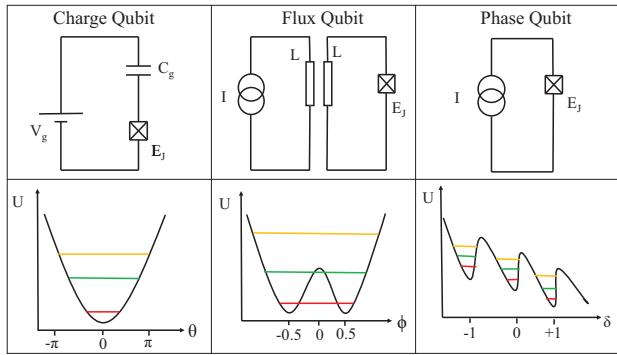


Figure 150: Simplified circuits for the different kinds of superconducting qubits, namely the charge qubit, flux qubit and phase qubit. Below each circuit are their respective energy level diagrams.

1. Charge qubits

A non-linear quantum circuit driven by voltage is referred to as a charge qubit, whose Hamiltonian takes the form,

$$\hat{H} = \frac{\hat{q}^2}{2C} - E_J \cos\left(\frac{2e}{\hbar}\hat{\phi}\right). \quad (42.65)$$

Here \hat{q} is the charge in the superconducting system and $\hat{\phi}$ is flux. The total capacitance of the circuit is given by C , and E_J is the Josephson energy. The Hamiltonian in Eq. (42.65) can be rewritten as,

$$\hat{H} = 4E_C(\hat{n} - n_g)^2 - E_J \cos(\hat{\phi}). \quad (42.66)$$

The variables \hat{q} and $\hat{\phi}$ are canonically conjugate and satisfy the commutation relation,

$$[\hat{\phi}, \hat{q}] = i\hbar. \quad (42.67)$$

In a truncated charge basis the Hamiltonian is,

$$\begin{aligned} \hat{H} = & 4E_C \sum_{n=-N}^N (\hat{n} - n_g)^2 |n\rangle\langle n| \\ & - E_J \sum_{n=-N}^{N-1} |n+1\rangle\langle n| + |n\rangle\langle n+1|. \end{aligned} \quad (42.68)$$

The energy eigenstates are the charge states $|n\rangle$, hence these qubits are referred to as charge qubits. In general the charge qubit (??) is operated in the region,

$$\frac{E_J}{E_C} \approx 1. \quad (42.69)$$

Charge qubits are highly sensitive to noise except at particular working points referred to as ‘sweet spots’. But it is experimentally difficult to control the voltage and current such that the qubit is maintained at these desired working conditions.

To overcome this, a special design of charge qubit known as the *transmission line shunted plasma oscillation qubit* or ‘transmon’ (?) with,

$$\frac{E_J}{E_C} \gg 1, \quad (42.70)$$

was suggested. The transmon is highly robust against external noise compared to the charge qubit. But the energy levels become more and more harmonic (i.e equally spaced) as we move away from the region,

$$\frac{E_J}{E_C} \approx 1. \quad (42.71)$$

Thus the charge qubits are designed by giving consideration to the trade-off between robustness against external noise and the anharmonicity between the levels. A 20-qubit prototype quantum computer developed by IBM employs transmon-type superconducting qubits (?).

2. Flux qubits

The flux qubit is popularly known as the RF SQUID (Radio Frequency Superconducting QUantum Interference Device), which uses an AC current. This qubit can be considered as the magnetic analogue of the charge qubit. In a charge qubit the Josephson junction is driven by a capacitor, but in a flux qubit, a superconducting transformer circuit generates the flux which drives the circuit. The Hamiltonian of the circuit is,

$$\hat{H} = \frac{\hat{q}^2}{2C_J} + \frac{\hat{\phi}^2}{2L} - E_J \cos\left(\frac{2e}{\hbar}(\hat{\phi} - \phi_{\text{ext}})\right). \quad (42.72)$$

Here we can observe that there are three energy scales namely,

$$\begin{aligned} E_J, \\ E_C = \frac{2e^2}{C}, \\ E_L = \frac{\phi_0^2}{2L}. \end{aligned} \quad (42.73)$$

The quantum properties of the qubits depend on the interplay between these parameters. The Cooper-pairs in a flux qubit are confined to a double well potential. The variables Q and the total magnetic flux Φ are the conjugate variables, satisfying the commutation relation,

$$[\hat{Q}, \hat{\Phi}] = i\hbar. \quad (42.74)$$

Flux qubits are very robust against charge noise (?), and hence have very long decoherence times, making them one of the most attractive qubit candidates for the construction of quantum computers. The early quantum computing devices developed by D-Wave employ flux qubits (?).

3. Phase qubits

Current-driven superconducting qubits are referred to as phase qubits (?). They are commonly known as DC SQUIDs, and operate in the regime of very high values of E_J/E_C .

The Hamiltonian of a phase qubit is,

$$\hat{H} = E_C \hat{p}^2 - I\phi_0 \hat{\delta} - I_0 \phi_0 \cos \hat{\delta}, \quad (42.75)$$

where $\hat{\delta}$ is the gauge invariant phase-difference operator and the charge on the capacitor is $2pe$. These operators are conjugate variables, satisfying the commutation relation,

$$[\hat{\delta}, \hat{p}] = i\hbar. \quad (42.76)$$

In the phase qubit, Cooper-pairs experience a washboard potential. Since their decoherence times are very small compared to flux and charge qubits, they are not that widely employed.

4. Quantum gates

To build useful quantum information processing devices, we require quantum gates to act upon our superconducting qubits. This is a field under active development (???). Below we provide a brief description of the operation of single- and 2-qubit quantum gates based on superconducting qubits.

a. Single-qubit gates A single superconducting qubit, which is coherently controlled using microwaves, can be used as a quantum gate. Let us consider a cavity with resonant frequency ω_r and drive frequency ω_d , where the difference $\Delta_r = \omega_r - \omega_d$ is the detuning between the cavity and the drive. When $\omega_d \approx \omega_r$ one can read the state of a superconducting qubit using microwaves. But when $\omega_d = \omega_q \ll \omega_r$ the microwave can be used to perform gate operations on the qubit without measuring its state.

A system comprising a superconducting qubit and a microwave can be described using the Jaynes-Cummings Hamiltonian,

$$\hat{H} = \Delta_r \hat{a}^\dagger \hat{a} - \frac{\Delta_q}{2} \hat{\sigma}_z + g(\hat{a}^\dagger \hat{\sigma}_- + \hat{a} \hat{\sigma}_+) + \xi(t)(\hat{a}^\dagger + \hat{a}), \quad (42.77)$$

where \hat{a}^\dagger (\hat{a}) is the creation (annihilation) operator corresponding to the microwave photon, and $\hat{\sigma}_+$ ($\hat{\sigma}_-$) is the spin raising (lowering) operator. The factors $\Delta_r = \omega_r - \omega_d$ and $\Delta_q = \omega_q - \omega_d$ are the detuning parameters. The factor g is the coupling between the microwave photon and the qubit, and $\xi(t)$ is the envelope of the microwave pulse.

The effective Hamiltonian is,

$$\begin{aligned} \hat{H}_{\text{eff}} = & \left(\Delta_r + \frac{g^2}{\Delta} \hat{\sigma}_z \right) \hat{a}^\dagger \hat{a} - \frac{1}{2} \left(\Delta_q - \frac{g^2}{\Delta} \right) \hat{\sigma}_z \\ & + \xi(t)(\hat{a}^\dagger + \hat{a}) - \frac{g\xi(t)}{\Delta} \hat{\sigma}_x. \end{aligned} \quad (42.78)$$

To perform an X -gate, we choose a drive frequency,

$$\omega_d = \omega_q - \frac{g^2}{\Delta}(2\bar{n} + 1), \quad (42.79)$$

which causes the $\hat{\sigma}_z$ term to disappear, leaving us with a pure $\hat{\sigma}_x$ rotation. Using a phase-shifted drive,

$$H_d(t) = \xi(t)i(\hat{a}^\dagger - \hat{a}), \quad (42.80)$$

one might obtain a pure $\hat{\sigma}_y$ rotation, yielding a Y -gate. Finally we note that using a drive,

$$\omega_d = \omega_q - \frac{g^2}{\Delta}(2\bar{n} + 1) - 2\xi(t)\frac{g}{\Delta}, \quad (42.81)$$

we may construct a Hadamard gate.

b. 2-qubit gates Quantum gates operating on two qubits can be realised in many different ways. But in terms of their construction and operation, they can be divided into two classes. In the first class of quantum gates the superconducting qubits can be tuned over a wide range of frequencies. A good example of this is the iSWAP gate, in which two Cooper-pair boxes are coupled via a transmission line resonator. In the rotating frame of reference, the effective Hamiltonian of the system is,

$$\begin{aligned} \hat{H}_{\text{eff}} = & \frac{g^2}{\Delta} \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right) (\hat{\sigma}_{z,1} + \hat{\sigma}_{z,2}) \\ & - \frac{g^2}{\Delta} (\hat{\sigma}_{+,1} \hat{\sigma}_{-,2} + \hat{\sigma}_{+,2} \hat{\sigma}_{-,1}). \end{aligned} \quad (42.82)$$

The parameters of the two qubits can be adjusted by tuning their flux. The interaction between qubits can be turned on and off by tuning the qubits in and out of resonance with one another. The advantage of the first class of quantum gates is that they can be operated in a region where the frequency of the two qubits differ from one another and the interaction between them is very strong. But the disadvantage is that they are sensitive to flux noise, hence requiring extra flux bias lines for tuning them properly.

The second class of quantum gates is built up of superconducting qubits with fixed frequencies, driven by microwaves. The cross-resonance gate, the bSWAP, and the MAP gate belong to this class. The effective Hamiltonian of the cross-resonance gate is,

$$\hat{H}_{\text{eff}} = - \left(\frac{\tilde{\omega}_1 - \tilde{\omega}_2}{2} \right) \hat{\sigma}_{z,1} + \frac{\Omega(t)}{2} \left(\hat{\sigma}_{x,1} - \frac{J}{\Delta_{12}} \hat{\sigma}_{z,1} \hat{\sigma}_{x,2} \right), \quad (42.83)$$

where,

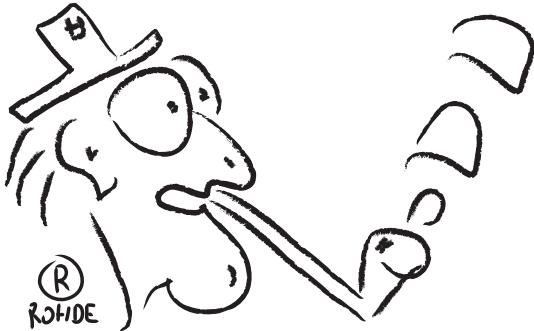
$$\begin{aligned}\tilde{\omega}_1 &= \omega_1 + \frac{J^2}{\Delta_{12}}, \\ \tilde{\omega}_2 &= \omega_2 - \frac{J^2}{\Delta_{12}}, \\ \Delta &= \omega_1 - \omega_2.\end{aligned}\quad (42.84)$$

The factors ω_1 and ω_2 are the frequencies of the first and second qubits, and Δ_{12} is the detuning. The first qubit is rotating with frequency $\frac{1}{2}(\tilde{\omega}_1 - \tilde{\omega}_2)$ around the Z -axis, with a little shift in the X -direction, yielding an X -gate. Similarly we can construct a microwave-activated CZ (MAP) gate using two transmons. The system of two transmons is modelled using a system of two coupled Duffing oscillators. The effective Hamiltonian in the two qubit space reads,

$$\begin{aligned}\hat{H}_{\text{eff}} = -\frac{1}{2} \left(\omega_{01} - \frac{\zeta}{2} \right) \hat{\sigma}_{z,1} - \frac{1}{2} \left(\omega_{10} - \frac{\zeta}{2} \right) \hat{\sigma}_{z,2} \\ + \frac{\zeta}{4} \hat{\sigma}_{z,1} \hat{\sigma}_{z,2}.\end{aligned}\quad (42.85)$$

Through this Hamiltonian we can realise a CZ gate, with gate time 514ns and high fidelity. The second class of quantum gates have a longer coherence time, since the superconducting qubits can be parked at the sweet spots of coherence where the effects of noise on the qubits are less substantial. But, control of the qubits is much harder, since we need to maintain them with the same qubit parameters for an extended period of time.

— *Acta deos numquam mortalia fallunt.*



Part IX

Cloud quantum computing

FROM the perspective of quantum computing, by far the most pressing goal for quantum networking is to facilitate *cloud quantum computing*, whereby computations can be performed over a network via a client/server

model. This will be of immense importance economically, allowing very expensive quantum computers to be accessible to end-users, who otherwise would have been priced out of the market. This economic model is critical to the early widespread adoption of quantum computation. Networking quantum computers is also of the immense importance to capitalise off the leverage associated with unifying quantum resources as opposed to utilising them in isolation (Sec. LVII).

There are several protocols necessary to facilitate cloud quantum computing. First of all, we must have a means by which to remotely process data prepared by a host on a server(s). At the most basic level, this simply involves communicating quantum and/or classical data from a client to a single server for processing, which returns quantum or classical information to the client — *outsourced quantum computation*. In the most general case, a computation may be processed by multiple servers, each responsible for a different part of the computation — *distributed quantum computation*.

Many real-world applications for quantum computing will involve sensitive data, in terms of both the information being processed and the algorithms being employed. This necessitates encryption protocols allowing computations to be performed securely over a network, such that intercept-resend attacks are unable to infer the client's data, and even the host itself is unable to do so — *homomorphic encryption* and *blind quantum computing*. These form the basic building blocks from which a secure cloud-based model for quantum computing may be constructed, and economic models based on the outsourcing of computations may emerge.

The consumer of cloud quantum computing will of course need to be convinced that their data was processed faithfully, according to the desired algorithm. This requires *verification protocols* to allow the server to prove to the client that their data was correctly and honestly processed.

XLIII. THE QUANTUM CLOUD™

WE begin by introducing the primitive building blocks for cloud quantum computing. These form the foundation for higher-level protocols to be discussed later in this part.

A. Outsourced quantum computation

Most simply, an outsourced computation involves Alice preparing either a quantum or classical input state, which she would like processed on Bob's computer. Bob performs the computation and returns either a quantum or classical state to Alice.

The algorithm, which Bob implements, could either be

stipulated by Alice, in which case she is purely licensing Bob's hardware, or by Bob, in which case she is licensing his hardware and software. In the case of classical input and classical output, such an outsourced computation is trivial from a networking perspective, requiring no usage of the quantum network whatsoever. In the case of quantum input and/or output data, the quantum network will be required.

Despite the model being very simple, there may still be stringent requirements on the costs in the network. When the result of the computation is returned to Alice, there may be fidelity requirements. An approximate solution to a problem, or a computation with any logical errors whatsoever, may be useless, particularly for algorithms, which are not efficiently verifiable. For example, if Alice is attempting to factorise a large number using Shor's algorithm, a number of incorrect digits may make the the correct solution effectively impossible to determine. Or if a large satisfiability problem is being solved, almost any classical bit-flip errors will invalidate the result, requiring additional computation by Alice to resolve (which may be exponentially complex to perform).

In the case of classical communication of input and output data, we can reasonably assume error-free communication, owing to its digital nature. However, in the case of quantum communication it is inevitable that at least some degree of noise will be present. Depending on the application, this may require the client and host to jointly implement a distributed implementation of QEC (Secs. XVI.A.6 & XLE), whereby Alice and Bob communicate encoded states with one another, to which syndrome measurement and error correction are applied upon receipt. This will necessitate a limited amount of quantum processing to be directly available to Alice. In the case where she is completely starved of any quantum processing resources whatsoever, this may be a limiting factor. Otherwise, this type of cooperative QEC may be plausible.

B. Distributed quantum computation

The elementary model described above is very limited, as many realistic data processing applications will require multiple stages of computations to be performed, potentially by different hosts. For example, a client may need data processed using multiple proprietary algorithms owned by different hosts, and the processing will need to be distributed across the network (?).

1. In-parallel & in-series computation

Classically, there are two main models for how a distributed computation may proceed – in *parallel*, or in *series* – whereby sub-algorithms are performed either side-by-side simultaneously, or one after another in a

pipeline. The two models are illustrated in Fig. 151.

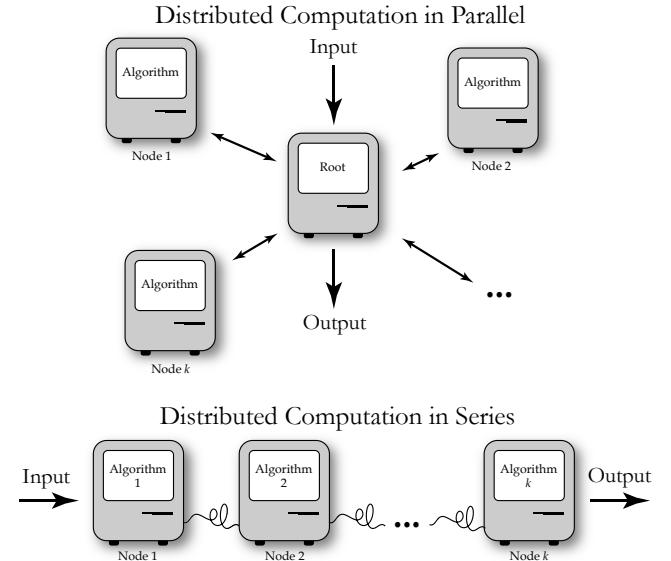


Figure 151: Models for distributed computation in parallel and in series. In parallel, a root node oversees the total computation, delegating tasks to child nodes, which process data independently of one another before being merged. In series the nodes sequentially process data in a pipeline of algorithmic stages.

Classical parallel processing typically involves a root node, which delegates tasks to be performed in parallel by a number of child nodes, and the results returned to the root node, which potentially applies an algorithm to merge the set of results, before returning a final result to the client. Classical models such as Google's MAPREDUCE protocol (?) are built on this idea.

In classical computing, parallel processing is widely employed to shorten algorithmic runtimes. However, the increase in clock-cycles scales only linearly with the number of nodes in the network: k -fold parallelisation yields an $\sim k$ -fold speedup. For time-critical applications, such a linear improvement may already be highly beneficial, albeit costly.

The alternate scenario is in-series distributed computation, in which a computation proceeds through a pipeline of different stages, potentially performed by different hosts. This model allows a complex algorithm comprising smaller subroutines, each of which may be proprietary with different owners, to be delegated across the network. The different stages may communicate classical and/or quantum data. As with the simple single-host model, if the different stages of the processing pipeline are sharing quantum data, distributed QEC will generally be necessary to protect the computation in transit. This necessarily introduces an (efficient) overhead in the number of physical qubits being communicated across the network, introducing additional bandwidth costs, which must be accommodated for in networking strategies.

2. Quantum enhancement

The attractive feature of quantum computing is the potentially exponential improvement in algorithmic performance of certain tasks over their classical counterparts as the size of the computer grows. This exponential relationship implies that computation in general no longer has a simple linear tradeoff as the number of participating nodes increases. In Sec. XLVI we quantify this via so-called *computational scaling functions* and study its economic implications in detail.

But not every effort at distributed quantum computation will automatically exhibit the holy grail of exponential speedup. The architecture and algorithm to which it is applied must be thoughtfully designed to fully exploit the computers' quantum power. A simple adaptation of in-series or in-parallel computation may not achieve this. Rather, we must cunningly exploit quantum entanglement between nodes to perform truly distributed computation, in the sense that no instance of an algorithm is uniquely associated with any given node, but is rather represented collectively across all of them.

Let us assume that we have such a carefully constructed distributed platform. Let t_c be the time required by a classical algorithm to solve a given problem, and t_q the time required to solve the same problem using a quantum algorithm. In the case of algorithms exhibiting exponential quantum speedup, we will have,

$$t_c = O(\exp(t_q)). \quad (43.1)$$

If we now increase the quantum processing power (i.e. number of nodes or qubits) k -fold, the equivalent classical processing time is (in the best case),

$$\begin{aligned} t'_c &= O(\exp(t_q k)) \\ &= O(\exp(t_q)^k) \\ &= O(t_c^k). \end{aligned} \quad (43.2)$$

Thus, k -fold quantum enhancement corresponds to a k th-order exponential enhancement in the equivalent classical processing time, which clearly scales much more favourably than the linear k -fold enhancement offered by classical parallelisation.

For this scaling to be possible, we expect that nodes will need to communicate via quantum rather than purely classical channels, so as to preserve inter-node entanglement and mediate non-local gates across nodes.

3. Quantum MapReduce

Designing native distributed algorithms is not trivial, and architectural constraints may physically limit the allowed set of inter-node operations available to us. Are there any simple constructions that allow us to achieve

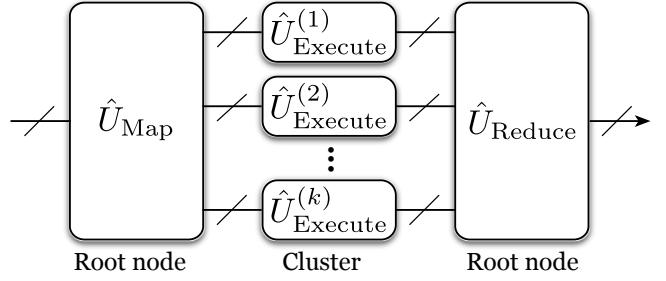


Figure 152: Structure of the quantum MAPREDUCE protocol. All operations are unitary, and the MAP and REDUCE operations may be entangling in general. The EXECUTE stage is separable into a tensor product of smaller EXECUTE operations that are executed in parallel by the k nodes.

this? We will propose an approach to parallelised quantum computation based on a direct quantum adaptation of the classical MAPREDUCE protocol.

MAPREDUCE, originally developed by Google for large-scale parallel processing, is simply an elegant formalism for parallelising classical computations. There are three stages to the protocol:

1. MAP: a root node generates k instances of an algorithm, each with different input data (or a different random seed).
2. EXECUTE: each of the k instances are executed independently on the k nodes in parallel.
3. REDUCE: all outputs are returned to the root node, collated and combined together according to some algorithm, yielding the final output of the computation.

Perhaps the simplest illustrative example is to consider the execution of a Monte Carlo simulation. Here we wish to execute a large number of instances of the same problem, each with a different random seed, and average the results to yield a statistical outcome. Here the MAP algorithm simply delegates out k copies of the same algorithm, assigning each node a different random seed, and the REDUCE algorithm needs only average their outputs. Note that the MAP and REDUCE algorithms are relatively simple, with the nodes operating in parallel doing all the hardcore number crunching.

Taking this model, one might intuitively follow a similar approach for quantum computation, where we simply replace all the operations with unitary processes, and replace the communication links with quantum channels. Now we have a model as shown in Fig. 152.

The goal in this construction is to make the MAP and REDUCE operations be relatively very simple, e.g. have low circuit depth, while the EXECUTE operations are more challenging to implement. Note that the MAP and REDUCE operations are now unitary processes, rather than being, for example, simple classical dispatch and collate

operations. This means that in general the MAP operation will prepare entanglement between the EXECUTE sub-computations, and REDUCE might similarly implement non-separable entangling measurements to measure collective properties of the joint system.

This architecture is merely a direct mapping of classical MAPREDUCE to the quantum setting. How might it be used? Consider quantum simulation, where we aim to simulate a Hamiltonian of the form,

$$\hat{H}_{\text{total}} = \sum_i \hat{H}_i, \quad (43.3)$$

where each of the \hat{H}_i terms are local Hamiltonians acting on orthogonal Hilbert spaces. This implies that all terms commute,

$$[\hat{H}_i, \hat{H}_j] = 0, \quad (43.4)$$

and therefore the unitaries they generate,

$$\hat{U}_j = e^{-\frac{i\hat{H}_j t}{\hbar}}, \quad (43.5)$$

have a separable tensor product structure,

$$\hat{U}_{\text{total}} = \bigotimes_i \hat{U}_i. \quad (43.6)$$

This separability lends itself directly to the tensor product structure of the EXECUTE unitaries. The MAP operation could now be a stage for preparing entangled initial states (entangled across the different subsystems), and the REDUCE operation might perform collective measurements or sampling.

4. Distributed quantum search algorithm

The Grover quantum search algorithm (Sec. XLI.B) can be easily parallelised by partitioning the search space, and allocating a different partition to each node.

Suppose we wish to search over the N -bit space x , to find a satisfying solution to some oracle function (e.g. when solving an **NP**-complete problem),

$$x \text{ s.t. } f(x) = 1. \quad (43.7)$$

Let there be M nodes available for computation, where for simplicity we assume M is a power of 2 (although the idea works generally for arbitrary M , albeit not as mathematically elegantly). We designate each of the nodes a $\log_2 M$ -bit identification number,

$$y = [0, M - 1]. \quad (43.8)$$

We now program each node to search over a smaller search space x' , which is $N - \log_2 M$ bits in length, concatenated with the node's identification number to produce the full

range of x . The input to each instance of the oracle is now,

$$x = x' \frown y, \quad (43.9)$$

where ' \frown ' denotes binary string concatenation.

For example, with four nodes the 2-bit identification numbers are,

$$y = \{00, 01, 10, 11\}. \quad (43.10)$$

If the input search space is N -bits in length, then each of the nodes are assigned the search-space $x' \frown y$, where x' is an $N - 2$ -bit number. Within each instance, the Grover search searches over only the reduced space x' , with y a constant of the instance. Fig. 153 illustrates the circuit schematic for the simple $M = 4$ example.

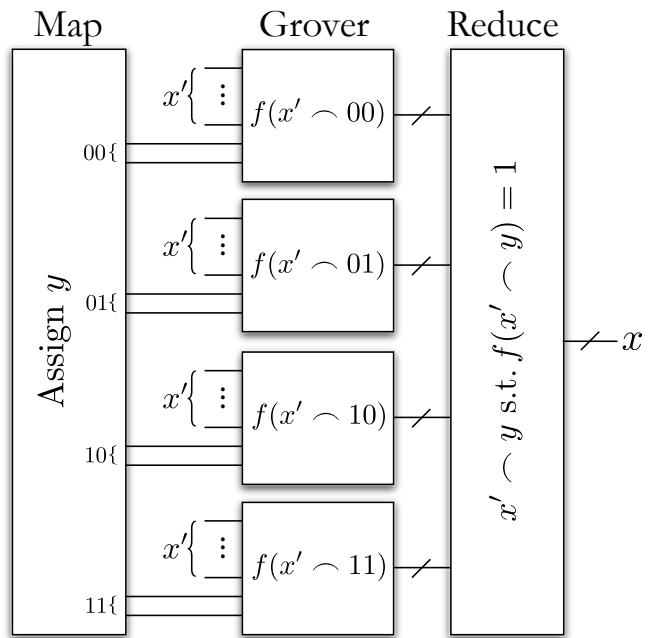


Figure 153: Example of a quantum MAPREDUCE protocol for implementing a distributed quantum search over four nodes. Each node performs a quantum search over the reduced space x' , concatenated with the identification number of the node, which recovers the full search-space x across all the nodes collectively. This effectively partitions and allocates the search-space across the nodes, which implement their reduced searches in parallel. The net speedup provided by M nodes operating in parallel scales as $O(\sqrt{M})$.

It can easily be seen that this approach is compatible with the general quantum MAPREDUCE formalism (Sec. XLIII.B.3), where the MAP function assigns the partitions denoted by the node identification numbers y , the EXECUTE functions implement the reduced searches associated with each instance, and the REDUCE function collects satisfying arguments from the instances,

$$x' \text{ s.t. } f(x' \frown y) = 1 \forall y. \quad (43.11)$$

From the runtime of the Grover algorithm, it follows that the time required to solve the search problem on the initial full search-space is $O(\sqrt{2^N})$, whereas the time required in the parallelised implementation is only $O(\sqrt{2^{N-\log_2 M}})$. Thus, the net speedup is,

$$O\left(\frac{\sqrt{2^N}}{\sqrt{2^{N-\log_2 M}}}\right) = O(\sqrt{M}). \quad (43.12)$$

Evidently, the net computational speedup scales as a factor of the square root of the number of nodes in the parallelised implementation. Note that this approach does not exploit entanglement between nodes, and does not offer a ‘quantum’ (i.e super-polynomial) speedup, since it’s really just brute-force partitioning of a problem into smaller, quicker, bite-sized chunks that are attacked completely independently of one another, much like classical parallelisation.

To the contrary, unlike most quantum algorithms, whose power grows exponentially with the number of qubits (increasing returns), the distributed quantum search algorithm exhibits diminishing returns with the degree of parallelisation – the computational gain from adding one additional node to the network scales as,

$$G = \sqrt{\frac{M+1}{M}}, \quad (43.13)$$

shown in Fig. 154, which in the large M limit asymptotes to,

$$\lim_{M \rightarrow \infty} \sqrt{\frac{M+1}{M}} = 1. \quad (43.14)$$

That is, increasing the number of nodes from 1 to 2 has far greater net gain than increasing them from 100 to 101. In fact, asymptotically, the gain from adding an additional node vanishes in the limit of a high degree of parallelisation.

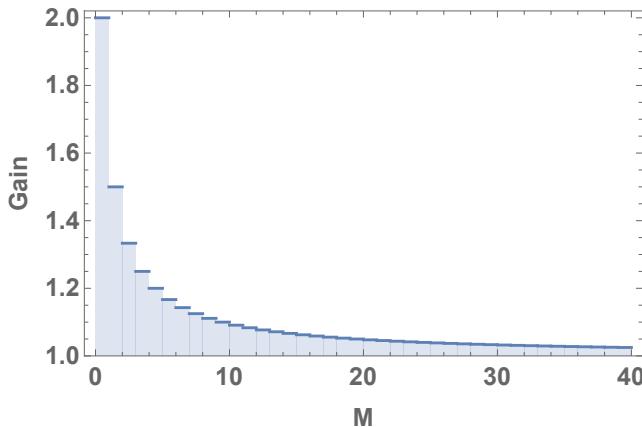


Figure 154: Computational gain from adding a single extra node to a parallelised implementation of the quantum search algorithm. Asymptotically, the computational benefit vanishes.

For this reason, parallelised implementation of a quantum search is not an example of a distributed quantum computation which achieves exponential gain with the addition of new nodes (i.e qubits). Rather, for this specific application it’s far more optimal to consolidate quantum resources into a single larger instance of a quantum search algorithm than using the quantum MAPREDUCE architecture to parallelise it.

5. Distributed unitary error averaging

In Sec. XVIII.B we introduced the unitary error averaging technique for minimising the errors associated with imperfect implementation of linear optics beamsplitter networks. This model is naturally of the form of QUANTUM MAPREDUCE, where the MAP and REDUCE operations implement the fan-in and fan-out respectively, and the independent instances of the noisy unitary are executed in parallel on different nodes, as shown in Fig. 155.

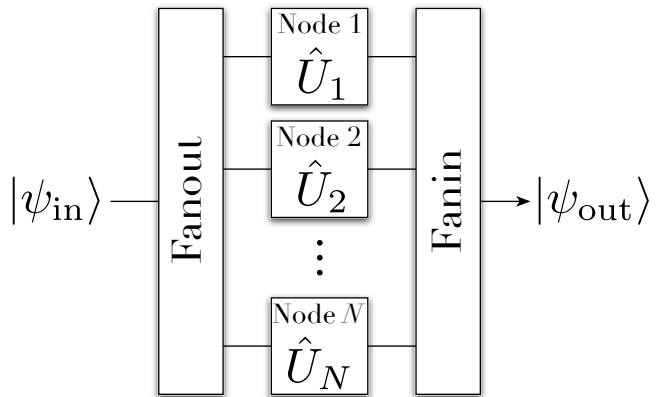


Figure 155: Representing the unitary error averaging technique for error-correcting passive linear optics in parallelised form, consistent with the QUANTUM MAPREDUCE structure.

Now the purpose of the parallelisation is not for computational gain, but rather for error minimisation. The more nodes involved in the parallelised execution, the smaller the final error rate.

6. Delocalised computation

The cluster state (Sec. XL.B), topological code (Sec. XVIII.A) and quantum random walk (Sec. XLII.D.4) models for quantum computation may find themselves to be particularly well-suited to distributed implementation, since they naturally reside on graphs, whose nodes needn’t be held locally by a single user, but could instead be shared across multiple hosts with the ability for graph nodes to intercommunicate. Then only classical communication is required to complete a computation and the quantum information is not localised to any particular node.

Additionally, the entangling gates which build cluster states all commute and may be implemented simultaneously in parallel. This enables a distributed cluster state to be constructed in a ‘patchwork’ fashion, as shown in Fig. 156. Now the computation is truly distributed in the sense that the computation resides collectively across the distributed cluster state, held by any number of users. No instance of an algorithm can be uniquely associated with any given node.

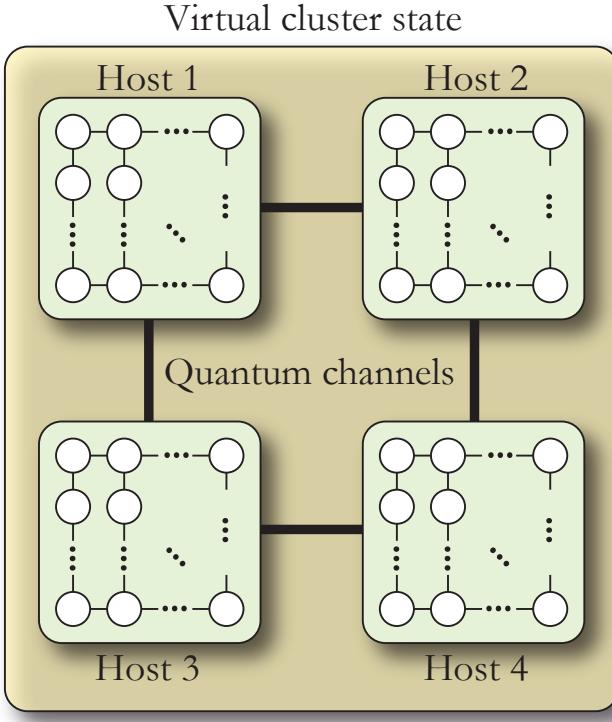


Figure 156: Approach for constructing distributed cluster states (or topological codes, or other graph states) across multiple nodes. The quantum channels allow neighbouring clusters in the topology to be fused together, constructing a large virtual cluster state for distributed computation. The nodes could be arbitrarily separated with optically-mediated interconnects to enable fusing nodes together.

This approach overlaps with the modularised approach for quantum computation discussed in the upcoming Sec. XIII.D, the difference being that in distributed cluster states the goal is to delocalise computations due to resource constraints, whereas for modularised computation the motivation is largely economical, driven by economy of scale.

C. Delegated quantum computation

Taking the notions of outsourced and distributed quantum computation to the logical extreme, we can envisage the situation where Alice has no quantum resources whatsoever (state preparation, evolution or measurement), but

knows exactly what the processing pipeline should entail, and who on the network has the different required quantum resources. We refer to this as *delegated quantum computation*, where the entire processing pipeline is outsourced to a series of hosts.

To illustrate this, let us consider a simple example – cat state quantum computation (Sec. XIII.E.2). There are three main elements to the protocol:

1. Cat state preparation.
2. Post-selected linear optics with feedforward.
3. Continuous-variable measurement.

Each of these stages present their own technological challenges, sufficiently challenging that one might wish to outsource all three stages. However, suppose there is no single host on the network with the ability to perform all three, but rather there are three hosts (B_1 , B_2 and B_3), each specialising in just one of those tasks. In this instance, it would be most resource savvy for the network to implement the pipeline,

$$A \rightarrow B_1 \rightarrow B_2 \rightarrow B_3 \rightarrow A, \quad (43.15)$$

without going back and forth to Alice after each step,

$$A \rightarrow B_1 \rightarrow A \rightarrow B_2 \rightarrow A \rightarrow B_3 \rightarrow A. \quad (43.16)$$

In fact, it may not even be technologically possible to implement back-and-forth to Alice if she has no capacity for handling quantum resources (i.e the $A \leftrightarrow B$ stages are purely classical). An example of such a pipeline is shown in Fig. 157.

This can be achieved by adding a PIPELINE field to the packet header prepared by Alice – a FIFO queue describing the entire processing pipeline that Alice’s packet (which initially contains only classical data) ought to follow through the network. Following completion of each stage of the pipeline we pop the stack and transmit the packet to the next specified host. Only at the very completion of the protocol is a packet (containing only classical data) returned to Alice.

Another good case study is quantum metrology using NOON states (Secs. XXIII.C & XXVII.H) for achieving Heisenberg limited precision. Preparing NOON states is extremely challenging, and additionally Alice may not possess the unknown phase to be measured, but rather wishes a NOON state, prepared by B_1 , to be provided to a third-party, B_2 , who applies the unknown phase, and passes the resulting state to B_3 , who implements the required high-efficiency parity measurements required to complete the protocol. In this case, the pipeline would take the same form as above, again with no back-and-forth communication to Alice.

Such delegated protocols will be very useful in quantum networks, where different hosts specialise in different tasks

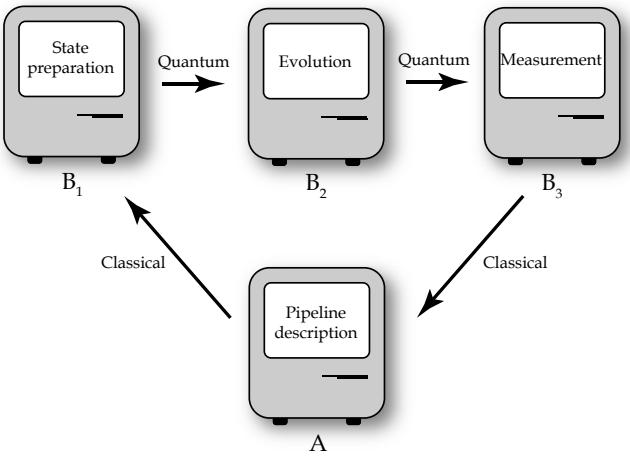


Figure 157: Delegated quantum computation, where each of the three computational stages (state preparation, evolution and measurement) are outsourced to the cloud without intermittent interaction with the client, *A*. *A* provides only a classical description of the processing pipeline to be implemented, each stage of which is delegated to a server specialised in that particular task. Thus, the total processing pipeline takes the form $A \rightarrow B_1 \rightarrow B_2 \rightarrow B_3 \rightarrow A$, where $A \rightarrow B_1$ and $B_3 \rightarrow A$ are classical, and $B_1 \rightarrow B_2$ and $B_2 \rightarrow B_3$ are quantum channels.

(which may be the most economically efficient model), but poor old Alice specialises in none of them, despite knowing exactly what needs to be done. This would allow an aspiring undergraduate student, who is poor (aren't they all?), to sit in his bedroom at his classical PC, and implement entire distributed quantum information processing protocols in the cloud, with no quantum resources or interactions whatsoever.

D. Modularised quantum computation

How does one build a large-scale quantum computer, given the extremely daunting technological requirements and high costs? In any industry, economies of scale allow the mass production, and rapid reduction in price of technology. To achieve this, we must find a way to make quantum technologies commodity items, which avoid all the hassle of customised cutting-edge labs. What we really desire is production-line ‘Lego for Adults™’, allowing ad hoc connection of *modules*, which implement small subsections of a larger computation (?).

We envisage that physically, a module is a black box with optical interconnects, that may be interconnected to form an arbitrary topology, yielding a physical platform as shown in Fig. 158. The user remains oblivious to the inner workings of the modules. The modules could all be identical, just patched together differently, paving the way for their mass production, and an associated quantum equivalent of Moore’s Law, allowing them to become off-the-shelf commodity items over time. Then the cost of a

quantum computer would simply scale linearly with its number of qubits.

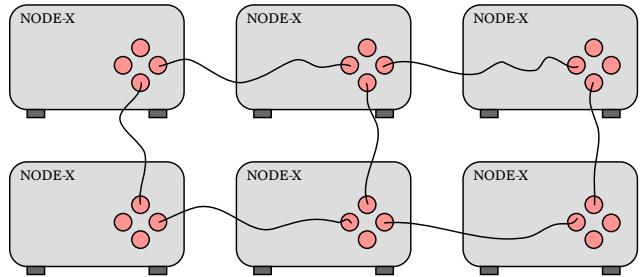


Figure 158: A possible physical realisation of commercially produced quantum modules, forming a 2×3 patchwork of cluster states. Each hosts a relatively small number of qubits. The nodes each have four optical interconnects, which are used to connect the modules via optical fibre. Entangling operations performed on photons shared via the interconnects create inter-module entanglement links, yielding a distributed virtual quantum computer with far more qubits. The computation is truly distributed and cooperative, in the sense that the entire computation is non-local, instead being collectively distributed across all the nodes, which coordinate their local operations via only classical communication. An alternate implementation is to replace the inter-node quantum links with Bell pair distributors. Then entanglement swapping can be employed to swap the entanglement into a link between nodes.

The modules forming a particular computation could either be all owned by a single well-resourced operator, or alternately might be shared across multiple hosts, who network them remotely using EOIs between emitted photons.

In Sec. XLIII.B we introduced the notion of distributed quantum computation. There the motivation was to enable a computation to be distributed across multiple servers, which either parallelise computation or process it as a pipeline in series.

An alternate direction, for economic reasons, is that it is unviable for a single server to host an entire computation. Rather, hosts will have limited capability, and performing large-scale computations will require employing a potentially large number of hosts cooperating and sharing resources with one another⁶⁶. This can be regarded as the most general incarnation of distributed computation.

This is not the same motivation as for in-series computation, where different servers in the pipeline have different

⁶⁶ Even some present-day massive-scale data processing and storage protocols are implemented virtually across multiple large-scale data-centres, which, for example, automatically handle geographically decentralised data redundancy and processing. Google and Amazon, for example, provide cloud services for this purpose, employed both internally, and licensed out to third parties, and the Apache Cassandra project provides an open-source equivalent. The key is for the underlying protocol to abstract this away from the user, such that they interface with the data as though it were a local asset.

proprietary algorithms as subroutines of a larger computation. And it also differs from in-parallel computation, where multiple servers implement the same algorithm on different data, which is subsequently merged by a root node, as per, for example, a MAPREDUCE-style protocol.

Instead, the motivation is one of economics. First, individual servers will have finite resources, but there may be many of them, which can be networked to cooperatively implement a larger algorithm virtually. Second, because the modules in the architecture are identical and lend themselves to mass production, one can expect more favourable economics than that offered by a provider who sells full-fledged, customised quantum computers, which do not lend themselves to the same level of mass production.

The concept of this model is best explained using the optical cluster state formalism (Sec. XL.B), which lends itself naturally to this approach. A rectangular lattice graph is sufficient for universal quantum computation, even if the cluster state graph is not local (but classical communication between nodes is allowed).

Let us first assume that we wish to construct a cluster state with n_{logical} logical qubits. We additionally allow each logical qubit to be the root node of a graph with a +-structure, where each branch comprises a chain of n_{ancilla} ancillary physical qubits. These are sometimes referred to as *micro-clusters* (?). A single micro-cluster collectively forms a single *module* in the topology. Our goal is to fuse modules via nearest neighbour entanglement to build up the desired distributed cluster state.

We arrange the modules to internally represent a +-topology where each node has neighbouring branches in each of the up/down/left/right directions. But we imagine the situation whereby each logical qubit, along with its respective ancillary branches, is held by a different server. Thus, the final cluster state is truly decentralised across all the servers, and in general entire computations cannot be performed locally.

Using the ancillary states in the respective directions, we attempt to fuse neighbouring clusters using EO_s, such as CZ gates (e.g a KLM CZ gate), linear optics *fusion gates* (i.e rotated polarising beamsplitters followed by photo-detection, implementing which-path erasure) (?), or atoms with a λ -configuration coupled to photons (?), which undergo which-path erasure (Sec. XLII.F). Importantly, using the fusion gate and which-path erasure approaches, only a single beamsplitter is required to perform the EO, which only necessitates high-visibility HOM interference, mitigating the need for far more challenging interferometric (MZ) stability (Sec. XXII). This is delightful, as current leading quantum optics experiments routinely achieve HOM visibilities well in excess of 99%.

An alternate fusion strategy is not to directly communicate qubits to be bonded, but instead rely off Bell pairs provided by a central authority. Each party then applies an EO between their half of the Bell pair and their target

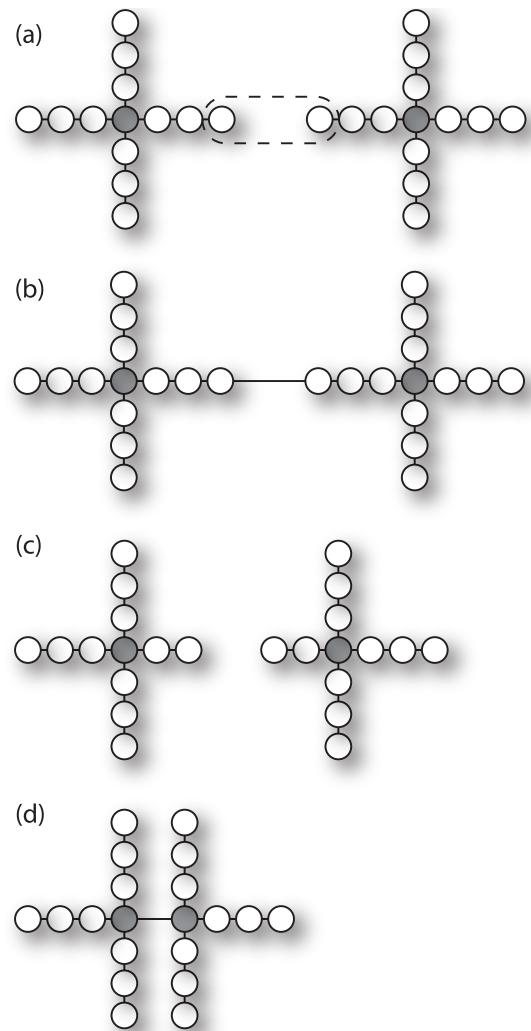


Figure 159: Several cluster state identities for modularised quantum computation. (a) Two cluster states with a +-topology are fused together using an EO (dashed). (b) Upon success, an edge is created between the respective qubits. (c) Upon failure, both qubits are effectively measured in the \hat{Z} basis, thereby removing them, and any associated edges, from the graph. (d) Following a successful EO, the unwanted ancillary qubits may be eliminated using measurements in the \hat{Y} basis, creating edges between their neighbours. If the grey qubits represent the desired logical qubits, this can be used to remove the remainder of the branches emanating from them, thereby distilling the irregular graph down to a regular lattice.

module qubit, which swaps the Bell pair entanglement onto the two respective module qubits (Sec. XXVII.E).

When an EO is successful, we have fused two modules together, albeit potentially with some leftover ancillary states between the logical qubits. When it fails, we have lost the respective ancillary states, and we attempt again using the next ancillary qubits in each of the the respective branches – a kind of REPEAT UNTIL SUCCESS strategy. The bonding only fails if all n_{ancilla} EO_s fail.

Note, however, that longer ancillary arms provide more

opportunity for errors to accumulate (?). Thus, despite its tolerance against gate failure, it is nonetheless highly desirable for EO_s to be as deterministic as possible, so as to minimise the required number of ancillary qubits.

Upon successful bonding, any remaining ancillary qubits between the respective logical qubits are measured in the \hat{Y} basis to remove them from the graph, whilst connecting their neighbours, leaving the two respective logical qubits as nearest neighbours in the graph. Now each module contains exactly one logical qubit, connected as desired to neighbouring modules. The relevant identities are shown in Fig. 159. Our goal is for the entire graph to have a lattice structure, once ancillary qubits have been measured out, as illustrated in Fig. 160.

This approach has been shown to be resource-efficient (??). Let us perform a rudimentary analysis of the resource scaling of this type of approach. The probability of successfully creating an edge between two modules is,

$$p_{\text{success}} = 1 - p_{\text{failure}}^{n_{\text{ancilla}}}, \quad (43.17)$$

where p_{success} is the probability of joining two modules, p_{failure} is the probability that a single EO fails, and n_{ancilla} is the number of ancillary qubits per chain. p_{success} can be made arbitrarily close to unity with sufficiently long ancillary chains, the required length of whom scales as,

$$n_{\text{ancilla}} = \frac{\log(1 - p_{\text{success}})}{\log(p_{\text{failure}})}. \quad (43.18)$$

Now, for simplicity we will consider the preparation of linear cluster states, although these ideas can easily be extended to more complex topologies, such as 2D lattice graphs.

Let us assume we have a ‘primary’ linear topology of modules, which we will incrementally attempt to ‘grow’ by tacking on new modules to the end. When we do so, with probability p_{success} we grow the length of the primary by 1, otherwise we decrement it by 1. This proceeds as a random walk, with on average $2p_{\text{success}} - 1$ new qubits added to the primary per time-step. Provided this number is positive, i.e. $p_{\text{success}} > 1/2$, which can always be achieved with sufficient n_{ancilla} , the length of the primary grows linearly over time, allowing efficient state preparation.

This is just a very primitive model for preparing linear cluster states, using an equally primitive INCREMENTAL strategy for constructing them using non-deterministic gates. As discussed in Sec. XL.B, much further work has been performed on the resource scaling of efficiently preparing cluster states of different graph topologies using different non-deterministic bonding strategies.

Of course, we have used the most simple model for modules, where each accommodates a single logical qubit. In due course, we would expect commodity modules to become far more capable, and resource scaling to improve. We might envisage that each module houses a small square lattice of logical qubits, as shown in Fig. 161, and the

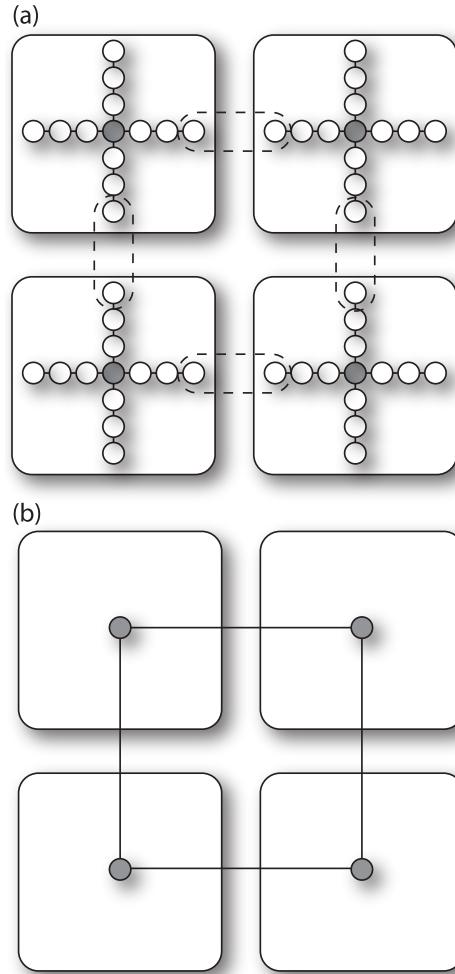


Figure 160: The modularised approach to scalable and economically efficient, distributed quantum computation using cluster states. The modules are all identical, and can be arbitrarily patched to one another, allowing the construction of arbitrary graph topologies. Because the modules are all identical, one might hope that mass production and economy of scale will drive down the cost of modules. We consider a simple 2×2 case where each module (rounded rectangles) comprises a single logical qubit (centre of each module in grey) and a number of ancillary qubits (white in each module), which facilitate bonding the logical qubits of nearest neighbours. The preparation of the modules is performed via nearest neighbour EO_s (dashed ellipses), beginning at the end of branches, and working towards the root node upon each failure, until (hopefully) an EO is successful. (a) A 2×2 lattice of modules with their respective ancillary qubits. We attempt to bond the endpoints of chains using EO_s. (b) Upon measuring the remaining ancillary qubits in the \hat{Y} basis, only the logical qubits remain, with nearest neighbour bonds between adjacent modules, creating a distributed cluster state.

interconnects between them glue them together like a patchwork quilt.

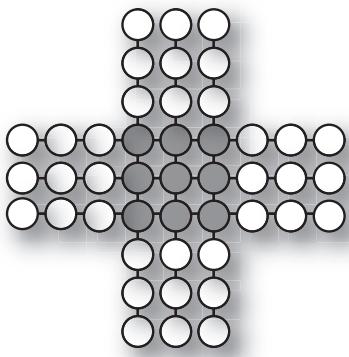


Figure 161: A larger cluster state module comprising a 3×3 lattice of logical qubits (grey), and dangling arms of ancillary qubits (white) in each direction for joining them to neighbouring modules. Fusing these modules enables the ‘patchwork’ preparation of large, distributed lattices.

E. Outsourced quantum research

Thus far we have focussed on computation as the key utility for outsourced quantum technologies, and certainly this is likely to be the dominant driving force behind quantum outsourcing. But of course not everyone wants to only solve complex algorithmic problems. Others may wish to study quantum systems themselves from the perspective of basic science research, or perform precise quantum metrology.

It is foreseeable that in the context of a true quantum internet, there will be a demand for not only the communication of bits and qubits, but more general ‘quantum assets’ (Secs. III & IV), involving all manner of state preparation, manipulation, evolution and measurement, potentially all performed by different interconnected parties, specialising in different aspects of quantum protocols. The demand for this will extend far beyond computation.

The availability of a globe-spanning quantum satellite network brings the opportunity for fundamental quantum mechanical experiments and unprecedented length scales and velocities in the future. Satellite-to-satellite photon transfer can allow for ultra-long distance quantum communications that are not possible on Earth due to atmospheric loss. Another unique aspect of space is that satellites move at high velocities – typically at 10^{-5} times the speed of light for LEO satellites. The combination of both of these effects gives a unique opportunity for performing relativistic quantum information experiments to test fundamental physics.

We anticipate that some of the first experiments will be extensions of what are already performed on Earth. For example, one can perform increasingly long space-based Bell violation tests at unprecedented distances (?). Another possibility is to examine the speed of influence of entanglement (?). In space, such experiments could be extended much further, giving tighter bounds. There

are demanding technical hurdles that must be overcome to succeed at such experiments, such as the necessity for synchronised clocks (Sec. XXVII.J).

In addition to examining extensions of existing experiments, the high satellite velocities can be used to perform relativistic quantum information experiments, such as entanglement tests in the presence of special and general relativity, Wheeler’s delayed choice experiment, and enhanced quantum metrology (??).

The QTCP protocol presented in Sec. XVI provides an extensible framework for facilitating these kinds of outsourced or delegated protocols using generic quantum assets. Bear in mind that, as designed, the payload of QTCP packets could encapsulate all manner of optical states, or mediate long-distance interaction between them.

