9th Conference on the Theory of Quantum Computation, Communication and Cryptography

TQC 2014, May 21–23, 2014, National University of Singapore, Singapore

Edited by

Steven T. Flammia

Aram W. Harrow

LIPIcs – Vol. 27 – TQC 2014 www.dagstuhl.de/lipics

Editors

Steven T. Flammia
Department of Physics
University of Sydney
steven.flammia@sydney.edu.au

Aram W. Harrow
Department of Physics
Massachusetts Institute of Technology
aram@mit.edu

ACM Classification 1998
E.3 Data Encryption, E.4 Coding and Information Theory, F Theory of Computation

ISBN 978-3-939897-73-6

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at http://www.dagstuhl.de/dagpub/978-3-939897-73-6.

Publication date
November 2014

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Digital Object Identifier: 10.4230/LIPIcs.TQC.2014.i

ISBN 978-3-939897-73-6 **ISSN 1868-8969** **http://www.dagstuhl.de/lipics**

LIPIcs – Leibniz International Proceedings in Informatics

LIPIcs is a series of high-quality conference proceedings across all fields in informatics. LIPIcs volumes are published according to the principle of Open Access, i.e., they are available online and free of charge.

ISSN 1868-8969

www.dagstuhl.de/lipics

Contents

9th Conference on the Theory of Quantum Computation, Communication and Cryptography.
Editors: Steven T. Flammia and Aram W. Harrow
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Preface

The 9th Conference on the Theory of Quantum Computation, Communication and Cryptography was held at the National University of Singapore, from the 21st to the 23rd May 2014.

Quantum computation, quantum communication, and quantum cryptography are subfields of quantum information processing, an interdisciplinary field of information science and quantum mechanics. The TQC conference series focuses on theoretical aspects of these subfields. The objective of the conference is to bring together researchers so that they can interact with each other and share problems and recent discoveries.

A list of the previous editions of TQC follows:

- TQC 2013, University of Guelph, Canada
- TQC 2012, The University of Tokyo, Japan
- TQC 2011, Universidad Complutense de Madrid, Spain
- TQC 2010, University of Leeds, UK
- TQC 2009, Institute for Quantum Computing, University of Waterloo, Canada
- TQC 2008, University of Tokyo, Japan
- TQC 2007, Nara Institute of Science and Technology, Nara, Japan
- TQC 2006, NTT R&D Center, Atsugi, Kanagawa, Japan

The conference consisted of invited talks, contributed talks, a poster session, a rump session, and a business meeting. The invited talks were given by Fernando G. S. L. Brandão (University College London, London), Vittorio Giovannetti (NEST, Scuola Normale Superiore, Pisa) and Yaoyun Shi (University of Michigan, Ann Arbor).

The conference was possible thanks to the financial support of the Centre for Quantum Technologies, Singapore.

We wish to thank the members of the Program Committee and all subreviewers for their precious help. Our warm thanks also go to the members of the Local Organizing Committee, for their considerable efforts in organizing the conference. We would like to thank Marc Herbstritt and Michael Wagner (Dagstuhl Publishing) for their technical help. Finally, we would like to thank the members of the Steering Committee for giving us the opportunity to work for TQC. And, of course, all contributors and participants!

November 2014

Steven T. Flammia and Aram W. Harrow

9th Conference on the Theory of Quantum Computation, Communication and Cryptography.
Editors: Steven T. Flammia and Aram W. Harrow
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Local Organizing Commitee

Rahul Jain
Centre for Quantum Technologies, NUS,
Singapore

Hartmut Klauck
Nanyang Technological University and
Centre for Quantum Technologies, NUS,
Singapore

Troy Lee (Chair)
Nanyang Technological University and
Centre for Quantum Technologies, NUS,
Singapore

Miklos Santha
Universite Paris Diderot – Paris 7, France
and Centre for Quantum Technologies, NUS,
Singapore

Evon Tan (Secretariat)
Centre for Quantum Technologies, NUS,
Singapore

Program Commitee

Koenraad Audenaert
University of London, UK

Michael Bremner
University of Technology Sydney, Australia

Jianxin Chen
IQC, University of Waterloo, Canada

Giulio Chiribella
Institute for Interdisciplinary Information Sciences, Tsinghua University, China

Steve Flammia (co-chair)
The University of Sydney, Australia

Sean Hallgren
The Pennsylvania State University, USA

Aram Harrow (chair)
MIT, USA

Masahito Hayashi
Nagoya University, Japan and Centre for Quantum Technologies, NUS, Singapore