This model for quantum research could be invaluable to less-well-resourced researchers, for example in developing nations or not-so-well-funded universities, opening up a field of experimental research previously inaccessible to them. Indeed, some private and university sector operators are making elementary, remotely programmable quantum information processing protocols available over the internet, bringing this type of research within reach of researchers and even curious hobbyists around the globe.

While such early implementations fall far short of being truly reconfigurable, outsourced or delegated quantum protocols, applicable to a broad range of applications, they certainly already demonstrate the interest such models for outsourcing is generating within the research community, and the viability of further extending it.

Examples of how this type of model might be applied could include, but not be limited to research into:

- Quantum information processing protocols, beyond only quantum computation, bits and qubits.
- Bose-Einstein condensates (BECs).
- Light-matter interactions.
- Quantum thermodynamics and quantum statistical mechanics.
- Quantum phase-transitions.
- Quantum optics, involving all manner of quantum states of light, beyond only those raised in Sec. XIII.
- Optical interferometry.
- Providing a practical platform for university teaching and education.

In some instances, such outsourced quantum protocols might be applicable to encryption protocols, like those discussed in the next section (Sec. XLIV), enabling highly valuable secrecy for the experiments being conducted by researchers and their hard-earned results and ideas⁶⁷.

⁶⁷ Note, however, that the upcoming protocols are designed for

F. The globally unified quantum cloud

In Sec. LXXXIV we argue that in the quantum era it will be optimal to unify the world's quantum computers into a single virtual, distributed device, rather than utilising smaller individual quantum computers in isolation. This owes to the super-linear scaling in the power of a quantum computer against its number of constituent qubits, a phenomena unique to quantum computers with no classical parallel.

This economic imperative implies that the world's many clients of quantum computing will all be interacting with a single vendor – the globally unified quantum cloud. This will create a competitive online marketplace for the licensing of timeshares in the utilisation of the unified device.

How this unified device will be managed, and by whom, is entirely open to speculation. Will a nation state or alliance of nation states monopolise it? Will a global consortium voluntarily emerge to manage the resources? Or will the whole thing be completely anarchic, potentially resulting in the fracturing of the unified device into several competing smaller ones? What policy and regulatory frameworks will emerge to oversee it?

The answers to these questions are entirely uncertain. But what is certain is that there will be an extremely high level of unification of quantum resources via the quantum internet, massively enhancing its collective computational power.

The Quantum CloudTM will be far more powerful than simply licensing compute-time from a single vendor. Its collective power will be far greater than the sum of its parts.

— *Timendi causa est nescire.*

XLIV. ENCRYPTED CLOUD QUANTUM COMPUTATION

— by Atul Mantri, Si-Hui Tan & Peter Rohde.

EXTRREMELY important to many high-performance data-processing applications is security, as proprietary or sensitive data may be being dealt with. To address this, there are two models for encrypted, outsourced quantum computation – *homomorphic encryption* (??) and *blind quantum computation* (??????).

In both cases, Alice has secret data[©], and wishes to not only ensure that an interceptor is unable to read it, but that even the server performing the computation isn't

application to particular optical states and protocols, and encryption schemes involving more generic quantum assets are likely to require some rethinking and adaptation (if possible at all, which isn't guaranteed!).

able to either – she trusts no one. That is, she wishes the data to be processed in encrypted form, without first requiring decryption.

The difference between the two protocols lies in the treatment of algorithms:

- Homomorphic encryption: Alice provides only the data, whereas Bob provides the processing and the algorithm it implements (which he would also like to keep to himself in general). When *any* circuit is allowed, the protocol is said to be a *fully homomorphic* encryption protocol (FHE). Otherwise, it is a *somewhat-homomorphic* encryption protocol. Although homomorphic encryption protocols have been around for a few decades in the form of privacy homomorphisms (?), classical FHE has only been described very recently (??).
- Blind quantum computing: Alice provides both the algorithm *and* the data, and wishes *both* to remain secret to her. It is known that universal blind *classical* computation is not possible, universal blind *quantum* computation is.

Both of these seem like very challenging goals, yet significant developments have been made on both fronts in the quantum world, with efficient resource overheads associated with the encryption.

In the usual circuit model, blind quantum computation has been shown to be viable, and optimal bounds derived. Equivalently, such protocols have been described in the cluster state model (Sec. XL.B). For universal computation, such protocols necessarily require classical interaction between the client and host. However, it was shown that in some restricted (i.e non-universal) models for optical quantum computation, specifically BOSONSAMPLING, quantum walks and coherent state passive linear optics, non-interactive, somewhat-homomorphic encryption may be implemented.

These encryption protocols induce a resource overhead in circuit size and number of qubits involved in the computation, with efficient scaling. They deliver (at least partially) information-theoretically secure (Sec. XXXII) data-hiding, enabling trustworthy outsourced processing of encrypted data, independent of the attack.

A. Classical computation

A universal QC can implement any classical algorithm. So QCs with homo/BQC give us the means by which to perform encrypted classical computations, bypassing limitations imposed by purely classical protocols. [Comment: The polynomial hierarchy is not contained in BQP. In fact, NP-complete problems are not contained in BQP.]

To set the stage for our upcoming treatment of encrypted quantum computation protocols, we begin by reviewing recent developments in *classical* homomorphic encryption, paying special interest to resource scaling and information-theoretic security.

The first FHE scheme was reported in Gentry's seminal paper (?). He showed that if a homomorphic encryption scheme can evaluate its own decryption circuit, and also slightly augmented versions of it—a feature he calls *bootstrapping*, one can construct a FHE scheme from it. Then he constructed a somewhat homomorphic encryption protocol using ideal lattices, and via a clever transformation that decreases the complexity of its decryption circuit, showed that it is bootstrappable with respect to a universal set of gates. For a security parameter λ , Gentry's scheme has a $\tilde{O}(\lambda^6)$ ⁶⁸ bit bound on complexity for refreshing a ciphertext corresponding to a 1-bit plaintext (?). This was subsequently reduced to $\tilde{O}(\lambda^{3.5})$ (?), $\tilde{O}(\lambda)$ (?), and $\text{polylog}(\lambda)$ for any width- $\Omega(\lambda)$ circuit with t gates (?).

A homomorphic encryption scheme is made up of four algorithms: a key generation algorithm, KeyGen, an encryption algorithm, Encrypt, an evaluation algorithm, Evaluate, and a decryption algorithm, Decrypt. The four algorithms have the following inputs and outputs:

- **KeyGen(λ):** Takes as input a security parameter λ , and outputs a public-key pk , and a secret-key sk .
- **Encrypt(pk, π_i):** Takes as input pk , and a plaintext π_i . It outputs a ciphertext ψ_i .
- **Evaluate(pk, C, Ψ):** Takes as input pk , a permitted circuit C , and $\Psi = (\psi_1, \dots, \psi_t)$. It outputs a ciphertext ψ .
- **Decrypt(sk, ψ):** Takes as input sk , and ψ and outputs $C(\pi_1, \dots, \pi_t)$.

The computational complexity of all these algorithms must be polynomial in λ , and in the case of the evaluation algorithm, polynomial in the size of the evaluation circuit C . The condition that $\text{Decrypt}(sk, \psi)$ outputs $C(\pi_1, \dots, \pi_t)$ is a condition known as correctness which we require of the homomorphic encryption scheme. Furthermore, we also require ciphertext size and decryption time to be upper bounded by a function of the security parameter λ , independently of C . This last

condition is known as compactness, and is necessary to exclude trivial schemes such as that which decrypts the ciphertexts first, and then apply C .

The specifics of these algorithm vary from scheme to scheme, and as is in the case of FHE, usually contains sub-algorithms within them. Making FHE practical is an active area of research. Much of the problem lies in the bootstrapping required in Gentry's scheme, and some of these efforts lies in reducing the overhead required in bootstrapping or removing the need for bootstrapping entirely. Although there exists a plethora of FHE schemes, they are based mainly on two types of problems in lattice-based cryptography: the Shortest Vector Problem (SVP), and Learning with Errors (LWE) Problem. Gentry's original FHE was based on SVP, but over time, the schemes have moved towards a LWE approach because they offer lower overhead and are conceptually simpler. An overview of advances, and applications of homomorphic encryption can be found in a recent review (?).

1. Homomorphic encryption

To do! Yes

2. Blind computation

To do!

B. Cluster states

Most simply, if Alice has the limited quantum resources required to perform single-qubit measurements, and she knows the algorithm she wishes to implement, then by outsourcing just the cluster state preparation stage, whilst performing the single-qubit measurements herself, she can obviously obtain *perfect* secrecy of both her data and her algorithm, since no one else is involved in the processing stage.

However, Alice may have access to no quantum resources whatsoever – even single-qubit measurements – requiring homomorphic encryption or blind quantum computing protocols that are native to the cluster state model. Both such protocols have been described, and in fact are conceptually more straightforward to understand in the cluster state formalism.

To do

Consider both BQC and homomorphic QC

1. Homomorphic encryption

⁶⁸ The tilde in the big-O notation means that we are ignoring logarithmic factors.

To do!

2. Blind quantum computation

To do!**C. Circuit model****To do**

1. Homomorphic encryption

To do!

2. Blind quantum computation

To do!**D. Passive optics**

The previously discussed schemes for encrypted universal quantum computation required a degree of client/server interaction via classical communication. But perhaps there are some restricted (i.e non-universal) models for optical quantum computation, which lend themselves to passive, non-interactive encryption? And perhaps these restrictions simplify the physical resource requirements for encryption?

Let us formalise some reasonable requirements for such a scheme. We will require that:

- Alice's encoding (state preparation) and decoding (measurement) operations are separable, single-mode operations (i.e she has no quantum power of entanglement at her disposal).
- Bob's computation is non-interactive, requiring no input from Alice beyond her input state.
- Bob's computation is passive, requiring no intermediate measurement and feedforward.
- Other than this, there are no constraints on the structure of the encoding/decoding operations, or the optical quantum computation being implemented (e.g it could encompass more than just linear optics).

We can express these requirements very generally and elegantly in terms of a commutation relation between the encoding (\hat{E}), decoding (\hat{D}), and computational (\hat{U}) operations. Furthermore, for the protocol to hide information, the plaintext basis states must not be invariant under the encoding operations. This enforces the criteria,

Def 8 (Encrypted passive optics) *Let*

$k = \{k_1, \dots, k_m\}$ be a partition of the key k into sub-keys $\{k_i\}$, one associated with each mode i . Let $\hat{E}_i(k_i)$ and $\hat{D}_i(k_i)$ be the encoding and decoding operations for the i th mode. \tilde{k} is a potentially transformed version of k , to accommodate that the encryption and decryption keys may be asymmetric, in which case we require that \tilde{k} be efficiently computable from k . Let \hat{U} be the computation. Then, separability of the encoding and decoding operations requires the following commutation relation to hold,

$$\hat{U} \left[\bigotimes_{i=1}^m \hat{E}_i(k_i) \right] = \left[\bigotimes_{i=1}^m \hat{D}_i^\dagger(\tilde{k}_i) \right] \hat{U}. \quad (44.1)$$

For the protocol to hide information, the plaintext basis states must not be invariant under the encoding operations,

$$\left[\bigotimes_{i=1}^m \hat{E}_i(k_i) \right] |\psi\rangle_{\text{plaintext}} \neq |\psi\rangle_{\text{plaintext}}. \quad (44.2)$$

The state observed by Bob is the mixture of Alice's plaintext over the complete set of encoding operations, implementing a quantum process \mathcal{E} , with Kraus operators $\hat{E}(k)$,

$$\begin{aligned} \hat{\rho}_{\text{encoded}} &= \mathcal{E}(|\psi\rangle_{\text{plaintext}} \langle \psi|_{\text{plaintext}}) \\ &= \sum_k \hat{E}(k) |\psi\rangle_{\text{plaintext}} \langle \psi|_{\text{plaintext}} \hat{E}^\dagger(k), \end{aligned} \quad (44.3)$$

where,

$$\hat{E}(k) = \bigotimes_{i=1}^m \hat{E}_i(k_i). \quad (44.4)$$

To minimise Bob's chances of guessing Alice's state, we would like to maximise the von Neuman entropy of Bob's state. For $S(\hat{\rho}_{\text{encoded}}) = 0$ we have no secrecy, whereas for maximal $S(\hat{\rho}_{\text{encoded}})$ we have maximal secrecy (for the given plaintext basis state).

Intuitively, this simply says that a tensor product of single-mode encoding operations commutes through the passive computation to yield a (potentially different) tensor product of single-mode decoding operations. This way, Alice's operations are all separable, requiring no entangling gates (after all, if she had access to entangling gates she might be able to do quantum computations herself!). This relationship can be illustrated as shown in Fig. 162.

Importantly, note that devising a scheme satisfying this commutation relation does not automatically imply that it is secure – it merely enforces the separability of Alice's encoding and decoding operations. An actual security proof is far more demanding, and will be highly state-

dependent.

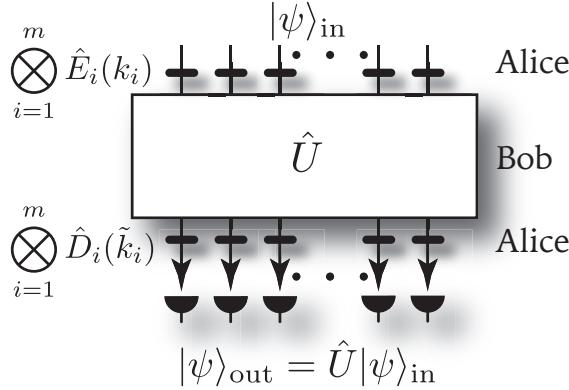


Figure 162: General structure for the relationship between the encoding (\hat{E}), decoding (\hat{D}), and computational (\hat{U}) operations in a passive, non-interactive, optical quantum computation, where Alice is restricted to non-entangling, single-mode encoding and decoding operations.

In the following sections we introduce non-interactive techniques for passive optical quantum computation based upon this general formalism. As encoding techniques compatible with the commutation relation from Eq. (44.1), we specifically introduce:

- *Polarisation-key encoding* (Sec. XLIV.D.1): a uniform random polarisation rotation is applied to each input mode, which we apply to photonic linear optics.
- *Phase-key encoding* (Sec. XLIV.D.2): a uniform random phase-shift is applied to each input mode, which we apply to the encryption of coherent states under evolution via linear optics and generalised non-linear phase-shift operations.
- *Displacement-key encoding* (Sec. XLIV.D.3): an arbitrary configuration of random phase-space displacements is applied to the input modes, which in principle applies to any optical encoding.

However, we leave it as an open question for future work to fully characterise the set of compatible encoding, decoding and computational operations, and to evaluate their security for different choices of input states.

1. Polarisation-key encoding

It was recently shown that processing photonic states using passive linear optics – i.e BOSON SAMPLING or quantum walks (Secs. XLII.D.1 & XLII.D.4) – may be trivially homomorphically encrypted with the addition of additional photons and randomised polarisation rotations on the inputs (?), so-called *polarisation-key encoding*. This encryption does not require any client/server interaction,

remaining completely passive, yet achieving near optimal secrecy, hiding $O(\log(m))$ bits of information in an m -mode interferometer. Furthermore, it does not impose an overhead in circuit complexity, only in the number of input photons.

For m modes, the resource requirements are:

1. m single-photons – one per input mode.
2. m classically controlled wave-plates, able to implement arbitrary polarisation rotations.
3. m polarisation filters.
4. m photo-detectors.
5. An $m \times m$ linear optics network.

The full protocol is described in Alg. XLVII and shown in Fig. 163.

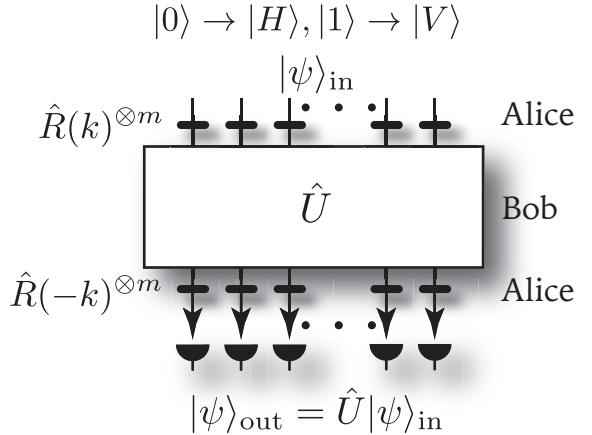


Figure 163: Protocol for implementing homomorphic encryption on photonic passive linear optics. Horizontal bars are wave-plates, implementing polarisation rotations $\hat{R}(\theta)$ – the encryption and decryption operations performed by Alice. $|\psi\rangle_{\text{in}}$ contains one photon per mode, polarisation-encoded such that vertically polarised photons belong to the desired computation, whilst the remaining horizontally polarised ones are dummies. The polarisation rotation angle, k , acts as Alice’s private-key. The photo-detectors are polarisation-resolving, discarding all dummy horizontally polarised photons at the output. The algorithm is described in detail in Alg. XLVII.

The key idea here is that orthogonal polarisations do not interfere with one another under linear optics evolution. Thus, by inserting additional orthogonally polarised ‘dummy’ photons, and applying uniform, random polarisation rotations, we can confuse any eavesdropper as to which photons belong to the computation, thereby hiding the secret data from them. Note that the encryption protocol does not affect the computation, since uniform polarisation rotations commute through linear optics circuits,

$$\hat{R}(k)^{\otimes m} \hat{U} \hat{R}(-k)^{\otimes m} = \hat{R}(k)^{\otimes m} \hat{R}(-k)^{\otimes m} \hat{U} = \hat{U}, \quad (44.13)$$

```

function PolarisationKeyEncoding( $S, k$ ):
    1. Alice meditates upon, but needn't actually
       prepare the state,
        $|\psi\rangle_{\text{number}} = |S_1, \dots, S_m\rangle,$  (44.5)
       where,
        $S_i \in \{0, 1\},$  (44.6)
       is the photon-number of the  $i$ th mode.
    2. Alice makes the substitutions from the
       photon-number basis into the polarisation
       basis,
        $|0\rangle \rightarrow |H\rangle,$ 
        $|1\rangle \rightarrow |V\rangle,$  (44.7)
       to obtain  $|\psi\rangle_{\text{pol}}$ , containing  $m$  photons in
       total, one per mode.
    3. Alice chooses a random private-key  $k$  as a
       real number from the uniform distribution,
        $k \in (0, 2\pi).$  (44.8)
    4. Alice prepares the encoded state by
       applying the same polarisation rotation
       (using wave-plates), of angle  $k$ , to each
       mode,
        $|\psi\rangle_{\text{enc}} = \hat{R}(k)^{\otimes m} |\psi\rangle_{\text{pol}},$  (44.9)
       where,
        $\hat{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$  (44.10)
    5. Alice sends  $|\psi\rangle_{\text{enc}}$  to Bob.
    6. Bob applies processing using his linear
       optics computer, to obtain,
        $|\psi\rangle_{\text{enc comp}} = \hat{U} |\psi\rangle_{\text{enc}}.$  (44.11)
    7. Bob returns  $|\psi\rangle_{\text{enc comp}}$  to Alice.
    8. Alice applies the inverse of the encoding
       operation,
        $|\psi\rangle_{\text{comp}} = \hat{R}(-k)^{\otimes m} |\psi\rangle_{\text{enc comp}}.$  (44.12)
    9. Alice applies polarisation filters to
        $|\psi\rangle_{\text{comp}}$ , discarding horizontally polarised
       photons.
    10. The remaining vertically polarised state
        is Alice's unencrypted output of the
        computation.
    11.  $\square$ 

```

Algorithm XLVII: Protocol for implementing homomorphic encryption on photonic passive linear optics, using polarisation-key encoding.

using the identity,

$$\hat{R}(-k) = \hat{R}^\dagger(k). \quad (44.14)$$

Practically, k could be chosen as some integer multiple of $2\pi/d$, where d is the number of distinct keys, since an infinite precision key would be equivalent to an infinitely long key, were it represented as a bit-string. In this case, the information security of the protocol increases with d .

(?) provided two relationships for the security of this protocol. First, the probability of Bob guessing Alice's input string approaches,

$$P_{\text{guess}} \leq \sqrt{\frac{8}{\pi m}}, \quad (44.15)$$

for sufficiently large m and d , which asymptotically (but unfortunately only polynomially⁶⁹) approaches 0.

Alternately, the mutual information between Alice and Bob, $I(A; B)$, can be upper-bounded using the Holevo quantity, χ (?). That is, $I(A; B) \leq \chi$. The Holevo quantity is defined as,

$$\chi = S(\hat{\rho}) - \sum_i p_i S(\hat{\rho}_i), \quad (44.16)$$

where,

$$\hat{\rho} = \sum_i p_i \hat{\rho}_i, \quad (44.17)$$

and $S(\cdot)$ denotes the von Neuman entropy (Sec. XV.A.9). Here $\hat{\rho}_i$ are the individual codewords, in our case the set of all polarisation-encoded basis states, and p_i are their respective probabilities, which are uniform here.

The upper-bound stipulated by the Holevo quantity is an information-theoretic bound, which holds under *any* choice of measurement bases by Bob. Thus, it is impossible for Bob to extract more information about Alice's state than allowed by this bound.

For this protocol the Holevo quantity scales with the number of modes as,

$$\chi(m) = m - \frac{1}{2} \log_2 \left(\frac{\pi em}{2} \right) + O\left(\frac{1}{m}\right), \quad (44.18)$$

for sufficiently large d . Since there are m bits of information in Alice's input state, this implies that the protocol hides at least,

$$\frac{1}{2} \log_2 \left(\frac{\pi em}{2} \right) + O\left(\frac{1}{m}\right), \quad (44.19)$$

bits of information from Bob.

⁶⁹ Note that an exponentially small bound is actually prohibited by no-go theorems for oblivious transfer and bit commitment (?)

Furthermore, because a single computation requires Alice to perform only a single call to Bob's algorithm, which we treat as a black box, Alice gains minimum knowledge about Bob's secret algorithm, which is optimal for Bob.

Note that while we have considered linear optics in the above discussion, we could in fact expand the list of ingredients available to the computation to include anything generated by a Hamiltonian that commutes with polarisation rotations,

$$[\hat{R}(\theta), \hat{H}] = 0. \quad (44.20)$$

This could include, for example, polarisation-independent non-linear operations.

Discuss follow-up paper by Fitzsimons group on using other photonic degrees of freedom to enhance security.

2. Phase-key encoding

As discussed in Sec. XLII.D.2, although not a *quantum* computation, a system comprising multi-mode coherent state inputs, evolved via passive linear optics, implements simple matrix multiplication on the vector of input coherent state amplitudes,

$$\vec{\beta} = U \cdot \vec{\alpha}, \quad (44.21)$$

for input,

$$|\vec{\alpha}\rangle = |\alpha_1, \dots, \alpha_m\rangle, \quad (44.22)$$

and output,

$$|\vec{\beta}\rangle = |\beta_1, \dots, \beta_m\rangle. \quad (44.23)$$

However, despite this being a classically efficient algorithm, it can experimentally be easily homomorphically encrypted with no computational resource overhead. This is in contrast to classical homomorphic encryption techniques, which incur a computational overhead.

The idea behind homomorphic encryption of coherent state linear optics is conceptually almost identical to the polarisation-space protocol for photonic linear optics (Sec. XLIV.D.1). The key difference is that the random rotations are no longer applied in polarisation-space, but in phase-space as phase-rotations (*phase-key encoding*). Specifically, the encryption/decryption operations are now given by the phase-shift operators,

$$\hat{R}(\phi) = \hat{\Phi}(\phi) = e^{-i\phi\hat{n}}. \quad (44.24)$$

Phase-shift operators acting on coherent states simply implement the transformation,

$$\hat{\Phi}(\phi)|\alpha\rangle = |e^{-i\phi}\alpha\rangle, \quad (44.25)$$

a simple rotation about the origin in phase-space.

Like polarisation rotations, uniform phase-shifts commute through linear optics networks, as per Eq. (44.13), and thus the protocol has similar mathematical structure to the photonic case. Now the phase-shift angle, ϕ , acts as Alice's private-key, which she applies uniformly to all input modes, applying inverse uniform phase-shifts after the computation to decrypt the state. The full algorithm is given in Alg. XLVIII.

```

function PhaseKeyEncoding( $\vec{\alpha}, k$ ) :
  1. Alice prepares the input multi-mode
     coherent state,
      $|\psi\rangle_{\text{in}} = |\vec{\alpha}\rangle = |\alpha_1, \dots, \alpha_m\rangle. \quad (44.26)$ 

  2. Alice chooses a random private-key  $k$  as a
     real number from the uniform distribution,
      $k \in (0, 2\pi). \quad (44.27)$ 

  3. Alice prepares the encoded state by
     applying the same phase-shift, of angle
      $k$ , to each mode,
      $|\psi\rangle_{\text{enc}} = \hat{\Phi}(k)^{\otimes m}|\psi\rangle_{\text{in}}, \quad (44.28)$ 
     where,
      $\hat{\Phi}(\phi) = e^{i\phi\hat{n}}, \quad (44.29)$ 
     is the phase-shift operator.

  4. Alice sends  $|\psi\rangle_{\text{enc}}$  to Bob.

  5. Bob applies processing using his linear
     optics computer, to obtain,
      $|\psi\rangle_{\text{enc comp}} = \hat{U}|\psi\rangle_{\text{enc}}. \quad (44.30)$ 

  6. Bob returns  $|\psi\rangle_{\text{enc comp}}$  to Alice.

  7. Alice applies the inverse of the encoding
     operation,
      $|\psi\rangle_{\text{comp}} = \hat{\Phi}(-k)^{\otimes m}|\psi\rangle_{\text{enc comp}}. \quad (44.31)$ 

  8. The resulting state is,
      $|\psi\rangle_{\text{comp}} = |\vec{\beta}\rangle = |\beta_1, \dots, \beta_m\rangle, \quad (44.32)$ 
     where,
      $\vec{\beta} = U \cdot \vec{\alpha}. \quad (44.33)$ 

  9.  $\square$ 

```

Algorithm XLVIII: Protocol for implementing homomorphic encryption on coherent state passive linear optics, using phase-key encoding.

In fact, this encryption technique applies to more than just linear optics, but extends to also include generalised non-linear phase-shift operations, generated by Hamiltoni-

ans that are polynomials in the photon-number operators,

$$\hat{H} = O(\text{poly}(\hat{n}_1, \dots, \hat{n}_m)), \quad (44.34)$$

where \hat{n}_i is the photon-number operator for the i th mode. This observation follows trivially from the observation that the phase-shift encoding operations (which are generated by Hamiltonians linear in the photon-number operators) commute with any polynomial in the photon-number operators,

$$[\hat{n}_i, \text{poly}(\hat{n}_1, \dots, \hat{n}_m)] = 0 \forall i. \quad (44.35)$$

This immediately significantly expands the class of operations available for the computation. In particular, while coherent states remain separable under linear optics evolution, the introduction of non-linear phase-shift operations enables quantum entanglement, and presumably a more powerful class of computations than simple matrix multiplication. It is unclear to us, however, exactly what this class of computations actually is.

(?) evaluated the security of this protocol in the case where the basis states were restricted to the binary $|\pm\alpha\rangle$ states. Thus, each input mode encodes at most a single bit (zero bits for $|\alpha|=0$, approaching one bit for $|\alpha|\rightarrow\infty$). Rather than employing the mutual information, they resorted to the alternative approach of calculating the distinguishability of codewords under the trace distance (Sec. XV.A.5). This has a direct operational interpretation as the probability of Bob guessing Alice's state in the best case. If the basis codeword states observed by Bob are indistinguishable, they are effectively decorrelated from the plaintext basis states, preventing him from guessing Alice's plaintext state, whereas if they are distinguishable, he can.

Let \vec{x} be the binary input string, and $\vec{0}$ be the special case of the all-zero string. Then, for unencrypted states, we have the trace distance,

$$D_{\text{unenc}} = \|\hat{\rho}_{\vec{x}} - \hat{\rho}_{\vec{0}}\|_{\text{tr}} = \sqrt{1 - e^{-4\text{wt}(\vec{x})|\alpha|^2}}, \quad (44.36)$$

where $\text{wt}(\vec{x})$ is the Hamming weight (number of 1s in the bit-string) of \vec{x} . On the other hand, for the encoded states, we have,

$$D_{\text{enc}} = \|\mathcal{E}(\hat{\rho}_{\vec{x}}) - \mathcal{E}(\hat{\rho}_{\vec{0}})\|_{\text{tr}} = \sum_{k=0}^{d-1} e^{-m|\alpha|^2} \frac{(m|\alpha|^2)^k}{k!} \sqrt{1 - \left(\frac{m - 2\text{wt}(\vec{x})}{m}\right)^{2k}}, \quad (44.37)$$

What's the limit as $d \rightarrow \infty$? where \mathcal{E} denotes the mixture over encoding operations observed by Bob, as per Eq. (44.3), and there are d distinct keys. We also define the ratio between these two distances,

$$R = \frac{D_{\text{enc}}}{D_{\text{unenc}}}, \quad (44.38)$$

as an indicator of data-hiding. $R < 1$ is indicative that information is hidden from Bob.

These relationships are illustrated in Figs. 164 & 165⁷⁰. Clearly, the encoded basis states exhibit greater indistinguishability than the unencoded ones, demonstrating that information is hidden from Bob. The trace distance does not have the elegant interpretation of 'number of bits hidden' that the mutual information does. Rather, it gives us the probability of Bob correctly guessing Alice's state, demonstrating the degree of partial hiding of information.

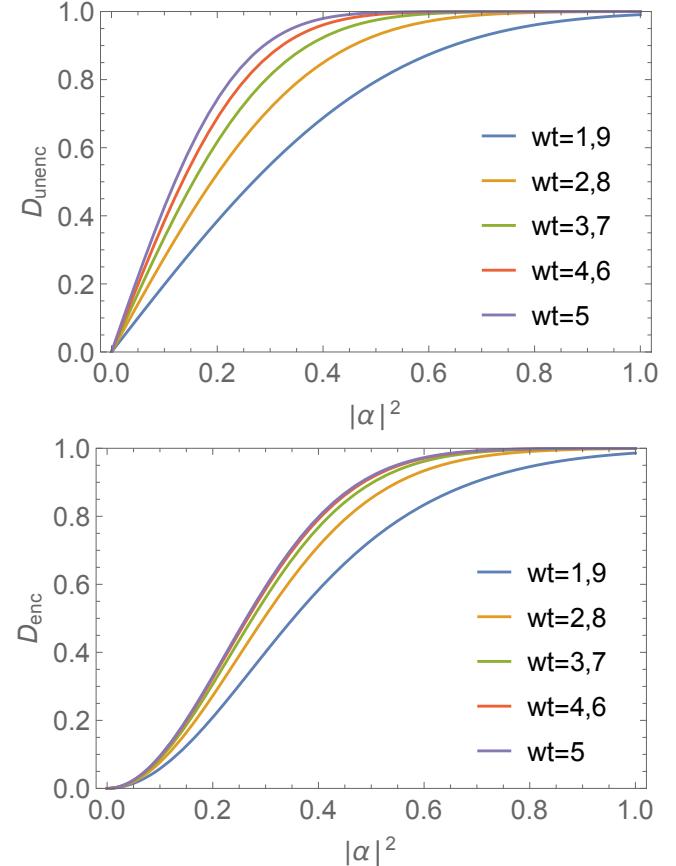


Figure 164: Trace distance between unencoded (top) and encoded (bottom) basis states for coherent state computation using phase-key encoding, with $d = 50$ keys and $m = 10$ modes. Each mode is inputted with one of two basis coherent states, $|\pm\alpha\rangle$. $\text{wt}(\vec{x})$ denotes the Hamming weight of bit-string \vec{x} . Two states are indistinguishable if their trace distance is 0, and distinguishable (orthogonal) if their trace distance is 1. Lower trace distance between encoded states implies a lower chance of Bob guessing Alice's plaintext input state.

It seems plausible that this approach to homomorphic encryption in phase-space could be extended to other

⁷⁰ Note that the distance between any arbitrary pair of codewords may be obtained by replacing $\text{wt}(\cdot)$ with their Hamming distance (in the simple case where one of the codewords is the $\vec{0}$ bit-string, the Hamming weight and Hamming distance are equivalent).

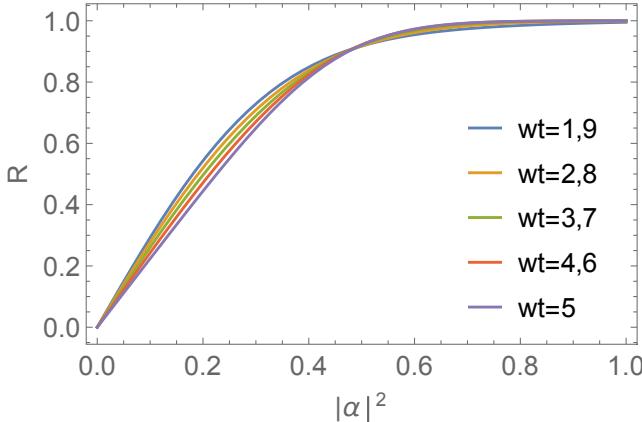


Figure 165: Ratio of the trace distance between unencoded and encoded basis states for coherent state computation using phase-key encoding, with $d = 50$ keys and $m = 10$ modes.

$R < 1$ implies information hiding from Bob.

quantum states of light. However, as is typically the case, performing entropic security proofs is notoriously difficult, and this remains an open problem. It should be noted, however, that this approach will definitely *not* work for any class of input states which are invariant under phase-shifts. This explicitly rules out employing this technique for, for example, photon-number states, which have no phase.

This protocol demonstrates that a simple optical system is able to homomorphically encrypt the classical computation of matrix multiplication, as well as the more general operations of non-linear phase-shifts, without incurring the computational resource overhead imposed by conventional classical homomorphic encryption techniques.

Include figures for asymptotics against m and d .

3. Displacement-key encoding

In the previous two sections we employed the encryption techniques of polarisation-key encoding and phase-key encoding. These are based on the observation that uniform polarisation- and phase-rotations commute through linear optics networks, as per Eq. (44.13). Are there any other types of encoding operations that observe this property?

The other obvious candidate is *displacement-key encoding*, whereby random displacements (not necessarily uniform across all modes) in phase-space (Sec. XXV.B) form the encoding operations. Displacement operators exhibit the property that a tensor product of displacements commutes through linear optics circuits to yield a different combination of tensor products of displacements, where the displacement amplitudes obey the same relationship

as for coherent states from Eq. (44.21). Specifically,

$$\bigotimes_{i=1}^m \hat{D}_i(\alpha_i) \rightarrow \bigotimes_{j=1}^m \hat{D}_j(\beta_j), \quad (44.39)$$

where $\hat{D}_i(\alpha_i)$ is the displacement operator for the i th mode, with displacement amplitude α_i , and the input and output displacement amplitudes are related according to,

$$\vec{\beta} = \hat{U} \cdot \vec{\alpha}. \quad (44.40)$$

Based upon this observation, if the unitary \hat{U} were known to Alice (i.e no secrecy for Bob's algorithm, unfortunately), she could efficiently encode and decode her state, since determining $\vec{\beta}$ from $\vec{\alpha}$ requires only classically-efficient matrix multiplication (residing in \mathbf{P}).

The algorithm for implementing displacement-key homomorphic encryption is shown in Alg. XLIX.

Because performing the decoding operation requires solving the matrix multiplication problem to determine the decoding displacement amplitudes, this technique would obviously be inapplicable to encrypting, for example, coherent states, or other states which can be as efficiently classically simulated as matrix multiplication. Instead, it would only be relevant to linear optics sampling problems, which offer an exponential quantum speedup – if performing the classical computation required for decryption is just as hard as performing the computation, Alice might as well do the computation herself!

Although currently no work has performed any security proofs for displacement-key encoding, this is a candidate approach that warrants future investigation, as displacements exhibit the right kind of commutation relations with linear optics that we desire. Furthermore, it is plausible that this approach might apply to a broad class of optical states, since, unlike phase-rotations, no optical states are invariant under non-zero displacements,

$$\hat{D}(\alpha \neq 0)|\psi\rangle \neq |\psi\rangle \forall |\psi\rangle. \quad (44.46)$$

Intuitively we expect displacement-key encoding to potentially offer better security than phase-key encoding for two reasons:

1. Phase-keys are constrained in the range $k = (0, 2\pi)$, whereas displacement amplitudes are effectively unbounded (nowadays we can make pretty big lasers!).
2. In phase-key encoding the encoding phase-rotation is uniform across all modes, yielding a mode-correlated encryption operation, which limits the entropy of Bob's perceived encoded state. On the other hand, for displacement-key encoding the encoding operations may be chosen independently for each mode, with no inter-mode correlations, potentially increasing the entropy of encoded codewords.

```

function DisplacementKeyEncoding(|ψ⟩, k):
    1. Alice prepares the  $m$ -mode state |ψ⟩.
    2. Alice chooses a set of independent
       complex displacement amplitudes as her
       private-key,
        $k = \{\alpha_1, \dots, \alpha_m\}$ . (44.41)

```

3. Alice applies the displacements to each mode, yielding her encrypted state,

$$|\psi\rangle_{\text{enc}} = \left[\bigotimes_{i=1}^m \hat{D}_i(\alpha_i) \right] |\psi\rangle, \quad (44.42)$$

where $\hat{D}_i(\alpha_i)$ is the displacement operator for the i th mode, with displacement amplitude α_i .

4. Alice sends the encrypted state to Bob.
 5. Bob applies the computation \hat{U} ,

$$|\psi\rangle_{\text{enc comp}} = \hat{U}|\psi\rangle_{\text{enc}}. \quad (44.43)$$

6. Bob returns the encrypted computed state to Alice.
 7. Alice calculates the inverse displacement amplitudes $\vec{\beta}$,

$$\vec{\beta} = U \cdot \vec{\alpha}. \quad (44.44)$$

8. Alice applies the inverse displacements to each mode,

$$|\psi\rangle_{\text{comp}} = \left[\bigotimes_{i=1}^m \hat{D}_i^\dagger(\beta_i) \right] |\psi\rangle_{\text{enc comp}}
 = \hat{U}|\psi\rangle, \quad (44.45)$$

9. $|\psi\rangle_{\text{comp}}$ is the unencrypted computed output state.

10. \square

Algorithm XLIX: Protocol for implementing homomorphic encryption using displacement-key encoding.

(?) presented a security analysis for displacement-key encoding applied to BOSONSAMPLING, which allows a direct side-by-side comparison with the polarisation-key encoded BOSONSAMPLING protocol discussed in Sec. XLIV.D.1. As expected from the intuitive arguments presented above, it was found that displacement-key encoding can outperform polarisation- or phase-key encoding. In fact, in the limit of large upper-bounds on the displacement amplitudes, it was found that the scheme asymptotically perfectly hides Alice's information. That is, the mutual information between Alice and Bob is zero, as is the trace distance between codewords.

To do! Insert figures and equations.

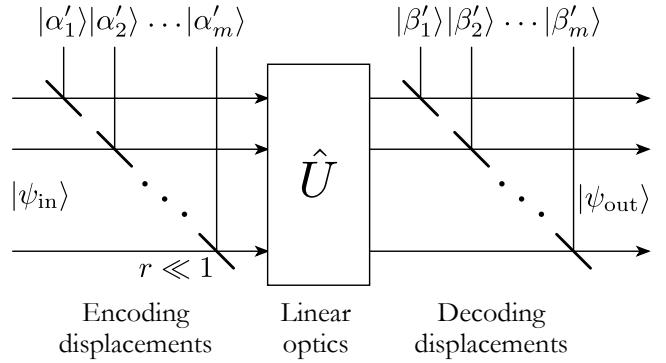


Figure 166: Circuit layout for displacement-key homomorphic encoding of passive linear optics, where $\alpha'_i = \alpha_i/r$ and $\beta'_i = \beta_i/r$. Mixing the input modes with respective coherent states on a very low-reflectivity (r) beamsplitter implements the displacement operations. Following decoding, the output state is simply given by $|\psi_{\text{out}}\rangle = \hat{U}|\psi_{\text{in}}\rangle$, the computed input state.

E. One-time quantum programs

To do! Si-Hui?

F. Authentication

To do

G. Digital signatures

H. Computing on shared sections

— *Audentes fortuna iuvat.*

XLV. VERIFICATION OF CLOUD QUANTUM COMPUTING

A. Randomised benchmarking

Insert

B. Zero-knowledge proofs

Ryan's links for NEXP ZKPs.

General results.

Ryan review.

A *zero-knowledge proof* (ZKP) is an interactive protocol between two parties – a *prover* Peggy, and a *verifier* Victor – where Peggy wishes to efficiently prove to Victor that she knows the solution to a problem, without actually revealing it. Thus, a ZKP can serve as a signature that a problem has been faithfully solved, without disclosing the solution.

ZKPs are useful in a number of cryptographic applications, most notably in authentication. In the case of classical computing, a variety of free software packages are available for compiling ZKPs for generic code. However, ZKPs become far more valuable in the case of cloud quantum computing, where there is an inherent complexity asymmetry between the client Victor (classical resources only) and the server Peggy (full quantum resources), whereby efficient ZKPs become a useful transactional tool during computational outsourcing. In a commercial context, a client can convince themselves that a server has actually performed an outsourced computation before finalising the transaction. The problem description for this scenario is shown in Fig. 167.

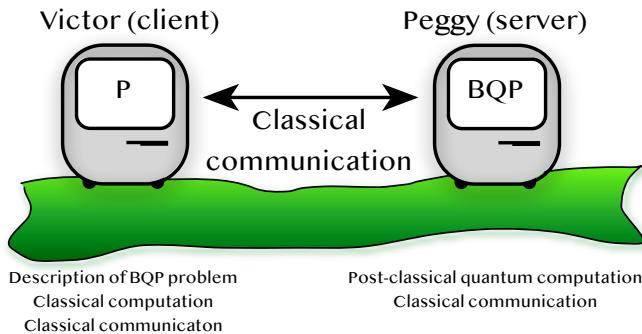


Figure 167: Problem description for using zero-knowledge proofs as a tool in the commercial outsourcing of quantum problems provided by a classical client (limited to both classical computation and communication) to a quantum-capable server. The proof provided by the server demonstrates that the client’s computation has been faithfully executed, without disclosing the outcome before finalising the commercial transaction.

A conceptually simple example for illustrating the operation of a ZKP protocol is the *graph isomorphism problem*. Two graphs are *isomorphic*, denoted $G_1 \sim G_2$, if there exists a permutation $\pi \in S_n$ ⁷¹ on their vertex labels that makes them equivalent, i.e. $G_1 = \pi \cdot G_2$ ⁷². The graph isomorphism problem is to find π for arbitrary G_1 and G_2 .

This problem is clearly contained in **NP**, since permuting vertices in graphs and directly comparing them are both computationally straightforward, making verification of the problem a polynomial-time affair. However, it is believed that explicitly determining the respective permutation is difficult in general. This is intuitively unsurprising, since for a graph with n vertices, there are $n!$ possible permutations to consider (which is super-exponential), and therefore a naïve brute-force approach would require

⁷¹ In group theory, S_n denotes the symmetric group, the set of all permutations on n elements, of which there are $|S_n| = n!$ (the order of the group).

⁷² Here we have employed the operator notation that $\pi \cdot G$ means ‘permutation π applied to graph G ’. Alternately, in matrix notation, where π is a permutation matrix and G is an adjacency matrix, this operation implies the matrix conjugation $\pi \cdot G \cdot \pi^\top$.

$O(n!)$ comparisons in the worst-case⁷³. This problem is not believed to be contained in either **P** or **NP**-complete, and therefore presumed to be **NP**-intermediate (Sec. III).

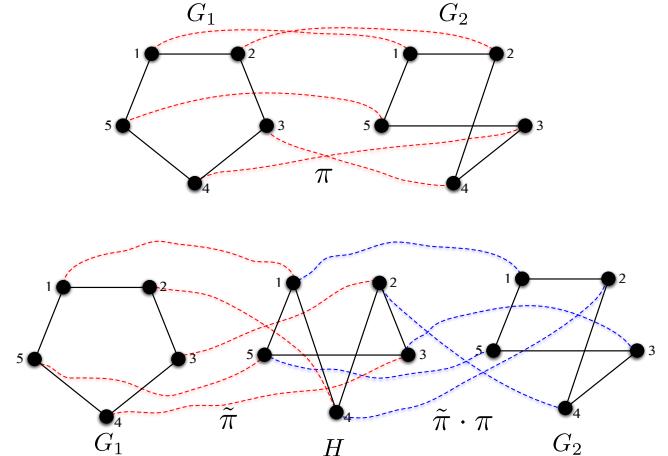


Figure 168: The isomorphisms $G_1 \sim G_2$ (top), and $G_1 \sim H \sim G_2$ (bottom). Coloured lines indicate the vertex relabelings associated with the respective isomorphisms. Knowing the latter two isomorphisms simultaneously implies knowledge of $G_1 \sim G_2$ via composition of the permutations. However, knowing only one of them does not, since H is chosen randomly. By repeatedly proving knowledge of one of the isomorphisms with H we achieve asymptotic certainty that the prover must have known $G_1 \sim G_2$, without actually revealing the associated permutation.

In the example isomorphism presented in Fig. 168(top), the vertex permutation mapping G_1 to G_2 could be expressed in vector form as,

$$\pi = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}_{G_1} \rightarrow \begin{pmatrix} 1 \\ 2 \\ 4 \\ 3 \\ 5 \end{pmatrix}_{G_2}, \quad (45.1)$$

where indices denote vertex labels, or equivalently via the permutation matrix,

$$\pi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (45.2)$$

To provide a ZKP for this, Peggy’s goal is to prove that she knows π , without explicitly revealing it. An efficient, randomised, classical ZKP protocol for achieving this is provided in Alg. L, and a specific example illustrated in

⁷³ This is the worst-case scenario. In many special cases, knowledge of underlying graph structure can simplify this enormously, yielding classically efficient runtimes.

```
function ZKP.GraphIsomorphism( $G_1, G_2$ ):
    1. Graphs  $G_1$  and  $G_2$  are known to both verifier Victor, and prover Peggy.
    2. Peggy knows the permutation  $\pi$  for the isomorphism  $G_1 \sim G_2$ ,
        
$$G_1 = \pi \cdot G_2. \quad (45.3)$$

    3. Peggy wishes to prove to Victor that she knows  $\pi$ , without disclosing what it is.
    4. Peggy chooses another random permutation  $\tilde{\pi}$ , and constructs the new permuted graph  $H$ ,
        
$$\begin{aligned} H &= \tilde{\pi} \cdot G_1, \\ H &= \tilde{\pi} \cdot \pi \cdot G_2. \end{aligned} \quad (45.4)$$

    5. Peggy shares  $H$  with Victor, randomly isomorphic to both  $G_1$  and  $G_2$ .
    6. Victor randomly ( $p = 1/2$ ) asks Peggy to prove either  $H \sim G_1$  or  $H \sim G_2$ .
    7. She accordingly reveals either  $\tilde{\pi}$  or  $\tilde{\pi} \cdot \pi$  to Victor. He can now efficiently verify either  $H \sim G_1$  or  $H \sim G_2$  respectively, by performing the inverse permutation,
        
$$\begin{aligned} G_1 &= \tilde{\pi}^{-1} \cdot H, \\ G_2 &= (\tilde{\pi} \cdot \pi)^{-1} \cdot H. \end{aligned} \quad (45.5)$$

    8. Victor is unable to determine  $\pi$  from either scenario alone, but could were he to know both isomorphisms simultaneously, since,
        
$$\pi = \tilde{\pi}^{-1} \cdot (\tilde{\pi} \cdot \pi). \quad (45.6)$$

    9. The above is repeated  $n$  times. Each time, Peggy chooses a new random  $\tilde{\pi}$ .
    10. If Peggy does not actually know  $\pi$ , the probability of fraudulently passing this test  $n$  times is,
        
$$P_{\text{deceive}} = \frac{1}{2^n}. \quad (45.7)$$

    11. With confidence  $1 - P_{\text{deceive}}$ , Victor knows that Peggy knows  $\pi$ , without knowing it himself.
    12.  $\square$ 
```

Algorithm L: A zero-knowledge proof for the graph isomorphism problem. Victor (verifier) provides two graphs to Peggy (prover), who can demonstrate with asymptotic certainty that she knows their isomorphism, without disclosing the associated permutation relating them.

Fig. 168. The key underlying principle here is to obscure π through randomisation, whereby instead of directly proving knowledge of π , it is implied via multiple proofs of isomorphisms with intermediate random graphs.

— Semper fidelis.



Part X

Economics & politics

“If you put the federal government in charge of the Sahara Desert, in five years there’d be a shortage of sand.” — Milton Friedman.

“It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest.” — Adam Smith.

A NY form of computation comes at an economic cost, but also brings with it a payoff. A key consideration in any model for computation is the tradeoff between the two. Because the computational power of quantum computers scales inherently differently than classical computers, we expect economic indicators to exhibit different scaling characteristics and dynamics also, thereby fundamentally altering the economic landscape of the post-quantum world.

We will now treat some of these economic issues in the context of a global network of unified quantum computing resources, which are then equitably time-shared. We argue in Sec. XLVIII that this time-shared model for quantum computation is always more computationally efficient than having distinct quantum computers operating independently in parallel, owing to the super-linear scaling in their joint computational power. While this section provides mathematical details of various economic models, Secs. LXXXIV & LXXXV provide a popular, high-level discussion surrounding these issues. Sec. LXV summarises the various economic models we present in this part.

XLVI. CLASSICAL-EQUIVALENT COMPUTATIONAL POWER & COMPUTATIONAL SCALING FUNCTIONS

L ET t be the classical-equivalent runtime of a quantum algorithm comprising n qubits – that is, how long would a given classical computer require to implement this

n-qubit quantum computation? We define a *computational scaling function* characterising this relationship,

Def 9 (Computational scaling functions) *The computational scaling function, f_{sc} , relates the number of qubits held by a quantum computer, n , and the classical-equivalent runtime, t , of the algorithm it implements,*

$$t = f_{sc}(n), \quad (46.1)$$

where f_{sc} is monotonically increasing, and depends heavily on both the algorithm being implemented, as well as the architecture of the computer, including the computational model and choice of fault-tolerance protocol.

The exact form of the scaling function will be specific to the algorithm being deployed⁷⁴, and the computational model (e.g cluster states vs the circuit model, as well as choices in error correction, amongst other factors). Most notably, different quantum algorithms offer different scalings in their quantum speedup – Grover’s algorithm (Sec. XLI.B) offers only a quadratic quantum speedup, compared to the exponential speedup afforded by Shor’s algorithm (Sec. XLI.G). Thus, the computational scaling function depends on both the hardware and software, and may therefore differ between different users operating the same computer. We abstract this away and assume all these factors and resource overheads have been merged into the scaling function.

A. Virtual computational scaling functions

If a network of quantum computers were combined into a single, larger *virtual quantum computer* (Sec. LXXXIII) using a distributed model for quantum computation (Sec. XLIII.B), we can define a computational scaling function relationship for the virtual device,

Def 10 (Virtual scaling function) *The joint classical-equivalent runtime of a distributed virtual quantum computation over a network is,*

$$t_{\text{joint}} = f_{sc}^{\text{virtual}}(n_{\text{global}}), \quad (46.2)$$

where,

$$n_{\text{global}} = \sum_{j \in \text{nodes}} n_j, \quad (46.3)$$

is the total number of qubits in the network, with j summing over all nodes in the network, each of which holds n_j qubits. f_{sc}^{virtual} is obtained from f_{sc} by factoring in network overheads and inefficiencies. With perfect network efficiency, $f_{sc}^{\text{virtual}} = f_{sc}$.

B. Combined computational scaling functions

Until now we have characterised the entire network by a single scaling function. Of course, the scaling functions observed by different market participants needn’t all be the same, as they are functions of not only the hardware, but also the participants’ different algorithmic applications (i.e software).

Consider taking a single unit of time (i.e we are ignoring cost discounting over multiple units of time) and dividing it amongst a number of nodes, n_{nodes} , each with their own scaling function, $f_{sc}^{(i)}$. The total classical-equivalent runtime of the computation is additive, given simply by a linear combination of the classical-equivalent processing times of the individual nodes. This yields the relationship for combining scaling functions,

Def 11 (Combined scaling functions) *The effective combined computational scaling function, $f_{sc}^{(\text{joint})}$, of a group of participants, each with their own scaling functions, $f_{sc}^{(i)}$, is given by,*

$$\begin{aligned} t_{\text{joint}} &= \sum_{i=1}^{n_{\text{nodes}}} \beta_i \cdot f_{sc}^{(i)}(n_{\text{global}}) \\ &= f_{sc}^{(\text{joint})}(n_{\text{global}}), \end{aligned} \quad (46.4)$$

where β_i characterise the share of processing time allocated to each node, and for normalisation,

$$\sum_{i=1}^{n_{\text{nodes}}} \beta_i = 1. \quad (46.5)$$

Thus, the joint scaling function of the entire network is simply given by a linear combination (weighted average) of the scaling functions of the different market participants.

⁷⁴ For example, the *circuit depth*, i.e number of gate applications in series, will heavily influence the number of classical steps required to simulate the circuit.

XLVII. PER-QUBIT COMPUTATIONAL POWER

ONE parameter that appears ubiquitously in the upcoming economic models and warrants a definition of its own is the computational power of a quantum computer per qubit. This relates the power and size of the computer. We define this as the *per-qubit computational power*,

Def 12 (Per-qubit computational power) *The per-qubit computational power is defined as the computational power per qubit,*

$$\chi_{sc}(n) = \frac{f_{sc}(n)}{n}. \quad (47.1)$$

This parameter acts as an overall, network size-dependent price scaling factor on:

- Quantum computational leverage (Sec. LVII).
- Cost of computation (Sec. LIV).
- Time-shared computational power (Sec. LV).
- Quantum computational leverage (Sec. LVII).
- Forward contracts (Sec. LIX).

This parameter lends itself to the elegant interpretation as a cost multiplier on qubit asset, dividend and derivative prices, which warrants investigation of its scaling characteristics, shown in Fig. 169.

Note that in the quantum context, the computational power per qubit is not intrinsic to the qubit itself, but depends on how many qubits it cooperates with, a phenomena which does not arise in the classical context.

The key observation is that this scaling factor is constant for classical computing, where the scaling function is linear, but monotonically increasing for any super-linear scaling function. For polynomial scaling functions, it has the effect of reducing the order of the polynomial by one. And for exponential scaling functions, it remains exponential.

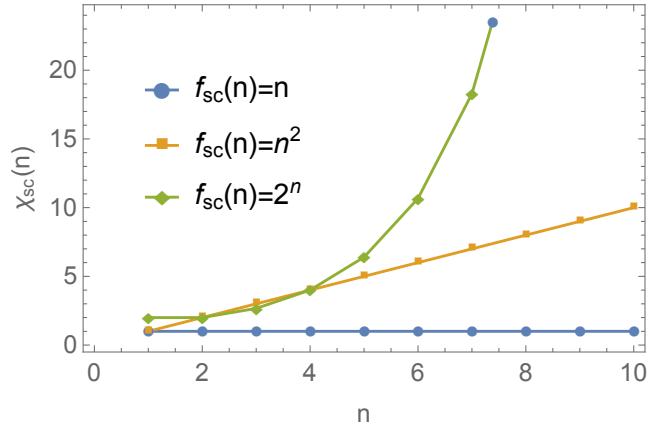


Figure 169: Per-qubit computational power, χ_{sc} , as a function of several representative computational scaling functions, f_{sc} , where n is network size.

XLVIII. TIME-SHARING

SUPPOSE Alice and Bob both possessed expensive classical Cray™ supercomputers, both identical. They're both connected to the internet, so does it make sense to unify their computational resources over the network to construct a more powerful virtual machine, which they subsequently time-share between themselves, or are they better off just using their own computers independently?

If there were an asymmetry in demand for computational resources, it would make perfect sense to unify computational resources, so as to mitigate wasting precious clock-cycles. However, if they were both heavy users, always consuming every last clock-cycle, it would make no difference: for a given computation, Alice and Bob could each be allocated half the processing time of the virtual supercomputer twice as powerful; or, each could exploit the full processing time of their half-as-fast computers. In either case, the dollar cost of the computation is the same. This simple observation follows trivially from the linear relationship between processing power and the number of CPUs in a classical computer.

More generally, in a networked environment where time-sharing of classical computational resources is applied equitably, proportionate to nodes' contribution to the network, the dollar cost per computation is (roughly, modulo parallelisation overheads) unaffected by the rest of the network. Instead, the motivation for networking computational resources is to improve efficiency by ensuring that clock-cycles are not wasted, but instead distributed according to demand by a scheduling algorithm, which could be market-driven, for example.

However, the computational power of a quantum computer generally doesn't scale linearly with its number of qubits, but super-linearly, often exponentially. This completely changes the economics, and market dynamics of networked quantum computers. Intuitively, we expect equitable time-sharing of unified quantum computational

resources to offer more performance to all nodes than if they were to exclusively use their own resources in isolation. That is, the cost of a computation is reduced by resource-sharing, even after time-sharing.

For this reason, henceforth we will assume an environment in which owners of quantum hardware network and unify their computational power, sharing the virtual quantum computer's power between them.

In Sec. LV we present an explicit model for equitable time-sharing, which is optimal from a market perspective.

XLIX. ECONOMIC MODEL ASSUMPTIONS

BEFORE proceeding with explicit derivations of economic models, we state some assumptions about the dynamics of a marketplace in quantum assets. These assumptions are largely based on historical observations surrounding classical technologies that we might reasonably expect to also apply in the quantum era. However, given that the quantum marketplace is one that hasn't been explored in detail until now, it may be the case that some of these assumptions will require revision. Nonetheless, the general techniques we employ could readily be adapted to some relaxations and variations in these assumptions.

A. Efficient markets

“In a dream it’s typical not to be rational.” — John Nash.

We make several assumptions about the efficiency of the quantum marketplace. These are largely based on the conventional efficient-market hypothesis (EMH) (?), readily taught in undergraduate ECON101 and subsequently summarily rejected upon entering ECON202. For ease of exposition, we will remain in the ECON101 classroom.

Some of these assumptions may reasonably turn out to be invalid, or require revision as we learn more about upcoming quantum technologies and the trajectories their marketplace will follow. However, for ease of exposition, and the purposes of presenting some initial rudimentary, *qualitative* analyses and thought experiments, these assumptions simplify our derivations and act as a good starting point for future, more rigorous treatment (which we highly encourage!).

Given that the quantum marketplace doesn't actually exist yet, it isn't immediately clear which assumptions are likely to be valid or not, and future, more sophisticated models will inevitably need to make more appropriate assumptions. Certainly it's no secret that in conventional settings the EMH is flawed in many respects, and some of its idealised assumptions break down in reality.

Post 1 (Efficient markets) *We make the following efficiency assumptions on the dynamics of the quantum marketplace:*

- *Qubits are a ‘scarce’ resource: there is always positive, non-zero demand for them.*
- *No wastage: quantum computational resources are always fully utilised, with no down-time.*
- *Transaction free: transaction costs are negligible, for both quantum assets and their derivatives.*
- *Negligible cost-of-carry: e.g storage and maintenance costs are negligible.*
- *High liquidity: it is always possible to execute transactions at market rates.*
- *Perfect competition: there are no monopolies gouging prices, which are in equilibrium.*
- *Arbitrage-free: market rates for different assets and derivatives are perfectly consistent, with no opportunity for ‘free money’ by trading on market discrepancies.*
- *Perfect information: all market participants have complete knowledge of all market variables, including one another.*
- *Rational markets: all market participants act rationally^a upon available information.*
- *Indefinite asset lifetime: there is no deterioration or death of quantum hardware over time.*
- *There is a risk-free rate of return (r_{rf}): the rate of growth exhibited by an investment into an optimal risk-free asset^b*

^a i.e with perfect economic self-interest — *Avaritia*.

^b Historically these risk-free assets are taken as being US government bonds, with the bond yield being the risk-free RoR.

B. Central mediating authority

In Sec. XLVIII we argued that because of the super-linear scaling in the computational power of networked quantum computers, it will be most economically efficient to unify the world's entire collective quantum computational resources over the network and time-share their joint computational power. For this reason, we will assume that global quantum computing resources are unified, and time-shared equitably (as will be described in Sec. LV),

overseen by a trusted central authority, congruent with our efficient market assumptions (Sec. [XLIX.A](#)).

The role of the mediating authority is to perform process scheduling – equitably allocating algorithmic runtime on the virtual computer to the different network participants. This could be in the form of a state-backed authority, or open market-driven alliances. In any case, the job of the authority is a relatively straightforward one, and we will assume it induces negligible cost and computational overhead, remaining largely transparent to the end-user.

However, as discussed in Sec. [LXXXIII](#), it may be the case that competing strategic interests will drive a wedge between the quantum resources of competitors and adversaries, partitioning them into a set of smaller networks, divided across strategic boundaries. In this instance, the arguments presented in the upcoming sections will apply to these smaller, isolated networks individually.

C. Network growth

We assume the number of qubits in the global network in the future is growing exponentially over time, i.e. the rate of progress of quantum technology will observe a Moore’s Law-like behaviour, as with the classical transistor.

This is a reasonable assumption based on the observation of this ubiquitous kind of behaviour in present-day technologies. Classical computing has been on a consistent exponential trajectory since the 1980’s, and although it must eventually asymptote, it shows no sign of doing so in the immediate future. Quantum technologies sit at the entry point to this trajectory, and we expect it to continue for the medium-term. Thus, we let the number of qubits in the network be,

Post 2 (Network growth) *The number of qubits in the global quantum internet is growing exponentially over time as,*

$$N(t) = N_0 \gamma_N^t, \quad (49.1)$$

where $\gamma_N \geq 1$ characterises the rate of exponential growth in the number of qubits available to the quantum network.

The exact value of the growth rate, γ_N , is obviously unclear at such early stages in the development of the market and will ultimately be determined empirically. Although in the case of classical computing we have seen a very consistent doubling of computational power roughly every 18 months. This may very well be different for quantum technologies, owing to their fundamentally different engineering requirements (which are far more challenging in general).

D. Hardware cost

Let the dollar cost of physical qubits follow Moore’s Law-like dynamics, decreasing exponentially with time,

Post 3 (Hardware cost) *The dollar-cost of a single physical qubit scales inverse exponentially against time as,*

$$C(t) = C_0 \gamma_C^{-t}, \quad (49.2)$$

where $\gamma_C \geq 1$ characterises the decay rate.

This is consistent with the observed evolution of classical hardware since the beginning of the digital revolution, and it is reasonable to think that technological progress in the quantum era will follow a similar trajectory.

L. NETWORK POWER

First and foremost, with a fully interconnected quantum computational network, what is the projection of its net computational power now and into the future? This is simply obtained via the joint computational scaling function applied to projected network size,

Post 4 (Network power) *The combined computational power of the entire network, measured in classical-equivalent runtime (i.e FLOPs), is given by,*

$$\begin{aligned} P(t) &= f_{\text{sc}}(n_{\text{global}}) \\ &= f_{\text{sc}}(N_0 \gamma_N^t). \end{aligned} \quad (50.1)$$

LI. NETWORK VALUE

THE simplest economic metric one might define is the collective dollar value of the entire network. That is, the product of the number of qubits on the network and the dollar cost per physical qubit at a given time.

Post 5 (Network value) *The dollar-value of the entire network is given by,*

$$\begin{aligned} V(t) &= C(t)N(t) \\ &= C_0 N_0 \left(\frac{\gamma_N}{\gamma_C} \right)^t. \end{aligned} \quad (51.1)$$

Note that the collective value of the network appreciates exponentially if the rate of network growth is greater than the rate of decay in the value of physical qubits, otherwise it depreciates. At $\gamma_C = \gamma_N$ the network’s dollar value remains constant over time, even if it continues expanding. This is shown in Fig. [170](#).

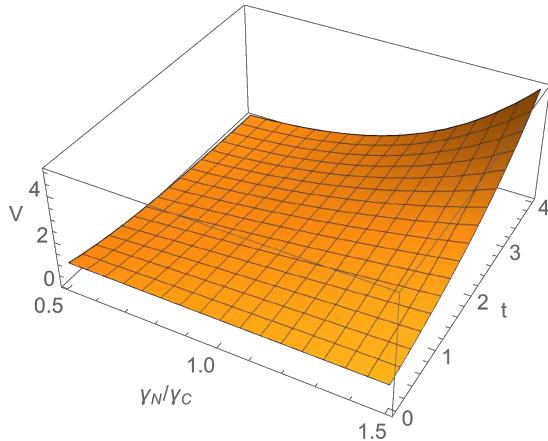


Figure 170: Dollar value (in units of $C_0 N_0$) of the network as a function of time, growth rate in the number of physical qubits, and rate of decay in the dollar value of physical qubits. When $\gamma_N/\gamma_C = 1$ the network's value remains constant over time. Above this the network's value appreciates exponentially, and below which it depreciates exponentially against time.

LII. RATE OF RETURN

THE execution of computations typically has monetary value to the consumer. After all, they are paying hard-earned money for access to the technology!

Suppose the owners of the quantum hardware are not running computations themselves, but rather are collectively licensing out their joint compute-time to end-users. The hardware owners will of course be demanding a profit from their enterprise. The rate at which they earn back their investment into hardware via the licensing of compute-time, we will refer to as the rate of return (RoR), γ_{ror} . We define this as,

Post 6 (Rate of return) *The RoR is defined as,*

$$e^{\gamma_{\text{ror}}(t)} = \frac{R(t)}{V(t)}, \quad (52.1)$$

where $R(t)$ is the profit made by licensing out the network's joint compute-power for a single unit of time, given a present-day network value of $V(t)$.

A higher γ_{ror} implies a faster payback rate on hardware investment⁷⁵.

LIID. MARKET COMPETITIVENESS

ECALL the risk-free RoR is r_{rf} . The difference between the RoR on our investment into qubit assets and the

risk-free rate effectively tells us our profitability relative to a baseline zero-risk asset. This difference in turn can be interpreted as an indicator of the competitiveness or efficiency of the market – more efficient and competitive markets exhibit narrower profit windows. This is yields the figure of merit,

Post 7 (Market competitiveness) *The competitiveness or efficiency of the qubit market is given by the difference between the risk-free RoR and that of our physical qubits,*

$$\xi_{\text{comp}} = \gamma_{\text{ror}} - r_{\text{rf}}. \quad (53.1)$$

There are three distinct regimes for market competitiveness:

- $\xi_{\text{comp}} = 0$: *The market exhibits perfect efficiency, since price competition is so strong that profit windows have narrowed to vanishing point. There is no profit incentive to buy into or sell qubits, since they have converged with the risk-free asset.*
- $\xi_{\text{comp}} > 0$: *The market is profit-making for qubit owners. There is a profit incentive to buy ownership of physical qubits and license them out on the time-share market.*
- $\xi_{\text{comp}} < 0$: *The market is loss-making for qubit owners. Purchasing of physical qubits is disincentivised, since it's more optimal to buy into the zero-risk asset than hold qubit assets.*

The $\xi_{\text{comp}} = 0$ limit is really a fairly hypothetical regime which ought not to arise in real markets, which necessarily exhibit inefficiencies. However, highly-competitive real markets will asymptote to the efficient regime, $\xi_{\text{comp}} \approx 0$.

LIV. COST OF COMPUTATION

As discussed in relation to combined computational scaling functions (Sec. XLVI.B), different market participants will be executing different software applications on their share of the quantum computing resources, with differing QCLs. Because the applications differ between users, as do their computational scaling functions (f_{sc}), QCLs, so too does the monetary value of the computations they are performing. This yields the distinction between *subjective* and *objective* value of computation:

- Subjective value of computation: the value to an end-user of a computation, measured in terms of their associated monetary profit from utilising its output, which is highly application-specific.
- Objective value of computation: the cost of the physical hardware and infrastructure, which is not

⁷⁵ We have parameterised the RoR as an exponential for convenience when performing derivations with compounding.

application-specific, but rather stipulated by technological and manufacturing progress.

This effectively implies that some users pay more for computation (in terms of return on investment) than others. While the objective cost of computation is conceptually simple to model (as performed in a rudimentary fashion in Sec. LIV), the subjective cost is a highly non-trivial one. It will depend heavily on the scaling function of the algorithm run by a user, and of course the economic objectives of their computation – a quantum simulation algorithm executed by an R&D lab is likely to be of greater monetary value than an undergrad using his university’s resources to execute the same task for completing an assignment!

A. Objective value

In the same scenario as before, where compute-time is being licensed out to end-users, the hardware owner’s return over a single unit of time equates to the cost of computation over that period.

Let $L(t)$ be the dollar-value of utilising the network’s computing resources for a single unit of time. This is obtained as the return made on the value of the network per FLOP,

Post 8 (Objective value of computation) *The efficient-market dollar-value of a computation for a single unit of time at time t , per FLOP is,*

$$\begin{aligned} L(t) &= \frac{e^{\gamma_{\text{ror}} t} V(t)}{P(t)} \\ &= \frac{e^{\gamma_{\text{ror}} t} C_0 \gamma_C^{-t}}{\chi_{\text{sc}}(N_0 \gamma_N t)}. \end{aligned} \quad (54.1)$$

which implies,

Post 9 (Spot price of computation) *The present-day ($t = 0$) spot price of a computation per FLOP is,*

$$L(0) = \frac{C_0}{\chi_{\text{sc}}(N_0)}. \quad (54.2)$$

That is, the value of computations simply approximates the return on initial hardware investment, scaled by its initial computational power, as is intuitively expected.

Note that if f_{sc} scales linearly, as per classical computation, we observe a regular exponential decay in the cost of computation, consistent with the classical Moore’s Law. On the opposing extreme, for exponentially quantum-enhanced f_{sc} , the cost of computation decreases super-exponentially with time, an economic behaviour unique to post-classical computation with no classical analogue.

The time-derivative of the cost of computation is strictly negative, assuming correctness of the growth and cost postulates (Post. 2 & 8),

$$\frac{\partial L}{\partial t} \leq 0, \quad (54.3)$$

which implies monotonic reduction in the cost of computation over time, unless network growth and cost completely freeze ($\gamma_N = \gamma_C = 0$), in which case the cost of computation remains flat.

Examples of the temporal dynamics of the cost of computation are shown in Fig. 171 for representative computational scaling functions.

B. Subjective value

With access to n qubits, let the subjective value extracted by user i from its execution for a unit of time be characterised by $f_{\text{sub}}^{(\text{joint})}(n)$. Then, the joint subjective value of computation (i.e total market subjective return) follows the same form as the joint computational scaling function given in Def. 11,

Def 13 (Subjective value of computation)

The joint subjective value of a time-share in the global network, $f_{\text{sub}}^{(\text{joint})}$, of a group of participants, each with their own subjective valuation functions, $f_{\text{sub}}^{(i)}$,

$$f_{\text{sub}}^{(\text{joint})} = \sum_{i=1}^{n_{\text{nodes}}} \beta_i \cdot f_{\text{sub}}^{(i)}(n_{\text{global}}). \quad (54.4)$$

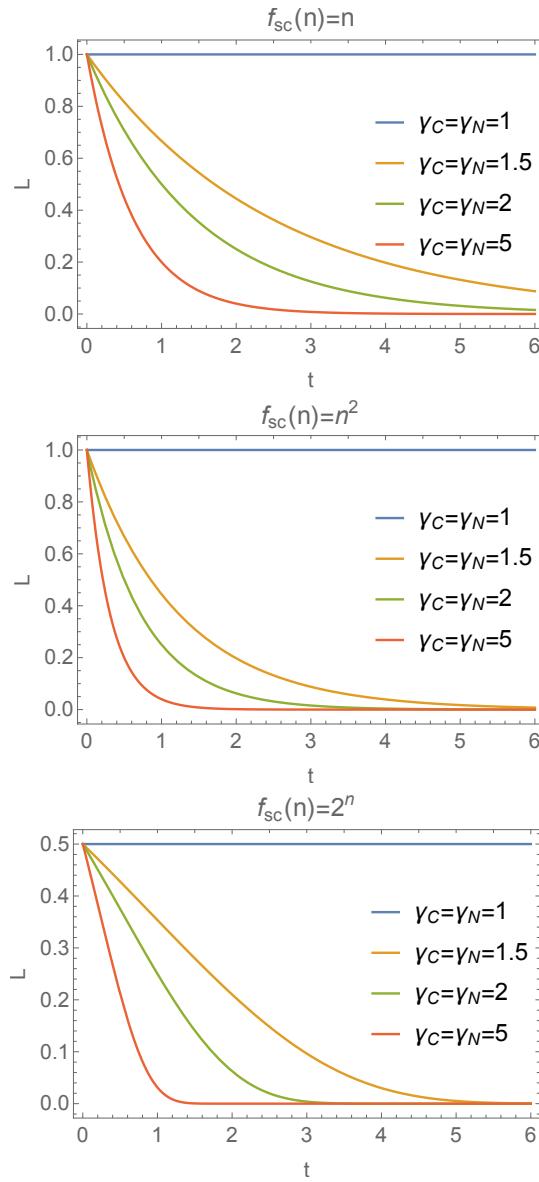


Figure 171: Examples of the temporal dynamics of the cost of computation for different scaling functions and exponential growth rates. Units are $C_0 = N_0 = n = 1$, with RoR $\gamma_{\text{ror}} = 0$. A non-zero RoR would simply scale these figures by a constant factor of $e^{\gamma_{\text{ror}}}$.

LV. ARBITRAGE-FREE TIME-SHARING MODEL

In the context of our time-shared global network of unified quantum computers (Sec. XLVIII), how do we fairly and equitably allocate time-shares between contributors? We now derive an elementary arbitrage-free model for equitable time-sharing in such a network.

Let,

$$0 \leq r_n \leq 1, \quad (55.1)$$

be the proportion of compute-time allocated to a node in possession of n qubits, in a global network of n_{global} qubits.

Arbitrage in the value of physical qubits will enforce the linearity constraint,

$$r_{n_1+n_2} = r_{n_1} + r_{n_2}. \quad (55.2)$$

This constraint effectively mandates that ‘all qubits are created equal’, and two qubits are twice as valuable as one — *Qubit aequalitatem*. Were, for example, a bundle of two qubits more expensive than two individual qubits purchased in isolation, a market participant could perform arbitrage and unfairly gain free compute-time by buying two qubits separately, unifying them, selling the bundle, buying them back individually, and repeating indefinitely until he seizes the entire network.

Additionally, we have assumed no compute-cycles are wasted – compute-time is always fully utilised, as per Post. 1. Then it follows that the time-share of the combined resources of the entire network should be unity,

$$r_{n_{\text{global}}} = 1. \quad (55.3)$$

$r_{n_{\text{global}}} < 1$ would imply inefficiency via wasted clock-cycles. Combining this with the linearity constraint implies the arbitrage-free time-sharing model,

Def 14 (Arbitrage-free time-sharing model)

In an efficient market for unified quantum computing time-shares, a network participant in possession of n of the entire n_{global} qubits in the network is entitled to the fraction of unified network compute time,

$$r_n = \frac{n}{n_{\text{global}}}, \quad (55.4)$$

where,

$$n_{\text{global}} = \sum_{j \in \text{nodes}} n_j, \quad (55.5)$$

is the total number of qubits in the network, and,

$$0 \leq r_n \leq 1. \quad (55.6)$$

$r_n = 1$ iff the node has a complete monopoly over qubits, i.e $n = n_{\text{global}}$.

Based on this equitable model for time-sharing,

Def 15 (Time-shared computing power) *The computing power allocated to each user under the arbitrage-free time-sharing model is,*

$$\begin{aligned} c_n &= r_n \cdot f_{\text{sc}}(n_{\text{global}}) \\ &= n \cdot \chi_{\text{sc}}(n_{\text{global}}). \end{aligned} \quad (55.7)$$

This model is intuitively unsurprising, since it is analogous to the case of classical computer clusters – nodes receive a time-share proportional to the proportion of the hardware they are contributing to the network. However,

it is important to point out that the arbitrage is taking place in the cost of physical qubits, but not in terms of the dollar value of their classical-equivalent processing power, since this is in general non-linearly related to the number of qubits. Arbitrage in computational power per se is complicated by the fact that it is a non-fungible asset that cannot be directly traded, or uniquely associated with a tangible, tradable asset – its computational value is a function of other assets.

LVI. PROBLEM SIZE SCALING FUNCTIONS

THE computational scaling function introduced previously expresses the power of a quantum computer in terms of its classical-equivalent runtime, or equivalently FLOPs. However, this may not be the metric of interest when considering a computer's algorithmic power. In many situations, of far greater interest is the size of a problem instance that can be solved in a given timespan. For example, the FLOPs associated with solving an instance of a 3-SAT problem grows exponentially with the number of clauses. When discussing the execution of this problem on a given computer, what we really want to know is how many clauses our device can cope with, rather than what the classical-equivalent runtime is.

This observation motivates us to re-parameterise the power of quantum computers in terms of the problem size of a given algorithm to be solved. Employing the same methodology as for computational scaling functions, we define the *problem size scaling function*, which relates the size of an algorithmic problem to its classical equivalent runtime. Then equating the computational and problem size scaling function yields,

Def 16 (Problem size scaling function) *The problem size scaling function relates the size of a problem instance (s), in some arbitrary metric, to its classical-equivalent runtime (t) under a time-shared network model,*

$$t = f_{\text{size}}(s). \quad (56.1)$$

Equating this with the time-shared computational power yields,

$$n \cdot \chi_{\text{sc}}(n_{\text{global}}) = f_{\text{size}}(s). \quad (56.2)$$

Isolating the problem size yields,

$$s = f_{\text{size}}^{-1}(n \cdot \chi_{\text{sc}}(n_{\text{global}})). \quad (56.3)$$

We now consider several choices of scaling functions.

First let us consider the classical case of linear scaling functions (for both the computational and problem size

scaling functions),

$$\begin{aligned} f_{\text{sc}}(n) &= \alpha_{\text{sc}} n, \\ f_{\text{size}}(s) &= \alpha_{\text{size}} s. \end{aligned} \quad (56.4)$$

Solving for the problem size simply yields,

$$\begin{aligned} s &= \frac{n}{\alpha_{\text{size}}} \\ &= O(1), \end{aligned} \quad (56.5)$$

where n is regarded as a constant, and n_{global} is a variable parameter of the network. That is, the problem sizes of solvable instances is independent of the size of the external network with whom we are time-sharing. This is to be expected, since these scaling functions are typical of classical computers.

For polynomial scaling functions,

$$\begin{aligned} f_{\text{sc}}(n) &= n^{p_{\text{sc}}}, \\ f_{\text{size}}(s) &= s^{p_{\text{size}}}. \end{aligned} \quad (56.6)$$

This yields problem size,

$$\begin{aligned} s &= (n \cdot n_{\text{global}})^{p_{\text{sc}}-1} \frac{1}{p_{\text{size}}} \\ &= O(\text{poly}(n_{\text{global}})), \end{aligned} \quad (56.7)$$

demonstrating polynomial scaling in our solvable problem size against the size of the network.

For exponential scaling functions,

$$\begin{aligned} f_{\text{sc}}(n) &= e^{\alpha_{\text{sc}} n}, \\ f_{\text{size}}(s) &= e^{\alpha_{\text{size}} s}, \end{aligned} \quad (56.8)$$

we obtain,

$$\begin{aligned} s &= \log \left(n \frac{e^{\alpha_{\text{sc}} n_{\text{global}}}}{\alpha_{\text{sc}} n_{\text{global}}} \right) \\ &= \alpha_{\text{sc}} n_{\text{global}} + \log(n) - \log(\alpha_{\text{sc}} n_{\text{global}}) \\ &= O(n_{\text{global}}), \end{aligned} \quad (56.9)$$

demonstrating that the solvable problem size grows linearly with network size. That is to say, waiting for a doubling in the external network's size will also double the size of a **BQP**-complete problem that can be solved in the same time.

LVII. QUANTUM COMPUTATIONAL LEVERAGE

IN Secs. [XLIII.B](#) & [XLIII.D](#) we introduced distributed and modularised quantum computation. Using this as a toy model, we will now investigate the market dynamics of uniting the quantum computational resources of multiple market participants, as per an equitable time-sharing model (Sec. [XLVIII](#)). We envisage a model whereby network participants are contributing modules to the networked quantum computer, thereby unifying their computational power.

The i th node is contributing the fraction of the hardware r_i , and receives this same proportion of compute-time under the arbitrage-free time-sharing model (Def. 14). This discounts his classical-equivalent processing time to,

$$\tau_i = t_{\text{joint}} \cdot r_i. \quad (57.1)$$

We are now interested in quantifying how much better off individual contributors are under this model than they were individually. Let us define the *quantum computational leverage* (QCL) of a node's quantum computer to be the ratio between their unified time-shared and individual classical-equivalent processing times,

$$\lambda_i = \frac{\tau_i}{t_i}, \quad (57.2)$$

yielding the QCL formula,

Def 17 (Quantum computational leverage)

For the i th node, and with scaling function f_{sc} , the QCL is defined as the ratio between the unified time-shared and individual classical-equivalent algorithmic runtimes,

$$\begin{aligned} \lambda_i &= \frac{\tau_i}{t_i} \\ &= \frac{n_i}{n_{\text{global}}} \cdot \frac{f_{sc}(n_{\text{global}})}{f_{sc}(n_i)} \\ &= \frac{\chi_{sc}(n_{\text{global}})}{\chi_{sc}(n_i)}, \\ \lambda_i^{\text{dB}} &= 10 \log_{10}(\lambda_i), \end{aligned} \quad (57.3)$$

where,

$$n_{\text{global}} = \sum_{j \in \text{nodes}} n_j, \quad (57.4)$$

is the total number of qubits in the network. The logarithmic version of the representation in decibels is simply a convenience when dealing with exponential scaling functions.

Effectively, the QCL tells us how much additional computational power we 'get for free' by consolidating with the network.

It is extremely important to note that the QCL is asymmetric, in the sense that the leverage achieved by a given node is larger than the leverage achieved by the network, upon the user joining the network (assuming the remainder of the network comprises more qubits than the respective user).

More generally, smaller users achieve higher computational leverage from their investment into quantum hardware than larger users. Specifically,

$$\lambda_i < \lambda_j \text{ for } n_i > n_j. \quad (57.5)$$

For any super-linear scaling function we have $\lambda_i > 1 \forall i$, and for any linear scaling function we have $\lambda_i = 1 \forall i$,

$$\begin{aligned} \lambda &= 1 \forall f_{sc}(n) = O(n), \\ \lambda &> 1 \forall f_{sc}(n) > O(n). \end{aligned} \quad (57.6)$$

For $\lambda_i > 1$ it is always computationally beneficial to all nodes to unify computational resources and time-share them equitably, as per the arbitrage-free time-sharing model. Similarly, the distributed network is better off accepting them into the network, albeit to a lesser extent for a large network.

This is in contrast to classical networks, where $\lambda \approx 1$, for any number of nodes in the network (i.e there is no leverage), and it makes no difference whether nodes unify resources or operate independently.

Finally, in the pathological case, where $\lambda_i < 1$, nodes are better off working in isolation, a situation which would only naturally arise as a result of algorithmic inefficiencies in parallelisation or distribution.

Def 18 (Single-qubit QCL) The single-qubit QCL is the leverage associated with adding a single qubit to the network, $n = 1$, defined as,

$$\begin{aligned} \lambda_{\text{qubit}} &= \frac{\chi_{sc}(n_{\text{global}})}{\chi_{sc}(1)}, \\ \lambda_{\text{qubit}}^{\text{dB}} &= 10 \log_{10}(\lambda_{\text{qubit}}). \end{aligned} \quad (57.7)$$

Using our postulate for network growth (Post. 2) yields the postulated time-dependent QCL,

Post 10 (Time-dependent QCL) The time-dependent QCL, based on the postulate of exponential network growth, is,

$$\begin{aligned} \lambda_n(t) &= \frac{\chi_{sc}(N_0 \gamma_N t)}{\chi_{sc}(n)}, \\ \lambda_n^{\text{dB}}(t) &= 10 \log_{10}(\lambda_n(t)). \end{aligned} \quad (57.8)$$

The initial ($t = 0$) time-dependent QCL reduces to the standard QCL formula.

Note that for any super-linear scaling function, the time-dependent QCL grows exponentially over time, unlike the classical case where there is no leverage, which does not change over time (i.e $\lambda_n(t) = 1 \forall n, t$).

The leverage is not merely a function of the hardware, but also of the software applications running upon it, each of which associated with a unique scaling function. Furthermore, it is to be reasonably anticipated that the size of the quantum internet will increase monotonically over time, yielding ever increasing leverage on the initial hardware investment by network contributors.

Examples of the temporal dynamics of the time-dependent single-qubit QCL are shown in Fig. 172.

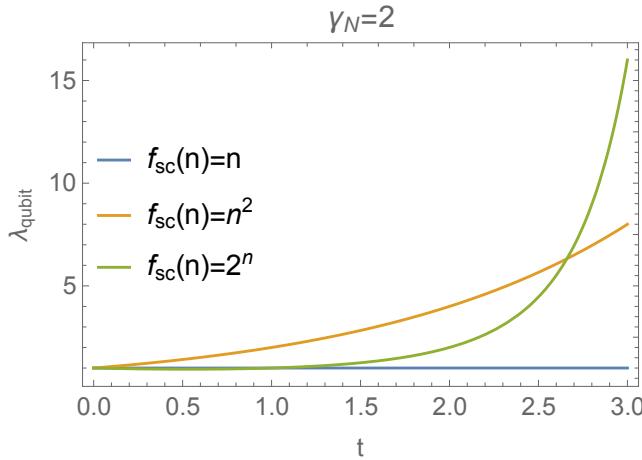


Figure 172: Time-dependent quantum computational leverage for a single qubit ($n = 1$) with different computational scaling functions, under the assumption of exponential network growth in units of $N_0 = 1$.

LXVIII. STATIC COMPUTATIONAL RETURN

THE computational leverage phenomena clearly implies that as the global quantum network expands over time, so too does the computational payback on investment into network expansion, or equivalently, the cost per unit of additional classical-equivalent processing time decreases.

Since existing network participants receive leverage upon *other* participants joining the network, an investment into contributing modules has monotonically increasing computational return over time as the network expands, even if that participant ceases making further investment into the network. This is in contrast to classical networks, whereby the computational return upon an investment is fixed over time.

To formalise this, consider the case where a user purchases an initial n qubits, while the global network expands over time as N_t . Then the classical-equivalent computational power of the user's fixed investment is,

Def 19 (Static computational return) *The static computational return is the classical-equivalent processing power of a user's time-share proportion (n/N_t), where the user has a fixed investment of n qubits, whereas the network is allowed to expand over time arbitrarily as N_t (e.g according to a quantum Moore's Law),*

$$\begin{aligned} r_{\text{static}}(t) &= \frac{n \cdot f_{\text{sc}}(N_t)}{N_t} \\ &= n \cdot \chi_{\text{sc}}(N_t), \end{aligned} \quad (58.1)$$

which intuitively follows as the computational power per qubit in the network, times the number of qubits in our possession.

Graphical examples for this relationship are equivalent to those presented earlier in Fig. 169. In particular, for linear classical scaling functions the return is constant, whereas for exponential quantum scaling functions the return is exponential in network size.

LIX. FORWARD CONTRACT PRICING MODEL

FORWARD contracts are immensely useful in conventional markets, as a means by which to secure future use or ownership of an asset at predictable points in time. For example, farmers make heavy use of forward contracts to lock in sale of their produce before it has been harvested, such that the value is locked in in advance and the sale guaranteed, providing a very valuable hedging instrument for managing risk.

We envisage similar utility in the context of quantum computing. A company engaging in heavy use of computing power might have a need to perform certain computations at predictable points in the future. In this instance, forward contracts could be very helpful in reducing exposure to risk and guaranteeing access to the technology when needed, at a pre-agreed rate.

Now let us price forward contracts on units of computation, whereby we wish to pay today for the future use of a block of runtime on the global network.

The key observation is that a unit of computation (FLOP) does not carry over time. It must be utilised immediately and cannot be stored for future use. This simplifies the forward price of a unit of computation to simply be the future spot price, discounted by the risk-free RoR, yielding the forward contract pricing model for quantum computing time-shares,

Def 20 (Forward contract pricing model)

The efficient market price for a forward contract in a unit of network runtime at future time T is,

$$\begin{aligned} F(T) &= e^{-r_{\text{rf}}T} L(T) \\ &= \frac{e^{(\gamma_{\text{ror}} - r_{\text{rf}})T} C_0 \gamma_C^{-T}}{\chi_{\text{sc}}(N_0 \gamma_N T)}. \end{aligned} \quad (59.1)$$

Note that in the limit of $T \rightarrow 0$ this reduces to the spot price of the asset,

$$F(0) = L(0), \quad (59.2)$$

as expected.

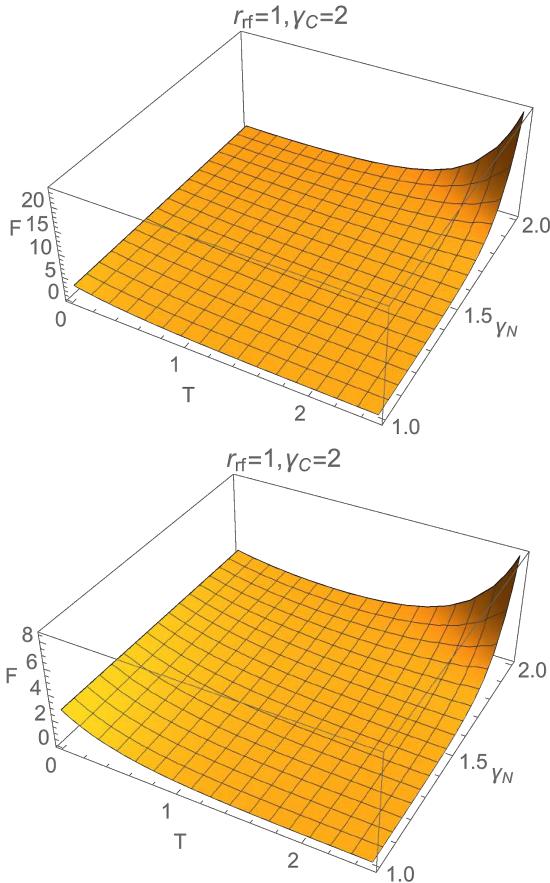


Figure 173: Forward price on a computation to be delivered at time T in the future, in units $C_0 = N_0 = 1$, where we are assuming an exponential scaling function, $f_{sc}(n) = e^n$.

LX. POLITICAL LEVERAGE

THE asymmetry in computational leverage observed by parties of different sizes – specifically, that parties possessing a smaller number of qubits observe greater leverage than those possessing a larger number of qubits – inevitably will bring with it some power politics, with potentially interesting geo-political implications.

This asymmetry implies that in a globally unified network, were a large party to expel a small party from the network, it would be far more devastating to the computational power of the small party than the larger one. This suggests that inclusion in the global network could be a powerful tool of diplomacy in the quantum era, where threats of expulsion or resistance to inclusion is the modern day era of gunboat diplomacy.

To quantify this we introduce the *political leverage* quantity – the ratio between the computational leverages observed by two parties belonging to the same network. This directly quantifies the power asymmetry between them.

Def 21 (Political leverage) *The political leverage is the ratio between the computational leverages of two parties residing on the same shared network,*

$$\begin{aligned}\gamma_{A,B} &= \frac{\lambda_A}{\lambda_B} \\ &= \frac{\chi_{sc}(n_B)}{\chi_{sc}(n_A)}, \\ \gamma_{A,B}^{\text{dB}} &= 10 \log_{10}(\gamma_{A,B}).\end{aligned}\quad (60.1)$$

We have the trivial identity that the leverage of A against B is the inverse of the leverage of B against A ,

$$\begin{aligned}\gamma_{A,B} &= \gamma_{B,A}^{-1}, \\ \gamma_{A,B}^{\text{dB}} &= -\gamma_{B,A}^{\text{dB}}.\end{aligned}\quad (60.2)$$

Note that when two parties are of equal size, there is no power asymmetry and $\gamma_{A,B} = 1$. Otherwise, when A and B are unequal, then $\gamma_{A,B} \neq 1$, indicative of power asymmetry. With linear (classical) scaling functions, the political leverage is always unity, $\gamma_{A,B} = 1$, regardless of any size asymmetry, whereas for super-linear scaling functions the political leverage diverges.

To the Machiavellian reader, this quantity can be thought of as answering the question ‘If I were to expel a party from the network, how much more would it hurt them than it would hurt me?’.

LXI. QUANTCOIN™ – A QUANTUM COMPUTATION-BACKED CRYPTOCURRENCY

As discussed in Sec. XXXV, the Bitcoin mining process involves finding bit-strings that hash under SHA256 to a value within some relatively small range. This so-called ‘proof-of-work’ principle associates computational complexity with the mining process, and since the hashing functions are one-way functions, they must be evaluated via brute-force trial-and-error to find hits.

However, what a waste this is! Our proof-of-work is nothing more than hashing a huge number of random bit-strings, computations which are of no intrinsic value to anyone. The market value in turn has nothing to do with any inherent value earned during the mining process. Rather it is based purely on the psychology of scarcity, since there is an upper bound on the number of Bitcoins that satisfy the legitimacy constraint.

What if we were to replace brute-force hashing of random data with computations of genuine monetary value? Then we would have a sounder currency, whose value derives from the monetary cost of executing useful computations. While it is not so easy to invent such a protocol for classical computation, the idea lends itself very naturally to quantum computation, owing to their ability

to undergo encrypted computation and be subject to verification protocols.

Building upon some of the pricing models introduced earlier in this section, there are two main candidates for backing a cryptocurrency with quantum compute-time:

- Spot market: we execute the computation immediately in exchange for a coin.
- Futures market: we own the right to utilise the computer at some fixed time in the future in exchange for a coin.

We consider the merits of both these candidates.

A popular-level essay on the future of quantum cryptocurrencies is presented in Sec. LXXXV.

A. Spot market model

In Alg. LI we provide a very rough sketch for how a protocol based on the spot market might be constructed. A corresponding graphical flowchart is shown in Fig. 174. We present the ideas in a very high-level manner, abstracting away the physical implementation details of the computation, encryption, and verification protocols, instead envisaging that we can interface with them using a very high-level API.

It is evident from the flow of Alg. LI that the mining process now comprises solving an actual quantum computation of intrinsic value to Alice, since she is exchanging assets (e.g dollars or already-existing QuantCoinsTM) in exchange for the computation. Completion of the computation followed by successful verification then further rewards Bob with a fresh QuantCoinTM courtesy of the distributed Blockchain algorithm.

The described protocol, in addition to mining a new coin, associated with the execution of a computation, acts as a currency converter for converting traditional assets (e.g dollars) into QuantCoinsTM. This ability to currency convert is necessary, and completely differs from the original Bitcoin mining process, where coins are fabricated out of thin air by anyone and everyone, independent of their pre-existing monetary wealth – Bitcoins are not created via conversion from any existing asset, they are an entirely new asset class of their own. The fact that the computation associated with each QuantCoinTM is of intrinsic value on the other hand, implies that Alice ought to be paying something for the service.

Note that we observe an expansion in the money supply with each successfully executed and verified computation – one additional unit of QuantCoinsTM is mined for every unit of computation implemented⁷⁶. Unlike Bitcoin, there

```

function QuantCoin( $\hat{U}_{\text{comp}}$ , Blockchain, data):
    1. Alice homomorphically/blindly encrypts
       data,
        $\text{encryptedInput} = \text{homoEncrypt}(\text{data})$  (61.1)

    2. Alice commits the  $\text{encryptedInput}$  to the
       public Blockchain,
        $\text{Alice.Blockchain.commit}(\text{encryptedInput})$  (61.2)

    3. Bob processes the  $\text{encryptedInput}$ ,
        $\text{encryptedOutput} = \hat{U}_{\text{comp}}(\text{encryptedInput})$  (61.3)

    4. Bob commits encrypted output to the
       Blockchain,
        $\text{Bob.Blockchain.commit}(\text{encryptedOutput})$  (61.4)

    5. Alice decrypts the output.
        $\text{output} = \text{homoDecrypt}(\text{encryptedOutput})$  (61.5)

    6. Alice and Bob execute a verification
       protocol,
        $\text{proof} = \text{verify}(\text{encryptedOutput})$  (61.6)

    7. If successful, alice commits a
       zero-knowledge proof of the result to
       the Blockchain,
        $\text{Alice.Blockchain.commit}(\text{ZKP}(\text{input}, \text{output}))$  (61.7)

    This signs off on the legitimacy of the
    mining entry previously committed by Bob.

    8. All users on the network can inspect the
       ZKP to validate the legitimacy of the
       execution, maintaining ignorance of the
       computational outcome.

    9.  $\square$ 
```

Algorithm LI: Sketch for how a quantum computation-backed cryptocurrency might be implemented. We have abstracted away the underlying Blockchain protocol, interfacing with it using a high-level API, since Blockchain technology is highly liable to evolve. We similarly call upon verification subroutines using a high-level implementation-independent API. The key technique to signing off on the legitimacy of a newly mined QuantCoin is for Alice to provide a zero-knowledge proof of the decrypted result (Sec. XLV.B), which maintains the secrecy of her data.

⁷⁶ To provide an analogy with the gold standard, think of the

physical quantum computer as the goldmine, and each unit of gold it produces as being a QuantCoinTM. The production of each unit of gold is associated with the utilisation of the mine for a particular amount of time, but the mine can in principle operate indefinitely, with no hard upper-bound on its total future gold yield.

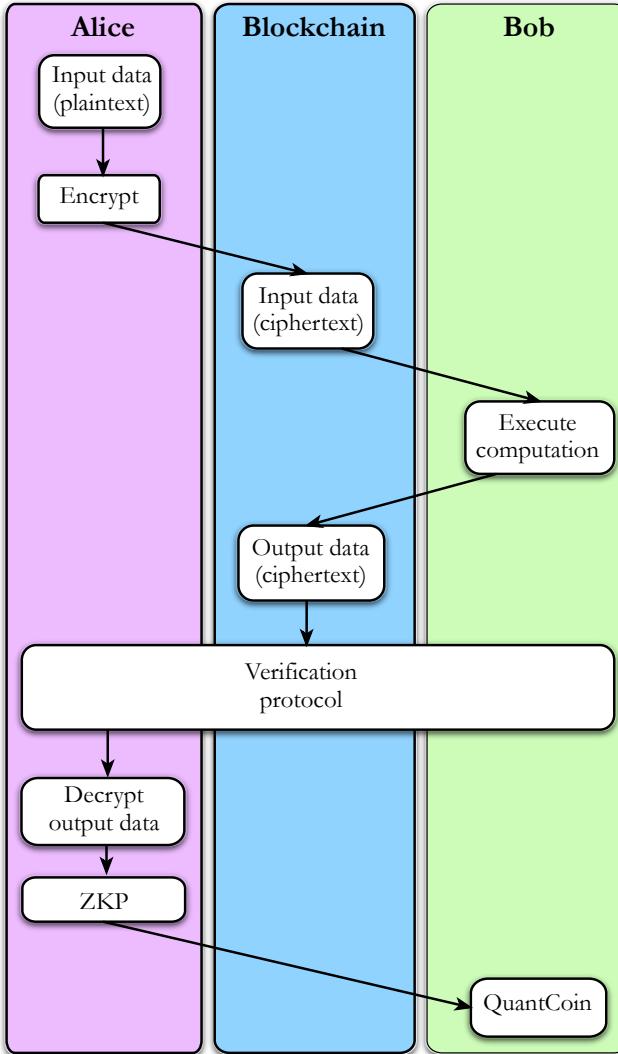


Figure 174: Flowchart for the QuantCoin™ protocol, introduced in Alg. LI.

is no inherent theoretical upper limit on the number of coins that can exist. However the QuantCoin™ money supply will be limited for the practical reason that mining each one is associated with a monetary transaction between Alice and Bob, and Alice will eventually run out of assets to exchange for computations.

What relationships characterise the value of QuantCoins™? First, in a perfectly efficient market (Sec. XLIIX.A) we have,

$$P_{\text{coin}} + P_{\text{reward}} = P_{\text{exec}}, \quad (61.8)$$

where, P_{coin} is the dollar value of a QuantCoin™, P_{reward} is the dollar value of the reward paid by Alice for execution, and P_{exec} is the dollar value of cost of execution for Bob.

Alternately, rather than Alice paying Bob's reward in dollars, she might pay for them in already-existing QuantCoins™. Suppose Alice pays λ QuantCoins™ as

Bob's reward. Then Eq. (61.8) reduces to,

$$P_{\text{coin}} = \frac{1}{\lambda + 1} P_{\text{exec}}, \quad (61.9)$$

providing us with a simple financial model relating the market price of QuantCoins™ and the monetary cost of execution of computations.

Note that with exception to the scenario where Alice buys into QuantCoins™ using dollar currency (or any non-electronic asset that cannot be committed to the Blockchain), the entire protocol is self-enforcing via programmed Blockchain transactions. Bob doesn't get paid his newly earned and freshly printed QuantCoin™ until the verification of the computation has completed and the proof committed to the Blockchain. He therefore cannot get paid until he has executed the computation he promised to, and proven to the network that he actually did.

The main security risk is that of Bob taking Alice's dollars and running, upon receiving the upfront reward in dollars, which necessarily don't reside on the Blockchain since they are not crypto-assets. This could be addressed by introducing trusted third-party escrow agents into the protocol, as method currently used in some online dark markets.

However, if the upfront payment were being made in pre-existing QuantCoins™, the Blockchain might be programmed to not release the reward until completion of the final verification stage of the protocol – effectively an escrow programmed directly into the Blockchain for self-execution. Such self-executing smart-contracts are already a feature in some cryptocurrencies such as Ethereum.

B. Futures market model

As described above, the cryptocurrency is effectively backed by the spot market in computation – we exchange currency for the execution of computations *immediately*. However, one might also envisage more complex cryptocurrencies being backed by the futures market in the licensing of quantum compute-time down the line.

Intuitively, one would expect such a form of cryptocurrency to be sounder than the one backed by the spot market. This is because our spot market-derived coins, once mined are not guaranteed to be convertible to anything of value, including computations. Recall that the execution of the computation takes place immediately when the coin is created.

On the other hand, a QuantCoin™ backed by a guarantee to access quantum compute-time at a designated point in the future necessarily has value, so long as the demand for compute-time does, and maintains value until the contract matures and converts into compute-time at which point it becomes worthless.

We leave explicit construction of a futures-based QuantCoin™ model as an exercise for the interested reader, primarily because we haven't thought it through properly.

LXII. ECONOMIC PROPERTIES OF THE QUBIT MARKETPLACE

"Economics is probably the weirdest academic discipline I've come across. I find myself constantly in a superposition of fascination and annoyance with how the field currently stands. How can smart people have come up with a collection of ideas that are simultaneously brilliant and ridiculous, insightful and delusional, pragmatic and useless?" — Andrew Ringsmuth.

— by Scott Harrison.

THE development and implementation of the quantum internet will give rise to a new tradable commodity – the qubit. The pricing mechanisms associated with a qubit market were explained earlier in this part. Here we provide a broader discussion on the economic properties of such a marketplace. There are two areas of particular interest we will focus on:

1. The responsiveness of the qubit market to price fluctuations, measured by elasticity.
2. The implications of qubit market properties for broader society in terms of pricing and taxation.

A. The concept of elasticity

Elasticity as a concept is measured through percentage changes. Starting with the demand for qubits as an example, the *elasticity of demand*, E_d , is the percentage change in the quantity demanded of a good, divided by the percentage change in the price of a good. Mathematically, the elasticity of demand is represented as,

$$E_d = \frac{\% \Delta Q_d}{\% \Delta P}. \quad (62.1)$$

From this relationship, inferences about the underlying commodity can be made, summarised as follows:

- $|E_d| > 1$: the percentage change in quantity is greater than the percentage change in price, and is therefore *elastic*. This indicates that demand for the asset in the market is responsive to small price changes.
- $|E_d| < 1$: there is a proportionally larger change in price, for a smaller shift in demand. This indicates that demand is less responsive to price changes, and considered *inelastic*.

- $|E_d| = 1$: we have *unit elasticity*, where the percentage change in quantity demanded is equal to the percentage change in market price.

Elasticity of demand is just one context where the concept of elasticity can be applied. Other contexts include: the elasticity of supply, measuring the supply side responsiveness to changes in price; income elasticity, which captures how the quantity of goods in the market change relative to changes in the income of consumers; and cross-price elasticity, which compares the percentage change in quantity of one good, relative to the percentage change in price of another good. An example that we will come back to is how changes in the price of quantum computing may have an effect on the quantity demanded of high-performance classical computing (i.e conventional supercomputing).

B. Elasticity of the qubit market

A number of factors will affect the elasticity of demand and supply in the qubit market. The most significant factor affecting the demand for qubits is the availability of substitutes. Given that quantum processing can efficiently solve unique problems that classical computing cannot, there are no close substitutes for qubits – a transistor is no substitute for a qubit! Consequently, elasticity of demand will be relatively inelastic: the quantity of qubits demanded will be relatively unresponsive to price fluctuations and changes, since there are no viable alternatives to substitute with.

The supply side of the equation is also initially going to be highly inelastic. However, as time progresses and technological advancements and enhancements increase the computational power of the quantum internet, the supply of qubits will become increasingly responsive to price fluctuations **Why?**. However, the extent to how elastic the supply becomes over time is also going to be affected by potential for excess capacity given the exponential trajectory of the manufacture of qubits as new and improved fabrication technologies emerge – the quantum Moore's Law.

LXIII. ECONOMIC IMPLICATIONS

OUR analysis thus far has been very theoretical. But our observations have very tangible implications in the real-world. This has implications for governments, regulatory authorities, fiscal and technology policy, national security, and any end users of the quantum cloud.

A. The price to pay for isolationism

In many traditional sectors of the economy there is an economic incentive to directly compete against other market participants. However in the quantum era the incentive is for owners of quantum computing hardware to cooperate and contribute their resources to the quantum internet rather than go it alone, as a direct consequence of super-linear leverage.

Only those hardware owners who unite with the global network will benefit from its leverage and remain competitive. Those who choose not to participate in the global network will be priced out of the market via exponentially higher cost per FLOP (assuming all other costs are equal).

This effectively taxes the cost of computation for those who fail to unify their assets with the network. And it is in the direct economic self-interest of all market participants to contribute their resources to the time-shared quantum cloud.

B. Taxation

Any asset, dividend, derivative or other financial instrument will inevitably be subject to taxation. Any form of taxation has multiplier effects as the cost markup is repeatedly handed from one market participant to the next, influencing the chain of supply and demand along the way. However, this multiplier and other economic consequences are highly dependent on the asset undergoing transaction – the economic implications of personal income tax are quite different to those of capital gains tax!

1. Computational perspective

We now consider the effect of taxation on quantum resources, specifically in the form of a *qubit tax* – a sales tax on the purchase of physical qubits. Although this model of taxation is unlikely to be implemented as we describe, it serves as an insightful test-bed for thought experiments into the qualitative implications of taxing quantum assets.

Imagine that consumers have an amount of capital available for the purchase of qubits. Let γ_T be the rate of taxation ($\gamma_T = 1$ represents no taxation, $\gamma_T > 1$ represents positive taxation, and $\gamma_T < 1$ represents subsidisation). Then the cost of physical qubits is marked up by γ_T , reducing the number of qubits that can be afforded by the consumers to (assuming fixed capital available for purchasing),

$$N_{\text{tax}} = \frac{N_{\text{no tax}}}{\gamma_T}. \quad (63.1)$$

We now wish to understand how this taxation influences the computational power of the network. We define the *tax performance multiplier*,

Def 22 (Tax performance multiplier) *The tax performance multiplier, is the ratio between computational scaling functions with and without qubit taxation,*

$$\begin{aligned} M(N_{\text{tax}}) &= \frac{f_{\text{sc}}(N_{\text{tax}})}{f_{\text{sc}}(N_{\text{no tax}})} \\ &= \frac{f_{\text{sc}}(N_{\text{tax}})}{f_{\text{sc}}(N_{\text{tax}}\gamma_T)}, \\ M^{\text{dB}}(N_{\text{tax}}) &= 10 \log_{10}(M(N_{\text{tax}})), \end{aligned} \quad (63.2)$$

where the consumers have purchased N_{tax} qubits, after taxation, at a markup rate of γ_T .

The tax performance multiplier effectively gives us a factor by which computational power is depreciated under taxation. We can accommodate for other models of taxation and regulation by choosing an appropriate relationship between N_{tax} , $N_{\text{no tax}}$, and the taxation and regulatory framework.