Zhengfeng Ji
IQC, University of Waterloo, Canada

Robert Koenig
IQC, University of Waterloo, Canada

Ashley Montanaro
University of Bristol, UK

Marco Piani
University of Waterloo, Canada

Beth Ruskai
Tufts University, USA

Peter Turner
University of Tokyo, Japan

Frank Verstraete
University of Vienna, Austria

John Watrous
University of Waterloo, Canada

Mark Wilde
Louisiana State University, USA

Xiaodi Wu
MIT, USA

Jon Yard
Microsoft Research, USA

9th Conference on the Theory of Quantum Computation, Communication and Cryptography.
Editors: Steven T. Flammia and Aram W. Harrow
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Steering Commitee

Wim van Dam
University of California, Santa Barbara, USA

Yasuhito Kawano
NTT, Japan

Michele Mosca
IQC and University of Waterloo, Canada

Martin Roetteler Microsoft Research, USA

Simone Severini
University College London, UK

Vlatko Vedral
University of Oxford, UK
National University of Singapore, Singapore

9th Conference on the Theory of Quantum Computation, Communication and Cryptography.
Editors: Steven T. Flammia and Aram W. Harrow
LIPICS Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

More Randomness From Noisy Sources*

Jean-Daniel Bancal[1] and Valerio Scarani[1,2]

1 **Centre for Quantum Technologies, National University of Singapore**
3 Science Drive 2, Singapore 117543

2 **Department of Physics, National University of Singapore**
2 Science Drive 3, Singapore 117542

Abstract

Bell experiments can be used to generate private random numbers. An ideal Bell experiment would involve measuring a state of two maximally entangled qubits, but in practice any state produced is subject to noise. Here we consider how the techniques presented in [1] and [2], i. e. using an optimized Bell inequality, and taking advantage of the fact that the device provider is not our adversary, can be used to improve the rate of randomness generation in Bell-like tests performed on singlet states subject to either white or dephasing noise.

1998 ACM Subject Classification J.2 Physical Sciences and Engineering

Keywords and phrases Randomness, Bell inequalities, Trusted provider assumption

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.1

1 Introduction

It is well known that the violation of a Bell inequality rules out the possibility for the outcomes of a Bell-type experiment to be known in advance [3]. Therefore, these outcomes are certifiably unpredictable. Recent works have shown that the uncertainty present in these outcomes can be quantified, thus allowing one to lower bound the number of random bits that can be extracted from a given Bell-type experiment [4, 5].

This possibility has given rise to a variety of randomness-related studies based on a similarly varied set of working assumptions. For instance, many works considered the case in which the adversary (the actor for whom outcomes are to be certifiably unpredictable) is allowed to distribute the quantum state measured by the authorized parties, and keep a purification of this state. Under this assumption, it was shown that *randomness expansion* is possible: if the user holds a secret string of finite length, he can expand it into a longer one [6], or, in principle, even an infinite one [8, 7].

Also, the outcomes observed by the user can be certified to contain some amount of randomness even when the adversary, in addition to distributing the state, holds partial information about the initial random string of the user [9]. This possibility, refered to as *randomness amplification*, was proved recently for initial randomness issued from a generic min-entropy source [10, 11] after a series of partial results [12, 13].

These results show the full power of quantum certification in principle. However, when it comes to realizing such protocols, a number of questions arise. For instance, in which practical situation would one wish to expand a random string if we already have access to a source that can produce initial random strings? Also, in the context of randomness

* This work was supported by the National Research Foundation (partly through the Academic Research Fund Tier 3 MOE2012-T3-1-009) and the Ministry of Education, Singapore.

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 1–6
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

amplification, it is unclear in which meaningful situation the dependence would exist at all but be bounded. If the adversary is allowed to tamper with the devices, for instance, or even to produce them, then he may have hidden some kind of emitter inside the boxes, in order to retrieve all numbers produced by the boxes (which otherwise work as expected). This simple possibility would compromise any certification of randomness.

For these reasons, in the design and assessment of practical realization of randomness protocols , it is very reasonable to work under the assumption that the adversary has no access to the devices used by the authorized partners. This *trusted provider* assumption was already introduced in the context of randomness protocols in [14], where it was shown that it restricts the adversary to hold only classical side information (i. e. he cannot hold a purification of the quantum state). Note that this contrasts with the case of quantum key distribution (QKD): practical QKD also requires the trusted provider assumption, for the same reason as mentione above, however the adversary can still hold a purification in this case since the quantum state passes in his hands. Another consequence mentioned in [14] is that the initial string used by the user to choose settings for his Bell test need not be private, but can be fully known in advance by the adversary. One thus speaks of *randomness generation* in this context.