Using our illustrative examples of computational scaling functions (linear, polynomial and exponential), the respective tax performance multipliers are given by,

$$\begin{aligned} M_{\text{linear}}(N_{\text{tax}}) &= \frac{1}{\gamma_T}, \\ M_{\text{poly}}(N_{\text{tax}}) &= \frac{1}{\gamma_T^p}, \\ M_{\text{exp}}(N_{\text{tax}}) &= e^{N_{\text{tax}}(1-\gamma_T)}. \end{aligned} \quad (63.3)$$

This demonstrates that the computational power of classical networks is simply inversely proportional to the rate of taxation, i.e a linear tax performance multiplier, as we intuitively expect. And for quadratic scaling functions the dependence is inverse quadratic in the taxation rate. In both cases the multiplier is a constant factor, independent of the network size. However, for exponential scaling functions we observe an exponential dependence on both the rate of taxation and the size of the network, shown in Fig. 175. Note that for large networks, executing computations with exponential scaling functions, there is enormous sensitivity to variations in tax rates, yielding very high leverage in computational return by tax rates.

This implies that as the quantum network expands over time, its joint processing power decreases exponentially with the rate of taxation, yielding an ever-decreasing performance multiplier. In Sec. LXXXIV we discuss some of the implications of this uniquely quantum phenomena.

However, taxation could also be negative, in the form of subsidisation. In Fig. 175 we focus on the region surrounding neutral taxation, showing small degrees of taxation and subsidisation on either side. Evidently, even small degrees of subsidisation have a very strong impact on the performance multiplier (more pronounced than the same rate of positive taxation!). This makes subsidisation of qubit expansion highly tempting.

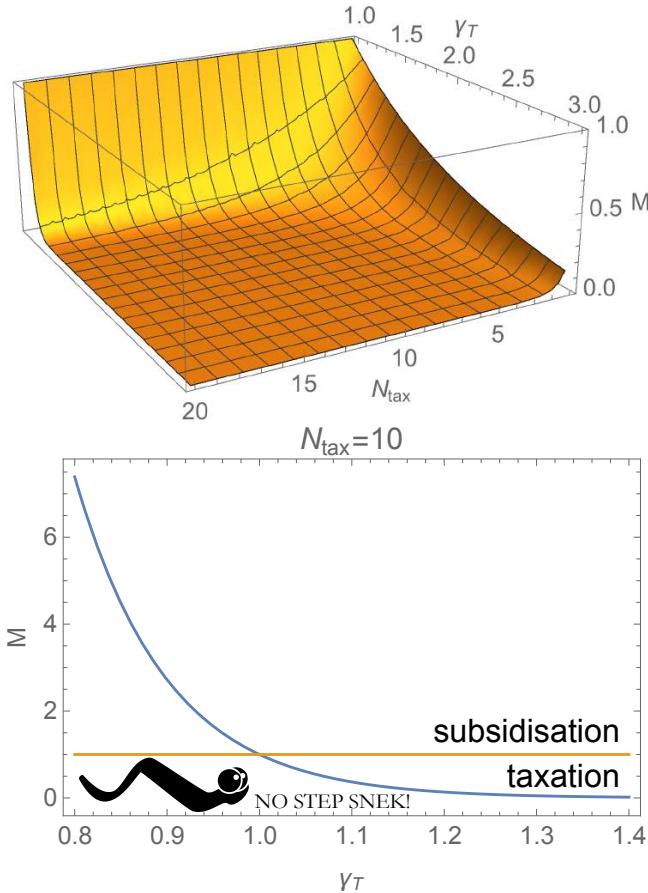


Figure 175: (top) Relationship between the tax performance multiplier, (positive) tax rate, and network size, assuming an exponential computational scaling function, in the regime of positive taxation, $\gamma_T > 1$. (bottom) For $N_{\text{tax}} = 10$, a zoom into the region around neutral taxation, where $\gamma_T \approx 1$, showing slight degrees of both taxation ($\gamma_T > 1$) and subsidisation ($\gamma_T < 1$). Neutral taxation, $\gamma_T = 1$, is shown in orange.

2. Policy perspective

— by Scott Harrison.

Irrespective of the magnitude of change, the elasticities, especially on the demand side, indicate that the qubit would be ripe for the application of a consumer-driven tax. Graphically the imposition of a tax within a perfectly competitive market would take on the form of Fig. 176(a).

The graph indicates a downward sloping demand line, indicating that the lower the price, the higher the demand. The upward sloping supply curve reflects financial incentive, the higher the price, the greater the incentive there is for firms to increase the quantity of qubits available to the market. Correlating this to elasticities, a steeper demand or supply curve indicates a higher degree of inelasticity. As such, in Fig. 176(a), the slope of both the demand and supply curves are relatively inelastic (compared to a 45° reference line). From this, the imposition of a consumer-based tax can be shown through the verti-

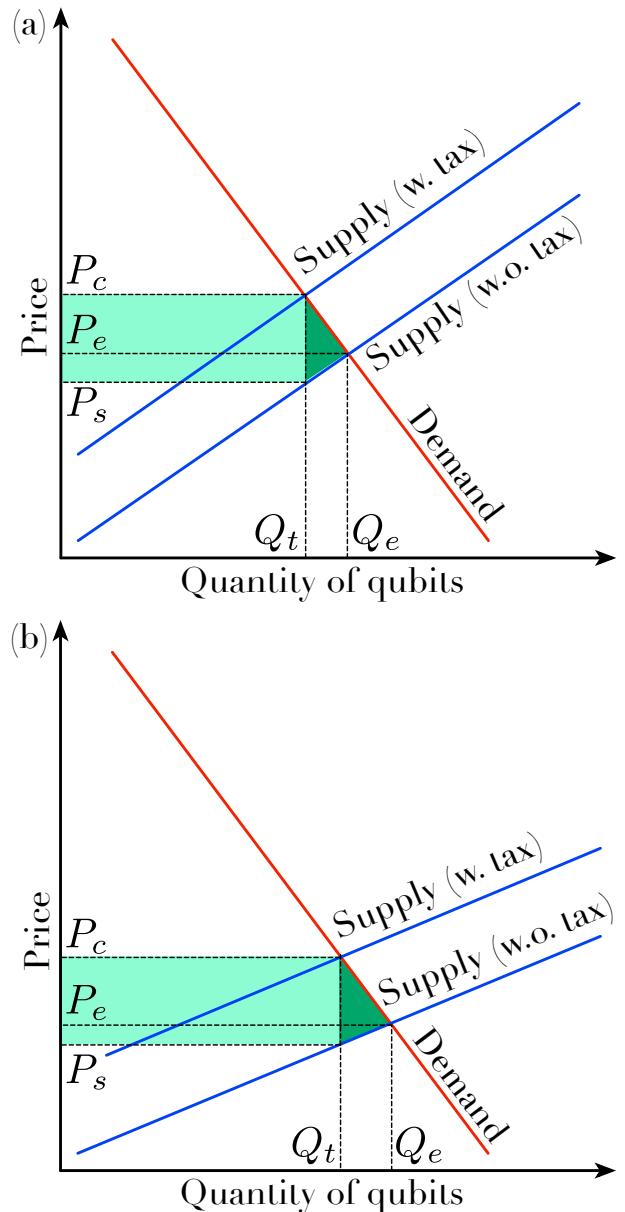


Figure 176: Hypothetical supply/demand curves, showing the impact of taxation on price and supply, for both inelastic (a) and elastic (b) market dynamics. Q_e and P_e are the efficient market quantity and price of the asset respectively. P_s is the price faced by suppliers, while P_c is the consumer price under taxation. Q_t is the quantity in a taxed environment. The net tax revenue collected is shaded in light teal, and the loss of market efficiency through the imposition of a tax is shaded in dark teal.

cal shift of the supply curve, with the magnitude of the shift, $P_c - P_s$, indicating the per-qubit value of the tax. Other important observations from the graph are the tax revenue collected (shaded in light teal), the loss of market efficiency through the imposition of a tax (shaded in dark teal), and a relatively small reduction in the quantity of qubits offered on the market from the efficient quantity, Q_e , to the quantity with the consumer tax, Q_t .

An important implication of the imposition of the tax is the share of the taxation burden. The change in price from the efficient price, P_e , to the new market price faced by consumers, P_c , is somewhat equivalent to the shift from P_e to the price point that the suppliers of qubits will receive, P_s . This means that the tax burden is likely to be equally shared between producers and consumers, which in the long term could act as a disincentive for increased production.

Fig. 176(b), however, shows a longer-term view of the qubit market, where the supply of qubits has become more elastic in nature. That is, the quantity supplied to the market becomes more sensitive to price fluctuations. This change in elasticity results in a shift in the tax burden, with a greater proportion of the tax now being paid by consumers, with only a small shift from the efficient price, P_e , to the price suppliers will face, P_s .

From a policy implications perspective, this means that governments wanting to cash in on the new technology need to be cautious with the imposition of taxes relative to market maturity. Imposing a significant tax early on may only act as a disincentive to the development of the industry. However, once the market matures further, the imposition of a qubit tax would make strong economic sense, as there will be minimal loss of market efficiency. Importantly, such a tax on computational power alone could serve to be a relatively stable revenue generation tool for governments.

C. The quantum stock market

In light of the distinction between subjective and objective value of computation, the question is how to reconcile this distinction in value, given the diversity of applications in the quantum marketplace. This will supersede the naïve models for cost of computation presented earlier, which were based entirely on objective value. Of course, subjective value is what people are actually willing to pay for in the real-world!

This will give rise to a marketplace for tradable units of quantum computation, where the underlying asset is time-shares in the global network. We refer to this as the *quantum stock market* – a marketplace subject to ordinary supply and demand, economic, and of course psychological pressures. In a scenario where a large number of users are executing computations with high return (think the R&D lab), asset values will be traded up. Contrarily, in a scenario of low-return computations (think our poor undergrad), they will be traded down. These market forces will be highly time-dependent, varying against many other factors in the economy, such as the emergence of new applications for quantum computation – the discovery of an important new algorithm could spontaneously distort the market leading to major corrections.

The relative market value of computation will subse-

quently drive the direction of investment into quantum hardware, with carry-over effects on future market prices. If investment stagnates, so too will growth in computational dividends, driving up market rates by limiting supply (assuming positive growth in demand). This will, after market adjustment, drive investment back into the system to satisfy increasing demand. Thus, despite the present uncertainty into the future dynamics of the quantum stock market, we expect this positive feedback loop to ensure consistent, ongoing investment into the quantum network, and at least some marginal degree of price stability.

What is likely to arise is that most owners of quantum hardware will not be consumers, but rather investors, potentially highly speculative ones, who float their resources on the quantum stock market, betting on changes in demand for computation and their associated subjective cost. This trading could involve transactions in the direct underlying asset, future contracts (Sec. LIX) for locking in required computational power at future points in time, or more complex derivatives. For example, an investor anticipating a surge in high-value computations is likely to invest more heavily into hardware with the expectation of an uptrend in market rates of their licensing. And their return on investment will reflect these market dynamics.

As all markets for tradable assets do, sophisticated derivative markets will inevitably emerge, whereby people can speculate on or hedge against market dynamics, taking long, short, or more complex market positions, potentially in a highly-leveraged manner. As discussed in Sec. LIX, derivatives such as future contracts can be extremely helpful in enabling consumers to lock in future prices, creating a stable and predictable business climate. Similarly, other derivatives will enable market participants to hedge other quantum-related investments. For example, suppose an investor held a stake in an R&D lab, highly reliant on quantum computing resources. By taking a leveraged long position on the market value for computation, he may limit losses on his R&D investment associated with the higher price (and hence lower profit) they will be paying for computation. No doubt, market manipulation and all the usual nonsense and shenanigans will ensue.

D. Geographic localisation

Because of the resource overheads associated with performing computations in a distributed manner, e.g via the resource costs associated with long-range repeater networks, there is an economic imperative to localise quantum infrastructure, so as to mitigate this – there is a clear economic benefit associated with housing qubits in close geographic proximity such that no long-distance quantum channels are required.

However, it's undesirable to *entirely* centralise infras-

tructure of *any* type, for two primary reasons:

- Geostrategic competition: competing nation states or enterprises may not want essential infrastructure to be located entirely offshore, placing them at the mercy of their strategic competitors.
- Geographical redundancy: to eliminate single points of failure (SPOF), which undermine network robustness, it's desirable for infrastructure to be geographically decentralised. In present-day large-scale distributed classical platforms, geographical redundancy is a key consideration. Even though it would be most efficient if all data were completely centralised, obviating communications overheads, it would be catastrophic if a single earthquake (or war!) could decimate the entire system. For this reason, it is desirable to distribute failure modes.

Thus, we can reasonably anticipate that the quantum internet will not evolve like the classical ‘internet of things’ (IoT), whereby a massive number of ultra-small computational resources are scattered across the globe and networked. Rather, a relatively small number of central ‘hubs’ are likely to emerge, which centralise enormous computational power, interconnected via the quantum internet to form the globally unified quantum cloud (Sec. XLIII.F).

Much like the classical internet, it's to be expected the network that will emerge will exhibit a very hierarchical structure, following a Pareto distribution in hub-size.

LXIV. GAME THEORY OF THE QUBIT MARKETPLACE

— by Scott Harrison & Peter Rohde.

“People are always selling the idea that people with mental illness are suffering. I think madness can be an escape. If things are not so good, you maybe want to imagine something better.” — John Nash.

EARLIER in this part, we established how a qubit market can function, and how the pricing mechanisms of various derivatives may work. One of the more interesting dimensions to the development of the quantum internet, is understanding *how* the cooperation between different suppliers will occur. Importantly, the expected high cost of quantum hardware means that there may be a limited number of competing vendors. For profit maximisation to occur, the most likely outcome is cooperation between them. The main question is, what can be learnt from applying game theoretic techniques to the strategies available to quantum computing vendors? And importantly, what are the implications associated with supply-side shifts, such as the imposition of taxation.

To analyse the decision making options for qubit suppliers, game theory is an analytical tool to understand ‘games’ between players, where the outcome of the game

is dependent on the various strategies employed by the players. The most well-known of these games is the prisoner’s dilemma (?), describing the potential risks and rewards for two prisoners who are being independently questioned about a crime.

Games can be classified based on their dimension, including the number of agents, the symmetry of the utility payoff, and whether they are cooperative. The Prisoner’s dilemma is an example of a *two-person, non-zero sum, non-cooperative game* (?). More detailed examples of game theory have explored many of the base assumptions of this scenario, such as what if the prisoners are able to cooperate from the outset? How does this then result in maximising utility, and is cooperation always the best answer?

In the case of the quantum internet, it should be clear from the outset that there is a strong benefit associated with cooperation between vendors. Cooperative games form an important subset of the game theory domain, and are the most applicable to quantum computing, where ‘cooperation’ translates to the unification of quantum computing resources into a larger distributed virtual quantum computer. As indicated previously, there will be exponential enhancements in computing power associated with unification, and as such, any qubit supplier will ultimately be able to produce excess computational power through networking and cooperating with others to exploit the computational leverage phenomenon (Sec. LVII). This idea is at the centre of the analysis when applying game theory to the decision making of suppliers.

A. Key concepts

For the uninitiated to economic analysis, particularly game theory, a few key concepts need to be established. This chapter by no means tries to cover these concepts in complete detail. For more detailed information we suggest referring to (???). Furthermore, the analysis at this point is only descriptive in nature, as a means of establishing the space where new research can be developed. Further more rigorous investigation is encouraged. The essential concepts that we rely on taken from game theory are:

- *Utility*: this can be generally defined as ‘the ability to assign a number (utility) to each alternative so that, for any two alternatives, one is preferred to the other if and only if the utility of the first is greater than the utility of the second’ (?). In this regard, utility is often seen as a representation of the overall benefit associated with a decision or preference. As utility is unobservable and may be defined essentially arbitrarily, it can become subjective in nature. However, the key is not whether the values assigned to any one preference are subjective, but rather whether the assigned values associated

with competing decisions are comparable, so that they can be quantitatively ranked. Thus, utility is a tool for the comparison of benefits associated with decision outcomes.

- *Utility payoff matrix:* Utility values assigned to any possible decision can be formulated into a matrix, which collates all the possible decisions from a given decision-maker's perspective. In a game with two participants, this would result in a 2-dimensional matrix. With n decision makers, there would be an n -dimensional matrix representation. The matrix also forms the basis for the graphical analysis undertaken throughout this section for the 2-player scenarios.
- *Negotiation set:* this defines a space of possible preferences. The negotiation set was first introduced in the seminal work of (?). At its most basic, the negotiation set forms a set of bounds for the payoff matrix. This limits possible solutions for a game to strategies that would actually see an improvement in the individual's utility above the base alternative of making no decision at all. This is also referred to as the *status quo*.

B. Strategies

To best develop an understanding of how a quantum internet game will be played out, we will begin by analysing the utility payoff space for classical computing. Currently we can easily define three key strategies for two market participants (X and Y), who both act as both suppliers and consumers of computational resources. The three strategies we compare are:

- **ISOLATION:** X and Y build their own systems in isolation, which they utilise independently. There is no cooperation or networking of computational resources between them. This can be considered the status quo (S) for the players, as it represents the autarky position and defines the von Neuman & Morgenstern utility space. This is represented in Fig. 177. Importantly, any changes to X 's computational capacity has no impact on Y 's. So X can take any position along the horizontal axis, and there is no change in Y 's utility, and vice versa. The resulting *utility point* **What's the utility point?** is described here as the intersecting minimum of these options, represented by the point (S_x, S_y) .
- **LICENSE:** Either X or Y build their own systems, but then licence unused compute-cycles to the other. This means that the other player may still have access to their required net computational resources, but essentially outsources the setup costs and ongoing infrastructure maintenance. This improves the

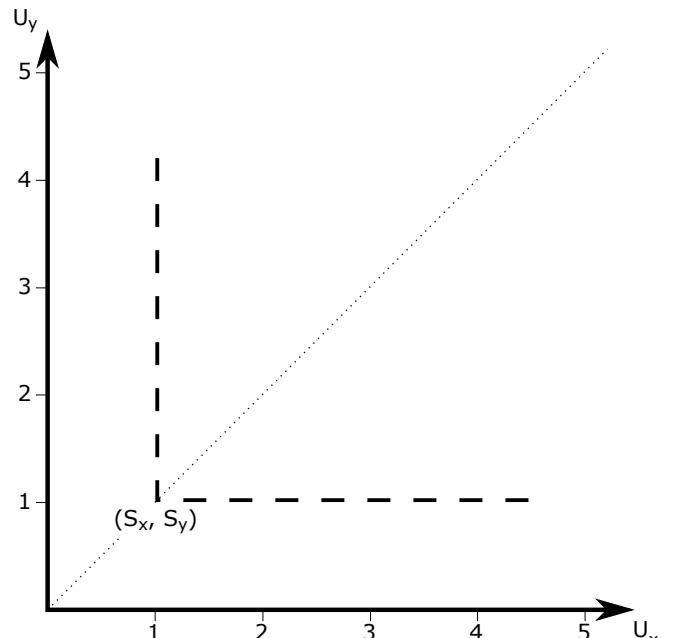


Figure 177: Utility space for two players with access to classical computing. The bold dashed lines show the minimum utility payoff for X and Y .

utilisation of the system for the player who licenses out, resulting in increased efficiency, profitability and subsequently utility (under the conditions of increasing economies of scale and assuming homogeneous system requirements).

- **UNIFY:** The two players consolidate their computational resources into a distributed cloud computing environment (Sec. XLIII.B), where the limitations of a single system are lifted, allowing for even better resource-sharing and improvement beyond what a single system solution can provide. In this scenario, both providers will be able to collaborate such that they can meet their individual computational needs, without having to fully build independent systems as before. Importantly though, while there may be a small loss in utility for one of the parties compared to the LICENSE strategy, unification allows an overall higher level of joint utility to be achieved, creating a Nash equilibrium at this point, C .

C. Utility payoff behaviour

In Tab. VII we present an example utility payoff matrix (numbers chosen arbitrarily) between two players engaging in the above three strategies with classical computing resources.

It's important to note that these values are totally arbitrary in nature, and do not have a 'real-life' interpretation, other than understanding the preferencing of the described strategies. Translating the utility payoff matrix

Player	<i>X</i>			
	Strategy	$X \rightarrow Y$	$X + Y$	$X \leftarrow Y$
<i>Y</i>	$X \rightarrow Y$	(3, 1.5)	(1, 1)	(1, 1)
	$X + Y$	(1, 1)	(2.5, 2.5)	(1, 1)
	$X \leftarrow Y$	(1, 1)	(1, 1)	(1.5, 3)

Table VII: Example of a payoff matrix for two classical computing vendors. ‘ \leftarrow / \rightarrow ’ indicates the LICENSE (from/to) strategy, ‘+’ indicates the UNIFY strategy, and the off-diagonal combinations are the status quo ISOLATION strategy. (X, Y) denotes the utility to players X and Y respectively. Note that there is some loss in net utility using LICENSE, owing to inefficiency through incurred transaction cost overheads.

into a graphical representation yields Fig. 178.

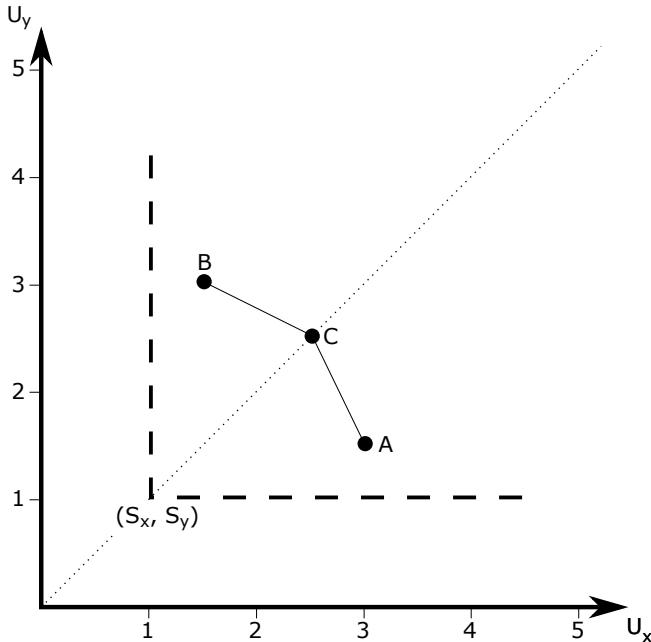


Figure 178: Utility payoff between two players in the classical computing environment. A indicates the utility combination where X builds the system and licenses out time to Y , and B represents the reverse. C is the point where both X and Y cooperate through distributed cloud computing, maximising the overall utility for both players. A and B derive identical utility values from the strategies, yielding symmetry about the diagonal axis. The lines AC and BC indicate possible solutions where mixed strategies are employed, combining components of LICENSE and UNIFY. C is also the Nash equilibrium, where cooperative bargaining would result in the best outcome overall.

Now, quantum computing uses similar strategies for possible solutions, with one key difference. Using classical computing, the relationship for the UNIFY strategy was described as additive in nature, where the computational resources of both players are accumulated additively when unified into a larger virtual computer. In the corresponding quantum UNIFY strategy, this effect is enhanced by their super-linear computational leverage. For the sake

of illustration, we will now assume this becomes multiplicative. Multiplicativity in computational power approximates behaviour under a UNIFY strategy when dealing with exponential scaling functions.

This is easily intuitively seen as follows. Let the computational scaling function be an arbitrary exponential,

$$f_{sc}(n) = O(\exp(n)). \quad (64.1)$$

Then it immediately follows that the scaling function obeys the identity,

$$f_{sc} \left(\sum_i n_i \right) = \prod_i f_{sc}(n_i), \quad (64.2)$$

yielding the multiplicative behaviour, shown diagrammatically in the context of a distributed quantum computation in Fig. 179. Note that in the classical case the product in Eq. (64.2) would become a sum, which is exponentially smaller in generally,

$$\prod_i f_{sc}(n_i) \geq \sum_i f_{sc}(n_i). \quad (64.3)$$

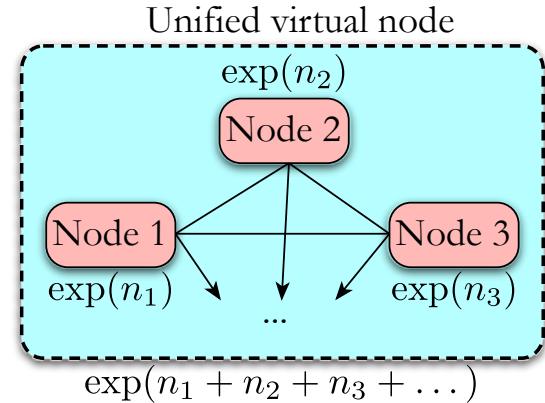


Figure 179: A distributed quantum computation across a number of nodes, each with n_i qubits. The computational scaling function is chosen to be exponential in form, yielding a classical-equivalent computational power of $\exp(n_i)$ for each node. However, the joint computational power of the network is given by $\exp(n_1 + n_2 + \dots)$, which is exponentially greater than the sum of the individual computational powers, $\exp(n_1) + \exp(n_2) + \dots$, in general.

As such, should X and Y build identical quantum computers, the hypothesised payoff matrix would become as shown in Tab. VIII.

In this scenario, the ISOLATION and LICENSE strategies are assumed to yield the same utility payoffs as in the classical case. The only difference arises when X and Y UNIFY. The effect of quantum computational enhancement is to therefore amplify the cooperative elements in the utility payoff matrix, potentially by very large factors. This has the generic effect that in the quantum realm, co-operation is more highly incentivised than in the classical one. The resulting graphical representation of this payoff matrix is shown in Fig. 180.

Player		<i>X</i>		
	Strategy	$X \rightarrow Y$	$X + Y$	$X \leftarrow Y$
<i>Y</i>	$X \rightarrow Y$	(3, 1.5)	(1, 1)	(1, 1)
	$X + Y$	(1, 1)	(5, 5)	(1, 1)
	$X \leftarrow Y$	(1, 1)	(1, 1)	(1.5, 3)

Table VIII: Example utility payoff matrix for two players with quantum computing resources. Note the enhancement in the diagonal $X + Y$ matrix element, compared to the classical case.

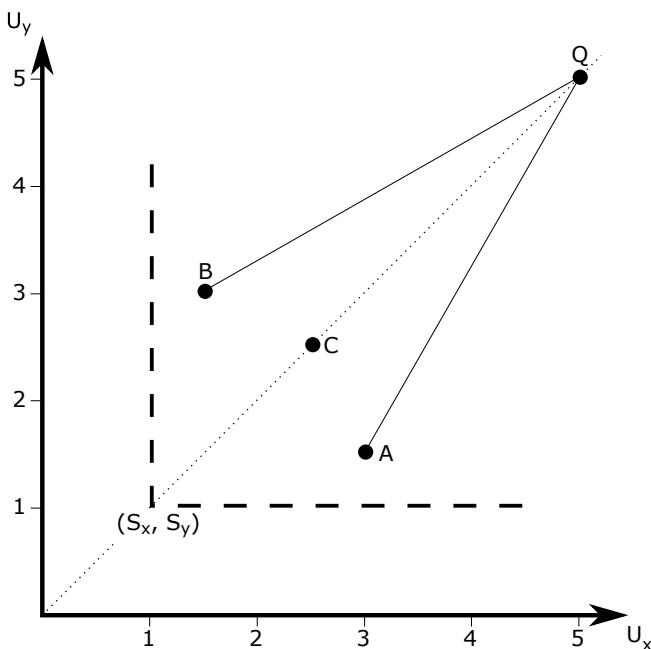


Figure 180: Tradeoff in utility between two players in the scenario of quantum computing. Q now represents the UNIFY strategy, where there is a quantum enhancement in the utility payoff for X and Y . Note that Q offers strictly greater utility than C , the corresponding classical point. Any mixed strategy combining LICENSE and UNIFY will result in a non-linear transition between points A/B and Q .

D. Cooperative payoff enhancement

An individual user, i , of a quantum computer operating on their own, observes computational power characterised by the appropriate computational scaling function acting on the number qubits in their possession, $f_{sc}^{(i)}(n)$. This stipulates the utility payoff for that individual, allowing for trivial construction (and efficient mathematical representation) of multi-player payoff matrices simply by characterising single-player payoffs independently.

Once cooperative strategies are introduced, the associated cooperative payoff matrix elements are transformed appropriately, according to the reallocation of resources and how they are collectively utilised.

We refer to this phenomena as *cooperative payoff enhancement*, one which, depending on what cooperative techniques are employed, drastically alters the economic landscape, and its associated game-theoretic analysis and outcomes.

We will present an elementary analysis of this concept for the three different cooperation strategies introduced earlier, initially in the standard 2-player context, subsequently generalised to an arbitrary multi-player environment.

1. ISOLATION strategies

When implementing quantum computations characterised by completely general computational scaling functions, $f_{sc}^{(X,Y)}$ (which in general can be distinct for the two players, X and Y), when using an ISOLATION strategy, the respective payoff matrix elements are simply given by,

$$\text{ISOLATION} = \begin{pmatrix} f_{sc}^{(X)}(n_X) \\ f_{sc}^{(Y)}(n_Y) \end{pmatrix}, \quad (64.4)$$

where payoff matrix elements are assumed to be in units of classical-equivalent processing time⁷⁷ (i.e FLOPs), and resource-sharing is based on the methodology for arbitrage-free time-share allocation presented in Sec. LV.

2. LICENSE strategies

Elements associated with LICENSE strategies are transformed as,

⁷⁷ There's nothing unique or special about using classical-equivalent computational power as our utility measure. Any other measure of 'payoff' could be equally well justified, depending on circumstance.

For example, one could instead represent utility in terms of the monetary value of computational power. In that case, we simply need to transform the payoff matrix elements using the cost of computation identity presented in Post. 8.

$$\begin{pmatrix} f_{sc}^{(X)}(n_X) \\ f_{sc}^{(Y)}(n_Y) \end{pmatrix} \xrightarrow{\text{LICENSE}} \begin{pmatrix} r_{X \rightarrow X} \cdot f_{sc}^{(X)}(n_X) + r_{Y \rightarrow X} \cdot f_{sc}^{(X)}(n_Y) \\ r_{Y \rightarrow Y} \cdot f_{sc}^{(Y)}(n_Y) + r_{X \rightarrow Y} \cdot f_{sc}^{(Y)}(n_X) \end{pmatrix}, \quad (64.5)$$

where $0 \leq r_{i \rightarrow j} \leq 1$ denotes the proportion of i 's compute-time licensed to j .

This is easily logically generalised to an arbitrary multi-player setting, in which case the transformation becomes,

$$f_{sc}^{(i)}(n_i) \xrightarrow{\text{LICENSE}} \sum_{j=1}^N r_{j \rightarrow i} f_{sc}^{(i)}(n_j), \quad (64.6)$$

where there are N players, all engaging with one another using licensing only. The $r_{i \rightarrow j}$ parameters are normalised for all users such that,

$$\sum_{j=1}^N r_{i \rightarrow j} \leq 1 \forall i. \quad (64.7)$$

With equality, this normalisation implies perfect licensing efficiency (i.e no overheads) and no wasted clock-cycles (full utilisation). Inequality implies either inefficiency or under-utilisation. Since under this strategy net computational power is conserved (at best), it might appear mindless to employ it at all, given that there is no net gain. Whilst this is true, there may be ulterior motives for employing it. For example, it might be employed for the purposes of load balancing across a distributed architecture, or implementing arbitrage between inconsistent market pricing of computational power between nodes.

Note that when $r_{i \rightarrow j} = \delta_{i,j}$ (i.e $r = I_N$ is the $N \times N$ identity matrix, and there is no inter-player licensing) the LICENSE strategy simply reduces back to the ISOLATION strategy.

3. UNIFY strategies

Elements associated with UNIFY strategies will undergo the quantum utility payoff enhancement,

$$\begin{pmatrix} f_{sc}^{(X)}(n_X) \\ f_{sc}^{(Y)}(n_Y) \end{pmatrix} \xrightarrow{\text{UNIFY}} \begin{pmatrix} n_X \cdot \chi_{sc}^{(X)}(n_X + n_Y) \\ n_Y \cdot \chi_{sc}^{(Y)}(n_X + n_Y) \end{pmatrix}, \quad (64.8)$$

from the definition for time-shared compute power given in Def. 15.

As before, we can logically generalise the payoff enhancement of a generalised UNIFY strategy to the multi-player scenario as,

$$f_{sc}^{(i)}(n_i) \xrightarrow{\text{UNIFY}} n_i \cdot \chi_{sc}^{(i)}(n_{\text{global}}). \quad (64.9)$$

Thus, it is evident that the enhancement in UNIFY strategies is highly dependent on:

- The total number of qubits held between the players,

$$n_{\text{global}} = \sum_{j=1}^N n_j. \quad (64.10)$$

- The proportion of the qubits held by each player,

$$r_i = \frac{n_i}{\sum_{j=1}^N n_j}. \quad (64.11)$$

- The respective algorithms to which the computational resources are being applied by each player, which influence the player-specific subjective scaling functions, $f_{sc}^{(i)}$, independently.

The final point is particularly noteworthy, since it implies that optimal game-theoretic outcomes are not objective, but subjective, and highly dependent on how players are employing their computational resources, which may be highly distinct and change dynamically over time. If one player is employing an exponential scaling function, whereas the other is only employing a polynomial one, this could completely distort the utility payoff dynamics of the game in favour of the player who would otherwise have been weaker under symmetric scaling functions. This in turn could completely alter the landscape of how users choose strategies to play optimally.

4. Strategic implications

“In savage countries they eat one another, in civilised ones they deceive one another; and that is what people call the way of the world!” — Arthur Schopenhauer.

It is clear that the UNIFY strategy, in which distinct quantum computing nodes are merged via the network into a larger distributed quantum computer, works to the (potentially exponential) benefit of all contributing parties. This distorts game-theoretic analysis of network participants compared to classical computing.

On one hand, the guaranteed mutual benefit of all players directly enhances their individual compute power. In a compute-centric world, where computation equates to productivity, this directly works in the self-interest of all.

However, taking a more strategic long-term perspective, despite self-enhancement, the associated enhancement of competitors may eventuate in outcomes that work against

self-interest to a sufficient extent that it outweighs this benefit. Thus, cooperative enhancement, in an appropriate strategic context, could equally be tantamount to ‘adversarial enhancement’, and be considered an overwhelming motivator to avoid cooperation with certain players.

Some non-technical discussion on the implications of these observations is presented in several of the essays in Part. XIV.

5. Inefficient markets

The utility payoff enhancement characterised by these transformations is based on the simplest of toy models, where unification is assumed to be perfectly efficient – there are no overheads (e.g transaction or communication costs) associated with cooperation, nor are there any externalities, such as taxes or regulations. In reality, these assumptions are of course completely unrealistic. Thankfully, such secondary effects can be relatively easily incorporated into the model by modulating them with additional layers of transformations capturing these features.

For example, consider the unification of computational resources between two players residing in different jurisdictions, which levy import/export tariffs against one another, an externality introducing inefficiency into cooperative strategies. When expressed in terms of the monetary cost of computation, this would effectively modulate the payoffs of UNIFY matrix elements by a tariff-dependent function.

E. Mixed strategies

How do mixed strategies affect cooperation? To perform this analysis, assume that X has chosen a mixed strategy comprising a combination of LICENSE and UNIFY. In the example shown in Fig. 181, this results in a utility value of 4 for X , represented by the vertical dotted line.

This results in two possible solutions, X' and X'' , that intersect the lines AQ and BQ respectively. X' represents a point where more compute-time is allocated to the LICENSE strategy, and less to UNIFY, while X'' is the converse. X is indifferent to both solutions, as they result in the same utility, but for Y , there is a clear choice. X'' will result in higher utility, placing an emphasis on cooperation. Once Y has chosen their strategy, X can further revise their original choice too, optimising their utility without undermining Y 's. This results in adaptive strategy revision, that asymptotes towards the optimal solution Q , as shown in Fig. 182.

The conclusion from this is that while mixed strategies comprising elements of LICENSE and UNIFY may be advantageous, there will always be an underlying pressure towards cooperation.

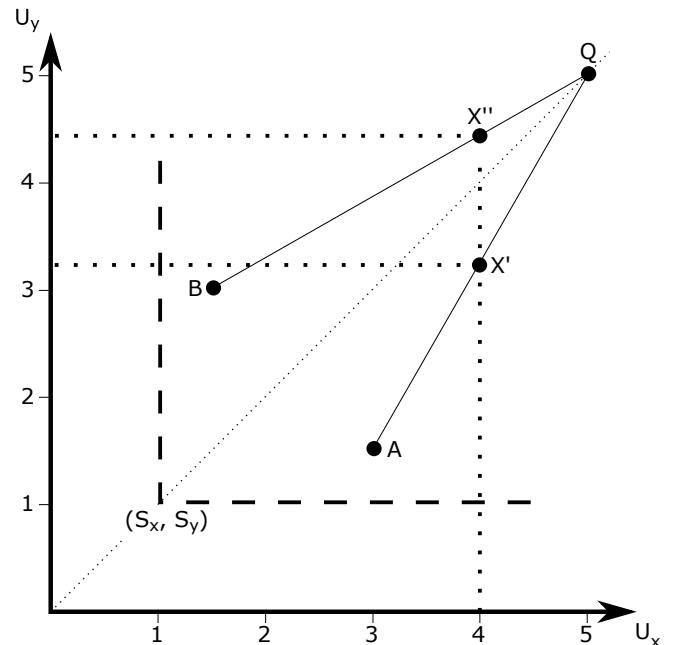


Figure 181: Utility payoff for two players with quantum computers, employing mixed strategies. X' and X'' denote two unique solutions for Y having a utility of 4. From Y 's perspective it makes no difference which strategy is chosen. However, X can locally optimise their utility by choosing X'' over X' .

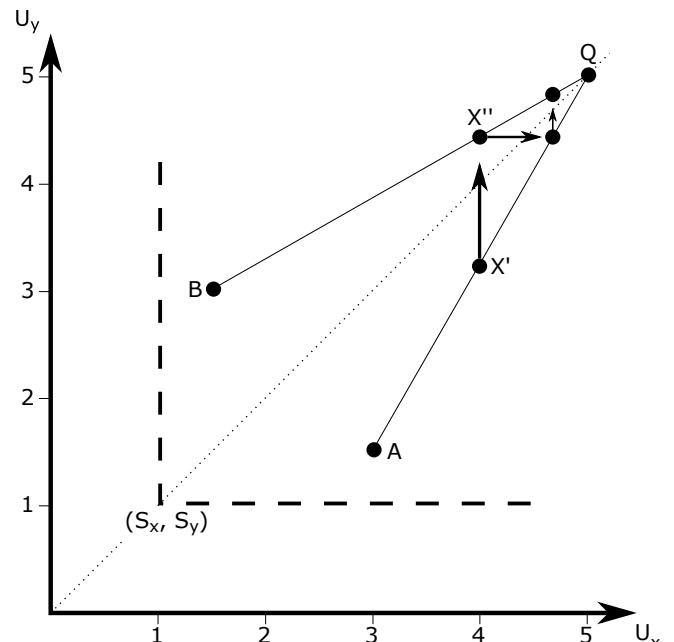


Figure 182: Adaptive strategy revision by X and Y , yielding incremental local utility improvements without affecting the other, progressively marches us towards the optimal point Q asymptotically.

F. Taxation

What happens to the utility payoff matrix when taxation is imposed? The assumption that X and Y are operating in similar regulatory environments can be relaxed. What happens when X is in a more heavily taxed environment than Y , or something impacts the utility achieved by the players in an asymmetric manner?

Economically, taxation operates by transferring utility from the supplier of a good to the government. This will result in some reduction in supply. However, in the UNIFY strategy, even small reductions in supply may be exponentially multiplied. The result is that all collaborative strategies will yield less utility for X , as it both reduces supply and transfers utility to government. The impact however is also felt by Y , as the overall joint computational power of the cloud is reduced. This results in a payoff matrix as shown in Tab. IX.

Player	X			
	Strategy	$X \rightarrow Y$	$X + Y$	$X \leftarrow Y$
Y	$X \rightarrow Y$	(2, 1.25)	(1, 1)	(1, 1)
	$X + Y$	(1, 1)	(3, 4.5)	(1, 1)
	$X \leftarrow Y$	(1, 1)	(1, 1)	(1.5, 3)

Table IX: Example utility payoff matrix in the presence of taxation on quantum computers. The effect is a net depreciation in achievable utility.

When X licenses compute-time to Y , there is a reduction in utility derived by X . The small reduction in supply will also mean that there is a reduction in the available compute-time available to be licensed to Y . For the LICENSE strategy, this means there would be a shift in utility from (3, 1.5) in Tab. VIII to (2, 1.25) in Tab. IX. Graphically, this is shown as the inward shift from A to A' in Fig. 181. We also assume here that the imposition of the tax itself doesn't completely abolish any utility gains from the LICENSE strategy over the status quo.

For the UNIFY strategy, there will be a depreciation in the final utility payoff. As such, the utility payoff shows a new combination of (3, 4.5), an asymmetric reduction from the previous position of (5, 5). Graphically, this is shown in Fig. 183 as a shift from Q to Q' .

Two things are clear from both the matrix and its resulting graphical representation. The imposition of taxation affects both the LICENSE and UNIFY strategies, but unequally. For X , the imposition of taxation results in a greater loss in net utility with the UNIFY strategy. This would ultimately undermine the likelihood for X to choose this strategy. Secondly, the rate at which utility is gained through a mixed strategy (moving from A' to Q') is decreased. This implies that, with a UNIFY strategy, X would be less motivated to cooperate with Y than Y would be to cooperate with X . Q' would still lead to the maximum utility payoff, but the negotiation process

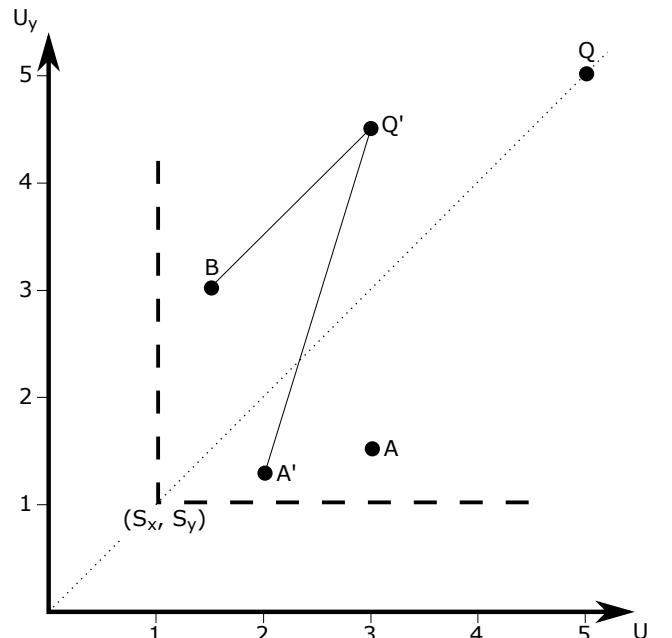


Figure 183: Utility payoff for two players with quantum computers, employing a UNIFY strategy, under the imposition of taxation.

to achieve it will be more difficult owing to the utility asymmetry. Thus, despite both players having identical systems, the regulatory asymmetry will strongly impact the likelihood of cooperation. This has important implications for future quantum policy-making.

G. Resource asymmetry

What happens with collaboration between players who have different systems? If this assumption is relaxed, what comes of the resulting utility payoff matrix and choice of strategies? The analysis for this assumption had already been presented in the previous discussion. The effective impact of taxation is that it will limit the amount of computational power delivered to market and capture governmental revenue from the sale of compute-time. This is analogous to what happens if X was to build a less powerful system than Y . Through collaboration, they could still attain point Q' , shown in Fig. 181, but this time the asymmetry arises as a design consequence rather than a regulatory imposition. Despite X now not contributing as much to the overall computational power of the quantum cloud, choosing the maximising strategy of offering all available computing power to the cloud does maximise individual utility. Furthermore, the subjective nature of the utility derivation means there doesn't have to be a change from the described payoff matrix shown in Fig. 180. Despite having a less powerful system, the proportional allocation of financial rewards relative to the contribution, means both X and Y could maintain

a symmetric utility payoff outcome, like in Fig. 180. In summary, the 2-player game shows that cooperation will result in the better outcome for both players.

— *Ceretis paribus.*

H. Multi-player games

What happens when the analysis moves beyond just two players? As established previously, the cost and scale limitations associated with quantum computing means there are likely to be limited vendors contributing to the quantum internet. However, it's also likely there will be more than just two. Formulating the n -person cooperative game opens up a plethora of possibilities that go beyond the scope of this introductory discussion, but one key takeaway point from the previous analysis is that in general there is a strong motivation for computational cooperation in the quantum world. This then introduces two possible scenarios:

- There is complete cooperation between all suppliers, i.e the global virtual quantum computer discussed in Sec. XLIII.F.
- Competing cartels develop, where for external reasons (e.g political, geographical, ideological, strategic), there is a benefit in cooperating with a limited number of players, and acting as a single supplier in direct competition with other cartels.

x The first scenario is by far the most attractive, and in a world where free negotiations may take place, this is clearly the option resulting in both attaining the greatest computational power for the quantum cloud, and also the greatest utility maximisation for all participants involved. The end result will be similar to that shown in Fig. 180. This implies that the quantum internet market structure would become like a natural monopoly, with diminishing long-run average costs as supply is increased. It would also mean that there is a strong case for some government involvement to prevent profit maximisation at the expense of efficiency.

However, history has shown us that another plausible outcome is the second scenario. While there are motivations to collaborate, there are also motivations to compete. Given the almost certain involvement of government intervention in the supply of quantum processing, the formulation of regional cartels due to external factors may also be a likely outcome. In such an environment, the cartels will internally operate as multi-player cooperative games, and externally towards other conglomerates, operate as separate multi-cartel competitive games (?). This result may still result in efficient Pareto optimal outcomes at a market level, but will always fall short relative to a model of complete cooperation. This scenario will naturally end with an oligopolistic market structure.

I. Conclusions

In summary, a game theory approach to understanding the quantum internet shows that there are strong motivations for quantum computing vendors to cooperate in order to globally maximise net utility. Furthermore, there is a strong potential to affect the possibility of cooperation through market distorting effects such as via the imposition of taxation or regulation. Finally, the two most likely market structures that will develop under the quantum internet are either a natural monopoly, where some form of regulation will be required to ensure economically efficient production, or an oligopoly, where a few cartels will compete with each other to maximise their productive output. Either way, the quantum marketplace is an extremely interesting and largely unexplored avenue for future research, a new interdisciplinary field sitting at the intersection between economics and quantum information theory. It is also one of great relevance and importance for when this technology becomes a reality.

— *Natura nihil frustra facit.*

LXV. SUMMARY OF ECONOMIC MODELS

In Tab. VII we summarise the economic models and parameters we developed, and applied them to several illustrative scaling functions of particular interest: linear (i.e classical computing), polynomial, and exponential (i.e best-case quantum computing).

— *Errare humanum est.*

Model	General form	$f_{\text{sc}}(n) = n$ (classical)	$f_{\text{sc}}(n) = n^p$ (intermediate)	$f_{\text{sc}}(n) = e^n$ (full quantum)
Per-qubit computational power (Sec. XLVII)	$\chi_{\text{sc}}(n) = \frac{f_{\text{sc}}(n)}{n}$	1	n^{p-1}	$\frac{e^n}{n}$
Network power (Sec. L)	$P(t) = f_{\text{sc}}(N_0 \gamma_N t)$	$N_0 \gamma_N t$	$(N_0 \gamma_N t)^p$	$e^{N_0 \gamma_N t}$
Network value (Sec. LI)	$V(t) = C_0 N_0 \left(\frac{\gamma_N}{\gamma_C}\right)^t$	$C_0 N_0 \left(\frac{\gamma_N}{\gamma_C}\right)^t$	$C_0 N_0 \left(\frac{\gamma_N}{\gamma_C}\right)^t$	$C_0 N_0 \left(\frac{\gamma_N}{\gamma_C}\right)^t$
Spot price of computation (Sec. LIV)	$L(0) = \frac{e^{\gamma_{\text{ror}}} C_0}{\chi_{\text{sc}}(N_0)}$	$e^{\gamma_{\text{ror}}} C_0$	$\frac{e^{\gamma_{\text{ror}}} C_0}{N_0^{p-1}}$	$\frac{e^{\gamma_{\text{ror}}} N_0 C_0}{e^{N_0}}$
Future cost of computation (Sec. LIV)	$L(t) = \frac{e^{\gamma_{\text{ror}}} C_0 \gamma_C^{-t}}{\chi_{\text{sc}}(N_0 \gamma_N t)}$	$e^{\gamma_{\text{ror}}} C_0 \gamma_C^{-t}$	$\frac{e^{\gamma_{\text{ror}}} C_0 \gamma_C^{-t}}{(N_0 \gamma_N t)^{p-1}}$	$\frac{e^{\gamma_{\text{ror}}} C_0 N_0 \left(\frac{\gamma_N}{\gamma_C}\right)^t}{e^{N_0 \gamma_N t}}$
Time-share computational power (Sec. LV); Cooperative payoff enhancement (Sec. LXIV.D)	$c_n = n \cdot \chi_{\text{sc}}(n_{\text{global}})$	n	$n \cdot n_{\text{global}}^{p-1}$	$\frac{n e^{n_{\text{global}}}}{n_{\text{global}}}$
Problem size scaling function (Sec. LVI)	$s = f_{\text{size}}^{-1}(n \cdot \chi_{\text{sc}}(n_{\text{global}}))$	$\frac{n}{\alpha_{\text{size}}}$	$(n \cdot n_{\text{global}}^{p_{\text{sc}}-1})^{\frac{1}{p_{\text{size}}}}$	$\alpha_{\text{sc}} n_{\text{global}} + \log(n) - \log(\alpha_{\text{sc}} n_{\text{global}})$
Quantum computational leverage (Sec. LVII)	$\lambda_n = \frac{\chi_{\text{sc}}(n_{\text{global}})}{\chi_{\text{sc}}(n)}$	1	$\left(\frac{n_{\text{global}}}{n}\right)^{p-1}$	$\frac{n e^{n_{\text{global}}}}{n_{\text{global}} e^n}$
Single-qubit leverage (Sec. LVII)	$\lambda_{\text{qubit}} = \frac{\chi_{\text{sc}}(n_{\text{global}})}{\chi_{\text{sc}}(1)}$	1	n_{global}^{p-1}	$\frac{e^{n_{\text{global}}-1}}{n_{\text{global}}}$
Time-dependent leverage (Sec. LVII)	$\lambda_n(t) = \frac{\chi_{\text{sc}}(N_0 \gamma_N t)}{\chi_{\text{sc}}(n)}$	1	$\left(\frac{N_0 \gamma_N t}{n}\right)^{p-1}$	$\frac{n e^{N_0 \gamma_N t - n}}{N_0 \gamma_N t}$
Static computational return (Sec. LVIII)	$r_{\text{static}}(t) = n \cdot \chi_{\text{sc}}(N_t)$	n	$n N_t^{p-1}$	$\frac{n e^{N_t}}{N_t}$
Forward contract price (Sec. LIX)	$F(T) = \frac{e^{\gamma_{\text{ror}} - r_{\text{rf}} T} C_0 \gamma_C^{-T}}{\chi_{\text{sc}}(N_0 \gamma_N T)}$	$e^{\gamma_{\text{ror}} - r_{\text{rf}} T} C_0 \gamma_C^{-T}$	$\frac{e^{\gamma_{\text{ror}} - r_{\text{rf}} T} C_0 \gamma_C^{-T}}{(N_0 \gamma_N T)^{p-1}}$	$\frac{e^{\gamma_{\text{ror}} - r_{\text{rf}} T} C_0 N_0 \left(\frac{\gamma_N}{\gamma_C}\right)^T}{e^{N_0 \gamma_N T}}$
Tax performance multiplier (Sec. LXIII.B)	$M(N_{\text{tax}}) = \frac{f_{\text{sc}}(N_{\text{tax}})}{f_{\text{sc}}(N_{\text{tax}} \gamma_T)}$	$\frac{1}{\gamma_T}$	$\frac{1}{\gamma_T^p}$	$e^{N_{\text{tax}}(1 - \gamma_T)}$
Political leverage (Sec. LX)	$\gamma_{A,B} = \frac{\chi_{\text{sc}}(n_B)}{\chi_{\text{sc}}(n_A)}$	1	$\left(\frac{n_B}{n_A}\right)^{p-1}$	$\frac{e^{n_B n_A}}{e^{n_A n_B}}$

Table VII: Summary of the dynamics of various economic models under several computational scaling functions (f_{sc}) of interest, where there are n qubits held by the respective node and $n_{\text{global}} = \sum_{j \in \text{nodes}} n_j$ qubits in the global network. N_0 is the initial number of qubits, undergoing growth rate γ_N . C_0 is the initial monetary cost per qubit, undergoing decay rate γ_C . γ_{ror} is the rate of return on the licensing of compute-time on qubits, and r_{rf} is the risk-free rate of return (typically the yield on US government bonds). γ_T is the rate of taxation, and N_{tax} is the number of qubits that would exist on the network in the absence of taxation.



Part XI

History & state-of-the-art

— by He-Liang Huang & Zu-en Su.

“We are not makers of history. We are made by history.”
— Martin Luther King, Jr.

WHEN will we have the quantum internet? This is a question with as many answers as there are people one asks. And the developments of the different components of such an internet will be staggered and under continual development – a quantum internet with the capacity for long-distance QKD will likely arrive far sooner than one supporting fully distributed, blind quantum computation!

In this section we present a brief history of recent major developments and the state-of-the-art of some of the most important quantum technologies and protocols relevant to the quantum internet, with a view to understand trends in the development of the field, so as to position us to gauge the rate of progress and make educated predictions about future developments – the timing of future developments is likely to have big strategic implications, and it's important we are able to anticipate technological transformations as radical as the quantum ones, so that we may be prepared for their fallout and consequences.

With this we aim to shed light on what quantum networking services could be readily implemented today, using present-day technology, and what is likely to be viable in the near and more distant future.

We will very succinctly summarise developments as timelines, without much elaboration on specific details. We provide detailed referencing such that the interested reader can follow up on specifics at their leisure. Effectively this section acts as a handbook for looking up important development events, milestones, and their references. We also provide links to many of the important review papers in the field, which are perhaps the best starting point for developing a deeper understanding as an outsider from the field.

The reader disinterested in a history lesson might comfortably skip this part, without serious risk of missing out on explanations of important topics (no new conceptual material is introduced in this part).

The tables summarising chronological timelines of major developments, and their referencing, are presented in bulk at the end of this part in Sec. LXXVI.

LXVI. QUANTUM TELEPORTATION & ENTANGLEMENT DISTRIBUTION

QUANTUM state teleportation (Sec. XXVII.C) has attracted broad experimental interest and been subject to widespread demonstration across many physical architectures, becoming one of the most well-investigated protocols in experimental quantum information science.

Tab. VIII summarises some of the notable recent developments in quantum teleportation and the entanglement distribution upon which it relies.

LXVII. ENTANGLEMENT SWAPPING & QUANTUM REPEATERS

CLOSELY related to entanglement distribution is entanglement swapping (Sec. XXVII.E), where the goal is to entangle two remote parties, each of whom have one half of two distinct Bell pairs. As discussed in Sec. XXVIII, entanglement swapping will be of fundamental importance in networks treating entanglement distribution as the most fundamental elementary resource, making it greatly applicable to quantum networking.

Tabs. IX & X summarise major recent developments in entanglement swapping and quantum repeater networks (in which entanglement swapping is also a basic primitive).

LXVIII. QUANTUM KEY DISTRIBUTION

WITH the development of quantum technology, QKD will gradually become increasingly practical, and economically accessible to consumers. Bennett (of BB84 protocol fame), first demonstrated a QKD protocol on an optical platform over a distance of just 30cm (?). Since then, experiments have developed rapidly from indoors to outdoors, over ever-increasing distances, and even recently between Earth-based ground stations and satellites in low-Earth orbit. Indeed, commercial QKD units are even available as off-the-shelf commodity products today, making QKD perhaps the first commercialised quantum technology.

Tab. XI summarises some major milestones in the development and deployment of QKD.

LXIX. ENTANGLEMENT PURIFICATION

ENTANGLEMENT purification has been very successful in optical systems. This forms an essential primitive in long-range quantum repeater networks, where entanglement must be recovered from states that have degraded during transmission over long distances.

Tab. XII summarises some of the major recent developments in entanglement purification.

LXX. STATE PREPARATION

A NY quantum protocol necessarily involves state preparation to create the inputs for the protocol. Different types of states come with their own unique challenges, and also their own unique utility.

In the context of the quantum internet, the most important class of quantum states is of course the optical ones, since light will almost inevitably be the primary information carrier in such a future network.

We now summarise some of the most common (and useful) optical states for quantum information processing applications.

A. Coherent states

Coherent states are well approximated by laser or maser light. Nowadays thousands of types of lasers are known with different power, temporal, spatial and spectral parameters, and the technology is already very mature and commercially available. Here we only introduce some of the basic concepts of lasers related to quantum networking.

A laser can be classified as operating in either a continuous or pulsed regime. Most laser diodes used in communication systems are continuous. But they can also be externally carved at some rate by modulators to create pulsed light. Usually, pulsed lasers are created by the technique of Q-switching or mode-locking.

Different applications require lasers with different output power. Typical output powers of single-mode laser diodes are some tens of milliwatts, up to at most a few hundred milliwatts. However, multiple transverse mode diode lasers can reach up to some tens of Watts, and can be used as pump sources for high-quality and high-power single-mode solid-state lasers. Such single-mode diode-pumped solid-state lasers can be further mode-locked to output femtosecond pulses, reaching as far as an average power of tens of Watts. Pulsed lasers can also be characterised by the peak power of each pulse. The peak power of a pulsed laser is many orders of magnitude greater than its average power.

Tab. XIII enumerates two types of lasers commonly used in quantum communication applications.

B. Single-photons

A highly attenuated laser can be used as a good approximation to a single-photon source when no more than one single-photon is used in an interferometric experiment, such as QKD. Otherwise, heralded or deterministic single-photon sources are required. This is because while the photo-statistics of a single-photon mimic those of a laser, the photo-statistics of multi-photon states do

not, irrespective of the laser's power – coherent light and multi-photon states obey fundamentally different rules of statistics.

Tab. XIV summarises some of the notable developments. A good review of solid-state single-photon emitters is (?). Different sources are compared in the review by (?).

C. Entangled states based on non-linear optics

Non-linear optics is one of the most convenient and cheapest methods to produce entangled states. Tab. XV summarises some of the major developments. Further material can be found in the review papers (??).

LXXI. NON-OPTICAL SYSTEMS

ALTHOUGH light is likely to be the information *carrier* of the future quantum internet, it may very well not win the race of becoming the preferred means for computation or storage. Here, there are countless other possibilities to consider, of which we now outline a few.

A. Atomic ensembles

A variety of techniques have been developed to create squeezing and entanglement in atomic ensembles. The main methods exploit atom-light interactions in cold gases, or interactions between particles such as atom-atom collisions in Bose-Einstein condensates, or combined electrostatic and ion-light interactions in ion chains. Atom-light interactions currently represent the most mature technology, and exhibit the highest squeezing of 20.1dB via an optical-cavity-based measurement (?).

Recent progress is summarised in Tab. XVI.

B. Single atoms

Many experimental methods have been developed for measuring and manipulating individual quantum systems, including single atoms in a cavity, trapped ions, and neutral atoms in an optical lattice.

Tab. XVII summarises some of the notable developments. More material can be found in the excellent review papers (???).

C. Quantum dots

In quantum dots, the recombination of an electron-hole pair, known as an *exciton*, emits a single photon. Photon sources based on this concept are rapidly gaining popularity in present-day labs, and a summary of developments in this area is provided in Tab. XVIII.

D. Nitrogen-vacancy centres

A nitrogen-vacancy (NV) centre in diamond refers to a nitrogen (N) atom replacing a carbon atom in the crystal lattice, neighbouring a single vacancy (V) defect (?). In such centres, both electrons and nuclear spins can exhibit long coherence times ($>1\text{ms}$ for the electron spin, and $>1\text{s}$ for the nuclear spin) even at room temperature (???).

Tab. XIX summarises some of the recent developments in NV centre technology. Further material can be found in the review papers (???).

E. Superconducting rings

Superconductors have become a promising candidate for processing quantum information, and much experimental progress has been made in recent years.

Tab. XX summarises some of the recent achievements in superconducting quantum technology. The review paper (?) provides excellent follow-up material for further details.

LXXII. MEASUREMENT

MANY quantum information protocols have been developed as two versions in parallel – a discrete-variable version, and a continuous-variable version.

Optical discrete-variable systems are typically measured using photo-detectors, whereas continuous-variable schemes are typically measured using homodyning.

A. Photo-detection

There are a variety of single-photon detectors, detailed descriptions of which can be found in several excellent review papers (??).

Tab. XXI summarises single-photon detectors that have good spectral response in the near-infrared regime.

B. Homodyning

Tab. XXII summarises some of the notable developments in homodyne detection, the most fundamental approach to quantum measurement in CV optical systems.

LXXIII. EVOLUTION OF OPTICAL STATES

QUANTUM states can be encoded in various degrees of freedom of a single photon. Tab. XXIII summarises some of the major developments in the controlled evolution of optical states.

There are two primary types of optical circuits: optical waveguides, and fibre-loops.

Optical waveguides can be miniaturised and integrated onto a small chip, which is alignment-free, compact, and phase-stable.

Fibre-loop schemes are extremely resource-frugal – resource requirements are essentially constant, and do not grow with the complexity of the underlying interferometer being implemented.

Some of these techniques are summarised in Tab. XXIV.

LXXIV. QUANTUM MEMORY

NUMEROUS excellent review articles exist (?????), providing detailed explanations of quantum memory schemes. Quantum memory is so ubiquitous and has such widespread utility that it has become a major research area in its own right.

The simplest approach to storing light is using an optical delay line, such as an optical fibre. Aside from loss, these also suffer that the storage time is fixed by the delay length, making on-demand output extremely challenging without first having access to high-speed ($\sim\text{GHz}$) optical switches.

Alternatively, light can be stored in a high-Q cavity. The light cycles back and forth between the reflecting boundaries, allowing it to be injected into and retrieved from the cavity (?????). Unfortunately, the storage of light in cavities suffers from a tradeoff between short cycle time and long storage time, which limits efficiency. Therefore, whereas optical delay lines and nano-cavities could be appropriate for obtaining on-demand single photons from heralded sources (??), they may not be suitable for quantum memory or quantum repeaters.

Tab. XXV summarises some of the notable developments in quantum memories.

LXXV. QUANTUM COMPUTATION

THE Holy Grail of scalable, universal quantum computation – perhaps the most exciting of future quantum developments – has been developing rapidly across various physical platforms simultaneously in parallel, including: trapped atoms/ions; nuclear magnetic resonance (NMR); photons; superconductors; spins in silicon; and many, many more. An excellent point of reference is the review (?).

Tab. XXVI summarises some of current records for the number of qubits and attained fidelities in different physical systems for implementing quantum computation. Most noteworthy, IBM and Google recently announced their record-breaking 50- and 72-qubit superconducting quantum processors respectively, a key step towards achieving the coveted goal of quantum supremacy (???).

— *Om mani padme hum.*

LXXVI. TIMELINES & REFERENCES

THE historical development timelines of the technologies referred to earlier in this part, and their references, now follow in table form.



Summary	References & years
A technique for the generation of high-intensity polarisation-entangled photon-pairs. For the partial Bell state projection a 50:50 beamsplitter was employed.	(??)
The first experimental demonstration of photonic quantum teleportation. Two photon-pairs were prepared by double-pumping a single non-linear beta-barium borate (BBO) crystal: one pair employed as the entanglement source; the other to prepare the state to teleport. The partial Bell state measurement was implemented using which-path erasure at a beamsplitter, with an efficiency of 25%.	(?)
Experimental demonstration of quantum state teleportation between two labs, separated by 55m, but connected by a 2km length of fibre, with photons at telecommunication wavelengths. This arouses the exciting prospect that future quantum networks might be able to piggyback off existing telecom infrastructure, which would be a boon to the quantum industry.	(?)
Quantum state teleportation was demonstrated between photonic and atomic qubits, a first step towards hybrid architectures, and an essential ingredient in interfacing optical and non-optical systems.	(?)
Quantum teleportation over a 16km long, noisy, free-space channel between distant ground stations was demonstrated. This distance is of especial interest as it is significantly longer than the effective thickness of the atmosphere, equivalent to 5-10km of ground atmosphere. This is an exciting benchmark as it suggests that free-space ground-to-satellite teleportation may be viable.	(??)
The teleportation distance in free-space was extended to 97km over Qinghai Lake, and 143km between the two Canary Islands of La Palma and Tenerife. These overcame the daunting challenges associated with source targeting and tracking, for long-distance, free-space quantum teleportation, and paved the way for future satellite-based quantum teleportation.	(??)
Accompanying the breakthrough of superconducting single-photon detectors with near-unit efficiency, 3-fold photo-detection for quantum teleportation was greatly enhanced by more than two orders of magnitude at telecom wavelengths, and the teleportation distance in optical fibre lengthened to 100km.	(?)
Quantum teleportation over fibre networks in Hefei and Calgary were demonstrated, with lengths of dozens of kilometres.	(??)
The first quantum satellite for entanglement distribution was launched in China. In addition to teleportation, this could facilitate intercontinental QKD. The team is aiming to achieve quantum teleportation between ground stations and satellite, and even between pairs of distant ground stations, separated by over 1000km, using shared entanglement provided by the satellite.	(?)
Using 5-photon entanglement, open-destination teleportation was implemented, whereby an unknown quantum state was teleported onto a superposition of 4 destination photons, which could be read out at any location – a type of ‘broadcasting’.	(?)
The state of a two-photon composite system was demonstrated – a breakthrough in the teleportation of a single particle onto a complex system comprising multiple particles.	(?)
Quantum teleportation over multiple degrees of freedom of a single optical mode was demonstrated.	(?)
Teleportation of CV optical states was demonstrated. The advantage of this teleportation protocol was that it could be deterministic in principle, overcoming the non-determinism inherent to partial Bell state projections using a beamsplitter.	(?)
The deterministic teleportation of photonic qubits was demonstrated using hybrid techniques.	(?)
Quantum teleportation has also attracted great interest in other physical architectures. Demonstrations have been performed in various physical systems, including atoms, ions, electrons, and superconducting circuits.	(?????)
Hybrid schemes combining different physical systems have been demonstrated, such as light-to-matter teleportation. Hybrid technologies are expected to play an important role in future quantum networks, where the underlying physical architecture for (say) a quantum computation is non-optical, but optics mediates the communication of quantum information.	(??)

Table VIII: Developments in experimental quantum state teleportation and entanglement distribution.

Summary	References & years
The first experimental demonstration of entanglement swapping. By pumping a BBO crystal in a double-pass configuration, two pairs of polarisation-entangled photons were generated to demonstrate the scheme. A visibility of 0.65 was observed, which clearly surpasses the classical limit of 0.5. This was later improved in 2001 to a visibility of 0.84, which violates the Bell inequality (for which the threshold is 0.71).	(??)
Aside from ‘event-ready’ mode, a delayed-choice mode of operation for entanglement swapping was proposed, where entanglement is produced <i>a posteriori</i> , after the entangled particles have been measured and may no longer even exist.	(?)
Delayed-choice entanglement swapping was designed and realised. This was performed by adding two 10m optical fibre delays of about 50ns for both outputs of the Bell state measurement.	(?)
A delayed-choice entanglement swapping experiment with vacuum-one-photon quantum states was realised.	(?)
By designing a special interferometer to realise active switching between Bell state measurement and separable state measurement, an entanglement swapping experiment with active delayed-choice was demonstrated.	(?)
By using three pairs of polarisation-entangled photons and conducting two Bell state measurements, multistage entanglement swapping was realised.	(?)
Multi-particle entanglement swapping using a three-photon GHZ state was demonstrated.	(?)
Entanglement swapping between photons that never coexisted was demonstrated. In their experiment, entangled photons are not only separated spatially, but also temporally.	(?)
Hybrid entanglement swapping between discrete-variables and continuous-variables optical systems was realised experimentally.	(?)
Entanglement swapping based on independent entangled photon sources has been experimentally verified.	(???)
Entanglement swapping using independent entangled photon sources separated by 1.3km in a real-world environment was achieved.	(?)
Entanglement swapping over 100km optical fibre with independent entangled photon-pair sources.	(?)
Starting with two entangled atom-photon pairs, we can project the two atomic qubits into a maximally entangled state by performing a Bell state measurement on the two photons.	(??)
Two trapped atomic ions separated 1m apart were entangled using entanglement swapping, exploiting interference between photons emitted by the ions. The fidelity of the states of the entangled ions was 0.63(3). In subsequent experiments, the ion-ion entanglement fidelity was improved to 0.81.	(??)
Two atomic ensembles were entangled, each originally with a single emitted photon, by performing a joint Bell state measurement on the two single photons after they had passed through a 300m fibre-based communication channel.	(?)
Robust entanglement (estimated state fidelity of 0.92 ± 0.03) between the two distant spins by entanglement swapping was generated. Such a high fidelity is sufficient to successfully perform loophole-free Bell inequality tests.	(?)
Quantum repeaters, combining entanglement swapping and quantum memory, which provides a potential solution to this problem, was proposed in 1998 by Briegel <i>et al.</i> And the first proposed practical quantum repeater architecture was proposed by Duan, Lukin, Cirac & Zoller (DLCZ) using atomic ensembles and linear optics.	(??)
To increase the repeater count-rate, various protocols have been proposed.	(?????????)

Table IX: Developments in entanglement swapping and quantum repeaters.

Summary	References & years
The concept of all-photonic quantum repeaters, based on flying qubits, which entirely mitigate the need for a matter quantum memory, was introduced.	(?)
Experimental demonstration of elementary segments of quantum repeaters were achieved.	(??).
In order to develop practical quantum repeaters, there are many experimental techniques that must be developed, such as multiplexing, and techniques based on non-degenerate photon-pair sources and quantum frequency conversion.	(???????????)
Quantum repeater techniques based on other physical systems have also been developed.	(????)

Table X: (continued) Developments in entanglement swapping and quantum repeaters.

Summary	References & years
Quantum cryptography using polarised photons in optical fibre over more than 1km.	(?)
QKD experiment over 10km using phase-encoding.	(?)
Outdoor experiment over 67km using a plug-and-play system to automatically maintain stabilisation.	(?)
QKD based on decoy-states over more than 100km, marking the beginning of long-distance QKD.	(???)
Decoy-state QKD over a 200km optical fibre cable through photon polarisation with a final key rate of 15Hz.	(?)
First realisation of a differential phase-shift (DPS) QKD protocol over a 42.1dB lossy channel and 200km of optical dispersion-shifted fibre.	(?)
The DPS protocol over 50dB channel loss and 260km of optical fibre using superconducting detectors was realised. This is the first implementation of QKD over more than 50dB channel loss.	(?)
The coherent one way (COW) protocol QKD system with a maximum range of 250km at 42.6dB channel loss using ultra-low-loss fibre, with secret bit-rates up to 15Hz was realised.	(?)
Zeilinger's group distributed entangled single-photons over a free-space quantum channel, demonstrating the viability of free-space quantum communication.	(?)
A complete experimental implementation of a QKD protocol over a free-space link using polarisation-entangled photon pairs.	(?)
The BBM92 QKD protocol based on polarisation encoding over 144km was realised.	(?)
To verify the feasibility of a quantum channel between space and Earth, a European Union group successfully received weak light pulses emitted from a ground station and reflected by a retroreflecting mirror on a low-Earth orbit satellite with orbital altitude of 1485km.	(?)
In the context of rapidly moving platforms, QKD over 20km from an airplane to the ground was realised.	(?)
Quantum communication with a hot-air balloon floating platform was accomplished successfully.	(?)
The first quantum satellite was launched in China, and opened a platform for satellite-to-ground quantum communication at an intercontinental level.	(???)
The network of the American Defence Advanced Research Projects Agency (DARPA) connected the three nodes – Harvard University, Boston University, and the BBN company – in 2005, later increasing this to 10 nodes.	(?)
Since 2006, the EU has established a 'SECOQ' network, combining the efforts of 41 research and industrial organisations from 12 countries, including the UK, France, Germany, and Austria. A typical network employing a trusted repeater paradigm, with 6 nodes and 8 links was demonstrated in Vienna.	(?)
National Institute of Communication Technology, together with Nippon Telegraph & Telephone Corporation (NTT), Nippon Electric Company, Mitsubishi Electric Corporation, Toshiba European company, Switzerland IDQ Company and an Austrian team constructed a Tokyo QKD network in a metropolitan area, demonstrating the world's first secure TV conferencing over a distance of 45km.	(?)
A 3-node network with a chained architecture, which demonstrated a cryptographically secure real-time voice call, was designed and constructed.	(?)
A metropolitan all-pass and intercity quantum communication network in field fibre for 4 nodes was designed. Any 2 nodes can be connected in the network, enabling QKD between arbitrary pairs of users.	(?)
A QKD network with wavelength division multiplexers, realising 4 and 5 nodes with a star topology was constructed.	(??)
The MDI QKD protocol in the laboratory over more than 80km of spooled fibre with time-bin encoding was demonstrated. They also tried outdoor experiments over 18.6km.	(?)
A Brazilian group demonstrated the MDI QKD protocol using a polarisation encoding scheme.	(?)
A full demonstration of time-bin phase-encoding MDI QKD over a 50km fibre link was reported.	(?)
Polarisation-encoded MDI QKD with commercial off-the-shelf devices over 10km, with a secure key-rate of 0.0047Hz was implemented.	(?)
The Chinese group continues to upgrade the performance of MDI QKD systems, making them viable over distances of up to 200km and 400km.	(??)
A QKD system over 421km is achieved, by using a 3-state time-bin protocol combined with a one-decoy approach.	(?)

Table XI: Developments in experimental QKD.

Summary	References & years
An experimentally viable purification scheme, requiring only PBSs and post-selection, was proposed and demonstrated. The scheme was demonstrated again, using mixed states with fidelity 0.75 (0.8), which they were able to purify to 0.92 (0.94), a major improvement.	(??)
A Bell experiment was performed using purified states. A state initially failing a Bell test successfully passed the test following entanglement purification. Unfortunately, the theoretical efficiency of the purification scheme is only 1/4.	(??)

Table XII: Developments in experimental entanglement purification.

Summary	References & years
As one kind of single-longitudinal-mode laser, distributed feedback (DFB) lasers have a stable wavelength since they depend on etched gratings, which can only be tuned slightly with temperature. Thus, they are widely used in optical communication applications, such as QKD or dense wavelength division multiplexing, where it is desired to have a tuneable signal and extreme narrow linewidth.	(??)
In recent years, several QKD schemes have been designed and implemented using multi-longitudinal-mode Fabry-Perot (FP) lasers, mainly because they are significantly cheaper than DFB lasers.	(??)

Table XIII: Two types of laser commonly used in quantum communication experiments.

Summary	References & years
Single photons can be easily created in pairs via SPDC. A state-of-the-art SPDC source at a wavelength of 788nm was reported simultaneously with a high brightness of ~ 12 MHz/W, a collection efficiency of $\sim 70\%$ and an indistinguishability of $\sim 91\%$.	(?)
An SPDC source at a wavelength of around 1550nm was reported with simultaneously 97% heralding efficiency and 96% indistinguishability between independent single photons.	(?)
Heralded sources were integrated via four-wave mixing (FWM) in waveguides or optical fibres. Despite being a probabilistic process, SPDC or FWM enjoys a warm popularity in the labs of quantum optics since they have been technically mature and very cheap. That is why so far almost all the quantum information experiments in quantum networks were based on SPDC or FWM sources.	(?????)
Truly deterministic single-photon sources will be indispensable for future large-scale quantum networks. One of the most promising single-photon sources is based on quantum dots, which have been developed simultaneously exhibiting a high purity of 99.1%, high indistinguishability of 98.5%, and high extraction efficiency of 66%.	(???????)

Table XIV: Some of the notable developments in single-photon state preparation.

Summary	References & years
In 1999, the world's first multi-particle entanglement – 3-photon GHZ entanglement – was generated. After that, Pan and his colleagues broke the records continuously, realising 4-, 5-, 6-, 8-, 10-, and 12-photon entanglement, and 10- and 18-qubit hyper-entanglement with multiple degrees of freedom.	(???)
8-photon optical cluster states have been realised, and used to demonstrate topological error correction against a single error on any qubit.	(?)
5-photon NOON states have been realised by mixing quantum and classical light.	(?)
6-photon Holland-Burnett states have been realised both at visible wavelengths and at telecom wavelengths.	(??)
Entangled photon-pairs were created with dimension up to 15×15 on a programmable silicon integrated circuit.	(?)
Zeilinger and his colleagues created orbital angular momentum (OAM) entangled photon-pairs differing by 600 in their quantum number. Four years later, they prepared 100×100 -dimensional OAM entangled photon pairs.	(?)
Furusawa <i>et al.</i> deterministically generate and fully characterise a continuous-variable cluster state containing more than 10,000 entangled modes. Three years later, they improved the system to more than one million modes.	(??)
Small amplitude cat states were generated via photon subtraction from a squeezed vacuum state.	(???)
Large amplitude cat states (i.e $ \alpha ^2 > 2.3$) were prepared by reflecting half of a photon-number state into a momentum quadrature homodyne detector.	(??)
Remote entanglement of two independent cat states was created by interfering small fractions of each pulse.	(?)

Table XV: Developments in entangled state preparation, based on non-linear optics.

Summary	References & years
Atom-light interaction currently represents the most mature method to create atomic squeezing, and owns the highest squeezing of 20.1dB via an optical-cavity-based measurement.	(?)
The detection of a single photon prepares almost 3000 atoms into an entangled Dicke state. The state was reconstructed with a negative-valued Wigner function – an important hallmark of non-classicality.	(?)
In 2010, two ensembles were entangled by storage of two entangled light fields; and ten years later, four quantum memories were entangled via spin waves.	(??)
Bose-Einstein condensates were used to create large ensembles of up to 10^4 pair-correlated atoms with an interferometric sensitivity – 1.61dB beyond the shot-noise limit.	(?)

Table XVI: Notable developments in atomic ensembles.

Summary	References & years
In 2000, Haroche and his colleagues generated entanglement between two atoms and a single-photon cavity mode. Later, entanglement was demonstrated between two photons sequentially emitted by the same single atom in a cavity. An entanglement fidelity of 86% between the two photons was obtained, and importantly, the entanglement generation efficiency was 15%, a drastic improvement compared to free-space experiments with single atoms.	(???)
In the atom-cavity system, one of the largest Schrödinger cat states was created, with a spin of 25 on a Rydberg atom.	(?)
In the recent year, Blatt <i>et al.</i> reported scalable and deterministic generation of multi-ion entanglement, including 8-qubit W-states and 14-qubit GHZ states.	(??)
Arbitrary coherent addressing of individual neutral atoms in a 3D optical lattice of up to 50 qubits was achieved.	(?)
Several groups reported generation and detection of two neutral-atom entanglement in optical lattices with a single-site-resolved technique.	(???)

Table XVII: Notable developments in single-atom technology.

Summary	References & years
Spontaneous emission from a trion (the simplest charged excitons) state prepares a single electron spin in a quantum dot, while spin-photon entanglement was generated from such decay.	(??)
Two-photon polarisation-entangled states can be generated from biexciton-exciton cascade radiative decay in a single quantum dot. The entangled photon pairs simultaneously exhibit a high purity of 0.004, high fidelity of 0.81, high two-photon interference visibility of 0.86, and on-demand generation efficiency of 0.86.	(?)
A family of pyramidal site-controlled quantum dots was used to construct an array of entangled photon emitters with fidelities as high as 0.72.	(??)
A cluster state of five sequentially-entangled photons was generated by periodic timed excitation of a matter qubit. In each period, an entangled photon is added to the cluster state formed by the matter qubit and the previously emitted photons.	(?)

Table XVIII: Major developments with quantum dots.

Summary	References & years
For NV centres, both electron and nuclear spins were observed with long coherence times ($>1\text{ms}$ for the electron spin and $>1\text{s}$ for the nuclear spin) even at room temperature.	(???)
Electron spin lifetime limited by phononic vacuum modes was observed. Negatively charged NV centres were proved to have exceptionally long longitudinal relaxation times, with T_1 -times of up to 8 hours.	(?)
Bipartite entanglement was created between two nuclear spins coupled to a single NV centre in diamond, and tripartite entanglement was generated via the electron spin of the NV centre itself as the third qubit.	(?)
Quantum entanglement was created between a polarised optical photon and an NV centre qubit. Because NV centres couple to both optical and microwave fields, they can be used as a quantum interface between optical and solid-state systems.	(?)
Room-temperature entanglement was created between two single electron spins over some 10nm distance in diamond. In the same year, heralded entanglement was created between two solid-state qubits located in independent low-temperature setups separated by 3m.	(??)

Table XIX: Advances in nitrogen-vacancy (NV) centre technology.

Summary	References & years
There are three basic types of superconducting qubits – charge, flux and phase. In 2007, a charge-insensitive qubit, called a transmon, was designed.	(?)
Transmon qubits in a waveguide cavity, called circuit QED, achieved coherence times on the order of 0.1ms.	(??)
In 2013, Martinis and his colleagues designed a cross-shaped transmon qubit, called an Xmon, which balances coherence, connectivity and fast control. In 2014, they constructed a 5-qubit GHZ state using five Xmons arranged in a linear array. In 2015, they reported the protection of states from environmental bit-flip errors in a 9-Xmon-qubit linear array.	(???)
Pan <i>et al.</i> reported the preparation and verification of genuine 10- and 12-qubit entanglement in a superconducting processor.	(??)
An 100-photon Schrödinger cat state was created mapping from an arbitrary transmon qubit state, and extended to a superposition of up to four coherent states.	(?)
In 2016, the lifetime of microwave photons was extended to 0.3ms with error correction in superconducting circuits.	(?)

Table XX: Developments in superconducting rings for quantum information processing.

Summary	References & years
<p><i>Single photon avalanche diodes (SPAD)</i> – SPADs are the most compact and common single-photon detectors in the world. Commercially, silicon SPADs are available with detection efficiency around 60% at 780nm, maximum count rates of 25MHz and dark-count rates as low as 25Hz. Several remarks are:</p>	(?)
<ul style="list-style-type: none"> • Higher detection efficiency is possible. However, it is often at the cost of a lower maximum count rate (typical dead-time is $1\mu\text{s}$, leading to 1MHz count-rates) and a larger dark-noise. • Telecom band SPADs are usually based on InGaAs/InP, which suffer from low efficiencies of around 25%. Another more efficient method is to convert the telecom photons into near infrared ones by up-conversion, and then detect them with a silicon SPAD. The detection efficiency can be improved to 30~40%. • SPADs are also commercially available in a multi-channel array with single power supply and individual inputs/outputs. Note that they are different from the multi-element SPAD arrays, which have either only one input or one output, and can be used to achieve photon-number resolution or high-speed single-photon detection. 	
<p><i>Superconducting nanowire single photon detectors (SNSPD)</i> – These are more efficient, but larger and more expensive than SPADs. Up until now, the highest detection efficiency of SNSPDs is 93% at 1550nm, demonstrated in 2014. SNSPDs are commercially available with peak detection efficiencies higher than 80% around 800nm, maximum dark-count rates of $100 \sim 300\text{Hz}$, and a dead-time of $10 \sim 70\text{ns}$. Similar performance is expected for optimised SNSPD at other wavelengths (780 ~ 1550nm). However, SNSPDs must be operated at cryogenic temperatures of a few Kelvin to maintain their superconducting state. This is the main reason why SNSPDs are much more cumbersome and expensive than SPADs.</p>	(?)
<p><i>Transition edge sensors (TES)</i> – These exhibit the highest detection efficiency, reported as high as 98% (95%) at wavelengths around 850nm (1556nm). Moreover, they are photon-number-resolving, owing to the linear response between resistance and temperature. However, the typical thermal recovery times of TESs are a fraction of a microsecond, limiting their applicability in high-speed quantum networks. Besides, TESs usually operate at ultra-low temperatures of 100mK, making it challenging to migrate the technique from lab to market.</p>	(??)

Table XXI: Some state-of-the-art single-photon detectors.

Summary	References & years
<p>The first time-domain homodyne detection was performed below 1KHz and achieved a shot-noise of 9dB.</p>	(?)
<p>Several groups have constructed high speed and pulse-resolved homodyne detectors in the near infrared and telecom regimes. State-of-the-art homodyne detectors have achieved a shot-noise to electronic-noise ratio of 7.5~14dB, and a bandwidth of $\sim 100\text{-}300\text{MHz}$, which enables quantum protocols with repetition rates ranging from tens of MHz to 100MHz.</p>	(?????)

Table XXII: Some state-of-the-art homodyne detectors.

Summary	References & years
<p>For polarisation encoding, single-photon states can be easily manipulated by using a series of wave plates. CNOT gates with probability 1/9 have been reported in the coincidence basis.</p>	(????)
<p>For path encoding, high-fidelity CNOT gate was reported with integrated optical waveguides.</p>	(?)
<p>For time-bin encoding, heralded controlled-phase gates was reported.</p>	(?)
<p>Assisted by two independent photons, CNOT gates have been realised in a nondestructive manner.</p>	(??)
<p>By using multi-level quantum systems to encode information, probabilistic 3-qubit and 4-qubit entangling gates were reported.</p>	(??)
<p>Ultra-low loss multi-mode optical bulk circuits have been assembled and used for boson-sampling machines beating early classical computers.</p>	(??)

Table XXIII: Some of the important developments in the evolution of optical states.

Summary	References & years
Silica-on-Silicon (SoS) planar light-wave circuit (PLC). For the state-of-the-art technique, SoS-PLCs can achieve ultra-low propagation loss (<0.01 dB/cm) and coupling loss (<0.1 dB/facet). Nowadays, small-scale quantum circuits have reached coupling loss ~0.4 dB/facet and each directional coupler (DC) loss ~0.1 dB.	(??)
Fused silica waveguides written by femtosecond laser direct written (FLDW) technology. Unlike PLC fabrication, the FLDW is a powerful and flexible technique for 3D rapid fabrication, requiring no masking procedure or cleanroom environment. FLDW waveguides exhibit losses on the order of 0.1 dB/cm.	(?)
Telecom silicon waveguides have higher refractive index, thus allowing a dramatic reduction in the size. For example, a 2×2 coupler in silicon ($27\mu\text{m}$) is 40 times shorter than the one in silica (1.1 mm). The best propagation (coupling) losses can reach 0.1 dB/cm (0.5 dB/facet).	(????)
A programmable eight-mode photonic time-bin circuit has been constructed to implement three and four photons boson-sampling. The controlled switch was realised using a bulk electro-optic modulator (EOM) with a transmission of 97.3%, and a single-loop efficiency was achieved with 83%.	(?)

Table XXIV: Some state-of-the-art programmable optical circuits.

Summary	References & years
One of the simplest way to storing light is to use an optical fibre. However, the storage time is hard to adjust and limited to tens of microseconds due to high transmission loss of ultra-long fibre.	(??)
An erbium-doped optical fibre was successfully used to store telecom-wavelength single photons and entangled photons.	(??)
An ultra-small high-Q photonic-crystal nano-cavity was used to trapping and delaying photons for 1.45ns.	(?)
A quantum light-matter memory has been realised with simultaneously a readout efficiency of 76% and a lifetime of 0.22s, which can support a sub-Hz entanglement distribution of up to 1,000km, for the first time going beyond the maximally achievable quantum communication distance using direct transmission.	(?)
Electromagnetically induced transparency (EIT) in cold atomic ensembles has been demonstrated to store entangled photon pairs and squeezed vacuum states.	(???)
EIT was used to store a qubit with storage times reaching the millisecond range.	(???)
EIT was used to store single-photon-level quantum images for 100 nanoseconds.	(?)
Nuclear spins of ^{31}P impurities in an sample of ^{28}Si were reported with a coherence time of as long as 192s at a temperature of 1.7K.	(?)
Ionised donors in silicon-28 was observed with a qubit storage time exceeding 39 minutes at room temperature.	(?)
Optically addressable nuclear spin in a solid have been demonstrated to achieve with remarkably long coherence times, and with a current record of six hours.	(?)

Table XXV: Some of the notable developments in quantum memory.

Summary	References & years
Linear optics: The first demonstration of an entangling, photonic, 2-qubit CNOT gate using post-selected linear optics.	(?)
Trapped atoms/ions: A register of 20 individually controlled qubits was generated and characterised; Single-qubit gates with an error per gate of 3.8×10^{-5} , and deterministic 2-qubit gate with a gate error of 8×10^{-4} ; 14-qubit entanglement with a 2-qubit fidelity of 98.6%.	(???)
Nuclear magnetic resonance: 12-qubit system with the simulated fidelity over 99.7% on each qubit; The average fidelity of 87.5% on a 7-qubit entangling gate.	(???)
Linear optics: A fully programmable 2-qubit linear optics quantum processor with an average quantum process fidelity of 93.2%.	(?)
Boson-sampling: Several five-photon boson-sampling machines were built and beat early classical computers.	(???)
Coherent Ising machine: An optical processing approach was used to model and optimise 2000-node optimisation problems based on a network of coupled optical pulses in a ring fibre.	(??)
Superconductors: 9-qubit system with an average single-qubit gate fidelity of 99.92% and a 2-qubit gate fidelity of up to 99.4%.	(??)
Spins in Silicon: A programmable 2-qubit quantum processor in silicon with state fidelities of 85~89%; 2-qubit CNOT gate; An addressable single-qubit with a control fidelity of 99.6%.	(???).
Nitrogen-vacancy centres: A programmable 2-qubit solid-state quantum processor at room temperature.	(?)
Adiabatic quantum computation: A 2048-qubit D-Wave quantum processor was used to predict phase-transitions and topological phenomena.	(??)

Table XXVI: Major milestones in the development of quantum computation.

Part XII

To the future & beyond

“The most exciting phrase to hear in science, the one that heralds new discoveries, is not ‘Eureka!’, but ‘That’s funny...’” — Isaac Asimov.

In this part we highlight some of – in the minds of the authors – the most exciting and transformative applications for the future quantum internet and its monstrous unified quantum computational power. None of the technologies discussed here have presently been realised on any practical scale, but all have revolutionary potential to transform the global economy, how we interact with the world, and how we live our lives.

However, these predictions on the future of world-changing applications for quantum technology are merely speculation. We believe it is far more likely that the most impactful and reverberating applications for quantum technology likely haven’t even been invented yet.

Currently, there is much ongoing research being conducted into new quantum algorithms and protocols, with major developments being made at a rapid pace. Most likely, mankind’s quantum future will be far more exciting than our speculations here, much as speculation surrounding early digital computers never anticipated the digital technological revolution and its impact on every aspect of our modern lives.

— *Excelsior.*

LXXVII. SPACE-BASED QUANTUM NETWORKS

— by Tim Byrnes.

“The Earth is the cradle of humanity, but mankind cannot stay in the cradle forever.” — Konstantin Tsiolkovsky.

ONE of the major hurdles that must be overcome before quantum networks achieve widespread commercial use is to span large distances between nodes distributing entanglement as a resource. The key technological and commercial centres of the world are decentralised, spread across the globe, hence connections over intercontinental distances are essential, with nodes separated potentially by thousands of kilometres. This distance is one of the major challenges facing the creation of large-scale quantum networks.

The most convenient way of transmitting quantum information is using optical fibres. However, it is well-known that due to the exponential scaling of loss with distance, the upper limit is of the order of several hundred kilometres. This shortcoming has inspired intensive research into extending these distances using quantum repeaters, discussed in detail in Sec. [XXIX](#).

An alternative to fibre optic communication is using terrestrial free-space links. Particular wavelength regimes exhibit low absorption, allowing the propagation of photons across long distances. Impressive experimental demonstrations of teleportation with free-space entanglement ([???](#)) over 100km have been achieved.

However, extending to larger distances has been problematic. Photon loss in free-space links is affected by weather conditions and other atmospheric effects, such as pollution, which degrade visibility. The Earth’s curvature provides an absolute upper limit, depending on the elevation of the source and receiver. For example, for an observer on a 30m tower, the horizon is at a distance of 20km, a relatively short distance. The observatory used in the experiments performed by the experiments of ([??](#)) were at an elevation of 2393m, allowing transmission over a distance of 143km.

In general, the distance to the horizon d (in kilometres) on the Earth’s surface at an altitude of h (in metres) is given by the approximation,

$$d \approx 3.57\sqrt{h}. \quad (77.1)$$

This relationship is shown in Fig. [184](#), covering the range of altitudes from sea level to the summit of Mount Everest (8,848m). It is evident that line-of-sight channels across the Earth’s surface are inherently limited by the curvature of the Earth to ~ 250 km, using altitudes that are practically realistic (i.e no ground stations on the summit of Everest thank you very much!).

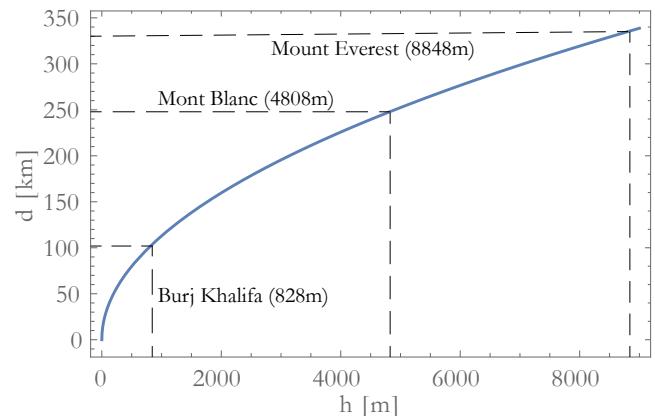


Figure 184: Distance to the horizon d (in kilometres), at altitude h (in meters), and several illustrative examples with their elevations. The range for h spans from sea level to roughly the height of Mount Everest, thereby covering the full range of altitudes possible for ground-based network nodes. However, using more practically realistic bounds on altitude, say on the order of 5,000-6,000m (already very optimistic, and only achievable in some parts of the globe), line-of-sight distance to the horizon will be limited to be on the order of 200-250km. This limitation necessitates the use of repeater networks to enable the efficient approaches for constructing quantum communication channels that bypass this constraint imposed by the Earth’s curvature.

One attractive possibility to overcome these issues is using space-based quantum communication. In this scenario, satellites orbiting the Earth would act as nodes of the quantum network, which could store, send and receive quantum information. At altitudes where satellites orbit the Earth, photon loss due to scattering is negligible, and photons can propagate across extremely long distances unhindered – despite various shortcomings of photonic quantum information, a key benefit is their robustness and preservation of coherence through empty space over very long distances. The main loss in this case is caused by diffraction, due to the finite diameter of the receiver and transmitter hardware. For example, using reasonable diameters it is in the region of 40–80dB loss for low Earth orbit (LEO) satellites (??).

Repeater network protocols (Sec. XXIX) are completely compatible in-principle with the space-based networks discussed here. In the context of a space-based network the key limitation we face is that the opposing hemisphere of the Earth is beyond direct line-of-sight, prohibiting a direct P2P communications channel. Instead, the long-distance channel must be broken down into shorter segments, within line-of-sight of one another, enabling, for example, entanglement to be incrementally swapped across neighbouring satellites and ultimately around the globe.

Involving ground-based quantum network nodes is not prohibitive as 80% of the atmosphere by mass is located within the first 12km in altitude, yielding low effective thickness of the atmosphere ($\sim 10\text{km}$). Despite this apparently low exposure to attenuation, the loss is nonetheless substantial, on the order of $\sim 40\text{dB}$ (depending on optical frequency) when the satellite is directly overhead. This significantly reduces effective count rates, a problem that does not represent any kind of engineering limitation, but rather an inherent physical one that cannot foreseeably be mitigated (it seems implausible that much can possibly be done to circumvent the atmosphere in satellite-to-ground communication!). This effective thickness of course changes with the azimuth of the satellite in the sky. Near the horizon the line-of-sight between satellite and ground-station is far greater than when the satellite is directly overhead. The light trajectory beyond this effective thickness regime effectively passes through vacuum with attenuation limited only by diffraction. Thus, when directly overhead, a satellite-to-ground channel is more favourable in terms of signal attenuation than long-distance ground-to-ground channels over free-space.

Ground-to-satellite and satellite-to-ground quantum communication have already been demonstrated, as will be discussed in detail in the next section. Such a satellite-based quantum communication system is naturally suited to various tasks, in particular QKD, which does not require quantum memories or impose onerous timing constraints. However, with the addition of quantum memories the capabilities of such a quantum network could be greatly enhanced, enabling many of the applications and

protocols described earlier (Part. V) at a global scale.

Naturally, such a space-based quantum network brings with it enormous engineering challenges that must be overcome prior to implementation. Creating even a short-distance quantum network is currently technologically challenging, let alone one that is loaded onto a satellite transmitting photons across distances potentially of the order of the diameter of the Earth. The cooperation of space agencies is necessary even for initial experiments.

As we describe more in the next section, several recent positive results have made the technology an immediate possibility in realising such a global level quantum network. In this section, we summarise the current technological state of the art internationally, and discuss some of the remaining major challenges.

There are two main ingredients for communication in a global satellite-based quantum internet: satellite-to-satellite, and ground-to-satellite, shown in Fig. 185. Each bring with them their own engineering challenges.

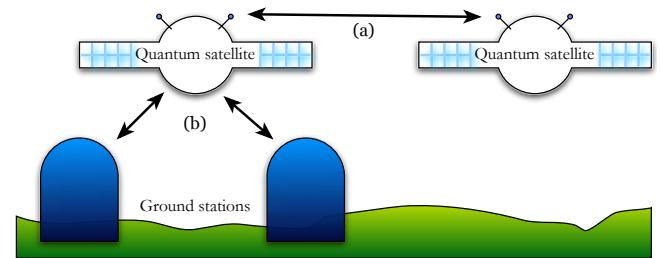


Figure 185: Various possibilities for space-based quantum communication. (a) Satellite-to-satellite quantum communication (?). (b) Ground-to-satellite quantum communication (?).

A. International efforts

We now briefly summarise some of the major international developments in the demonstration of satellite-based quantum communication. Some of these developments are extremely recent at the time of writing this work. It is a certainty that major developments will follow in the near future, with increasingly capable satellites, enabling the demonstration of ever more sophisticated quantum protocols.

1. China: QUESS

In 2016, the group at the University of Science & Technology China (USTC), led by Jian-Wei Pan, launched QUESS (Quantum Science Experiment Satellite), a quantum communication satellite (??). The main feature of the satellite is an ultra-bright source of polarisation entangled photons at a wavelength of 810nm, generated by SPDC. The source is capable of emitting 5.9×10^6 photon-pairs per second with a fidelity of ~ 0.91 .

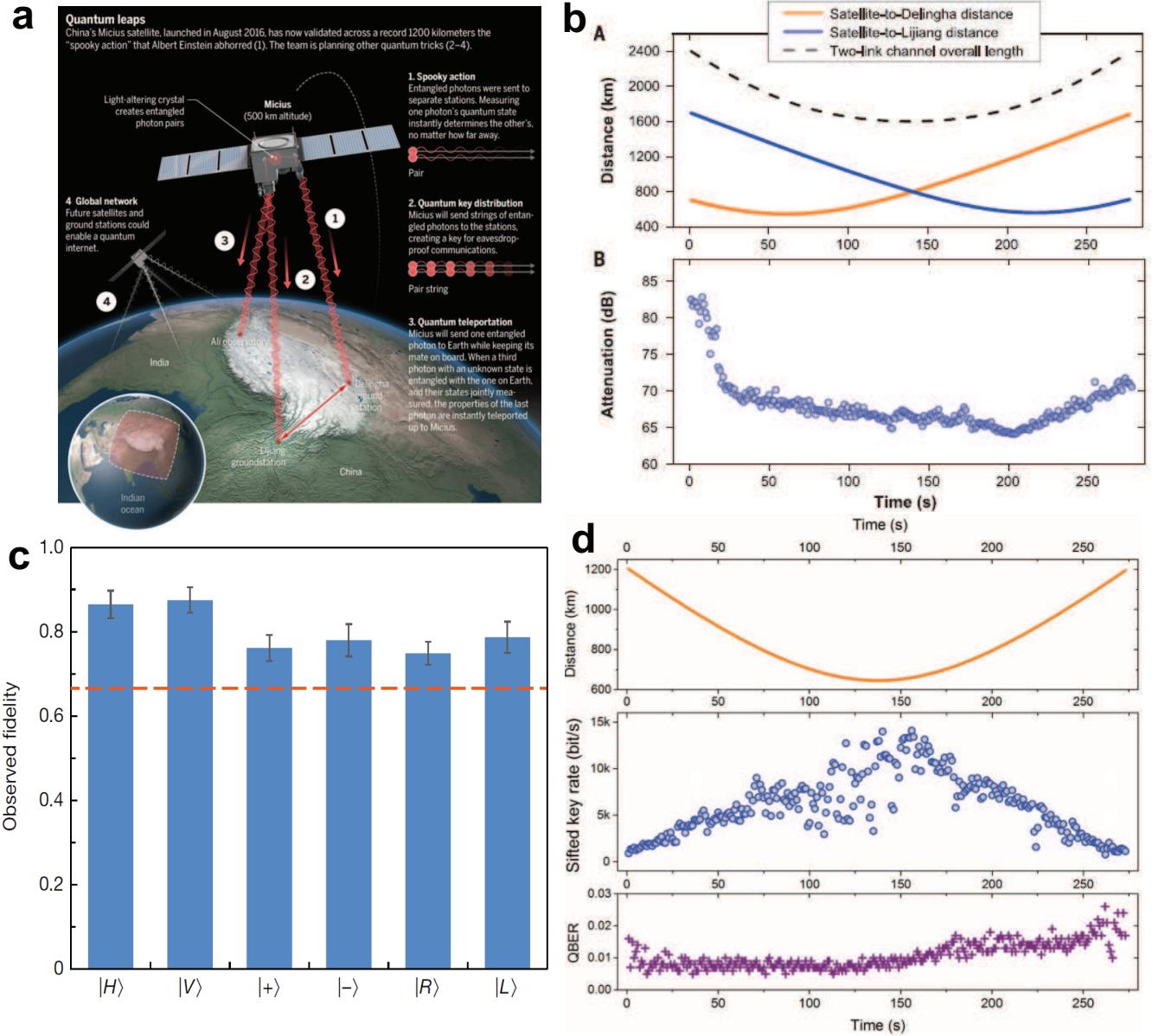


Figure 186: The Chinese Micius quantum communications satellite. (a) Schematic of the satellite and ground stations used to observe the entangled photons (?). [TB: Took image from Science News, may need permission]. Picture is watermarked. (b) Attenuation during entanglement distribution from (?). (c) Fidelities achieved for teleportation of various states as marked from (?).

The other important piece of technology is the acquisition, pointing, and tracking technology, which allows the ground stations to track the location of the satellite as it moves across the sky (which is very rapid for LEO), and vice versa. This is achieved using lasers at separate wavelengths, and allows the photon transmission and the ground-based receivers to be pointing to each other to within $\sim 1.2\text{mrad}$ (?). The major aims of the project are to:

- Perform QKD between Xinglong and Urumqi (a ground distance of 2,500km).

- Test Bell's inequality over a distance of 1,200km.
- Perform quantum state teleportation between the satellite and Ali in Tibet (see Fig. 186).
- (Ultimate goal) Perform QKD between Beijing and Vienna (a distance of 7,500km that extends well beyond the Earth's curvature).

At the time of writing, three main results have been reported. The first is an entanglement distribution experiment where a Bell violation was observed between ground stations at Delingha and Lijiang (separated by 1,203km)

and Delingha and Urumqi (separated by 1,120km) (?). The satellite is in LEO at an elevation of $\sim 500\text{km}$, and is in view from the observatories for a narrow duration of $\sim 275\text{s}$.

The attenuation of the photons during the downlink transmission is shown in Fig. 186(b). Taking into account photon transmission distances, the attenuation rates are far better than the best performance of optical fibres at 0.16dB/km (?) and even theoretical loss limits. The Bell violation recorded in this experiment was at the level of $2.37 \pm 0.09 > 2$, and the non-locality of the entanglement was confirmed.

The second experiment performed ground-to-satellite QST (?). Here, a single polarisation-encoded photon defines the qubit to be teleported, and the satellite acts as the receiver. The entangled photons are generated on the ground, at the observatory in Ngari, Tibet. As with the first experiment described above, the satellite is in LEO and sweeps across the sky, creating a narrow time-window during which the teleportation must be achieved. The longest distance for successful teleportation was $\sim 1400\text{km}$, when the satellite emerges from the horizon. When directly overhead, the satellite is at a distance of $\sim 500\text{km}$. As Fig. 186(c) shows, the fidelity of the types of states teleported are all above the theoretical classical bound of $2/3$, with an average of 0.80 ± 0.01 .

A third experiment performed QKD in a satellite-to-ground configuration, where the ground station was at Xinglong (?). The protocol that was used was the decoy-state BB84 protocol, which is robust against a photon-number splitting attack. A key-rate of between $1\text{--}12\text{kbit/s}$ was achieved while the satellite was visible, depending upon the relative location in the sky, shown in Fig. 186(d).

2. Japan: SOCRATES

SOCRATES (Space Optical Communications Research Advanced Technology Satellite) is a micro-satellite developed by NICT (National Institute of Information & Communications Technology) (?). This is a 48kg satellite of volume 50^3cm^3 that demonstrates multi-purpose (i.e both classical and quantum) optical communication. Its primary mission is to demonstrate laser communication in space. As one of its subgoals, the on-board equipment is adapted to perform QKD experiments. To this end, the satellite is equipped with a photon source, where the polarisation can be classically switched between non-orthogonal angles, necessary for BB84 for example.

Initial results indicated polarisation preservation of photons emitted from the satellite, a preliminary step towards performing protocols such as BB84 (?). More recent results showed quantum-limited satellite-to-ground communication was possible, attaining a quantum bit error rate below 5% (?) during the closest approach of the satellite to the ground at a distance of 744km . The

downlink was established over distances of over $1,000\text{km}$, although for obvious reasons the error rate was higher for longer distances.

3. Singapore: Cubesat

The group at the National University of Singapore (NUS) also has taken the approach of using nano-satellites, with the capability of generating correlated photon-pairs on-board (?). The nano-satellite weighs just 1.65kg and has all the components required for creating and detecting SPDC photon pairs. The nano-satellite approach greatly reduces the cost of space-launch, and the launch itself was performed by the Indian space agency [What's it called](#). As photon generation and detection are both on the same satellite, no quantum communications channel (to other satellites or Earth) was established in the experiment, although it demonstrated in-principle the space-worthiness of the basic hardware components.

4. Europe: Retroreflector satellite

European groups used laser-ranging satellites fitted with corner-cube retroreflectors to demonstrate quantum communication through the atmosphere (?). A ground-based photon source transmitted photons towards the satellite, which reflected them back down again to an Earth-based observatory. This confirmed the ability of photons to travel through the atmosphere into space and back, with a bit error ratio at the level of 5%. Although no optical sources or detectors were placed on the satellite, there has been long-standing active interest in a European space-based quantum communication project, led in particular by the group of Anton Zeilinger at the University of Vienna (?). The aims of such projects are to perform space-based QKD, with terrestrial free-space experiments (?) a part of the overall research effort.

5. Canada: QEYSSat

The QEYSSat (Quantum EncrYption & Science Satellite) proposes a micro-satellite that incorporates a quantum receiver (?). The satellite acts as a trusted node, to which photons would be transmitted from Earth-based sources. This could be employed to perform QKD between two locations on Earth separated by large distances. No photon sources or quantum memories would be included on the QEYSSat.

B. Future challenges

While early demonstrations of satellite-based quantum protocols have been successfully performed, future more

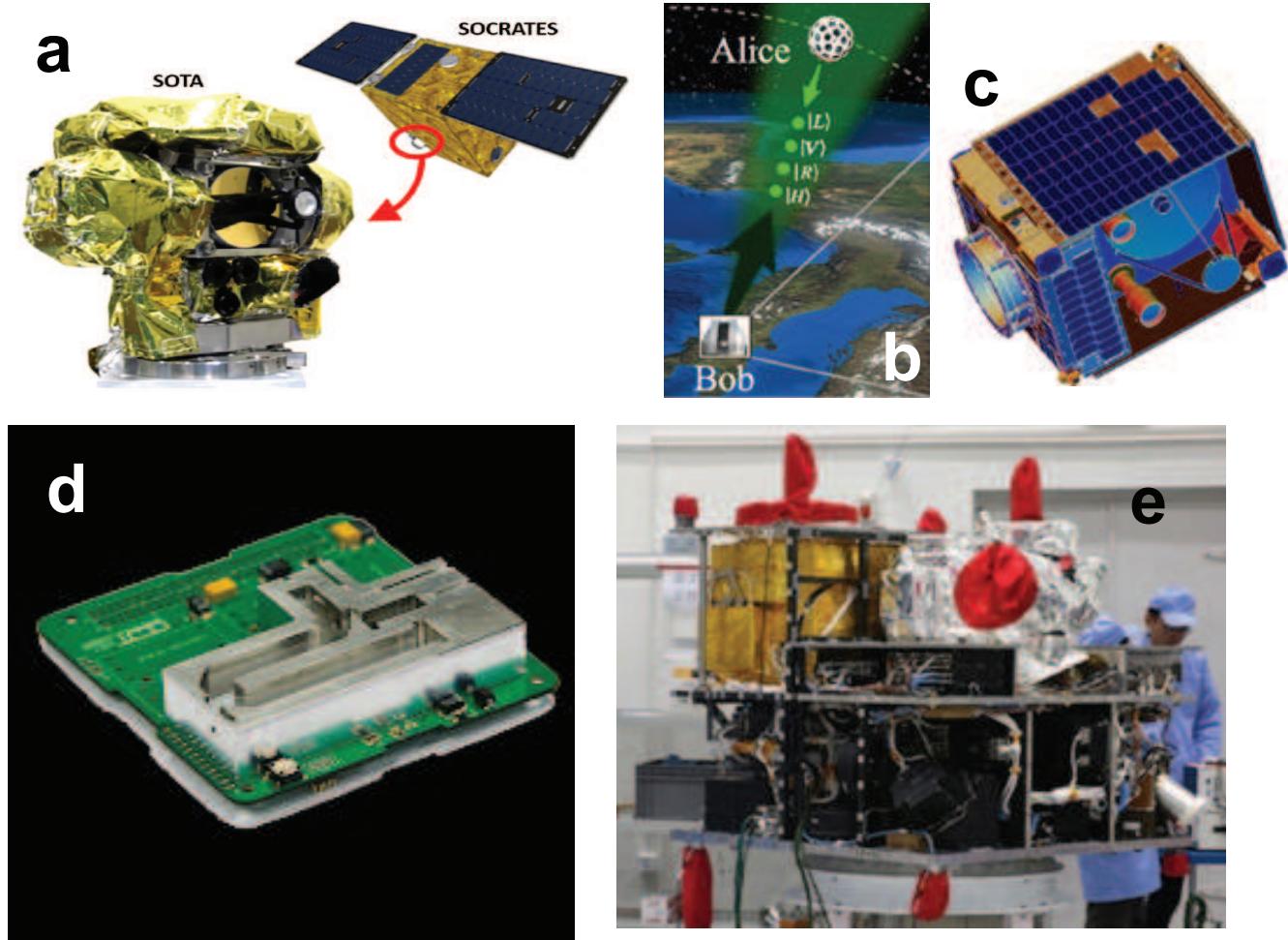


Figure 187: Satellites employing quantum technologies from various groups across the world: Japan (?), Italy (?), Canada (?), Singapore (?), and China (?). **Tim: do we have figure permission for this?**

sophisticated protocols will require more capable satellite hardware, which will inevitably begin to emerge in the immediate-term.

We identify two of the most pressing challenges facing a space-based quantum internet as:

- Coverage: how much of the surface of the globe will always be within line-of-sight of at least one satellite in the constellation network. Visibility of at least one satellite from any point on Earth will in-principle allow universal access to the quantum network, in the same way that the GPS constellation serves all points on the Earth's surface.
- Precision: many optical quantum protocols require strict synchronisation of photons undergoing interference. This becomes extremely challenging when those photons are originating from different satellite nodes, which must then be temporally synchronised on the scale of the photons' temporal wave-packets (i.e temporal mode-matching, discussed in Sec. XIV.G). Complicating things further, at satel-

lite velocities relativistic effects become relevant to synchronisation. This is significant enough that the existing GPS satellite positioning network must compensate for relativistic effects to maintain positioning accuracy.

1. Coverage

Currently there are only two satellites with the capability of sending and/or receiving photons through space: the Chinese QUESS, and the Japanese SOCRATES satellites. In order to achieve world-wide coverage multiple communicating satellites will be necessary such that there be a constellation of satellites that are visible at all times, anywhere on Earth, in a similar manner to the GPS satellite network, thereby bypassing line-of-sight constraints, as per Fig. 188.

In the case of the QUESS satellite, the orbit is such that it passes over the observatories once per day, with a visibility window of ~ 275 s. The coverage of a LEO

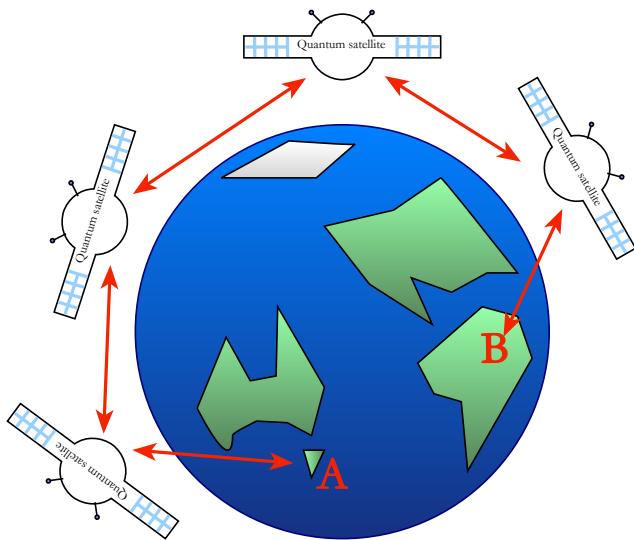


Figure 188: Satellite constellation network for bypassing line-of-sight constraints in long-distance quantum communication. Neighbouring satellites connect via line-of-sight in space, and there are a sufficient number of satellites to enable global coverage such that all points on the Earth's surface are visible to at least one satellite at any given time, giving them indirect quantum access to any other point on the globe. In this simple example, entanglement swapping between satellites shares a Bell pair between Columbia and Tasmania, enabling the teleportation of certain Columbian products from Bogotá to the dance parties featuring at DJ Rohde's holiday home.

satellite of elevation 500km has a maximum radius of $\sim 2,400\text{km}$ from the point where the satellite is directly overhead, a limitation imposed by the curvature of the Earth. In practice this will be reduced, due to engineering imperfections and atmospheric influences, to distances on the order of $\sim 1,000\text{km}$ (see Fig. 186). By placing satellites into higher orbits, coverage could be greatly increased, but this will introduce additional engineering challenges in establishing the more sophisticated satellite links, and effects such as dispersion which equates to loss. To this end, some of the main technologies that will need to be developed further are:

- Larger-sized telescopes with bigger apertures to increase collection efficiencies.
- Better tracking systems, to maximise the overlap between detection apertures and photon spot-size.
- Wave-front correction through adaptive optics (?).

Currently, communications are performed at night due to the laser wavelengths that are employed, but plans for using telecommunication wavelengths pave the way for daytime quantum communication. For obvious reasons, this will be an essential ingredient in a global network of satellites forming a constellation capable of communication between opposing points on the Earth where night and day are reversed.

From the point of view of security, we anticipate that initial quantum networks for QKD will be based on trusted nodes. Due to the inherent practical difficulty of tampering with space-based trusted nodes, it is likely this will offer a sufficient level of practical security for most realistic applications. Thus, using a constellation of satellites employing QKD can achieve long-distance cryptography by simply relaying the information, either by multiple satellite-to-ground or satellite-to-satellite configurations. For ultimate security, not relying on the use of trusted nodes, one would employ alternative protocols that do not require line-of-sight quantum communication, such as the E91 protocol (?). As this requires entanglement distribution, storage, and purification, this would most likely be a second-generation technology after the trusted node QKD network is fully established and successfully demonstrated.

2. High-precision applications

For applications such as quantum clock synchronisation (Sec. XXVII.J) and experiments testing fundamental physics at large length-scales and velocities, it is likely that extremely high fidelities of quantum operations will be required. For example, for clock synchronisation, even current atomic clocks used as national standards have an accuracy at the level of 1 part in 10^{-14} . Thus, such high-precision experiments will almost certainly follow the less demanding QKD experiments.

As the precision of technology improves, other sources of error may appear, which may require correction. It is well-known from existing GPS satellites, that it is crucial to account for relativistic effects, due to time-dilation and the gravitational red-shift. Not accounting for these effects would seriously compromise the GPS system, with inaccuracies equating to an error on the order of 10km/day. This is due to the high velocities of LEO satellites, traveling at speeds of $\beta = 10^{-5}$ times the speed of light. Such relativistic effects can affect entanglement in the presence of diffracting photons (?). Even for single-photon transmission, polarisation-encoded photons can give rise to corrections at the order of β (?) **Not sure what this means? What correction?** Relativistically invariant entanglement distribution protocols have been proposed to avoid such effects, which would otherwise require correction (???).

— Volens et potens.

LXXVIII. QUANTUM GENETIC MEDICINE

“I have all these great genes, but they’re recessive. That’s the problem here.” — Bill Watterson.

“The laws of genetics apply even if you refuse to learn

them.” — Allison Plowden.

UNTIL now, medicine has largely relied on one-size-fits-all diagnosis and treatment options. Needless to say, everyone’s different, manifest in their unique genotype. However, with access to individuals’ unique genetic makeup, the next generation of medicine will become highly personalised, catering for individual genetic differences. People with different genetic predispositions, mutations, or traits will be able to have treatment options tailored to them. Different cancer types, which are genetically distinct from one another, may be genetically targeted.

With the ability to sequence individuals’ genomes extremely cheaply, we will open up entirely new medical possibilities for genetically-personalised treatments – the era of *genetic medicine*, the next revolution in medicine.

Doing so will require highly complex processing of massive amounts of genetic and drug information, with demanding computational resource requirements. We foresee quantum computing as playing a central role in the processing pipeline of genetic drug development, some initial ideas for which we sketch here.

A. The human genome project

The first complete mapping of the human genome was a major scientific achievement (?), opening up entirely new avenues for medical research that were previously never possible. It was, however, an extraordinarily expensive undertaking, costing on the order of \$1b.

Since then, genetic sequencing tools have undergone a massive technological transformation and are now available as commodity hardware at price-points accessible to any well-resourced bioscience lab. This now enables sequencing the human genome orders of magnitude cheaper than the first attempt. It is now possible to map a genome for $\sim \$1,000$ ’s, and less sophisticated consumer-grade handheld devices are even becoming available.

Following its own technological Moore’s Law, one can reasonably anticipate this process becoming sufficiently cheap that in the near future it will become economically viable (and desirable) for every individual to have their own personal genome fully sequenced and available for medical use.

— *Benedictus benedicat.*

1. Short-read sequencing

The major technological transition that has enabled this rapid progress is the adoption of next-generation sequencers (NGS) based on *short-read* technology. Using this process, a DNA sequence is not mapped exhaustively from beginning to end, but rather is chemically decon-

structed into an enormous number of *short-reads* – small genetic segments, each typically on the order of ~ 50 base-pairs in length. Having prepared an enormous pool of such short-reads, they are then sequenced in parallel, yielding a large database of N short strings, corresponding to small segments of the larger genome. The task then is to reconstruct a complete genome from this data.

Typically the end result is not a complete human genome from start to finish, but rather segments that capture regions of interest in the respective genome. In particular, certain genes or single-nucleotide polymorphisms (SNPs)⁷⁸ may be targeted as regions of interest during sequencing.

To achieve this, there are two approaches that are most commonly employed – *de novo* assembly, and *mapping*.

2. *De novo* sequencing

In *de novo* sequencing we treat the short-read data as a jigsaw puzzle that must be reassembled. We define an overlap threshold, n , and upon comparing every string against every other, look at whether their ends overlap consistently by at least n base-pairs. When a match is found, the respective short-reads are merged into a larger *contig*. This process continues until no further sufficiently-overlapping strings are found. At this point we should, at least in principle, have a fully reassembled genome, or at least very large contigs belonging to it (assuming a sufficiently large pool of short-reads to begin with).

Open-source software for implementing *de novo* assembly of short-read data is available. The well-known *Velvet* package (?) does this graph-theoretically using de Bruijn graph (?) representations for contigs/reads, with very efficient computational resource requirements.

The number of comparisons between short-reads scales only as $O(N^2)$, and employing hash table representations for short-reads, the lookups for each individual comparison can be efficiently implemented in $O(1)$ time. The confidence that two contigs/reads actually overlap increases exponentially with n , but with increasing n comes a reduction in the number of matches that will be identified – the tradeoff between *sensitivity* and *specificity* in contig reconstruction.

The *de novo* approach is extremely powerful, as it does not require any genomic reference. Rather, we can reconstruct a genome *ab initio*, with sufficient read data.

Fig. 189 provides a graphic example of *de novo* sequencing.

⁷⁸ SNPs are positions in a genome whose nucleotides may differ between individuals within a species. Of course, the vast majority of the genetic makeup of individuals within a species is identical. Thus, the SNPs are sufficient to characterise the genetic differences between individuals and identify their genotype.

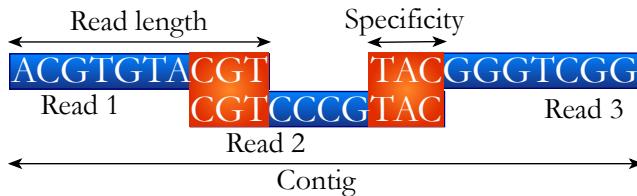


Figure 189: Three short-reads, each of 10 base-pairs in length (blue), which we wish to assemble into the largest possible contig belonging to the underlying genome. The specificity defines the number of endpoint base-pairs that must overlap for them to be considered a match (orange). In this case the specificity is 3, and 3 short-reads have been assembled into a single 24 base-pair contig. The greater the specificity the more accurate will be the matching, but the lower the probability of finding matches. A specificity too low will create false matches, whereas a specificity too high makes finding sufficiently overlapping short-reads unlikely. This makes optimum choice of specificity an important consideration, creating a direct trade-off between reconstructed contig length and accuracy. This sequencing technique is best applied when there is no known reference genome for the data under analysis. For the technique to work effectively there must be a sufficient quantity of short-read data that with high probability overlaps between reads will exist. If the data is too sparse few overlaps are likely to be found. Thus, average reconstructed contig lengths grow with the amount of short-read data available to the engine.

3. Mapping

The failing of *de novo* sequencing is that the pool of short-read data must be sufficiently large that all the pieces in the jigsaw puzzle are present, such that there are no gaps between neighbouring pieces, whereby contig ‘islands’ remain isolated from other contigs. However, resources are of course always finite, as too is the number of short-reads available for reconstruction.

The other approach, *mapping*, is to use an existing genome as a reference. Rather than piece short-reads together, we compare them against this reference to infer their relative locations in the genome. When doing so, we allow some error threshold in the matching. Specifically, we require the Hamming distance between a short-read and an equally-long segment of the reference be below some threshold. The flexibility offered by this threshold accommodates for mutational differences between the reference and the short-reads (i.e the SNPs), which is ultimately what we wish to characterise.

To illustrate the power of this approach, consider the following. Only a minuscule fraction of the genome differs from one individual to the next – almost all of it is identical. Thus, mapping allows us to identify the mutational differences between individuals. Importantly, unlike *de novo* sequencing, short-reads can be mapped to the reference even with incomplete data – contig ‘islands’ needn’t remain so. Non-overlapping short-reads can still be successfully mapped, making this approach applicable (but potentially incomplete) even with datasets insufficiently

large for *de novo* sequencing to be effective.

In the case of cancer diagnosis, for example, a cancer cell’s genetic makeup is virtually identical to that of its host, modulo some small number of mutations that characterise the cancer. Using the mapping approach we can quickly identify these mutations, hence understanding the genetics of the underlying cancer, potentially opening opportunities for genetically-targeted treatment.

As with *de novo* sequencing, open-source software for efficiently implementing mapping using commodity hardware is available (?).

Fig. 190 provides a graphic example of mapping in the presence of a SNP.

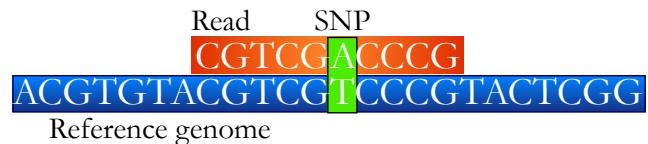


Figure 190: In genome mapping we begin with a known reference genome (blue) to act as a comparison for short-reads (orange). The short-reads are slid along the reference until a location with sufficient overlap is found, at which point we record a match. By allowing some degree of single nucleotide mismatches (green) between the reference and short-read, we are able to characterise the SNPs in the short-read data. This sequencing technique is appropriate when we have a known genome for the species to which the short-read data belongs, but we wish to identify minor genetic mutations between the samples. For example, mapping short-reads belonging to human cancer cells to a human reference genome characterises the genetic mutations associated with that particular cancer. The technique works with any amount of short-read data, even just a single read, provided the reference is known.

B. Genetic medicine

At its simplest, the design of genetically-tailored medical treatments will require understanding the interactions between drug compounds and genetic processes at the molecular level. For example, we might desire that a drug modifies gene expression for specific genes, or the transcription of DNA to specific proteins. In the case of cancer treatment, we may wish to inhibit the function of biochemical processes reliant on genes exhibiting particular cancerous mutations, whilst not affecting healthy, unmutated ones.

The drug design process for conventional medicines is an incredibly tedious one, requiring massive latitudinal and longitudinal studies on the effects of drugs on physiological symptoms. It is not uncommon for the entire drug development process, from design, to testing, to approval, to cost on the order of $\sim \$100m$ s for a single drug. However, in the case of individually tailored medicine such studies are clearly not possible. This will require replacing human studies with accurate biomolecular-level

simulations of these chemical processes. At this scale, interactions are inevitably quantum mechanical in nature, which must be accommodated for in simulations.

The key goal is to properly simulate the interaction of candidate drug compounds, ligands, or functional groups with biochemical processes at the molecular level. Classical techniques for such simulations include estimating electron densities using density functional theory (the potential energy surface) of compounds, from which useful properties such as bonding affinities may be estimated.

C. Quantum chemistry

Despite the existence of classical estimation techniques, molecular-level interactions are necessarily quantum mechanical in nature. For this reason, classical simulation techniques are limited to approximations that ignore realistic and important quantum effects, since this would require exponential classical computational resources in general (i.e in general, simulating quantum systems is **BQP**-complete).

Using standard quantum chemistry techniques, one can construct accurate Hamiltonians describing the evolution of molecular systems at the quantum level. Efficiently implementing such simulations can then be performed on quantum computers using standard, efficient Hamiltonian simulation quantum algorithms (Sec. XLI.F) for simulating the Schrödinger equation,

$$\hat{H}|\psi(t)\rangle = i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle, \quad (78.1)$$

where \hat{H} is the system's Hamiltonian, describing the time-dependent evolution of the state as,

$$|\psi(t)\rangle = e^{-\frac{i\hat{H}t}{\hbar}} |\psi(0)\rangle. \quad (78.2)$$

These Hamiltonians can be constructed *ab initio*, combining relevant one- and two-body interactions, such as Coulomb potential, kinetic energy, spin- and spin-orbit coupling, and magnetic and electric dipole terms appropriately. A net interaction Hamiltonian may be obtained by summing over the different one- (i) and two-body (j, k) terms for all particles comprising the system,

$$\hat{H}_{\text{int}} = \sum_i \hat{H}_i + \sum_{j>k} \hat{H}_{j,k}. \quad (78.3)$$

This structure is of the form of Eq. (41.31) and thus lends itself to the quantum Hamiltonian simulation algorithm (Sec. XLI.F).

The dominant one-body interactions are kinetic energy terms, of the form,

$$\begin{aligned} \hat{H}_i^{(\text{kinetic})} &= \frac{\hat{p}^2}{2m} \\ &= -\frac{\hbar^2}{2m_i} \nabla_{\vec{r}_i}^2, \end{aligned} \quad (78.4)$$

where \hat{p} is the momentum operator, m_i is mass, \vec{r}_i is the position vector, and,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \quad (78.5)$$

The dominant two-body terms are Coulomb (electrostatic) interactions,

$$\hat{H}_{j,k}^{(\text{Coulomb})} = \frac{q_j q_k}{4\pi\epsilon_0 |\vec{r}_j - \vec{r}_k|}, \quad (78.6)$$

where q_j denotes charge.

D. Quantum drug trial simulations

These approaches borrowed from quantum chemistry allow simulation of the interaction between drug compounds and molecular biochemical interactions. The next step is to perform this process against a huge library of candidate drug compounds, ligands or functional groups, from which, we hope, some candidates will exhibit desired characteristics, such as strong bonding affinities.

To achieve this, let us construct a classical algorithm for exhaustively constructing and enumerating organic drug molecules or functional groups, up to some cutoff size. The length of this list grows exponentially with the cutoff size. Next we implement this algorithm unitarily (which is always possible, writing the classical circuit as a reversible one), and construct it as a quantum oracle such that every input state (which acts as a pointer to an element in the list) yields as output a qubit representation of the corresponding drug compound. Specifically, the oracle implements a transformation of the form,

$$\hat{U}_{\text{oracle}} |i\rangle_{\text{pointer}} |0\rangle_{\text{drug}}^{\otimes n} \rightarrow |i\rangle_{\text{pointer}} |\psi_i\rangle_{\text{drug}}^{(n)}, \quad (78.7)$$

where i is the index (pointer) to the drug element in the enumeration, and n is the number of qubits encoding the representation of the drug molecule⁷⁹ $|\psi_i\rangle_{\text{drug}}^{(n)}$.

This oracle may be constructed by writing a classical algorithm that generates a search tree for exhaustively enumerating physically allowed drug compounds via brute-force. Writing this as a reversible quantum circuit then yields the oracle, where the output pointer register is entangled with the output drug description register, uniquely mapping them to one another.

Having constructed this oracle, we feed the output qubit molecule description into the Hamiltonian simulation algorithm, with exponential speedup, yielding as output a ‘score’ characterising some desired property of the interaction, such as bonding affinity. This implements the

⁷⁹ This may be implemented unitarily, which is easy to see via its preservation of orthonormality.

simulate-and-score operator,

$$\hat{U}_{\text{sim}}|\psi_i\rangle_{\text{drug}}|0\rangle_{\text{score}} \rightarrow |\psi_i\rangle_{\text{drug}}|s_i\rangle_{\text{score}} \quad (78.8)$$

The joint oracle→simulation algorithm is finally embedded within a quantum search subroutine, which searches over the input space enumerating the set of drug candidates for scores above some desired threshold. This yields a quadratic enhancement in the number of drug candidates that can be simulated in parallel. Successful search results, where a compound achieves the threshold score, are then tagged for further investigation as treatment drug candidates.

Combining the Grover search and quantum simulation subroutines, the total time-complexity of the entire (unencrypted) algorithmic pipeline is,

$$O\left(\frac{N\sqrt{Mt^2}}{\epsilon}\right), \quad (78.9)$$

where t is simulation time, ϵ is the simulation accuracy, N is the number of local terms in the global Hamiltonian, and M is the number of drug compound candidates in the algorithmic drug search space.

The complete algorithmic pipeline is shown in Fig. 191.

E. Cyborgs – The era of genetic editing

In very recent years, new technologies have emerged for performing highly targeted gene editing, allowing genetic manipulations to be performed on living organisms. Using CRISPR technology, for example, one can enable and disable entire functional genes, or perform even more intricate manipulations, right down to the single nucleotide level.

This exquisite control over the genomes of living organisms, something that would never have been dreamed possible only years ago, opens up our ability to enhance or suppress genetic characteristics of species, potentially facilitating everything from the development of highly improved agricultural crops, all the way to more ethically questionable motives such as genetic enhancement of the human race.

Using the genetic medicine pipeline discussed previously one might envisage even more ambitious goals in the era of genetic editing. Rather than test algorithmically-generated databases of drug compounds for their genetic effects, one might instead trial genetic variations for their outcomes at a molecular biological level. This might provide guidance and insight into genomic variations worth experimenting with for desired practical outcomes.

F. In the cloud

The quantum genetic medicine processing pipeline lends itself well to several important cloud-based protocols. The

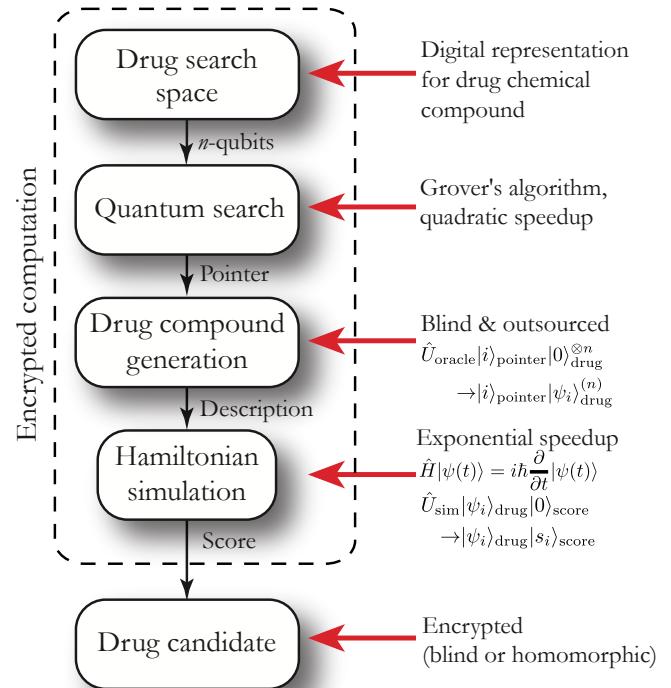


Figure 191: Pipeline for quantum-enhanced genetic drug development. A quantum search algorithm searches over algorithmically-generated drug compounds or functional groups, for each performing Hamiltonian simulation, to determine if its score is above a threshold. The quantum search subroutine yields a quadratic enhancement in the number of drugs that can be simulated in parallel, and the Hamiltonian simulation subroutine yields an exponential enhancement over classical simulation techniques.

pipeline comprises several distinct subroutines, which might be outsourced or distributed independently of one another.

An R&D organisation might specialise in the algorithmic generation (or hard-coding) of candidate drug compounds, which they would like to retain as a trade secret, but desire to license to third-party drug designers, who wish to employ them in their computational pipeline. The quantum search and Hamiltonian simulation subroutines might similarly be outsourced or distributed over the cloud to vendors specialising in the implementation of these particular subroutines.

In an era where the genetic composition of every individual is fully characterised, data security will be of utmost importance – medical confidentiality will be lifted to an entirely new plane when people’s genes are at stake, as the nefarious uses (and misuses) for obtaining other people’s genomes are immense. This provides a perfect example for the value of encrypted quantum computation (Sec. XLIV). If a medical lab is outsourcing some aspects of a computation involving clients’ genetics, it is paramount that this be obscured from third-parties performing the computations in the interest of medical

confidentiality. Computational efficiency aside, this consideration on its own already justifies implementing such simulation pipelines quantum mechanically, as encrypted classical computation is effectively unviable in general.

— *Annus mirabilis.*

LXXIX. THE QUANTUM MIND™ – QUANTUM MACHINE LEARNING

“If computers get too powerful, we can organise them into committees. That’ll do them in.” — Unknown.

“My unshakeable belief in the stupidity of the animal man has never disappointed me and has often helped me throughout my life.” — Arthur Schopenhauer.

— by Nana Liu.

Tidying, editing, proofing, structure. Section titles.

In recent years, the implementation of artificial intelligence via machine learning techniques has become a forefront area of research, transforming many software industries and those that depend upon them. An obvious question to ask is to what extent machine learning techniques translate across to the quantum computing environment, and whether upon doing so there is an associated quantum enhancement. While this field of *quantum machine learning* (QML) is far too extensive to summarise here, we will instead present a brief overview of how machine learning might apply in a networked quantum environment.

There are two distinct avenues to consider here. First, how does machine learning influence the operation of quantum networks? Second, how do quantum networks affect the implementation of machine learning? We will outline some of the key considerations here from both these opposing angles of consideration.

A. Overview

Today, classical machine learning is affecting the understanding and better regulation of the classical networking infrastructure underpinning the modern internet. This includes network pattern recognition, security and fault management, routing and traffic management, resource management, and distributed computation. In these applications, it may not solely be processing power that is of importance, but additionally reliability and security may also be paramount.

The large intersection between machine learning and network systems is perhaps unsurprising. Firstly, machine learning relies on access to data, and in many real-world scenarios, data naturally emerges from distributed sources.

Secondly, especially for complex systems like large networks, the information to process is complex, containing many uncertainties, and subject to errors. Exactly solvable models in these regimes are rare, a scenario where estimation techniques based on machine learning are often helpful.

In the coming quantum era we can envision three distinct ways where quantum resources might be introduced: quantum communication; quantum processing at individual nodes; and, data that is inherently quantum in nature. To begin, we first make a classification of the four foreseeable network types (summarised in Tab. XXVII). These can be classified as:

- CC: classical data and processing over a classical network (e.g the present-day internet).
- CQ: classical data and processing over a quantum network.
- QC: quantum data and processing over a classical network.
- QQ: quantum data and processing over a quantum network – a fully-quantum internet.

Now the inevitable question arises: how do these different scenarios relate to machine learning?

This is not yet an active research area in its own right. However, there are some preliminary toolkits that are starting to be developed in the new field of QML. By first summarising the intersection between the classical internet and classical machine learning, we gain insight into the kinds of tools required to start examining their quantum counterparts. For instance, we will see how aspects of quantum processing of quantum data might be aided by machine learning, how machine learning may be enhanced by quantum resources, and how machine learning may be implemented in these distributed quantum settings. A summary is presented in Tab. XXVIII.

B. Classical machine learning in classical networks

To create and maintain efficient classical networks, one requires efficient and reliable processing at individual nodes, security, efficient routing and data transmission, efficient use of resources, and a means for distributed processing (??). We now present a brief overview of how machine learning techniques apply in these areas.

1. Machine learning basics

Machine learning algorithms allow us to make predictions about a current or future dataset without requiring explicit instruction on how to do so. Since the aim is directed more towards *prediction* than purely *estimation*,

Data/Network	Classical	Quantum
Classical	Current internet (CC)	Classical data in quantum network (CQ)
Quantum	Quantum data in classical network (QC)	Fully quantum internet (QQ)

Table XXVII: Classical and quantum data in a network.

Network concerns	Classical data	Quantum resources and classical data	Quantum data
Individual computing	Classical machine learning	Quantum-enhanced machine learning	Quantum learning
Security and faults	Machine learning for anomalies and faults: detection and prediction; Adversarial machine learning	Adversarial quantum machine learning	Anomaly detection and change point detection for quantum data
Routing and traffic	Machine learning for traffic prediction, classification, congestion control and routing	Open problems	Open problems
Distributed computing	Distributed machine learning	Distributed quantum machine learning	Distributed quantum learning
Communication	Data compression and machine learning	Open problems	Data compression and quantum machine learning

Table XXVIII: Classical and quantum machine learning applications in classical and quantum networks. Almost all of the categories here are very new and open to exploration in the quantum domain.

it differs from the field of statistical estimation, although they share many techniques.

There are three main paradigms for machine learning:

- Supervised learning: relies upon training data from which inferences and predictions about new test data can be extracted.
- Unsupervised learning: makes inferences from the data at hand without training.
- Reinforcement learning: operates using a different framework, and aims to find the best action to take to maximise a given reward in a particular environment.

Machine learning is used regularly for data collection, feature engineering, and model learning. There are many excellent introductory texts on this topic (?????).

2. Security & fault management

There are two primary ways in which machine learning applies to managing security and faults in networks. The first is *using* machine learning techniques to predict and detect security breaches and faults in the network, including anomaly detection. The second is in studying the security vulnerabilities of machine learning algorithms themselves, as the presence of adversaries is natural in real-world networks – so-called ‘adversarial machine learning’.

a. *Anomaly detection & fault management* When there are security breaches in a classical network, one desires the

ability to predict and detect them, as well as a method for making protocols more robust against them. Machine learning is often used in anomaly and intrusion detection. These algorithms seek out unusual data or changes within it.

Broadly, there are three classes of anomalies: point, contextual, and collective, which refer respectively to single datum anomalies, unusual data with respect to a specified context, and clusters of data which suggest unusual behaviour. Both supervised and unsupervised algorithms are employed in these settings (??). One of the prime challenges here is determining the presence of anomalies when limited data is available, and the associated rates of false identifications.

Fault management in a network is also extremely relevant, especially for complex networks more exposed to errors. We desire the prediction, detection and localisation of faults. Most applicable machine learning methods here employ supervised algorithms. However, the paucity of real training data (as opposed to synthetic data generated via simulation) means that algorithms might be poorly trained, especially in newly established networks (??). This is to be reasonably anticipated with the deployment of a future quantum internet. To accommodate for this, new methods have arisen where unsupervised machine learning techniques are used instead to detect changes in the network rather than relying on labelled fault data (?).

In particular, to identify and localise unusual network behaviour, either due to natural faults or adversaries, network anomaly detection methods can be employed (??). Since results can be sensitive to the employed training data, it is important to examine which datasets are most appropriate for a given application (?). Particularly, there

have been many proposals for utilising anomaly detection in network intrusion. However, this approach has been criticised for its use in real-world scenarios, where it's often difficult to distinguish anomalies related to intrusions from those attributed to other factors, and the complexity of real-world networks may make it too difficult to define what even constitutes a 'normal' signal (?).

b. Adversarial machine learning Machine learning algorithms themselves exhibit security vulnerabilities (?). There are two main types of attacks to which they are vulnerable:

- Evasion: directed at the test data.
- Poisoning: directed at the training data and machine learning models.

In real-world scenarios, data often originates from different sources, making adversarial attacks more likely. It has been discovered that many machine learning algorithms are in fact vulnerable to adversarial attacks, the first discovered in (?). A large proportion of the literature focuses on the details of specific algorithms: the detection of adversaries; their different methods of attack; and, the particular defences against them (?). However, recently, more foundational work has emerged, explaining the origins of this vulnerability as arising from the high dimensionality of the underlying data (???).

3. Traffic management & routing

The effective operation of large-scale networks requires automated management protocols. This includes efficient means for traffic prediction, traffic classification, routing, and congestion control. Machine learning algorithms have been developed for all of these.

Predicting network traffic is becoming increasingly important, especially in diverse and complex networks. This is commonly addressed using time-series forecasting (TSF) methods. This can make use of either statistical analysis techniques, or supervised machine learning methods (???). Non-TSF methods also exist (??).

The most commonly used technique for traffic classification is the so-called flow feature-based technique. This takes into account information about unidirectional packet transmissions. Here, supervised machine learning techniques have been found to be accurate. However, unsupervised techniques have been found to be more robust. Their joint application is a very powerful tool (??).

Machine learning is most applicable to dynamic routing problems, requiring rapid updating of optimal routes. Since such settings require frequent reevaluation, reinforcement learning algorithms are most appropriate. In particular, Q-learning has performed well in various networks (???).

Network congestion control is important to ensure stability and the minimisation of packet loss. Well-known congestion control methods like queue management already exist. However, machine learning can be used to enhance the effectiveness of congestion control in various scenarios, especially for packet-based TCP/IP networks (???).

4. Distributed machine learning

Distributed machine learning is simply the fusion of distributed computation with machine learning, where the learning algorithm is distributed across a network. This becomes highly relevant in several notable scenarios:

- Training and/or testing data originates from different sources. This is the naturally distributed setting.
- There is too much data to store locally on a single device.
- When fault-tolerance becomes important (e.g for high-value data), decentralised storage provides enhanced data integrity.

The toolbox and infrastructure for distributed machine learning is rapidly developing, and there are many known algorithms (??). Existing platforms catering for distributed machine learning include MLbase (?), Hadoop (?), and Spark (?).

Caution is required, however, as there are cases when one *shouldn't* employ distributed machine learning, such as when:

- Communication and synchronisation between distributed parties presents a bottleneck for computation.
- Developing and executing distributed software is too complicated.
- One can run the same algorithm on a multi-core machine. This is possible with smart data-sampling, offline schemes, and efficient parallel codes.

C. Machine learning on classical data with quantum resources

There are at least three broad ways in which we can employ quantum resources for classical data over a network, specifically they:

- Enhance data-processing at individual nodes.
- Improve security.
- Enhance communication.

In a classical network, the first question is whether or not quantum resources can assist in any of the relevant algorithms. These belong to the class of quantum-enhanced machine learning algorithms.

In a quantum network with only classical data, communication complexity improvements are possible (?). It's unclear whether machine learning has utility in this setting, although there are some promising hints in this direction (??).

1. Quantum machine learning overview

Quantum(-enhanced) machine learning (QML) algorithms are quantum algorithms performing machine learning tasks, exhibiting super-classical enhancements. They have so far mostly concentrated on quantum speed-ups with respect to the dimensionality of the underlying data.

a. Fully-quantum algorithms The first of these algorithms relied on fully quantum devices, maintaining coherence throughout computation, requiring full fault-tolerance. For supervised learning algorithms claiming exponential quantum enhancement (?), the HHL algorithm (?) for matrix inversion is often employed. However, HHL exhibits a number of shortcomings, making it impractical for near-term quantum devices:

- The ability to efficiently encode classical data into quantum states and memory (?).
- Effective quantum state read-out (?).
- They generally require high circuit-depth.
- There are restrictions on the sparsity and conditioning of the matrices to which the algorithm is applied.

Although subsequent developments have attempted to circumvent sparsity restrictions, and rather focus on low-rank matrices (e.g quantum principal component analysis for low-rank matrices (?)), recent work on quantum-inspired classical algorithms has demonstrated that previously undiscovered, efficient classical algorithms can exist (??). In fact, classical sampling methods (?) for quantum-inspired machine learning algorithms suggest that classical methods for linear algebra problems in low-dimensions are likely to have efficient classical algorithms. Although these classical sampling methods are not yet more practical than existing classical sampling methods, they are still more practical than their quantum counterparts.

Another set of approaches, relying on amplitude amplification and Grover's search algorithm, can provide up to quadratic runtime enhancement. These include quantum algorithms for reinforcement learning (?), and training of

quantum perceptrons (?). While theoretically appealing as long-term objectives, viable near-term proposals are absent.

b. Hybrid algorithms To find algorithms practically viable in the the near future, research is increasingly devoting its attention to hybrid classical-quantum algorithms. These algorithms, which include variational methods for optimisation (?), exhibit low circuit-depth, where the optimisation process is performed iteratively and classically. There are roughly two varieties: one that attempts to enhance classical algorithms with classical input data; and another where the quantum advantage lies in efficient quantum state preparation, thus relying on quantum input data. Prominent examples of the former include quantum approximate optimisation algorithms (QAOA) (?), and the latter includes variational quantum eigensolvers (VQE) (?), which we return to in the subsequent section.

Both QAOA and VQE can be considered as belonging to the same broader framework, and their optimisation component (which may be considered only as a component, not the entirety of machine learning) is performed classically. One begins with an ansatz quantum state. A unitary operation with classically-tuneable parameters is then applied to this state, and an observable whose expectation value represents the problem's cost function is subsequently measured. The classical parameters of the unitary are then iteratively adjusted until cost function minimum is reached (i.e a Hamiltonian ground state), for instance using the classical gradient-descent algorithm.

In QAOA, the respective ground state encodes the classical solution to a classical optimisation problem, like MAXCUT, exhibiting efficient polynomial runtime. Thus, it is not a quantum-enhanced algorithm for a classical machine learning problem, but rather exploits a classical machine learning algorithm. It remains to be seen if optimisation problems more directly relevant to networking applications can be solved in this way.

Alternate frameworks have been developed to find quantum-enhanced algorithms that not only take advantage of classical optimisation algorithms, but also enhance classical machine learning algorithms. These new proposals include quantum circuit learning (?), quantum generalisations of neural networks (?), and Born machines (?). Theoretical demonstration of quantum enhancement in such settings remains an important open problem.

2. Security & other applications

a. Anomaly detection The chief machine learning method for detecting and averting faults and security breaches in classical networks is in anomaly detection. However, for anomalies in classical data, it appears unlikely that currently available QML algorithms can enhance detec-

tion speed or reliability. One of the primary reasons is the necessity for encoding classical data into quantum states, which can be very costly (?). Thus, even if there are QML algorithms for anomaly detection in the computational stage of processing, state preparation and readout overheads may be prohibitive. However, this is no longer the case if we instead begin with quantum data, to be discussed in Sec. LXXIX.D.

b. Adversarial quantum machine learning Just as classical machine learning algorithms are vulnerable to attacks, so is QML. This is a very new field, known as *adversarial QML*. As with adversarial machine learning, the aim is to find more robust QML algorithms, and some robust algorithms have indeed been proposed (?). In addition to finding more robust algorithms, it's also important to understand the respective limitations on robustness, currently an open problem. A recent result suggests that the same quantum resource requirements may be necessary for detecting adversaries in higher dimensions as compared to quantum tomography (?). Thus, it remains unclear what the total resource cost of QML is in the presence of adversaries. However, there is the tantalising yet unexplored prospect that quantum resources may enhance the security of machine learning algorithms, in a similar way that information-theoretic security is afforded by quantum cryptographic protocols.

c. Other applications Whether or not there exist helpful QML techniques for traffic and routing management is currently very unclear, and may even appear unlikely. There may be some quantum-enhancements for machine learning algorithms applicable to traffic and routing management. However, the obstacle of efficient quantum encoding/decoding of classical data remains. The no-cloning theorem forbids state replication, and in general the overheads associated with encoding classical information into quantum states are very high (??), potentially outweighing any computational gain.

3. Distributed quantum machine learning

The motives for considering distributed QML are similar to those for distributed classical machine learning. Suppose one wishes to perform distributed machine learning, either because the given data is naturally distributed or there is limited processing power on any given device. Then there are existing protocols for implementing general distributed quantum algorithms that might be helpful in delegating QML algorithms (?).

Secure delegated quantum computational protocols (?) can also be modified and applied to QML (??). However, the same problem with state preparation could persist,

for the server rather than the client. Alternatively, hybrid classical-quantum algorithms for distributed QML have been devised (?). Here, the quantum state preparation assumptions can be obviated by using a hybrid gate that takes in classical input data and implements classically-controlled unitary evolution.

D. Machine learning with quantum data

Suppose our data is inherently quantum, in the form of quantum states or channels – *quantum data*. We might additionally face restrictions in the number of copies we have access to, imposed by the no-cloning theorem.

In these cases, it has been found that classical machine learning methods may be helpful over traditional methods in dealing with quantum data. Another approach is to use quantum protocols to directly process quantum data. Learning protocols in the latter case belong to the field of quantum learning.

It is possible to process quantum data over both classical and quantum networks. Techniques from classical machine learning for quantum data may assist in the communication of quantum data over classical networks, while quantum learning protocols may be more appropriate over quantum networks. This is an exciting new research direction, as it is presently unclear whether such methods find utility.

1. Classical machine learning for quantum data

a. Tomography For classical processing of quantum data over a classical network, the first step is to find its classical description. The canonical methods for this are quantum state tomography and quantum process tomography. However, tomography is in general extremely resource intensive. Recent work has provided efficient methods for state tomography using classical machine learning techniques (??).

b. Separability While tomography provides complete classical descriptions for quantum data, sometimes it may be sufficient to first classify data in terms of quantum characteristics. For instance, methods for classifying quantum states directly in terms of separability have been devised using classical machine learning (??). Here there are empirical demonstrations of some advantage compared to the CHSH inequality. However, accumulating sufficient training data may still remain problematic for higher-dimensional states.

c. Automated experiment design In a future quantum internet, it is desirable to find optimal methods for generating

inter-node entanglement. It's also desirable for this process to be automated. Recently, such automated methods based on classical reinforcement learning (?) have been proposed to experimentally create a variety of entangled states, providing an exciting starting point for automated design in future quantum internet protocols.

d. Variational quantum eigensolvers We saw that variational quantum eigensolvers (VQE) rely on classical optimisation. When applied to quantum data, they have found success mostly in quantum chemistry (??). In the context of quantum networks, the most promising developments are perhaps in its applicability to quantum data compression (?), which may improve quantum data communication.

2. Quantum learning protocols

a. Template matching The first quantum algorithms for processing quantum data most relevant to machine learning were quantum template-matching algorithms (??). These are classification algorithms, where each class is represented by a quantum state: a ‘template’. The task is to find the class to which a given test quantum state belongs, where this state is not identical to any of the template states. It is unclear whether quantum template matching is directly applicable to quantum networks. However, the ideas introduced provide the key foundations for supervised learning of quantum data, which can be used in the quantum counterparts to supervised algorithms in traffic prediction, classification, and anomaly detection.

b. Learning quantum processes Suppose we want to transmit just enough information about a quantum process over a quantum network in order for the other parties to replicate it. For quantum data, we don't have access to the classical description *a priori*. Instead, we are only allowed to query the process a finite number of times. For a unitary operation, this problem is addressed in (?), in a problem called the *quantum learning of unitary operations*. A very interesting observation here is that the optimal strategy is semi-classical rather than fully-quantum, meaning it's sufficient for the classical data encoding the estimation of the unknown unitary to be stored. It remains an open question as to whether this extends to more general quantum processes.

c. Security In a future quantum network communicating quantum data, it becomes important to detect unusual behaviour in the incoming data-stream. These may present the first signs of a security breach or fault in the network. For dynamic time-series data, this is addressed by change

point detection. This has been extended to the quantum domain (??), where the optimal methods for detecting changes in quantum data are found using methods from state discrimination. For static data, anomaly detection methods based on machine learning become more appropriate as the definition of unusual behaviour is based on *a priori* training data. Classical anomaly detection algorithms have been applied to quantum data for the purpose of error detection (?), in the case where the classical description for quantum data is known. However, for cases where this classical description is unknown (as expected over a quantum internet), it is instead far more efficient to directly apply quantum algorithms. Examples of this include several quantum algorithms for anomaly detection (?).

— *Grandescunt aucta labore.*

LXXX. OPTIMISING THE WORLD

“*Premature optimisation is the root of all evil.*” — Donald Knuth.

EVERY aspect of our modern world is heavily optimised to improve efficiency, affordability, profitability, and every other kind of desirability. Even long before the digital revolution, which enabled the widespread implementation of modern optimisation-theoretic techniques, humans have always striven to make life as easy as possible – the least effort for the greatest possible reward. Our modern application of optimisation theory is merely a logical extension of this to the many other facets of the 21st century economy that now exist.

A. Optimisation in everyday life

Broadly speaking, optimisation problems can all in some sense be thought of as *resource allocation* problems – given finite resources, that we would like to utilise most efficiently and for the greatest reward, what is the best way to allocate and distribute them within a complex system?

To lay the context for the remainder of the section, and illustrate the importance and transformative potential for improved optimisation approaches, we summarise several familiar everyday applications for optimisation theory that we all rely on, usually unwittingly, executed somewhere in the background up in the cloud, the end user oblivious to the gargantuan calculations being performed behind the scenes merely to update a barely noticeable on-screen logo. There are countless more examples than we can't possibly have the time and space to comprehensively summarise here.

The examples we present are all examples of *satisfiability problems* – the problem of trying to simultaneously

satisfy a large number of potentially competing constraints in a system, with the best possible outcome.

It is well known that the most complex such satisfiability problems are **NP**-complete in general, with no known efficient classical solutions (in fact even very trivial constructions of satisfiability problems are already **NP**-complete, e.g see Fig. 193).

Here are a mere few notable everyday applications for optimisation theory.

1. Traffic networks

Given our highly interconnected road networks, what is the best route to take to get home, given the complex and often unpredictable dynamics of road traffic? And how does that answer change when potentially millions of travellers are simultaneously competing to minimise their individual travel times?

To implement a straightforward **GREEDY** strategy, a simple application of Dijkstra's shortest path algorithm (Sec. XI.B) will yield the optimal route for an individual user – and very quickly too since it has only $O(n^2)$ complexity (i.e a **P** algorithm)!

However, in such multi-user systems, with so many users, **GREEDY** strategies, based on optimising users individually, are highly sub-optimal, and globally optimised algorithms must be pursued. For example, if there are a thousand drivers competing to get from Town Hall to Redfern, it makes no sense to optimise them individually, since then they will all be identically directed down George St, which will of course quickly saturate and congest. Obviously load-balancing across the various routes benefits everyone collectively *and* individually.

The complexities of a global multi-user optimisation are intuitively evident – with countless competing interests, finding a set of routes that minimises net transit time yields an enormous space of possibilities and variations to consider, far beyond individual optimisations performed independently. Alas, Dijkstra's tempting algorithm is not very useful, and we must employ harder algorithms, many of which are **NP**-complete. Some of these were introduced in the context of network packet routing algorithms in Sec. IX.D, a conceptually identical problem to road traffic routing.

Fig. 192 provides a simple example illustrating how individual local optimisations can yield suboptimal routing, with a global optimisation yielding a better overall outcome, both collectively and individually.

2. Public transport scheduling

Given a fixed network of train lines, and a fixed number of trains, but the ability to schedule them freely, how

does one schedule a roster that minimises average waiting times?

This might seem straightforward in simple test-cases. As before, optimising a single route, or several independent routes is trivial. But once complex, conflicting interdependencies are in place, conflict resolution becomes impossible to eliminate. Scheduling the Epping train to leave 5 minutes earlier will allow its passengers to catch the 5pm Hornsby train. But by leaving those few minutes earlier, passengers arriving from the Blue Mountains train will miss their connection and have to wait for the next one.

We are once again in a situation where we are overwhelmed with exponentially growing combinatorics to try and minimise the countless possible schedule conflicts that can occur. It is obvious upon inspection that finding a global optimum to this problem will not be possible via independent local optimisations, which do not accommodate for competing interdependencies.

The importance of this problem is obvious. Minimising the resources required to operate public transport effectively could save enormous amounts of money in state budgets. Not to mention, passenger waiting time is of value too. If a million passengers lose just a few minutes of productivity per day to increased commute times, this can amount to billions of dollars a year in lost productivity to the broader economy.

3. Economics

Governments inevitably have finite budgets⁸⁰, and must therefore allocate tax revenue optimally. Public spending is associated with many of the same combinatorial difficulties as the previous examples.

The countless sectors of the economy are all interdependent with one another, and raise their own demands (constraints), which must be mutually satisfied. A minor misallocation of funding to one project could have flow-on effects to other interdependent programs, yielding another resource allocation nightmare.

Of course, these ideas don't just apply to government budgets, but also budgets at smaller scales, say within a large corporation with many distinct spending programs.

4. Supply chain networks

The modern economic infrastructure for the production and distribution of goods is built upon supply chains – complex, interdependent networks of the exchange of resources between entities as goods pass through their many stages of production before reaching the consumer.

⁸⁰ Except in Venezuela.

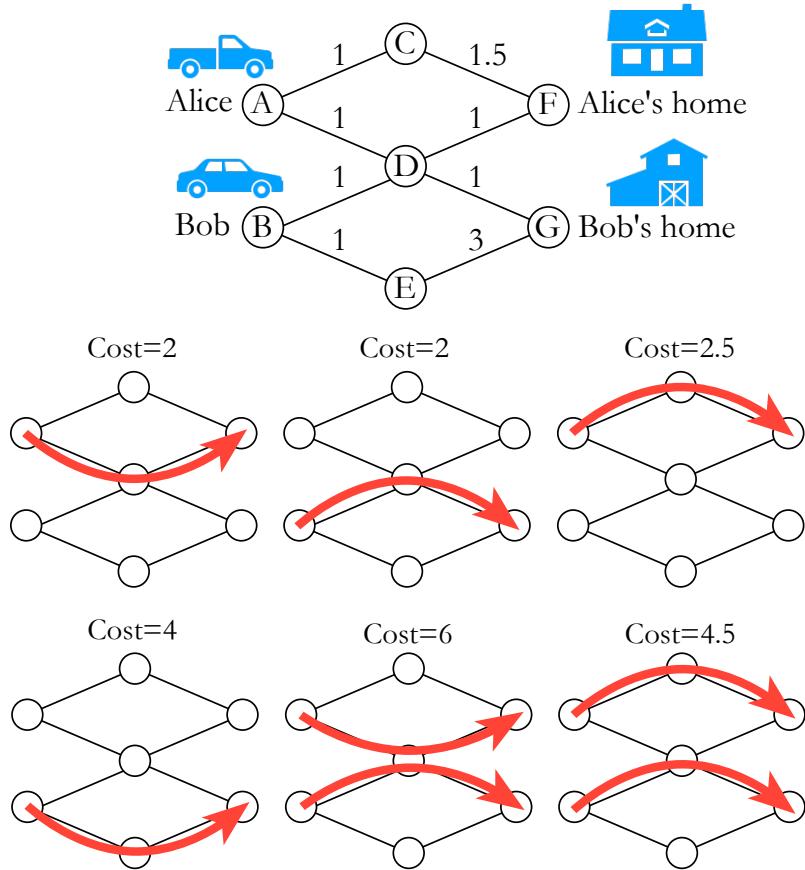


Figure 192: A simple road network, where vertices represent intersections and edges represent roads. Edge weights indicate their associated travel time. The goal is for Alice and Bob to both drive home optimally, i.e. minimising net commute time. Let us assume that the intersections do not delay single cars, but if they become congested (i.e. more than one car simultaneously) they induce a time delay of 1 on all drivers present at the intersection. If Bob were to stay at work, and only Alice was attempting to drive home, her optimal route would be $A \rightarrow D \rightarrow F$, with a commute time of 2. Conversely, if Alice stayed at work and Bob was driving home, his optimal route would be $B \rightarrow D \rightarrow G$, also with a commute time of 2. However, if both Alice and Bob were driving home at the same time there would be congestion at intersection D , and their combined travel time would be 6, not the intended 4. Thus, by optimising their routes independently of one another their competition for intersection D would penalise both of them by 1. On the other hand, a global optimisation would show that a more optimal routing would be for Alice to take $A \rightarrow C \rightarrow F$ (time 2.5), with Bob remaining on $B \rightarrow D \rightarrow G$ (time 2). This would penalise Alice by only 0.5 for taking the longer road, but both Alice and Bob would no longer be penalised at the congested intersection, reducing their collective transit time to 4.5, a saving of 1.5. Note that Alice and Bob are not only better off collectively (i.e. their joint travel time), but also individually. That is, Alice making the sacrifice of taking the longer route home actually benefits her individually, with an individual saving of 0.5. In this trivial example there is only one point of conflict in the network. In general, as the number of users scales up, the distinct ways in which combinations of conflicts could emerge grows exponentially.

Each exchange of resources will typically be subject to its own constraints, such as strict time of arrival demands, or age for perishable goods. Effectively this yields a massive instance of a complex satisfiability problem, where all, or as many as possible of the demands of the units in the chain must be simultaneously met. Conceivably, a single unsatisfied demand may break the functionality of the entire supply chain!

Bearing in mind that modern supply chains may be dealing with billions of dollars in resources at any given time, even minor improvements to their optimisation could be of enormous monetary value.

B. Classical optimisation techniques

Classical techniques for attacking optimisation problems, such as the ones presented above, typically come in the following flavours:

- Heuristics: efficient algorithms that find suboptimal, but satisfactory *approximations* to the solution to a problem. These algorithms are fast and can be efficiently implemented classically, but can have highly variable accuracy in the solution they provide.
- Brute-force: use raw computational power to exhaustively work through all the combinatorics of

a problem. This is guaranteed to find the optimal solution, but is typically extremely slow and limited to small instances of the problem.

- Efficient classical algorithms: in the most ideal scenario, we may not need either of the above, since efficient classical optimisation algorithms may exist. Perhaps the best-known example of this is *linear programming*, whereby the relationships between entities in a system are defined via linear transformations. This class of problems is classically efficient to solve exactly. Clearly this won't solve **NP**-complete optimisation problems, but there are nonetheless many problems of interest in this category.

These techniques present us with a tradeoff between the optimality of a solution and the computational resources required to obtain it. Can quantum technology improve this tradeoff?

C. Quantum enhancement

We now provide a non-comprehensive overview of some of the better-known quantum optimisation algorithms. The speedups offered by these algorithms vary enormously, from simple quadratic enhancements all the way to exponential enhancement.

1. Satisfiability problems

Alas, satisfiability and **NP**-complete problems are not believed to be efficiently solvable on quantum computers. This rules out a quantum future where the world's resources are perfectly allocated and optimised.

However, such problems can be *quadratically* enhanced in runtime by cunningly employing Grover's search algorithm (Sec. XLI.B). To see this, note that all **NP**-complete problems can be efficiently mapped to one another with polynomial resource overhead (see Fig. 1). Thus, we can restrict ourselves to considering the satisfiability problems discussed above – the archetypal **NP**-complete problems. An example of the 3-SAT problem, which is **NP**-complete, is shown in Fig. 193.

By defining an oracle that implements a polynomial-time algorithm on n qubits, a Grover search over the input space of $O(2^n)$ configurations will determine the satisfying input to the oracle for a given desired output, which acts as the tagged element within the search algorithm. The Grover search yields a quadratic speedup for this search compared to a brute-force classical search, therefore requiring only $O(2^{n/2})$ oracle calls. While this is short of the exponential speedup one might hope for, a quadratic speedup can nonetheless be very significant for large problem instances, where even minor improvements could be of enormous value.

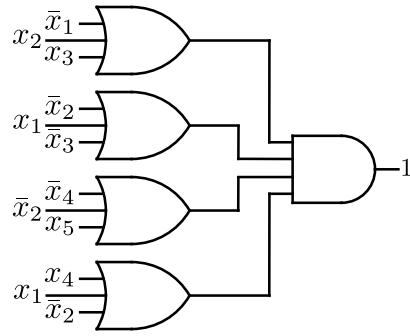


Figure 193: Digital circuit for an instance of the 3-SAT problem, with 4 clauses acting on input variables $\{x_i\}$. Each of the OR gates is input with some combination of 3 of the input bits or their compliments (i.e with a NOT gate). Each of these is referred to as a ‘clause’, of which there are 4 in this example, but could be any number in general. The final AND gate requires that all clauses be simultaneously satisfied in order to yield a final output of ‘1’. The goal of the problem is to find an input bit-string x that yields an output of ‘1’. In general, this may require exhaustively searching over the entire space of input states via brute-force, which exhibits time-complexity exponential in the length of the bit-string and associated number of clauses. This problem is proven to be **NP**-complete. Note that the similarly-defined 2-SAT problem (i.e clauses each contain 2 input bits) is **P**, making 3-SAT the simplest model to consider in the study of **NP**-complete problems. For this reason 3-SAT is often used as the computational model when studying **NP**-complete problems.

2. Graph colouring

One particularly useful approach to representing various optimisation problems is to relate them to the *graph colouring problem*. This problem is known to be **NP**-complete, and when operating on a graph with n vertices has $O(2^n n)$ classical runtime using the best known algorithm (?). Since this approach is **NP**-complete in general, and can therefore be mapped to a satisfiability problem, the quadratic enhancement offered by Grover is directly applicable.

The goal of the graph colouring problem is to take a graph G , and a set of k colours, and assign a colour to each vertex such that no two vertices connected via an edge share the same colour. If such an assignment of colouring exists, we say the graph is k -colourable. The *chromatic number* of G , denoted $\chi(G)$, is the smallest number k such that the graph is k -colourable. An example is shown in Fig. 194(top).

This problem relates closely to the *map colouring problem*, where the goal is to assign colours to states within a map, such that no two neighbouring states share the same colour, e.g Fig. 194(bottom). The map colouring problem is a subset of the graph colouring problem, since a map is merely a planar graph. As opposed to graph colouring, where the chromatic number could be arbitrarily high, for map colouring the chromatic number is provably at most

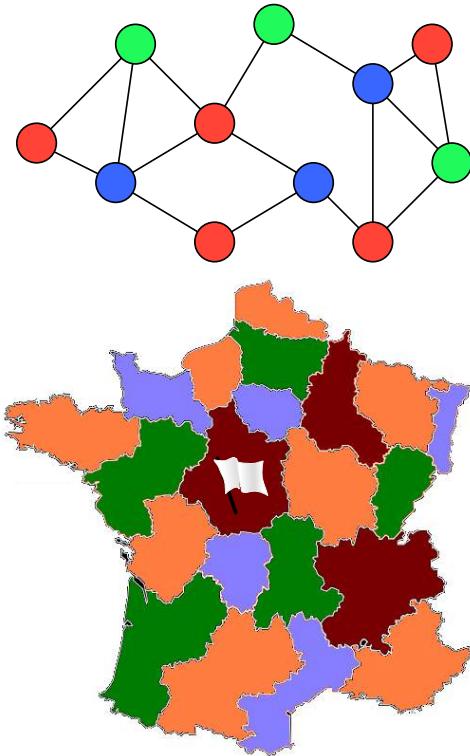


Figure 194: (top) Example of a graph with 3-colouring, such that no two vertices connected by an edge share the same colour. For this particular graph, 3-colouring is optimal (in the sense that the number of colours is minimised). It can easily be seen upon inspection that no 2-colouring is possible for this graph. Thus, this graph's chromatic number is $\chi(G) = 3$. (bottom) A 4-colouring of the states of France^a. Any planar map such as this has a chromatic number of at most 4 – the ‘four colour theorem’.

^a “I say to you that you are better than a Frenchman. I would lay even money that you who are reading this are more than five feet seven in height, and weigh eleven stone; while a Frenchman is five feet four and does not weigh nine. The Frenchman has after his soup a dish of vegetables, where you have one of meat. You are a different and superior animal – a French-beating animal (the history of hundreds of years has shown you to be so).” — William Thackeray.

4 – the ‘four colour theorem’.

How does this seemingly abstract mathematical problem relate to optimisation? Here are two simple examples illustrating the usefulness of the graph colouring problem in optimisation theory.

a. Time-scheduling Our goal is to schedule the execution of a large number of tasks, so as to minimise total execution time. However, some tasks may conflict with others and cannot be executed simultaneously – for example, they compete for access to a shared resource, which can only serve one task at a time.

Let the vertices in the graph represent tasks, and colours represent distinct points in time. If two tasks are in com-

petition with one another and cannot be executed simultaneously, we designate an edge between them. Thus, the set of edges in this ‘conflict graph’ fully characterises all conflicts in resource allocation between tasks.

Now finding a valid scheduling of tasks amounts to solving the graph colouring problem on the conflict graph, where the colouring specifies the temporal order in which tasks ought to be executed. The graph's chromatic number, $\chi(G)$, now tells us the minimum time in which all tasks can be executed subject to their constraints – the optimised scheduling.

b. Broadcast networks Consider a country comprising different regions, where each region is served by its own local radio station, operating at its own carrier frequency. But the radio signal within a region is strong enough to penetrate its immediate neighbouring regions also. For this reason, the frequency at which a regional station transmits should never be the same as that of its neighbours, so as to avoid signal interference.

In this incarnation of the map colouring problem, a valid colouring specifies a set of transmission frequencies that avoid interregional signal interference, and the chromatic number specifies the smallest possible radio spectrum that must be allocated⁸¹.

The same principles could apply to all manner of broadcast networks, such as the allocation of spectrum in cellular phone networks, where now map regions represent cellular transmission towers and their respective coverage.

3. Non-satisfiability-based optimisation algorithms

The optimisation problems discussed until now have been satisfiability problems residing in **NP**-complete, limited to only quadratic quantum enhancement owing to their utilisation of Grover's algorithm. Are there any other optimisation problems to which quantum computers lend themselves, potentially with the cherished exponential enhancement. There are, although the examples are far less ubiquitous than the plethora of satisfiability-based optimisation problems.

4. Curve fitting

A quantum-enhanced algorithm for performing least squares curve fitting was described by (?), potentially offering exponential enhancement, depending on the sparsity of the data.

⁸¹ Spectrum is a scarce resource, in high demand, and worth a lot of money!

As input it takes a set of data-points, (x_i, y_i) , and functions, f_j , outputting a set of coefficients, λ_k , defining a linear combination of those functions that fits the data, as well as a parameter characterising the quality of the fit, E . That is, we find the λ that defines the fitting function,

$$f_{\vec{\lambda}(x)} = \sum_k \lambda_k f_k(x), \quad (80.1)$$

such that the quality estimator is minimised according to a sum-of-squares metric,

$$E = \sum_i |f_{\vec{\lambda}}(x_i) - y_i|^2. \quad (80.2)$$

This can be thought of as an optimisation problem in the sense that our goal is to find a linear combination of functions that maximises the fit quality. This problem finds widespread use across many fields, particularly in statistics where it is ubiquitous.

5. Semidefinite programming

Semidefinite programming is a technique for finding solutions to constrained linear systems. Although the technique is already classically efficient, running in polynomial time, a quantum-enhanced version has been described (?), which promises exponential speedup in certain parameter regimes.

D. An example of a poorly constructed joke

“How many apples grow on a tree? All of them.” — Dad.

A number of years ago there was a popular quantum computing joke circulating the internet. It went along the lines of:

Q: What happens when two quantum computers play chess?

A: White e3, black resigns.

This joke is of course based on the assumption that quantum computers can efficiently and optimally solve the chess problem, foreseeing the entire future of a game, and determine the perfect optimal next move. Thus, after making the optimal first move ‘e3’, black concedes that he cannot win.

However this is a poorly-designed joke, since it has been proven that chess is an **EXP**-complete problem⁸²,

believed to lie well outside of **BQP**, that which is efficiently solvable on quantum computers. Therefore this is a bad joke and it isn’t funny.

However, although quantum computers are unlikely to be able to efficiently play chess *optimally*, there may still be some room for improvement in *approximating* optimal moves, to be discussed next in Sec. LXXX.E.

E. Approximate optimisation

Some problems in even harder classes than **NP**-complete can in some instances be *approximated* using the same approach as for **NP**-complete problems. The key to solving such problems is to define an oracle that attributes a *score* to a given input, rather than a yes/no answer to perfect satisfiability, and answers ‘yes’ or ‘no’ depending on whether that score is above some defined threshold for approximation. As an illustrative example, consider the optimisation of, say, a complex traffic network, where the goal is to maximise flow through the network. Then we might define our score to be some flow metric for the network’s graph.

We then apply a Grover search repeatedly, each time incrementing this threshold until the algorithm outputs ‘no’. Then we know that the last input yielding the ‘yes’ outcome had the highest score. The reason this approach is *approximate* rather than *exact* is that defining such a score-oracle mightn’t be always efficiently implemented, or maybe it mightn’t make sense at all to define score measures for a given problem. Alternately, maybe a particular problem inherently requires perfect satisfiability, and any imperfect approximation is insufficient – if even a single constraint isn’t satisfied, the system is broken.

An intuitive example of a problem taken from a far harder complexity class than **NP**-complete, that can be approximately optimised using the Grover approach is the game of chess. Generalisation of the rules of chess to an $n \times n$ board yields a game proven to reside in **EXP**-complete – the class of all problems requiring exponential runtime on a classical computer.

How do current classical computers play chess? They construct a *search tree* (see Fig. 195), which is explored exhaustively up to a given depth (bounded by computational resources), and assign a score to each board position, indicating its relative strength. The branch in the tree with the highest scores determines the next move.

This approach lends itself ideally to approximation via Grover. We take the same classical board-position-scoring algorithm, and implement it as an oracle. Grover is then able to search through a quadratically larger search tree, querying the oracle for scores above threshold.

Indeed this intuitive example is a very powerful one – since all **EXP**-complete problems are by definition equivalent, and **EXP** contains many other lesser (but nonetheless hard) classes, the approximate optimisation technique

⁸² The complexity of chess is defined in the context where the game is generalised appropriately to an $n \times n$ board, since complexity classes are all about computational scaling against problem size, and obviously any *fixed* size instance of a game can be solved in $O(1)$ time (albeit with perhaps a very large constant!).

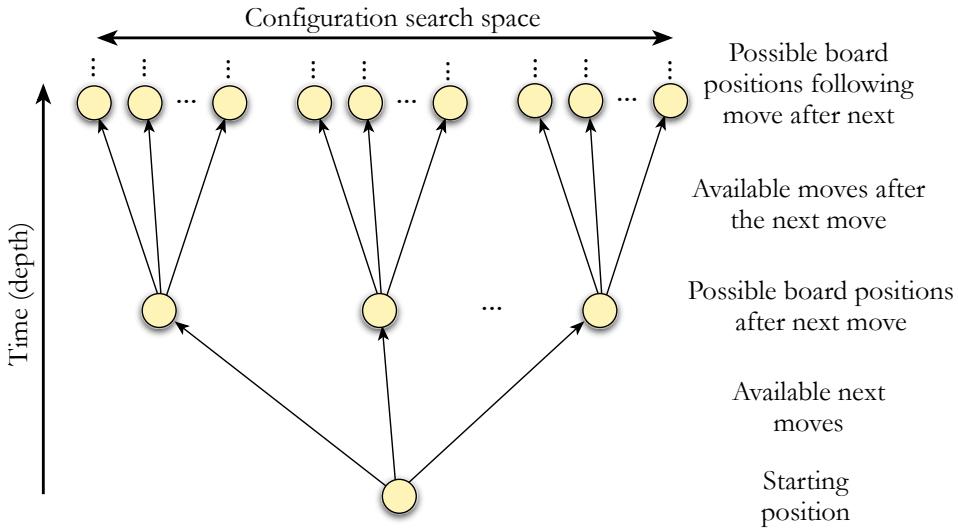


Figure 195: A search tree is an outwardly-directed tree graph in which each vertex represents a state of the system (i.e ‘board position’ in chess). Edges represent actions (or ‘moves’) available from the position from which the edge emanates. The root node represents the current position, and the depth of a vertex (i.e its distance from the root node) represents its future time. A path through the tree therefore represents a sequence of moves up to some point in future time. If the tree had maximum depth, a route (from root to leaf) would represent an entire single game, and the complete set of routes would represent all possible ways in which the game can be played. A complete analysis of a complete search tree with maximum depth would therefore allow us to exhaustively analyse all possible ways in which a game could be played to completion, enabling optimal gameplay. However, since the number of vertices/edges in a tree graph grows exponentially with depth, $|V| = O(\exp(d))$, the configuration search space in a game grows exponentially with how many moves into the future we wish to explore, thereby strongly limiting tree depth in practical implementations.

immediately captures a vast array of interesting problems.

F. How beneficial is quantum-enhanced optimisation?

The utilisation of Grover’s algorithm as the basis for improving optimisation problems may seem a little depressing, given that quantum search algorithms only confer a modest quadratic enhancement. By comparison, so many other quantum algorithms offer exponential speedup! How significant is such a quadratic enhancement actually?

In the context of a search tree, the size of the search space (i.e number of leaves on the tree) is $s = b^d$, where b is the tree’s branching parameter (assumed uniform here for simplicity) and d is tree depth. A quadratic enhancement in the searchable search space to $s' = s^2$ effectively doubles the search tree depth that can be accommodated by a given number of oracle calls to $d' = 2d$.

Needless to say, doubling one’s foresight in planning for the future is highly advantageous and could lead to huge improvements in competitive advantage and efficiency in resource allocation. Doubling the number of moves a competitive chess player could look ahead would result in their complete and utter dominance of the game!

Despite all this optimism, there remains one major concern, one shared by any quantum algorithm offering only polynomial enhancement. In any real-world quantum computer fault-tolerance is necessarily required to enable the algorithm to be executed faithfully in the inevitable

presence of noise. However, fault-tolerant codes necessarily induce a time and space overhead (i.e in terms of the number of physical qubits, and algorithmic runtime). If this overhead is too large it could well overpower the enhancement offered by the underlying algorithm. For this reason, real-world implementations of these types of algorithms will require very careful consideration of the construction of their associated error correction circuitry, such that there is still some net tangible algorithmic gain at the end. The quadratic enhancement in algorithmic runtime is actually a bound based on the ideal-case where there are no errors. In a fault-tolerant construction the actual achievable enhancement will necessarily be less than this. Specifically, for a given runtime, if the enhancement in the searchable search space is a polynomial of order p , this equates to an effective increase in search tree depth by a factor of p . For a useful quantum enhancement following fault-tolerance overhead we obviously want $p > 1$.

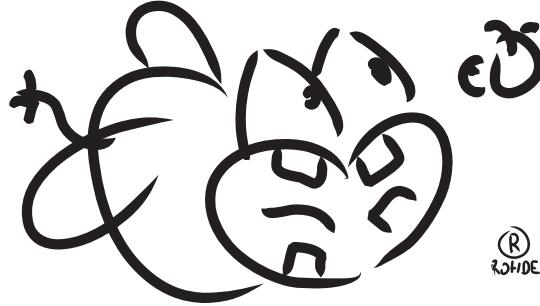
Note that for the quantum search method for enhancing **NP**-complete problems, a quadratic improvement is known to be optimal. Therefore, strictly $p \leq 2$.

G. Implications for the future

While the dream-goals of perfectly optimised resource allocation and perfect economic efficiency are fantasy, we should not underestimate the impact that even modest enhancements in the efficiency of large-scale economic

systems could have. While quantum computers will never perfectly solve humanity's immense resource allocation problems, even the more modest improvements they offer could be of immense value to the world economy, with enormous implications monetarily, socially, and environmentally. Indeed, optimisation problems are so ubiquitous that pursuing quantum-enhanced optimisation will be of central importance in the quantum era.

— *Ceteris paribus.*



Part XIII Roadmap – The way forward

“You got to go down a lot of wrong roads to find the right one.” — Bob Parsons.

“The further a society drifts from the truth, the more it will hate those that speak it.” — George Orwell.

LXXXI. THE TECH-WEB FOR THE QUANTUM FUTURE

We have in this work presented a vast number of ideas, protocols, applications and implications for future quantum technology, mediated by the quantum internet. The relationship between all these concepts is highly complex, with a large number of interdependencies. In Fig. 196 we express this as a mind-map timeline, capturing these relationships to show the logical flow of developments that will need to take place to reach various technological milestones and destinations.

Note that the evolution of future quantum technologies proceeds as a *tech-web* rather than as a strictly linear progression. This necessitates that many quantum technology research projects will need to execute in parallel to achieve multiple desired endpoints in a timely manner without bottlenecks.

— *Aut inveniam viam aut faciam.*



Part XIV Essays – A new frontier

“Only the very weak-minded refuse to be influenced by literature and poetry.” — Cassandra Clare.

“There is no harm in doubt and skepticism, for it is through these that new discoveries are made.” — Richard Feynman.

In this part we provide a non-technical outlook on the future quantum internet and its implications, for the benefit of the technically disinterested reader, who merely wishes to grasp some of the ‘big issues’. This section is in the form of a collection of short essays, requiring little or no technical background knowledge in quantum computation, quantum mechanics, or mathematics.

We acknowledge that while parts of these essays are certainly highly plausible, if not certain, others are highly speculative, but nonetheless based on believable although somewhat futuristic (perhaps even bordering science fiction) reasoning. We can't predict the future. But at the very least we hope to stimulate the exchange of ideas, and promote their exploration and development. After all, the great ideas of the future always begin speculatively! We encourage the reader to critically question the ideas presented in these essays, and put forth their own thoughts and predictions for what the quantum future may bring and the implications it will have for humanity.

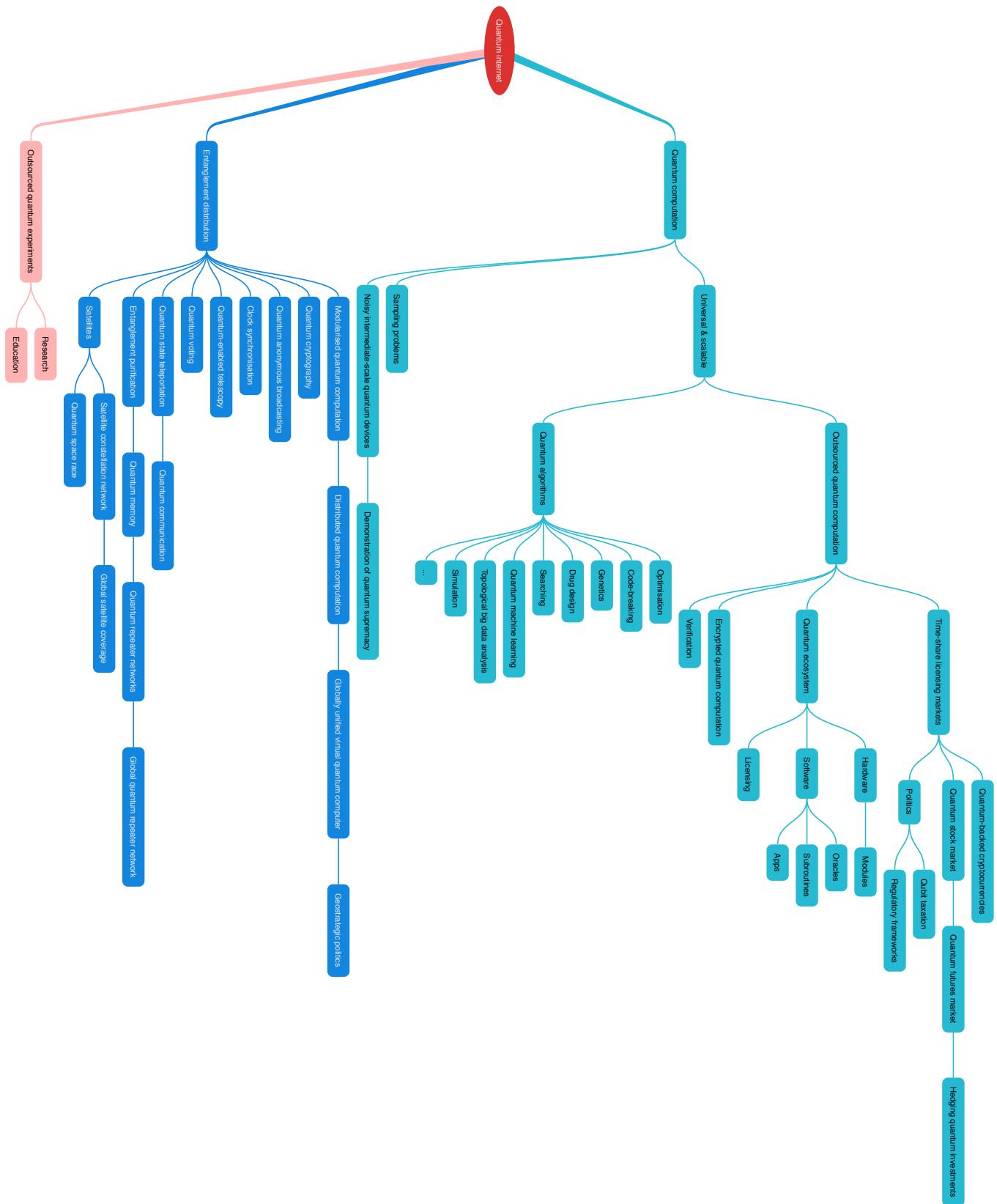


Figure 196: Relationships and dependencies in the development and deployment of the major concepts in the quantum internet – the quantum *tech-web* – showing the logical flow of developments that will need to take place to reach various technological milestones and destinations. The progression is highly non-linear, mandating many research programs operating in parallel to prevent technology bottlenecks.

LXXXII. THE ERA OF QUANTUM SUPREMACY

“No matter how tiny you look, you can lead huge men if you have what the huge men don’t have.” — Michael Bassey Johnson.

A pertinent question to ask is ‘What is the timescale for useful quantum technologies? When will they be viable?’ The correct answer is likely very soon.

From the perspective of classical computing, Moore’s Law (observation!) for the exponential growth trend in classical computing power has proven to be a very accurate one. In Fig. 197 we illustrate the historical evolution in classical computing power, and extrapolate 5 years into the future.

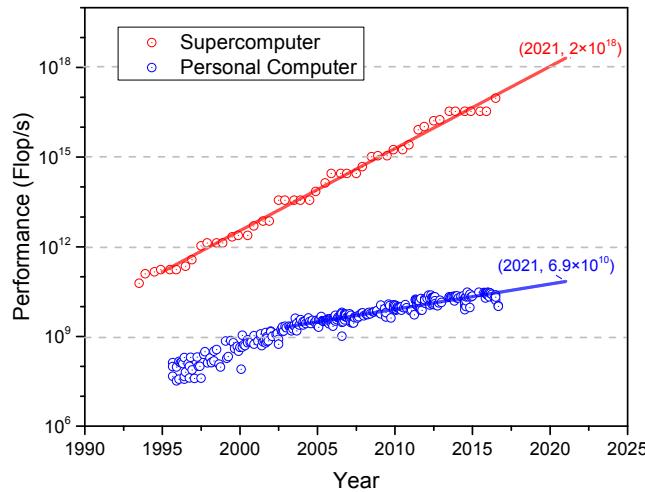


Figure 197: Historical trends in classical computing power for both PCs and top-end supercomputers, with an extrapolation 5 years into the future. The close fit to consistent exponential growth in performance over time is apparent from the logarithmic scale.

To put this into context, current day microprocessors contain on the order of billions of single transistors. Current day experimental quantum computers, on the other hand, contain fewer than 100 qubits. We sit at the mere very beginning of Gordon Moore’s adventure through the quantum era.

While the power of classical computers scales at most linearly with the number of transistors, the classical-equivalent power of quantum computers scales exponentially with the number of qubits (in the best-case scenario). The classical Moore’s Law is close to saturation – we simply can’t make transistors too much smaller than they already are⁸³! We therefore envisage a new Quantum

Moore’s Law, which follows a far more impressive trajectory than its classical counterpart. The point of critical mass in quantum computing will take place when the classical and Quantum Moore’s Law extrapolations intersect, signalling the commencement of the *post-classical era of quantum supremacy*. Estimating this is more challenging than it sounds, since although the classical Moore’s Law is extremely well established with an excellent fit to an exponential trajectory, there aren’t yet enough data-points to make a confident prediction about a Quantum Moore’s Law, to what trajectory it best fits, and at what rate it progresses, not to mention unforeseeable black swans.

Aside from quantum computing, theoretically unbreakable quantum crypto-systems, in the form of quantum key distribution (QKD), are already technologically viable, and are in fact commercially available off-the-shelf today, as end-to-end units connectable via fibre-optics. Recently, satellite-based QKD was demonstrated, enabling direct intercontinental QKD over thousands of kilometres. Although only a single such satellite has been demonstrated, its success implies that constellations of interconnected such satellites are inevitable in the near future, enabling point-to-point QKD between any two points on Earth. It is likely the next space-race will be the one for quantum supremacy.

As the era of post-classical quantum computation edges closer, the importance of QKD networks will intensify, and along with it the demand for quantum networking infrastructure.

It is clear that humanity already sits at the precipice of harnessing quantum technologies, and must act quickly to enable them to be fully exploited as they emerge in the near future.

— *Ad astra per alas porci.*

LXXXIII. THE GLOBAL VIRTUAL QUANTUM COMPUTER

“No generation has had the opportunity, as we now have, to build a global economy that leaves no-one behind. It is a wonderful opportunity, but also a profound responsibility.”
— Bill Clinton.

FROM the uniquely quantum phenomena that computational power can scale exponentially with the size of a quantum computer, as opposed to the linear relationship observed in classical computing, emerges an entirely new paradigm for future supercomputing. Rather than different quantum hardware vendors competing to have the biggest and best computers, using them independently in isolation, they are incentivised to unite their resources over the network and leverage (‘piggyback’) off one another, forming a larger and exponentially more powerful *virtual quantum computer*, which could then be time-shared between them, to the benefit of all parties.

⁸³ Current transistor feature sizes are on the order of several hundred atoms. Under a Moore’s Law prediction, we are likely to hit fundamental physical barriers in transistor size within a decade. Presumably, we can’t make a transistor smaller than an atom!

The key observation is that *all* contributing users to the network gain leverage from other users unifying their assets, irrespective of their size. In fact, this computational leverage is greater for smaller contributors than larger ones, making the benefits of this phenomenon disproportionately benefit the less-well-resourced. “*The only thing that will redeem mankind is cooperation.*” — Bertrand Russell.

Users who make an initial fixed investment into quantum computing infrastructure, which they contribute to the network, but are then unable or unwilling to finance further expansion of, will nonetheless observe exponential growth in their computing power over time. That is, the computational dividend yielded by a fixed investment increases exponentially over time. This creates a very powerful economic model for investment into computational infrastructure with no classical parallel, which could be particularly valuable in developing nations or less-wealthy enterprises.

It follows that in the interests of economic efficiency, market forces will ensure that future quantum computers will *all* be networked into a single *global virtual quantum computer*, providing exponentially greater computational power to all users than what they could have afforded on their own.

Vendors of quantum compute-time who do not unite with the global network will quickly be priced out of the market, owing to their reduced leverage, rendering the relative cost of their computations exponentially higher than vendors on the unified network. “*United we stand, divided we fall.*” — Matthew Walker.

This might have very interesting implications for strategic adversaries – government or private sector – competing for computational supremacy, but nonetheless individually benefitting from jointly uniting their competing quantum resources. Bear in mind that using encrypted quantum computation all parties could maintain secrecy in their operations. Despite this secrecy, will the KGB and NSA really cooperate, to the benefit of both, or will the asymmetry in the computational leverage incentivise them to not unify resources and instead construct independent infrastructure?

The leverage asymmetry will be a key consideration in answering this question, since although both parties benefit on an absolute basis from unification, on a relative basis the weaker party achieves the higher computational leverage. For this reason, it is plausible the global virtual quantum computer will fracture, dissolving into independent smaller virtual quantum computers, divided across geostrategic or competitive boundaries, with the stronger parties seceding from the union – even though they would individually benefit computationally from unification, they may not wish the weaker ones to piggyback

off them, achieving greater leverage than themselves⁸⁴.

— *Kyrie eleison.*

Discuss adversarial enhancement, or in geostrategy section.

LXXXIV. THE ECONOMICS OF THE QUANTUM INTERNET

“Underlying most arguments against the free market is a lack of belief in freedom itself.” — Milton Friedman.

“Either we believe in free speech for those we despise or we don’t believe in it at all.” — Noam Chomsky.

Q UANTUM computers are highly likely to, at least initially, be extremely expensive, and affordable outright by few. Client/server economic models based on outsourcing of computations to servers on a network, will be essential to making quantum computing widely accessible. The protocols we have presented here pave the way for this type of economic model to emerge. It is paramount that the types of technologies introduced here be fully developed in time for the deployment of useful quantum computing hardware, such that they can be fully commercialised from day one of their availability, enabling widespread adoption, enhanced economy of scale, and rapid proliferation.

A key question regarding the economics of the quantum internet is the extent to which it will be able to piggyback off existing optical communications infrastructure, given that networking will almost inevitably be optically mediated. We have an existing intercontinental fibre-optic backbone, as well as sophisticated satellite networks. To what extent will this existing infrastructure (or future telecom/satellite infrastructure) be able to be exploited so as to avoid having to rebuild the entire future quantum internet infrastructure from scratch? This is a question worth billions of dollars. We also need to factor in that given the massive driving force behind telecom technology, its cost is following a Moore’s Law-like trajectory of its own, and what costs a billion dollars today might cost a million dollars in a decade’s time. In light of this, telecom wavelength quantum optics is being hotly pursued.

Technology should benefit humanity, not only an elite few — *Homo sum humani a me nihil alienum puto*. In light of this, who exactly will benefit from the quantum internet? Its beauty is that it doesn’t create a system of winners and losers. Rather, it establishes a technological infrastructure from which all can benefit, rich or poor. Well-resourced operators who can afford quantum computers, for example, will benefit from being able to

⁸⁴ Insert jokes about Greece and Germany here — *Im Wandel der Zeiten – Eine Geschichte der Zivilisation.*

license out compute time on their computers, ensuring no wasted clock-cycles and maximising efficiency. The less-well-resourced will benefit in that they will have a means by which to access the extraordinary power of quantum computing on a licensed basis, facilitating access to infrastructure by those who otherwise would have been priced out of the market. This is essentially the same model as what is employed by some present-day supercomputer operators, enabling small players access to supercomputing infrastructure. The quantum internet is critical to achieving the same goal in the quantum era. This could have transformative effects on the developing world in particular. And many emerging industries, for whom access to quantum computation will be critical, but who cannot afford them, will benefit immensely from the client/server model.

Already today, even before the advent of useful post-classical quantum computers, we are seeing the emergence of the outsourced model for computation. IBM recently made an elementary 16-qubit quantum computer freely available for use via the cloud. Interested users can log in online, upload a circuit description for a quantum protocol, and have it executed remotely, with the results relayed back in real-time. Although still very primitive, this simple development already makes experimentation with elementary quantum protocols accessible to the poor layman, undergrad, or PhD student in a developing country, people who just a few years ago would never have dreamt of being able to run their own quantum information processing experiments! This effectively opens up research opportunities to people who otherwise would have been priced out of the market entirely, unable to compete with established, well-resourced labs. Evidently, the market already recognises the importance of outsourced models for quantum computation. We encourage the impatiently curious reader to log onto the ‘IBM Quantum Experience’ (<http://www.research.ibm.com/quantum/>) and take a shot at designing a 16-qubit quantum protocol, without even needing to be in the same country as the quantum computer.

The quantum internet will facilitate the communication and trade of quantum assets beyond just quantum computation and cryptography. There are many uses for various hard-to-prepare quantum states, for example in metrology, lithography, or research, where outsourcing complicated state preparation would be valuable. Alternately, performing some quantum measurements can be technologically challenging, and the ability to delegate them to someone better-equipped would be desirable. The quantum internet goes beyond just quantum computing. Rather, it extends to a full range of quantum resources and protocols.

To commodify quantum computing, if constructing large-scale quantum computers were a simple matter of plug-and-play, where QuantumLegoTM building blocks are available off-the-shelf and straightforward to assemble

even for monkeys, mass production would rapidly force down prices. By arbitrarily interconnecting these boxes, large-scale quantum computers could be scaled up with demand, with a trajectory following a new Quantum Moore’s Law, with potentially super-exponential computational return.

We envisage that each of these commodity items is a black box, within which a relatively small number of qubits are held captive. Then, to build a larger quantum computer, we don’t need to upgrade our boxes. Rather, we simply purchase more boxes to interconnect over the network – modularised quantum computation. This notion is tailored to graph states in particular – because a graph state can be realised by nearest neighbour interactions alone, and since all preparation stages commute with one another, they naturally lend themselves to modularised, distributed preparation.

Such an approach lends itself naturally to distributed computation, where modules may be shared across multiple users, with the economic benefit of maximising resource utilisation, and the practical benefit of the end-user effectively having a much larger quantum computer at their disposal.

By having a standardised architecture for optically interconnecting modules, we also somewhat ‘future-proof’ our hardware investment – if interfacing modules is standardised, existing hardware can be fully compatible with newer, more capable module versions. We might envision the emergence of open standards on optical interconnects and fusion protocols.

On the other hand, if quantum computers were only ever sold as specialised, room-sized, all-in-one solutions (think D-WaveTM), such mass production would not experience the driving force of commodified, off-the-shelf building blocks, each of which is cheap, yet frugal in its computational power alone.

Essential to existing financial markets are pricing models for physical assets. Furthermore, derivative markets increase trading liquidity, market efficiency, enhance price discovery, and importantly, allow risk management via hedging and the ability to lock in future prices. This is invaluable to traders of conventional commodities, and it is to be expected that it will be equally valuable to consumers of quantum resources. We have made initial steps in deriving pricing models for quantum assets and derivatives, which although they may require revision in the future real-world quantum marketplace, provide an initial qualitative understanding of quantum market dynamics.

Networked quantum computing will present new challenges for policy-makers, whose fiscal policies strive to maximise economic efficiency and optimise resource allocation. Devising policies of taxation and a regulatory framework in the quantum era will require careful deliberation.

It is evident that taxation of qubits has far deeper

economic implications than the taxation of other typical financial assets or classical technologies, owing to their exponential scaling characteristics. Generally speaking, taxation of an asset disincentivises its growth. But if the computational return on quantum assets grows exponentially with network size, so too will sensitivity to taxes that stifle it. This will require extremely prudent consideration when designing fiscal policies in the quantum era, so as to avoid exponential suppression of quantum-related economic activity.

Conversely, the exponential dependence on the rate of taxation could be exploited for leverage via subsidisation. It may be economically beneficial to subsidise quantum infrastructure, reaping its exponential payback, via taxation of other economic sectors, less sensitive to taxation.

The future quantum economy might be made more efficient by artificially transferring capital from low-multiplier sectors to high-multiplier quantum technologies. Or maybe the market will do this on its own accord⁸⁵? This is a uniquely quantum consideration that never previously applied to conventional supercomputing. The onset of the quantum era may redefine our entire economic mindset and fiscal policy-making, to adapt to the unique economic idiosyncrasies of this emerging technology.

— *Deus ex machina.*

LXXXV. THE QUANTUM FUTURE OF CRYPTOCURRENCIES

“Bitcoin is the most stellar and most useful system of mutual trust ever devised.” — Santosh Kalwar.

“By 2030, some form of Crypto will become the global reserve currency but it will not be based on what exists today. Existing cryptos need to transform or will disappear. Also around 2030 or so, the first Nobel Prize in Economics will be awarded to a Cryptoeconomist.” — Tom Golway.

THE advent of cryptocurrencies (Sec. XXXV) places the death of fiat currency firmly on the horizon — *Deo gratias*. Central banks around the world have been consistently inflating and devaluing national currencies, destroying their integrity through loose print-on-demand monetary policies to finance ever-increasing debt. National currencies and currency unions sit at the brink of crumbling. Can we opt out? Is there an alternative? Let us discuss an alternative!

What makes a sound currency? First of all, it must exhibit scarcity and be difficult or impossible to counterfeit – it should not be possible to forge unlimited quantities

out of thin air, a quality most certainly not inherent to the fiat currencies maintained by today’s central banks. Second, its abundance and demand should exhibit relative stability and predictability over time, so as to create a stable money supply and inflationary/deflationary rate.

For these reasons, gold for thousands of years was almost universally accepted as the legally recognised form of tender, since it is naturally scarce and much work must be invested into its production. For the same reasons, emerging cryptocurrencies like BitCoin have become widely adopted and even the norm in contemporary hyper-inflating economies like Venezuela where fiat currency has lost all integrity, as the cryptocurrencies exhibit these desired traits, immune to government. But rather than scarcity of a natural resource, we are dealing with artificial scarcity of bit-strings, cryptographically enforced to satisfy certain mathematical properties and constraints that cannot be easily counterfeited.

We propose that units of quantum computation meet these criteria very well. The only way to forge new computations is via investment into infrastructure, which has direct monetary cost and cannot be mitigated. Recent history has shown us that Moore’s Law has made the growth in classical computing power highly predictable and relatively stable over time, and it is to be expected that a quantum Moore’s Law will hold.

Time-shares in unified computing power over the quantum network, via licensing out qubits from hardware owners, would provide all these essential desired qualities for a sound currency. It can be envisaged that forward contracts in compute-time (Sec. LIX) would act as a good basis for backing a currency. Since these are nothing but simple forward contracts, they lend themselves to highly fluid, low-overhead trading on international markets.

Existing Blockchain-based cryptocurrencies like BitCoin (Sec. XXXV) are actually examples of computation-backed currencies, where the mining process requires brute-force computation of a large number of SHA256 hash functions, to discover hashes satisfying certain constraints. Unfortunately, however, in the case of BitCoin these computations are perfectly wasted, since they are not solving any problems of merit. Rather miners are made to perform them purely for the sake of imposing artificial scarcity via ‘proof-of-work’⁸⁶.

QuantCoin™ (Sec. LXI) on the other hand backs the currency with real-world computations of value, as determined by market participants at the time, a far better utilisation of computational power, with far greater confidence in its objective monetary value. Such a currency is no longer backed purely by the psychology of scarcity, but also the economic value of executing useful quantum

⁸⁵ Have faith in the invisible hand.

⁸⁶ This proof-of-work notion was originally borrowed from the Hashcash protocol for preventing email spamming.

algorithms on real-world data. Thanks to homomorphic encryption and blind quantum computing, users' data may be protected from eavesdropping end-to-end during computation, whilst still allowing the computation to be associated with a unit of cryptocurrency.

Such currencies could be either commissioned and backed by nation states, or operate entirely in the private sector, leading us on a path to free banking, devoid of nation-backed currencies altogether.

Because future contracts have predetermined times until maturity, they also serve the very helpful role of being hedging instruments, an important tool for end-users of computation who may wish to lock in prices in advance to mitigate exposure to market risk.

Were a computation-backed currency to emerge, it would immediately further incentivise investment into expansion of quantum computational hardware, as it would be directly convertible to currency with zero overhead. The implications for compute-intensive industries would be immense, as there would be negligible transaction costs associated with the purchase of computation – since contracts in computation *are* the accepted currency – thereby driving forward investment into the next technological revolution.

Consider the time-share future contract model as a basis for a currency. Unlike fiat currency, this monetary system would not be inflationary since the commodity backing the currency is one which cannot be easily counterfeited – the only way to make more currency is to provide more genuine, functional, online qubits, which increases the money supply over time in tandem with the underlying asset backing it. This would in effect be a full-reserve banking system, where the direct one-to-one convertibility between currency (forward contracts on computation) and its backing asset (time-shared access to physical qubits) eliminates the money multiplier, a system essentially immune to bank runs.

Because the currency is forward contracts in computing time-shares, not ownership of the physical underlying qubits, the qubits needn't change hands upon being utilised in monetary transactions. The currency could reside entirely on a distributed digital ledger recording transactions in the future contracts, independent of trading in physical qubits, who owns them, or where they reside.

In a strategically fractured world, where multiple, independent quantum internets may exist in isolation to one another, partitioned along geostrategic boundaries, each with their own local QuantCoin™ currencies, there would be an enormous monetary incentive to breaking down trade barriers and globalising the network by unifying smaller ones. This is contrary to nationalised fiat currencies, where there is little incentive for, yet much to lose by unifying currencies. Greed on behalf of those owning QuantCoins™ would therefore directly incentivise harmony and integration amongst all the world's leading

players in the technological realm. Well-financed market participants would have much to lose from fracturing of the network. This could make QuantCoin™ a major driver towards international peace and prosperity in the quantum world of tomorrow.

Importantly, a computation-backed currency would be largely immune to political interference. Politicians would have zero ability to directly manipulate the money supply, short of suicidally self-destructive policies like shutting down or curtailing the quantum internet.

Having a sound monetary system, immunised against political interference, and incentivised to integrate, will play an important role in constraining the power of government and spreading economic liberty across the globe.

— *End the Fed!* *Libertas justitia veritas. Libertas perfundet omnia luce.*

LXXXVI. SECURITY IMPLICATIONS OF THE GLOBAL QUANTUM INTERNET

“Truth is treason in the empire of lies.” — George Orwell.

“I don’t even know why any of us are here. This is the worst job I’ve ever had.” — John Kelly.

WITH any new technology comes ethical considerations. Who will have access to it, and how do they plan to use it? For this reason, many developed nations have export bans or restrictions in place on ‘dual-use’ technologies – those which have clearly legitimate and morally justifiable uses, but also nefarious ones by competitors and criminals. Nuclear technology is the obvious archetype. Quantum technologies (in particular quantum computing and quantum cryptography) are particularly vulnerable to dual-use, and for this reason are becoming subject to dual-use technology legislation, such as export controls, in some nations. In Australia, for example, legislation is being introduced criminalising the transfer of knowledge on certain quantum and cryptographic technologies to foreign nationals of certain target countries.

With a global QKD infrastructure in place, any person on Earth with access would have uncrackable encryption at their fingertips. Whilst this might be welcomed by the populace of a despotic regime (or the libertarians in a democratic one), it would clearly be unwelcome for that level of secrecy and protection to be awarded to the regime itself. Similarly, criminal and terrorist organisations would be immune to government surveillance. With widespread global adoption of QKD technology, the signals intelligence agencies of nation states would become at least partially obsolete, leaving the NSA and its Five Eyes™ partners furious.

Quantum computing also has dual-use potential. In fact, given their ability to compromise some existing cryptographic protocols, it appears highly likely that the first

useful, post-classical quantum computers will find their way into the hands of national SIGINT agencies. Of course, it doesn't take much imagination to see that many other quantum algorithms could be employed for sinister purposes. For this reason we are likely to see export limitations placed on quantum computer technology in the future.

Much as the internet has eliminated national electronic borders, a quantum internet employed for distributed or outsourced computation, would make quantum computer technology available to hackers, criminals, terrorists, and strategically competing nations. And a distributed model for computation as unregulated as the classical internet would make it near impossible to prevent.

Many falsely argue that once quantum computers become available, capable of cracking current classical cryptographic codes, the world will have transitioned to quantum-proof QKD as a replacement encryption standard, and therefore that the security implications of quantum computing will not be relevant. In terms of an individual citizen's private online banking, this is largely true – who wants to read a 10 year old bank transaction? However, when it comes to national security things aren't quite so rosy. This is because major national security agencies like the NSA of the United States systematically vacuum up astronomical amounts of internet traffic and store it for future reference, knowing that one day they may be able to crack it. Thus, if the KGB had at some point in the distant past electronically communicated sensitive information that was intercepted by the NSA but unable to be cracked at the time, when sufficiently sophisticated quantum computers become available, those messages may simply be pulled from the NSA's treasure trove and trivially cracked.

Bear in mind that as recently as the late 70's the Data Encryption Standard (DES) was a US government-approved encryption standard. However, its mere 56-bit key-length is no match for a universal quantum computer. Therefore anything stored using this encryption standard in the past had better contain information that is irrelevant by the time the quantum computers arrive, since no doubt the NSA will immediately put them to good use cracking their entire historical catalogue of stored encrypted messages.

Combined with encrypted quantum computing protocols, no one would even know what they were up to when using this awesome computing power, and what they learn, they could keep to themselves.

Alg. X describes a typical protocol for a particularly nefarious application for this, with the end result shown in Fig. 198.

These are all legitimate concerns. But they are very much the same ones that detractors expressed about the classical internet and strong encryption. Nonetheless, it can be said that encryption and the internet have on balance been overwhelmingly beneficial to mankind, en-

```
function MakeAmericaGreatAgain(Putin):
```

1. A Russian bedroom hacker with no direct access to quantum technology, delegates a factorisation algorithm to the cloud using homomorphic encryption.
2. The computation is physically executed on a server in the United States.
3. The result is returned to our Russian comrade.
4. The Russian uses the obtained private RSA key to hack Hillary's emails.
5. The emails are strategically leaked during the next Presidential election.
6. This swings the election in favour of Trump™.
7. The NSA and FBI have no clue who was behind it, since it was homomorphically encrypted.
8. They blame Edward Snowden.
9. Fox News™ calls for his execution.
10. They'd kick themselves if they found out the algorithm was actually executed on US soil.
11. Unless the NSA switches off the entire quantum internet, they can't prevent it from happening again in subsequent elections.
12. return(America is Great Again™).
13. □

Algorithm X: A typical example of a nefarious use for cloud quantum computation. See Fig. 198 for the outcome.

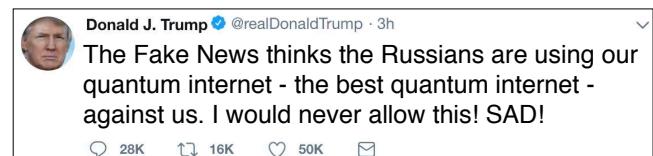


Figure 198: POTUS responds to Alg. X.

abling unprecedented rates of technological and economic progress. Any attempts to eliminate or undermine them could be economically catastrophic.

We take the view that the same ethical stance ought to be applied in the quantum era. While quantum technologies clearly have dual-use potential, the magnitude of the implications they will have for scientific and technological progress overwhelms the discussed proliferation issues. No doubt, politicians will nonetheless attempt to regulate and restrict the quantum internet – that's what governments like to do. But this will inevitably fail for the same underlying reasons that it failed for the classical internet – no tech-savvy Chinese person can actually say they are hindered by the Great Firewall of China™.

Talk about QKD, post-classical crypto, halving

private-key lengths, hacking stored public-key encrypted data from past is a security threat even once we have transitioned to post-classical crypto. NSA probably has mass storage of collected, but as yet uncracked data. Now they can work back through it.

— *Annus horribilis.*

LXXXVII. GEOSTRATEGIC QUANTUM POLITICS

“The people can always be brought to the bidding of the leaders. That is easy. All you have to do is tell them they are being attacked and denounce the pacifists for lack of patriotism and exposing the country to danger. It works the same way in any country.” — Hermann Göring.

“What the United States fears the most is taking casualties. The loss of one super carrier would cost the US the lives of 5000 service men and women. Sinking two would double that toll. We’ll see how frightened America is.” — Admiral Lou Yuan.

COMPUTATION is a commodity – perhaps the most valuable of all in the 21st century economy – and with any valuable, sought after commodity comes geostrategic powerplay. World powers fight wars, apply sanctions and use political leverage against one another to secure access to traditional commodities essential to economic progress and competitive advantage. It is to be expected that computation will be no different.

In conventional international relations, political leverage between conflicting parties is achieved through alliances, shared common interests, threats of military action, and even more sinister possibilities. How will this differ in the quantum era?

The central point to note is the computational leverage phenomena associated with the quantum internet – unification of resources is better for all. However, it is important to be cognisant that the leverage gained by parties unifying their resources with the cloud is asymmetric, biased in favour of (or against in an adversarial context) the weaker parties. That is, despite the fact that all players benefit from unification, smaller players relatively have more to gain. While this asymmetric computational leverage may seem favourable for the weaker parties, it also places them in a compromised situation whereby the threat of a major player expelling the smaller one from the network⁸⁷ creates asymmetric political leverage in the opposite direction. A major player will have relatively little to lose under the expulsion of a smaller player. But the smaller player could suffer immensely in the relative power of their computational assets.

This observation leads to the foreseeable possibility that future trade-wars may be for computational power, with stronger parties exploiting their huge leverage over weaker parties for geopolitical objectives. Sanctions and political punishment in the quantum era may very well employ computational isolation of nation states or organisations.

It is foreseeable that the future quantum internet may become fractured along geostrategic boundaries, with players (particularly stronger ones) unwilling to provide computational leverage to strategic competitors, even though on an absolute scale they would themselves benefit, since the leverage the competitor gains may compromise their own position, for example in cryptographic applications.

A further consideration is that the unification of quantum resources may very well require some form of central authority or marketplace to mediate the distribution and allocation of resources globally. Who will fill this role, and what strategic significance it will have is hard to predict. Certainly in the case of the United Nations, the Security Council, comprising a handful of self-declared world leaders, has immense geopolitical clout, with substantial power to influence international relations across the globe. Will the United Nations, under the supervision of the Security Council or some other politicised mediating authority, oversee the international quantum marketplace, or will some self-regulating, laissez-faire, libertarian utopia emerge under the guidance of the invisible hand.

— *Magnus est mundus.*

LXXXVIII. THE QUANTUM SPACE RACE

“We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard, because that goal will serve to organise and measure the best of our energies and skills, because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one which we intend to win, and the others, too.” — John F. Kennedy.

AT the time of writing this book the world’s first quantum-capable satellite was very recently launched into low-Earth orbit by Chinese scientists (?). The key capability of the satellite was to distribute entangled pairs of photons between ground stations thousands of kilometres apart. Using these entangled pairs, quantum key distribution was demonstrated, allowing theoretically unbreakable cryptography between the ground stations that no eavesdropper could compromise, guaranteed by the laws of physics.

However, entanglement distribution has many additional applications that are perhaps even more exciting than cryptography, most notably distributed quantum computation, enabling the world’s future quantum computers to be networked into a virtual device with exponentially greater power than the sum of the parts.

⁸⁷ Quantum interneXit.

The first-generation satellite that was recently developed merely contained an entanglement source, and two satellite-to-ground optical links via telescopes armed with laser tracking (Fig. 199). However, this prototype is strictly restricted to distributing entanglement between two ground stations, both simultaneously in line-of-sight of the satellite.

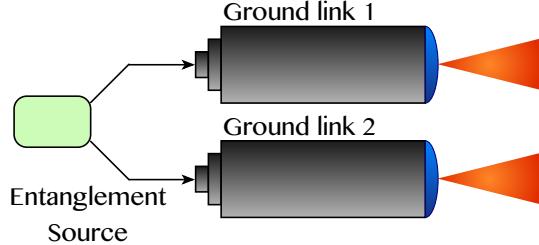


Figure 199: First-generation satellite for entanglement distribution. The on-board entanglement source couples to two telescopes, which lock onto independent ground stations using laser tracking.

To facilitate a true global network, next-generation satellites will need to form a constellation sufficiently dense that every point on the Earth's surface is always within line of sight of at least one satellite (Fig. 200).

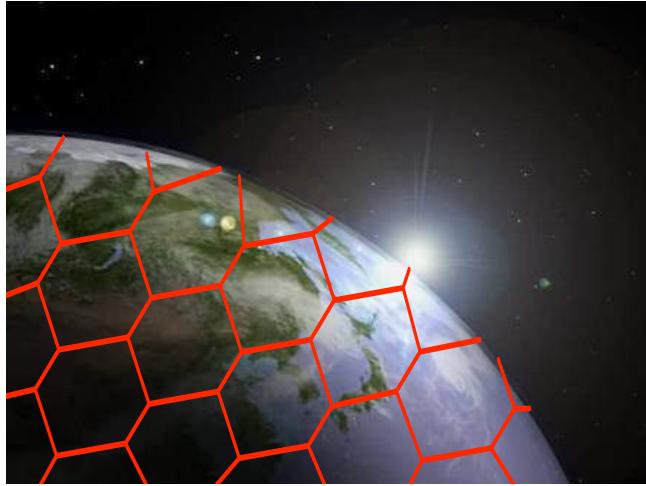


Figure 200: A honeycomb lattice is the lowest order two-dimensional lattice that could be employed to construct a satellite constellation network covering the Earth. Edges represent quantum communications channels, and their intersections are where the satellites reside. Such a network will require next-generation satellites with satellite-to-satellite links.

To enable a constellation, the satellites will need satellite-to-satellite links in addition to the satellite-to-ground links, such that they can relay the entanglement around the curvature of the Earth to overcome line-of-sight limitations. Additionally, they will need to do more than just prepare entangled states, but also perform entangling measurements, such that they can be configured as a quantum repeater network. A concept model for a next-generation satellite with these essential capabilities

is shown in Fig. 201.

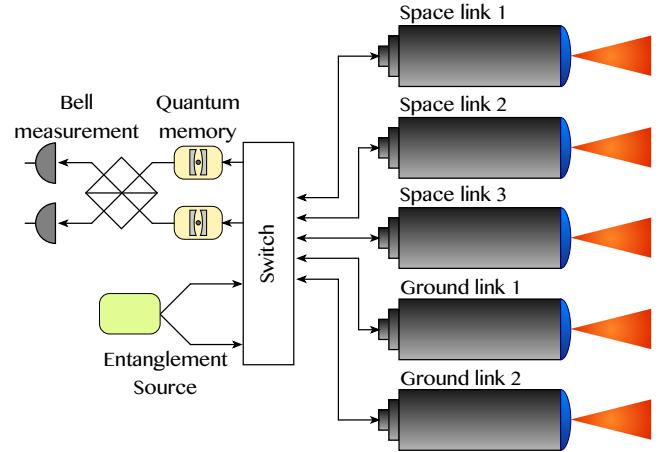


Figure 201: A basic layout for how next-generation quantum satellites might be constructed. Each satellite is capable of both entangled state preparation, as well as entangling measurements. There are three space links for communicating with neighbouring satellites so as to enable a honeycomb lattice configuration, as well as two ground links, as per the first-generation satellite. The switch at the centre must be universal to enable arbitrary pairs of telescopes to couple with either the entanglement source or entangling measurement. The quantum memories prior to the entangling measurement facilitate synchronising distinct photons with different arrival times such that they can interfere.

While the next-generation satellite may appear only incrementally more complex than the first-generation one, it is in fact far more technologically challenging. The main obstacle is that when performing entangling measurements, photons must arrive at the detector simultaneously. Obviously this is hard to enforce in space over long distances on fast-moving objects. Therefore quantum memories will be required, such that the first of two arriving photons is held in memory until the second one arrives, at which point it is read out from memory and the two photons are jointly measured. Unfortunately, such quantum memories are still very much in their infancy, and not reliable enough or of sufficient quality that they are ready for prime-time applications like a global space-based repeater network. It is unclear how far-off these technologies are, despite being under intense investigation.

A global constellation network may require hundreds or thousands of individual satellites. The key to deploying such a network will be via economies of scale. We must design a single standardised satellite (for example along the lines of that shown in Fig. 201), rather than a variety of more specialised models, make it as minimalistic as possible, and then mass produce them on a large scale. With this approach we can hope for economical deployment of a true space-based point-to-point global network.

The Chinese have successfully launched and demonstrated the first quantum satellite. This marks the beginning of the quantum space race. Who will respond?

For he who achieves a global network first will wield a huge competitive technological advantage in the upcoming era of the quantum internet and all its foreseeable and unforeseeable applications.

— *Carpe futurum.*

LXXXIX. THE NEAR FUTURE: NOISY INTERMEDIATE-SCALE QUANTUM TECHNOLOGY (NISQ)

— by Zixin Huang.

“The future belongs to those who believe in the beauty of their dreams.” — Eleanor Roosevelt.

In the near- to medium-term we are unlikely to make sufficient technological advances to realise fully scalable, fault-tolerant, universal quantum computation. But that doesn’t mean we will have no quantum capabilities at all! Noisy intermediate-scale quantum technology (NISQ) refers to quantum processors which may be available in the next few years, with around 50 to a few hundred qubits. These are going to be noisy and will not have full quantum error-correcting capabilities. They are likely to be special-purpose devices targeted at specific applications, possibly yielding only approximate answers owing to the absence of fault-tolerance (?).

Although fully universal, fault-tolerant quantum computers are still somewhat distant, with advances in quantum control, we are now in the position to explore a new frontier of physics, where we have quantum entanglement as part of our computational toolbox. We may not yet have *all* quantum capabilities, but we at least have some!

Scalable quantum computers, unlike classical ones, will be able to efficiently simulate any process that physically occurs in nature, enabling us to study the properties of complex molecules and new materials. This confidence is based on quantum complexity arguments, and our eventual capabilities to perform quantum error correction (which is admittedly very challenging and a potentially long-term vision). Both are based on quantum entanglement, a type of correlation between systems uniquely quantum mechanical, with no classical analogue. We have strong evidence that quantum computers have capabilities beyond classical computation. To illustrate this, consider the following:

- Quantum complexity: we have strong reason to believe that some tasks efficient on quantum computers may be computationally difficult classically. The best-known example is Shor’s algorithm (?), allowing us to factorise large numbers exponentially faster than using the best classical methods. Whilst we do not have a proof that an efficient classical algorithm doesn’t exist, the brightest of mathematicians have been trying to find one for decades to no avail. Integer factorisation has significant implications for cryptography, where the security of some codes is underpinned by the believed computational complexity of this particular problem.

avail. Integer factorisation has significant implications for cryptography, where the security of some codes is underpinned by the believed computational complexity of this particular problem.

- Complexity theory arguments: computer scientists have shown that quantum states which can be easily prepared with a quantum computer have super-classical properties. For example, given single photons input into a multi-mode interferometer, it’s hard for a classical computer to sample the probability distribution at the output, the so-called BOSON-SAMPLING problem. On the other hand, a quantum computer can trivially implement this experiment.
- No known classical algorithm can efficiently simulate a universal, fault-tolerant quantum computer, or simulate general quantum systems.

As we see, there is a clear distinction between what is hard classically and quantum mechanically. Intense research efforts are being dedicated to understanding which problems exactly are hard for a classical computer but easy for a quantum one.

The huge obstacle that lies between us and building a scalable quantum computer is the need to keep the system isolated from the environment to minimise noise (environmental noise is the arch-enemy of quantum computation!), at the same time being able to control it with extraordinary precision. Eventually, we expect to be able to protect quantum systems using quantum error correction. However, in order to perform quantum error correction, we currently believe that perhaps 10^3 - 10^4 physical qubits will be required to encode each logical qubit (depending on the physical architecture and its associated error rates). This adds huge overheads to the number of physical qubits needing to be individually prepared, manipulated and measured, all with extremely high fidelity. Therefore, reliable fault-tolerant quantum computers with quantum error correction are not likely going to be available in the near future.

In terms of the number of qubits, 50 is a significant number because it approximates the number of qubits we can still simulate by brute-force with our most powerful existing classical computers (?) – a benchmark for the meaning of the term *quantum supremacy*. The main question is: when will quantum computers be able to solve useful problems faster than classical ones? This leads us onto several potential uses for limited quantum computation in the NISQ era:

A. Quantum optimisers

For many problems, there is a big gap between the approximation achieved by classical algorithms and the barrier of exact-case NP-hardness. We do not expect

quantum computers to efficiently solve worst-case **NP**-hard problems, however, quantum devices may be able to find better *approximate* solutions to such problems, or at least find such approximations more quickly. The vision for using NISQ to solve optimisation problems is a hybrid quantum-classical algorithm. In this scheme we use the quantum device to produce and manipulate an n -qubit state, measure the qubits, then process the measurement outcomes classically. This then is utilised as feedback for the next round of quantum state preparation and evolution. The cycle is repeated until convergence is obtained to a quantum state from which the approximate answer can be extracted. Two such algorithms are known as *quantum approximate optimisation algorithms* (?), and *variational quantum eigensolvers* (?).

B. Quantum machine learning

“Many do not lose their mind because they do not have one.” — Arthur Schopenhauer.

Much of the quantum machine learning (QML) literature builds on algorithms which speed up problems in linear algebra (?). One of the potentials for QML rests upon QRAM – quantum random-access memory. For classical data processing, by using QRAM we may be able to represent a large amount of classical data, N -bits, using only $O(\log N)$ qubits, an exponential improvement in resource efficiency. However, the bottleneck may be in the encoding/decoding of the QRAM, which may seemingly mitigate potential gains, owing to the fact that measurements yield only one element at a time, not the full exponentially-large structure. QML may find applications in a more natural setting where both the input and output are quantum states, for example, to control a quantum system, or in learning probability distributions where entanglement plays an important role.

C. Quantum semidefinite programming

Semidefinite programming is the task of optimising a linear function, given some matrix inequality constraints. Classically, the problem can be solved in time polynomial in matrix size, and the number of constraints.

A quantum algorithm has been shown to find an approximate solution to this problem with an exponential speedup (??). In this algorithm, the initial state is a thermal state that is a function of the input matrices for the semidefinite program. The success of the implementation depends on whether the particular thermal state can be efficiently prepared. The output is a quantum state, which approximates the optimal matrix. The quantum state can be measured to extract (via sampling) features of this matrix.

The crucial feature in the quantum algorithm is the preparation of a thermal state of non-zero temperature, suggesting the algorithm may be intrinsically robust against thermal noise – this would be a fantastic trait to exhibit in the NISQ era of no fault-tolerance. It’s therefore entirely possible that a quantum solver for semidefinite programs might be achievable with near-term NISQ technology.

D. Quantum dynamics

As was stressed previously, quantum computers are very well suited to studying highly entangled, multi-particle systems. It’s the natural platform to simulate entangled states, where quantum computers appear to have a clear intrinsic advantage over classical ones.

With a universal quantum computer, we anticipate that studying quantum chemistry (especially noisy quantum chemistry) will be enabled. Ideally, if the noise model in the quantum computer can be cleverly mapped to be isomorphic to the noise present in the physical system being simulated, then noise becomes a feature not a bug! This could be used in the design of new pharmaceuticals, for example, as well as catalysts for improving the efficiency of nitrogen fixation or carbon capture. We may be able to find new materials with better resistive properties, leading to more efficient transmission of electricity. However, these promises may not be fulfilled with NISQ, because algorithms to accurately simulate large molecules and materials may not succeed without quantum error correction.

We do know that classical computers are particularly inefficient at simulating quantum dynamics, i.e how highly entangled quantum states will evolve over time. Here quantum computers have a particularly obvious advantage, and one example would be quantum chaos. In these systems entanglement spreads very rapidly. Insights might be gained using noisy devices on the order of only 100’s of qubits, a perfect regime for the NISQ era.

We’ve barely had a glimpse of the promises of NISQ. But it’s clear that although near-term devices will be limited, they may nonetheless open up exciting new prospects and computational applications, beyond the capabilities of present-day classical machinery.

— *Carpe futurum.*

XC. THE FUTURE OF QUANTUM CRYPTOGRAPHY

— *by Zixin Huang.*

“How long do you want these messages to remain secret?... I want them to remain secret for as long as men are capable of evil.” — Neal Stephenson.

QUANTUM cryptography is the first field in quantum information task to reach commercialisation. At its early stages, quantum cryptography was almost synonymous with quantum key distribution (QKD), but has since branched and became one of the fastest growing areas in quantum information. The purpose of QKD is to distribute a secret-key between two trusted parties who share a quantum channel, as well as a classical channel for authentication. Unlike current cryptography systems, which are secure based on the presumed limitations of an adversary's computer (*computational security*), the security of QKD is based on the laws of quantum mechanics, providing guaranteed security unless our understanding of quantum physics is inherently wrong (*information theoretic security*). In this section we discuss some of the challenges in QKD, as well as other aspects of quantum cryptography beyond traditional QKD.

The typical setting of QKD is as follows. There are two trusted parties who want to establish a secret-key, Alice and Bob. They share two channels: a quantum channel, which allows them to send quantum states (encoded in photons or other states of light) to one another; and, a classical channel, with which they can send classical messages. The communication over the classical channel is assumed to be public and completely insecure, and the eavesdropper, Eve, has full anonymous access to it. However, Eve cannot modify messages shared over the classical channel.

The quantum channel is subject to possible manipulation by Eve. The task of Alice and Bob is thus to guarantee security against an adversarial eavesdropper. The typical protocol assumes that Alice and Bob do not share any secret to begin with. The origin of the security of QKD springs from the fundamentals of quantum mechanics, that is, any act of measurement by an observer on a quantum state necessarily induces a change in the state – measurement collapse. This means that in combination with classical communication, actions of an eavesdropper cannot go undetected, ruling out intercept-resend attacks by Eve.

The ultimate goal of a QKD network is long distance secure quantum communication with imperfect sources.

Despite the significant advances in both the theoretical and experimental development of QKD, a number of challenges remain for it to be widely adopted in securing everyday communications (??). Experimentally, much effort is being invested into improving the performance of QKD systems. On the theoretical side, showing the security of a QKD system with finite key-size is also a challenge, because information-theoretic security is achieved only when immunity against the most general (coherent) attack is proven (?).

A. Performance

Some of the criteria for assessing the performance of a QKD scheme include key-rate, range, cost and robustness.

1. Key-rate

Currently, a strong disparity exists between classical and quantum key distribution rates. Classical optical communication delivers on the order of \sim 100Gbits/s per wavelength (for a frequency-multiplexed implementation), whereas communication rates only on the order of \sim Mbit/s are achievable using current QKD implementations.

The obtained key-rate depends on the performance of the detector used for measurement. For QKD based on single-photon detection techniques, to achieve a high bit-rate, one requires true single-photon states, in combination with detectors with high efficiency and short dead-time, both of which effectively induce loss, mandating more trials. Current developments are promising, with a reported quantum efficiency of 93% at telecom wavelengths (?).

For continuous-variable QKD systems, increasing the bandwidth of the homodyne/heterodyne detectors whilst keeping the electronic noise low is essential.

2. Range

Extending the range of QKD systems is a major challenge and driving factor for QKD in terms of future network applications. Two approaches are being pursued – free-space and quantum repeaters. A quantum repeater, similar to its classical analogue, is a device that can extend the range of quantum communication between sender and receiver. However, one cannot amplify the signal that contains the quantum information, owing to the no-cloning theorem, which prohibits making copies of unknown quantum states. The fact that an intercept-resend attack by Eve must disrupt the state of the system is the basis for the security of QKD – one of the major limitations imposed by quantum mechanics works to our advantage!

A quantum repeater effectively needs to restore the quantum information without measuring it directly, and is extremely technologically challenging. Over optical fibre networks, the standard loss for 1550nm wavelength light is 0.2dB/km. Over long enough distances, this unavoidable loss will reduce the key-rate to a level of little practical relevance, therefore a ground-based solution would be to divide the entire channel into segments, where two partners exchange pairs of entangled photons and store it in a quantum memory (??).

The second is to use free-space quantum communication techniques via satellite links. Satellite QKD is achievable

with present-day technology. Here satellites are used as intermediate trusted nodes for communication between locations on the ground. Direct links can be established between ground stations and the satellite, thus enabling communication between parties separated by long distances, potentially relaying across a satellite constellation network to overcome line-of-sight limitations from the Earth's curvature. Satellite QKD suffers comparatively very low loss between satellites in orbit, but the satellite-to-ground links, which cannot be avoided at the endpoints, suffer around 40dB loss when propagating through the effective atmospheric thickness of $\sim 10\text{km}$ when the satellite is directly overhead (and worse for satellites with lower azimuth). The atmospheric loss is a major hurdle, since distribution of a Bell pair between two ground stations effectively incurs 80dB inefficiency, meaning that only 1 in every 100,000,000 Bell pairs are successfully distributed — *Stupor*. Nonetheless, in China, satellite QKD over 1200km has been demonstrated (?), sufficient for sharing a secret-key for private-key cryptography with guaranteed key secrecy.

3. Cost & robustness

For QKD systems to be consumer-friendly, low cost and robustness are crucial features. Preferably QKD systems should make use of existing data fibre-optic infrastructure, since the use of dark fibres are not only expensive, but often unavailable (?), and there is a big economic incentive to reuse existing infrastructure rather than rebuild it from scratch. Single-photon detectors at room temperatures are also desirable, because this can remove the requirement for cryogenic cooling, hence reducing power consumption and making consumer systems far more practical.

Integrated photonic platforms are being explored to reduce cost, since miniaturisation can lead to light-weight, low-cost QKD modules that can be mass-manufactured, essential for economies of scale.

Currently, two platforms are being explored: silicon (?), and indium phosphide (?). A reconfigurable QKD system employing an In-P transmitter and silicon detectors has been demonstrated in the laboratory (?).

B. New protocols

In parallel to hardware development, research efforts are being directed towards finding new QKD protocols which can outperform existing ones. Two of these are high-dimensional (HD) QKD and the Round-Robin differential phase-shift protocol (RR-DPS).

HD QKD aims at encoding more than one bit in each detected photon, which can increase the information capacity when the photon rate is limited. Security proofs against collective attacks are being developed, and an

experiment has demonstrated an information capacity 6.9 bits per coincidence rate at 2.7Mbit/s over 20km (?).

The RR-DPS protocol (?) removes the need to monitor signal disturbance. In a conventional QKD protocol, the noise parameter needs to be estimated; and if high precision is required, the portion of the signal that is sacrificed increases, thus decreasing the efficiency of the protocol (?). This protocol has a high tolerance to the qubit error rate ($< 50\%$) (?), and makes it attractive for implementation when high systematic errors are unavoidable.

However, currently, neither of the protocols out-compete the more mature decoy-state BB84.

C. Challenges in security

Although QKD protocols are provably information-theoretically secure, physical implementations often contain imperfections which are not considered in the theoretical model — no experiment ever perfectly matches its design! Attacks can be designed to exploit such imperfections, on either the source or the detector side.

Tab. XXIX, taken from (?), summarises some attacks against certain commercial and research systems.

To regain security, a number of solutions have been proposed:

1. QKD with imperfect sources

The source is typically less vulnerable to attacks because Alice can prepare her quantum states in a protected environment, and we expect that she can characterise her source. Therefore, flaws in state preparation can be easily incorporated into the security proof.

Loss-tolerant protocols have been proposed (?), and further developed by (?), where decoy state QKD with tight finite-key security has been employed. A wide range of imperfections with the laser source have been taken into account (?), including intensity fluctuations. A security proof has shown that perfect phase randomisation is also not necessary (?).

This provides strong evidence that secure quantum communication with imperfect sources is feasible (?). Intuitively, QKD with imperfect sources is viable because by assuming that states prepared by Alice are qubits, Eve cannot unambiguously discriminate Alice's states (?) — quantum measurement collapses quantum states.

2. Measurement-device-independent QKD

To prove security for Bob's measurement device is more problematic, since Eve has complete access to the quantum channel and she can send any signal.

Attack	Targeted component	Tested system	References
Time shift	Detector	Commercial	(???)
Time information	Detector	Research	(?)
Detector control	Detector	Commercial	(??)
Detector control	Detector	Research	(?)
Detector dead-time	Detector	Research	(?)
Channel calibration	Detector	Commercial	(?)
Phase remapping	Phase modulator	Commercial	(?)
Phase information	Source	Research	(?)
Device calibration	Local oscillator	Research	(?)

Table XXIX: Summary of various attacks against some commercial and research QKD systems.

One candidate for a long-term solution to side-channel attacks⁸⁸ is device-independent (DI) QKD (?). This relies on the violation of a Bell inequality (?), and the security can be proven without knowledge of the implementation. However, the expected secure key-rate is low even over short distances. A more practical approach is measurement-device-independent (MDI) QKD (?), which is immune to side-channel attacks against the measurement device. Here the device is treated as a black box, and can be untrusted. However, an important assumption for MDI QKD is that Eve cannot interfere with the state preparation process, which is practically reasonable.

Another candidate is detector-device-independent (DDI) QKD (?), which has been designed to take advantage of both the strong security of MDI-QKD, with the efficiency of conventional QKD. However, DDI-QKD has been shown to be vulnerable to certain attacks (?).

The MDI-QKD protocol has been extended to the continuous variable framework. However, this system requires homodyne detectors with efficiency $> 85\%$, and a reliable phase reference between Alice and Bob.

We have discussed some significant remaining challenges in QKD. These range from theoretical security proofs to hardware developments. Advances in QKD will not only enable point-to-point quantum communication, but have implications for a range of network applications, such as quantum secret sharing (???), blind quantum computing (?), quantum anonymous broadcasting (?), and many more.

As remarked in (?), “*Determining the exact power and limitations of quantum communication is the subject of intense research efforts worldwide. The formidable developments that can be expected in the next few years will mark important milestones towards the quantum internet of the future.*”

XCI. THE QUANTUM ECOSYSTEM

“The most dangerous worldview is the worldview of those who have not viewed the world.” — Alexander von Humboldt.

ASSOCIATED with any new computer platform comes a hardware/software *ecosystem* that evolves around it. If we consider the release of the original iPhone and its iOS operating system, it wasn’t just the product itself that was revolutionary, but the third-party software industry that emerged surrounding it, and it wasn’t until this software ecosystem became established on the App Store that the product realised its full potential and became truly transformative.

From the hardware perspective, it wasn’t until interfacing standards such as USB and Wi-Fi emerged, allowing the plethora of competing hardware products to arbitrarily interconnect and interface with one another, that the hardware realised its full potential.

In the quantum era we anticipate the same phenomena to arise. What will this *quantum ecosystem* look like? Here are some of the elements that vendors might specialise in, providing compatible components for the quantum ecosystem:

- Quantum operations: vendors selling the capacity for non-trivial state preparation (e.g Bell, NOON and GHZ states), or measurements (e.g complex entangling syndrome measurements).
- Software subroutines and libraries: much like classical code, many quantum computations (and other quantum protocols) can be decomposed into pipelines of subroutines. There are many quantum operations that arise repeatedly (such as quantum Fourier transforms and syndrome calculations), which vendors might specialise in for outsourcing.
- Oracles: as an essential quantum software building block, oracles will become a fundamental unit for outsourcing. These oracles will store hard-coded or algorithmically-generated databases, or mathematical functions. For example, for use in genetic

⁸⁸ A side-channel attack is one which exploits knowledge of the imperfect implementation of a system (e.g details of source or detector characteristics) to compromise security, rather than a weakness in the theoretical model underpinning it (normally approached using cryptanalysis).

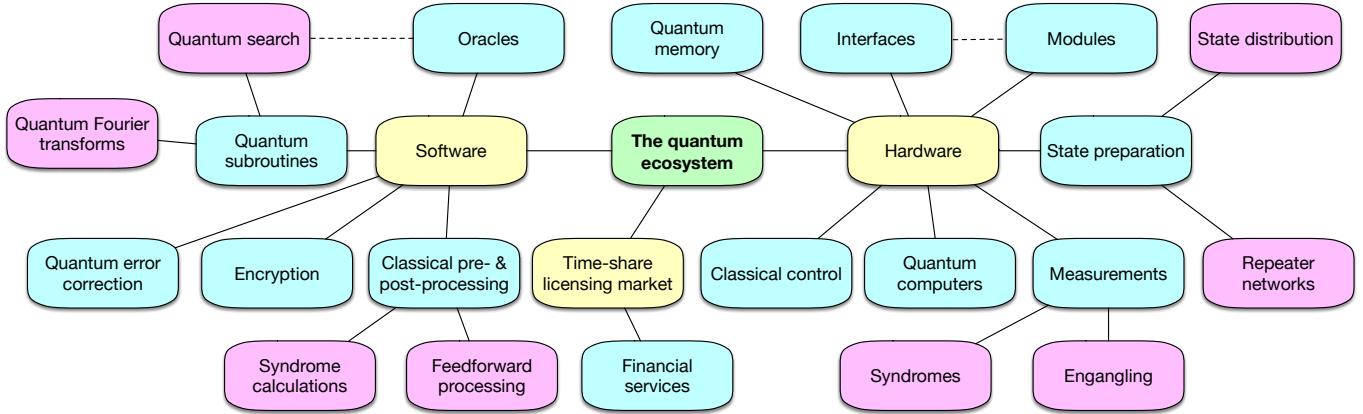


Figure 202: Map of just a few of the elements of the quantum ecosystem that are likely to arise with the advent of the quantum internet. The distinct units could become areas of specialisation for quantum vendors, which might be licensed out or sold to customers as discrete units, or as complete integrated processing pipelines, all outsourced and distributed over the quantum network.

medicine (Sec. LXXVIII), such databases could algorithmically generate tables of candidate drug compounds, or they could implement mathematical functions whose input space is to be searched over when quantum-enhancing the solving of NP-complete problems.

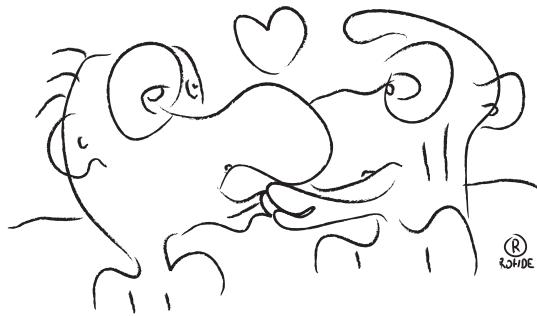
- Interfacing: *de facto* standards will emerge for interconnecting quantum hardware units. Most notably, standards for optical interconnects will arise.
- Modularisation: arbitrarily-interconnectable units will develop, allowing quantum hardware to be constructed in an ad hoc, Lego-like manner. These modules could implement small elements of a larger quantum computation, such as housing a small part of a larger graph state (Sec. XLIII.D), communications building blocks (such as transmitters or receivers of Bell pairs), or algorithmic building blocks such as quantum Fourier transforms (Sec. XLI.D).
- Classical pre-, post- or intermediate-processing: quantum computation, and other quantum protocols, typically require some degree of classical pre- or post-processing. These classical operations can be highly non-trivial. For example, a novel topological quantum error correcting code (Sec. XVIII.A) might require complex encoding, decoding and feed-forward operations. Determining and implementing these operations may require complicated optimisation protocols. These might be outsourced to a specialised provider.
- Classical control: many quantum protocols require intermediate classical control, i.e feedforward. For example, in a quantum repeater network we must control the order of entangling operations and track

the ‘Pauli frame’, a tally of the corrections accumulated by the final entangled Bell pair.

- Quantum memory: storing qubits with long decoherence lifetimes is extremely challenging using today’s technology, and it is foreseeable that vendors might specialise in this particular operation, especially once error correction is built into the memory.
- Quantum error correction: any given quantum error correcting code follows a well-defined recipe for encoding, correction, and decoding. Thus, it might become an example of a subroutine specialised in by a dedicated quantum error correction vendor, and licensed out as a building block for embedding into larger, fault-tolerant protocols.
- Time-share licensing market: a market will emerge for the trade and allocation of time-shares on the global virtual quantum computer (Sec. XLIII.F). Associated financial services industries will emerge around this marketplace, including secondary markets, derivative markets, managed funds in quantum infrastructure, and IPO markets.

A map of just a few of the potential major hardware and software elements to emerge in the quantum ecosystem is presented in Fig. 202.

— *Disiecti membra poetae.*



Part XV

The end

“When something is important enough, you do it even if the odds are not in your favour.” — Elon Musk.

“Be nice to nerds. Chances are you’ll end up working for one.” — Bill Gates.

“We are just an advanced breed of monkeys on a minor planet of a very average star. But we can understand the Universe. That makes us something very special.” — Stephen Hawking.

XII. CONCLUSION – THE VISION OF THE QUANTUM INTERNET

“We will either go down as the world’s greatest statesmen, or its greatest villains” — Hermann Göring.

Include ref

QUANTUM technologies, particularly quantum computing, will truly revolutionise countless industries. With early demonstrations of key quantum technologies – such as QKD, long distance quantum teleportation, and quantum computing – becoming a reality, it is of utmost importance that networking protocols be pursued now.

We have presented an early formulation and analysis of quantum networking protocols with the vision of enabling a future quantum internet, where quantum resources can be shared and communicated in much the same way as is presently done with digital assets. Whilst it’s hard to foresee exactly how future quantum networks will be implemented, as there are many unknowns, many of the central ideas presented here will be applicable across architectures and implementations on an ad hoc basis.

There are a number of schools of thought one might subscribe to when quantum networking. One might demand perfect data integrity and best-case network performance. But that would come at the expense of necessitating an

all-powerful central authority to oversee all communications, ensuring that scheduling was absolutely perfect – a potentially very challenging optimisation problem. Or one might tolerate lost data packets or suboptimal performance, at the expense of limiting applicability, but with the benefit of improved flexibility and reconfigurability. Or maybe some arbitrary compromise between different metrics and attributes is best. These are open questions that needn’t have concrete, one-size-fits-all answers. They certainly needn’t be answered right now.

The QTCP framework we presented is sufficiently flexible and extensible that these questions can be answered and enforced independently by different subnets, depending on their individual characteristics and requirements, in much the same way that every organisation connected to the classical internet today is free to structure their own LAN as they please, enforcing their own internal network policies.

The quantum internet will allow quantum computation to become distributed, not just outsourced. In the same way that many present-day classical algorithms are heavily parallelised and distributed across large clusters, CUDA cores, or even across the internet itself (e.g the SETI project), quantum networks will allow the distribution of quantum computation across many nodes, either in parallel, in series, or in a modularised fashion. This will be pivotal to achieving scalability. Keeping in mind that the classical-equivalent power of a quantum computer may grow exponentially with the number of qubits, it is highly desirable to squeeze out every last available qubit for our computations – every qubit is worth more than the last!

Combined with recent advances in homomorphic encryption and blind quantum computation, commercial models for the distribution of quantum computation will emerge, allowing computational power to be outsourced, with both client and server confident in the security of their data and proprietary algorithms. This is a notion that is challenging on classical computers, but will be of utmost importance in quantum computing, where it is expected sensitive or valuable data and algorithms will often be at stake.

From the security perspective, the global quantum internet will enable an international QKD communications network with perfect secrecy, guaranteed to be information-theoretically secure by the laws of physics. This will be of immense economic and strategic benefit to commercial enterprises, governments, and individuals. Classical cryptography is already a multi-billion dollar industry worldwide. Quantum cryptography will supersede it, and be of especial importance in the era of quantum computers, which compromise some essential classical cryptographic protocols, such as RSA, which forms the basis of most current internet encryption, digital signatures, and the Blockchain/Bitcoin protocols. Not only is quantum cryptography being pursued optically, but even credit cards with embedded quantum circuitry are being actively de-

veloped to prevent fraud. Inevitably, this will require the communication between bank automats and servers to be mediated by a quantum network.

Already, off-the-shelf QKD systems are available as commodity items, from vendors such as MagiQ and ID Quantique, which may be simply interconnected via an optical fibre link, thereby implementing end-to-end quantum cryptography in a modularised fashion. This is of a similar flavour to, and first technological step towards, modularised quantum computing, which would greatly enhance the economic viability and scalability of general purpose quantum computing by paving the way for the mass production of elementary interconnectable modules as commodity items.

We have focussed our attention thus far on the application of quantum networking to quantum information processing applications, such as quantum computing and quantum cryptography. However, with plug-and-play quantum resources available over a network, one might envisage far greater applicability than just these.

Of particular interest are the implications of quantum networking to basic science research. Presently, experimental quantum physics research is limited to well-resourced labs with access to state of the art equipment. With the ability to license these assets over a network, and dynamically interconnect them on an ad hoc basis, the ability to construct all manner of quantum experiments could be extended to all. An undergraduate laboratory would now have the ability to approach a host to politely borrow their state engineering technologies, send it to another with the ability to perform some evolution to that system, and to yet another to perform measurement and analysis of the output – all from an undergrad lab equipped with nothing more than desktop PCs. This has broad implications for basic science research, opening it up to aspiring researchers across the globe, regardless of their direct access to cutting-edge tools. This will greatly expand the intellectual base for conducting quantum experimentation to the entire global scientific community, decimating the scientific monopolies controlled by a handful of world-leading, highly-resourced experimental teams.

The reality is that we are only just beginning to understand the full potential for quantum technologies, and as we learn more we will inevitably find new uses for networking them. The full potential of digital electronics was never fully realised (or anticipated) until the emergence of the internet. It is to be expected the same will hold in the quantum era, an era only in its inception.

Large-scale quantum computing may still seem a formidable, and somewhat long-term challenge. But it isn't likely to remain so. Once we have mastered the technological art of preparing qubits and implementing high-fidelity entangling operations between them, it's just a matter of sitting back and watching Gordon Moore perform his witchcraft, and scalability of quantum technology, and its rapid market-driven reduction in cost, will

quickly ensue. The quantum internet will drive this rapid development by expanding both the supply and demand for access to this technology, and through unification of computational resources allow them to massively enhance their collective computational power, beyond their individual capabilities.

It is essential for the adoption and development of quantum technology, that quantum networking infrastructure be sufficiently well developed that it is ready to be deployed the minute the first useful, post-classical hardware becomes available. The proliferation of the defining technology of the 21st century depends upon it.

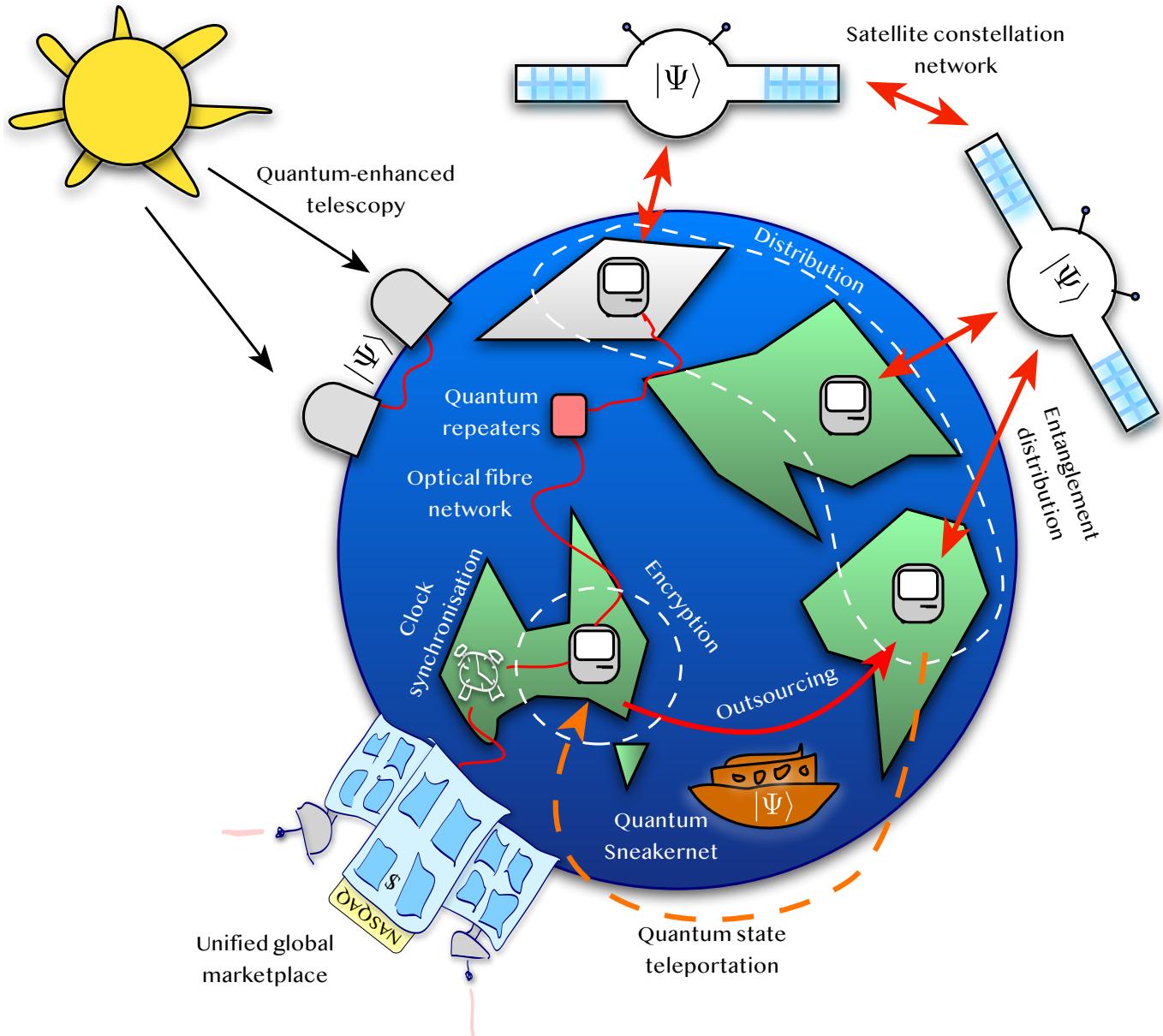


Figure 203: Overview of some of the essential services integrated into a future globally-unified quantum internet ecosystem.

XIII. HOW TO LEARN MORE & GET YOUR HANDS DIRTY

FOR the reader interested in delving deeper into quantum technology with a hands-on approach, there are a number of freely available software tools for simulating various quantum technologies, including quantum computing and quantum networks. We highlight some of the major ones below, all of which are freely available:

- SimulaQron (<http://www.simulaqron.org>): Developed by QuTech Delft (?), a simulator for designing software for the quantum internet. The package includes its own quantum internet programming language for deploying networked quantum algorithms.

The institute runs competitions for designing the best quantum internet app.

- Google Cirq (<https://github.com/quantumlib/Cirq>): Google's Python-based SDK for designing quantum software, specifically targeted at development for Google's own in-house quantum computing hardware platform.
- Microsoft Quantum Development Kit and the Q# language (<https://www.microsoft.com/en-au/quantum/development-kit>): a full SDK for writing code executable on quantum computers. The Q# language is used to describe quantum operations, which classical code written in C#

can interface with and control. The platform is extremely versatile and allows highly complex quantum operations to be coded in a high-level, platform-independent manner. The platform includes a simulator for executing code for which the hardware doesn't yet exist.

- IBM Quantum Experience (<http://www.research.ibm.com/quantum/>): IBM's cloud-based platform for remotely executing simple quantum protocols, as written and uploaded to the cloud by the user. The platform includes QISKit (<https://qiskit.org>), a Python-based SDK for executing quantum algorithms. It is based on the Open Quantum Assembly Language (OpenQASM) language.
- Rigetti Quantum Cloud Services (QCS) (<https://www.rigetti.com/qcs>): Rigetti's cloud platform for accessing and deploying algorithms to their hardware over the cloud. Development is powered by the Forest SDK, their own in-house development platform.
- Keep looking, as the list is rapidly growing with new participants entering the market regularly...

XCIV. EPILOGUE

Translate to Sanskrit

The biological evolution of the human race is reaching a plateau. The only mechanism for us to advance further is via technology. The 20th century was that of the transistor, gifting us unprecedented progress as a species, on an exponential trajectory. However, that trajectory is rapidly approaching saturation. The 21st century will usher in entirely new technologies, far beyond just incremental developments in our current ones. Quantum technology will be one of them, transforming the world as profoundly as the transistor did. Our mission is to contribute to the realisation of this, propelling the human race into the future, benefiting all of humanity, blind to demographics and national borders. Together we will make the future — Seize the future. To the stars. Knowledge is power.

— *Opus dei.*

ACKNOWLEDGMENTS

- We acknowledge Gavin Brennen, Chris Ferrie, Joseph Clark, Klaus Rohde, Eric Cavalcanti, Charlotte Bridault, Marco Tomamichel, Ryan Mann, Jonathan Olson, Samuel Marks, Nicholas Umashev, Elija Perrier, Kate Hrayssi & Alexis Shaw for helpful discussions.

- We thank the authors of QCircuit (<http://physics.unm.edu/CQuIC/Qcircuit/>) for providing this very helpful package.
- Peter P. Rohde is funded by an Australian Research Council Future Fellowship (project FT160100397).
- Tim Byrnes is supported by the Shanghai Research Challenge Fund; New York University Global Seed Grants for Collaborative Research; National Natural Science Foundation of China (Grant No. 61571301); the Thousand Talents Program for Distinguished Young Scholars (Grant No. D1210036A); and the NSFC Research Fund for International Young Scientists (Grant No. 11650110425); NYU-ECNU Institute of Physics at NYU Shanghai; and the Science & Technology Commission of Shanghai Municipality (Grant No. 17ZR1443600).
- Jonathan P. Dowling would like to acknowledge support from the US Air Force Office of Scientific Research, the Army Research Office, the National Science Foundation, and the Northrop-Grumman Corporation.
- We thank almost unethical quantities of 1,3,7-Trimethylxanthine for aiding in the motivation for this work.

— *Dum spiro, spero.*

STATEMENT OF CONTRIBUTIONS

- Peter P. Rohde conceived, directed and edited this project, developed the original ideas, and prepared most of the manuscript.
- Zixin Huang prepared Secs. XXVII.K, XXXVII.A.4, XL.C, XLII.E, XC, XXXVIII, XLI.K & LXXXIX.
- He-Liang Huang & Zu-en Su prepared Part. XI.
- Simon Devitt prepared Sec. XXXI.
- Rohit Ramakrishnan prepared Sec. XXI.B.
- Si-Hui Tan & Atul Mantri contributed to Sec. XLIV.
- Nana Liu prepared Secs. XLV & LXXIX.
- Scott Harrison prepared Secs. LXII, LXIII.B.2 & LXIV.
- Darryl Vietch contributed to Part. III.
- Chandrashekhar Radhakrishnan prepared Secs. XX.A.4 & XLIII.I.
- Time Byrnes prepared Secs. XXVII.J & LXXVII.
- Jonathan P. Dowling contributed to Sec. LXXVII.

- William J. Munro prepared Sec. **XXIX**.
- Jonathan P. Dowling contributed to editing.
- All authors played active roles in the discussions that developed this work.

RECOMMENDED WINE PAIRINGS

L IKE any good meal, a good book chapter is best paired with a suitably-chosen wine that complements its unique notes and characteristics. Ancient tradition tells us that this enhances the quality of the reader's interpretation, making a given text be perceived as far more remarkable and ingenious than it ordinarily would be. Here we suggest several contemporary, budget-friendly, Australian wines, hand-selected by the authors, that perfectly match the subtle undertones of each part in the book.

- Part. **I**: Howard Park Flint Rock Shiraz, 2015, Great Southern.
- Part. **II**: Helen & Joey Inara Pinot Noir, 2016, Yarra Valley.
- Part. **III**: Leconfield Merlot, 2016, Coonawarra.
- Part. **IV**: Lou Miranda Leone Cabernet Sauvignon, 2015, Barossa Valley.
- Part. **V**: Kaesler Stonehorse Grenache Shiraz, Mourvedre, 2014, Barossa.
- Part. **VI**: Ministry of Clouds Tasmania Chardonnay, 2016, Tamar Valley.
- Part. **VII**: Piccadilly Valley Pinot Noir, 2016, Piccadilly Valley.
- Part. **VIII**: The Story R. Lane Vintners Westgate Vineyard Shiraz, 2015, Grampians.
- Part. **IX**: d'Arenberg Artisans The Money Spider Roussanne, 2017, McLaren Vale.
- Part. **X**: Payten and Jones Major Kong Shiraz, 2015, Yarra Valley.
- Part. **XI**: Two Rivers Stones Throw Semillon, 2018, Hunter Valley.
- Part. **XII**: Penfolds Reserve Bin 17A Chardonnay, 2017, Adelaide Hills.
- Part. **XIII**: Watershed Senses Sauvignon Blanc, 2016, Margaret River.
- Part. **XIV**: Peter Lehmann Wines Wigan Riesling, 2015, Eden Valley.

- Part. **XV**: Giant Steps Yarra Valley Pinot Noir, 2017, Yarra Valley.

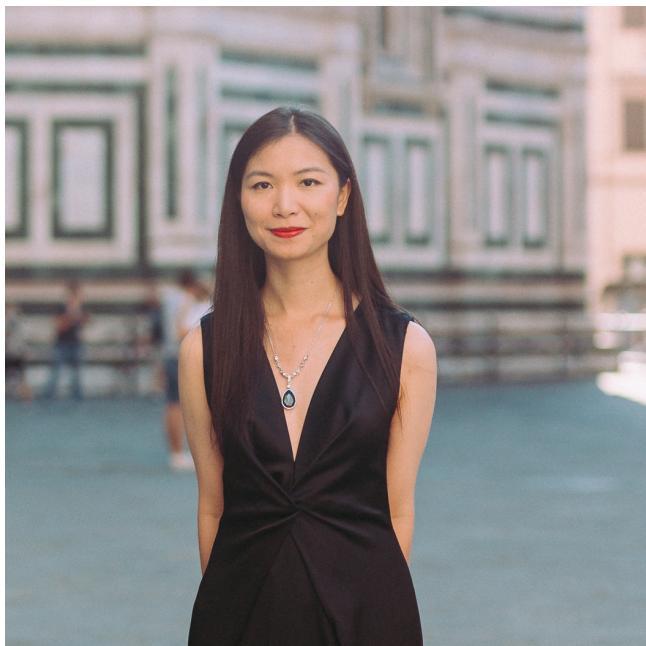
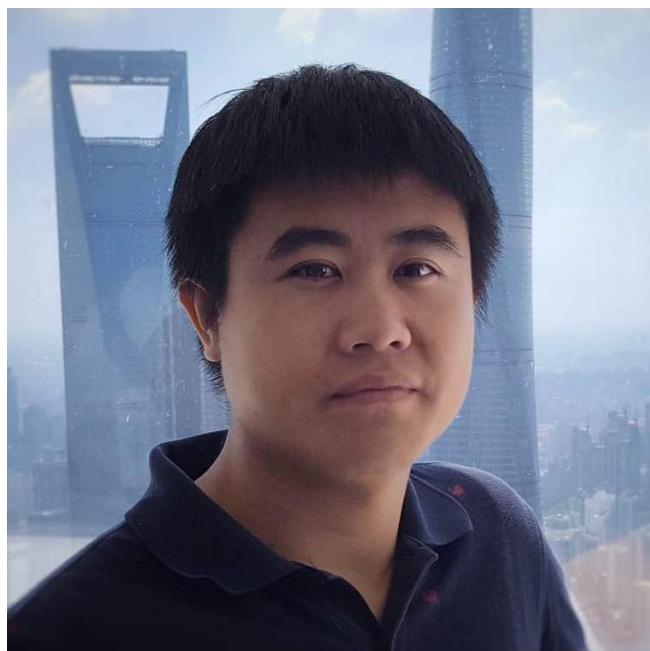
— *Don't drink and derive.*

ABOUT THE AUTHORS



PETER ROHDE is an Australian Research Council Future Fellow in the Centre for Quantum Software & Information at the University of Technology Sydney, Australia (UTS:Q|SI), and associate member of the Hearne Institute for Theoretical Physics at Louisiana State University. He obtained a Bachelor of Computer Systems Engineering with First Class Honours, and a PhD in theoretical physics at the University of Queensland. He has worked at highly-acclaimed international institutes, including the University of Oxford, University of Queensland, Institute for Molecular Biosciences and Max-Planck Institute for the Science of Light, with over 60 publications and 1,500+ citations in quantum optics, quantum computing, quantum information theory, ecology, and politics. His theoretical proposals have inspired several world-leading experimental efforts, including time-bin encoded BOSON-SAMPLING and sub-shot-noise limited quantum metrology.

He is a collaborator in China's world-first quantum satellite program, aiding in the design of quantum protocols for space-based demonstration. In his spare time he is a musician, composer, public speaker, libertarian political activist, charity worker, mountaineer, and adrenaline junkie. He regularly speaks at prominent political and scientific outreach events, features in newspaper and magazine articles, and conducts radio and podcast interviews. He is a very naughty boy and was expelled from preschool.

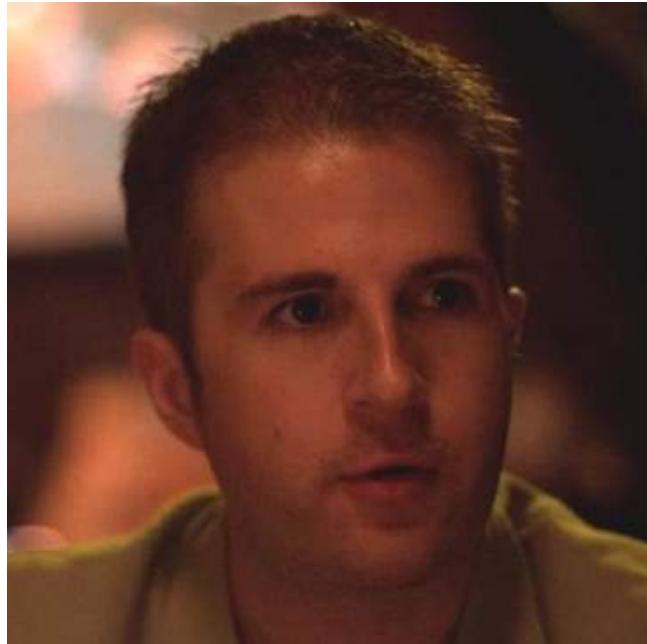


HE-LIANG HUANG is a Postdoctoral Fellow at the University of Science & Technology China (USTC). His research interests include secure cloud quantum computing, quantum big data analysis, and the physical implementation of quantum computing architectures, in particular using linear optical and superconducting systems. As an experimentalist he has been a part of several groundbreaking experimental efforts, most notably in the demonstration of world-leading multi-photon optical quantum information processing protocols. This includes the first experimental demonstration of topological data analysis on a photonic quantum computer, an optical demonstration of blind quantum computing, and a loophole-free Wheeler-delayed-choice experiment.



ZIXIN HUANG completed her PhD in physics at the University of Sydney, Australia. She is currently a postdoctoral researcher at the University of Sheffield, UK. She has worked on a wide range of topics ranging from quantum metrology and cryptography to experimental quantum optics.

ZU-EN SU is a Postdoctoral Fellow at the Technion – Israel Institute of Technology. He obtained his PhD at the University of Science & Technology China. His research interests currently include the physical implementation of quantum computation and quantum metrology, primarily based on linear optics and solid-state quantum photonics systems.



Complete
SIMON DEVITT



ROHIR RAMAKRISHNAN is a researcher at the Indian Institute of Science.
[Complete this section](#)

ATUL MANTRI is [Complete this section](#)



SI-HUI TAN is a Research Scientist at the Singapore University of Technology & Design. She obtained a Bachelor of Science with Honours from Caltech, and subsequently a PhD from MIT. She has received numerous awards, including the MIT Presidential Fellowship.

[Complete this section](#)



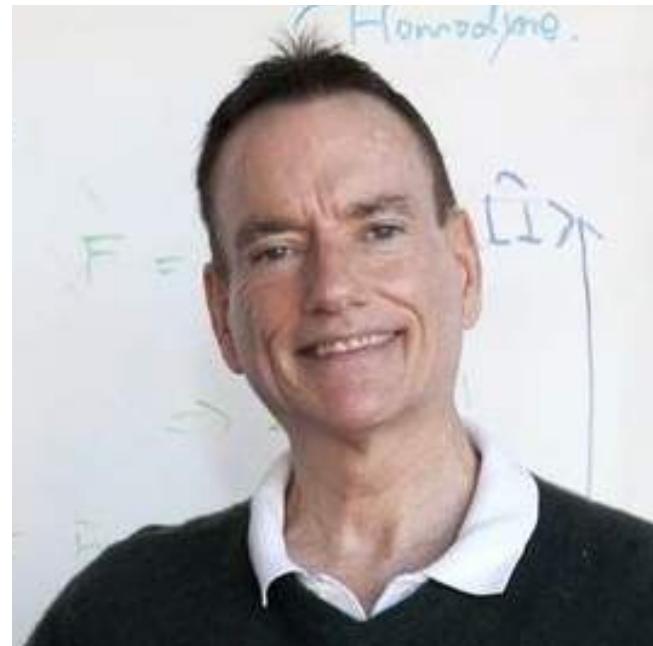
NANA LIU is a Postdoctoral Research Fellow at the Singapore University of Technology & Design.
[Complete this section](#)

Scott Harrison is[Complete this section](#)

Darryl Veitch is[Complete this section](#)



Chandrashekhar Radhakrishnan is a Postdoctoral Fellow and Bright Overseas Global Corporation Scholar at New York University, Shanghai, China. He is a recipient of the National Natural Science Foundation of China, International Young Scientist research grant. Before Joining NYU Shanghai he worked as a Postdoctoral Fellow at the Institute of Mathematical Science, India and the National Chung Hsing University, Taiwan. He obtained his Master of Science in Physics from the Indian Institute of Technology Madras and a PhD in theoretical physics from University of Madras. He has worked on a variety of fields, including statistical mechanics, mathematical physics and quantum information theory. His fundamental interest in quantum information theory lies in investigating quantum resources alternative to entanglement, which can be used for quantum communication and quantum computation. He has active collaborations with experimentalists, through which he tries to verify his theoretical results. He also has some interest in the application of statistical mechanics to biological problems.



Jon Dowling is Professor and Hearne Chair of Theoretical Physics, and Co-Director of the Hearne Institute for Theoretical Physics at Louisiana State University. He has a highly-acclaimed research career, with over 450 publications and 15,000 citations. He is author of the popular book *Schrödinger's Killer App: Race to Build the World's First Quantum Computer*.

[Complete this section](#)



Tim Byrnes....
[Complete this section](#)



BILL MUNRO obtained his PhD at the University of

Waikato, and has since held numerous prestigious fellowships and positions. He is currently Senior Research Scientist and leader of the theoretical quantum physics research group at NTT Japan.

[**Complete this section**](#)



"We all die. The goal isn't to live forever, the goal is to create something that will." — Chuck Palahniuk.