It was shown in [1] that additional randomness can be certified under the trusted provider assumption compared to that granted by randomness expansion protocols, by extracting randomness from all the settings. Moreover, this same paper as well as [2] demonstrated that Bell-like inequalities that certify more randomness than usual Bell inequalities (like e. g. CHSH) can be derived from knowledge of the full correlations. In this paper, we analyse the advantage provided by these techniques when the quantum state measured by the user is a singlet states mixed either with white or dephasing noise. White noise typically describes the effect of many small errors in a setup whereas dephasing noise is the dominant noise in SPDC-based sources when the pump power is low. The case of white noise was already partially studied in both [1] and [2]. The analysis given here gathers the information presented in both studies and provides a comparison with the dephasing noise case.

Even though our analysis relies on the trusted provider assumption, it is worth noting that some of the results obtained here could also apply to more general adversaries; we refer to [15] for a concise review of adversarial classes relevant to randomness protocols.

For the present paper, we assume that the source emits exactly one pair of particles per unit time and that these are detected with certainty. The case of finite detection efficiency was studied in [1], in absence of noise; when the emission is not heralded, more effects come into play, see e. g. [16].

Another assumption that we make here is that the devices used by the user are i.i.d. and that he can use as many of them as he wants. We thus focus on the rate of randomness generation, defined as the number of random bits generated in each use of the devices.

2 Randomness analysis

We consider here a usual Bell-type experiment performed by a user [3]. At each round, the user chooses some inputs x, y for his two devices to use as measurement settings, and observes their outcomes a, b. The i.i.d. behavior of the boxes follows the quantum conditional probability $P(a,b|x,y) \in \mathcal{Q}$.

In general, these correlation can admit a decomposition $\{q_\lambda, P_\lambda\}$ such that

$$P(ab|xy) = \sum_\lambda q_\lambda P_\lambda(ab|xy) \tag{1}$$

with $q_\lambda \geq 0$, $\sum_\lambda q_\lambda = 1$, $P_\lambda(ab|xy) \in \mathcal{Q}$. When this decomposition is not trivial, by knowing in each round which value of λ corresponds to the realization of the box, the adversary can hold a more precise decription of the box's behavior for that run, as given by $P_\lambda(ab|xy)$.

Following [1, 2], we thus define the adversary's guessing probability on the outcomes observed by the user when using settings x, y and in presence of the decomposition $\{q_\lambda, P_\lambda\}$ as

$$G_{x,y}(\{q_\lambda, P_\lambda\}) = \sum_\lambda q_\lambda \max_{a,b} P_\lambda(ab|xy). \tag{2}$$

The average guessing probability when settings are chosen with probability $p(x, y)$ is then the maximum of

$$G(P) = \sum_{xy} p(x,y) \sum_\lambda q_\lambda \max_{a,b} P_\lambda(ab|xy) \tag{3}$$

over all decompositions (1) compatible with the correlations $P(ab|xy)$.

It was shown in [1] that this quantity can be upper bounded by considering an SDP (Semidefinite Program) relaxation of the set of quantum correlations [17]. In the following section, we thus use this program to evaluate the rate, as given by the min entropy

$$H_{\min}(P) = -\log_2(G(P)), \tag{4}$$

at which random bits are generated in the experiment.

Note that the particular case of this optimization where randomness is extracted from a fixed choice of settings ($p_{xy} = \delta_{x,x_0}\delta_{y,y_0}$), or where the outcomes of different settings are allowed to by guessed with different decomposition, was also presented independently in [2].

In the following we compare three quantities:

1. The rate of randomness obtained from a fixed set of settings as certified by a CHSH violation.
2. The rate of randomness obtained from a fixed set of settings as certified by an optimized Bell-type expression.
3. The rate of randomness obtained when using all settings with the same probability as certified by an optimized Bell-type expression.

Note that here we consider extracting randomness from the pair of outcomes (a, b) rather than from the outcome of a single party. A similar computation could be done by taking only one party's outcome into consideration, but would result in a lower rate. Also, for the first two quantities, the fixed set of settings is chosen as to maximize the rate of randomness.

For all results presented next, the numerical computations were performed using the relaxation of the SDP hierarchy at local level 2 [18].

2.1 White noise

First, let us consider the case in which the measured state is

$$\rho(V) = V|\phi^+\rangle\langle\phi^+| + (1-V)\mathbb{1}/4, \tag{5}$$

for some visibility V. The settings which provide the largest violation $2\sqrt{2}V$ of the CHSH inequality

$$S = \langle A_0B_0\rangle + \langle A_0B_1\rangle + \langle A_1B_0\rangle - \langle A_1B_1\rangle \leq 2, \tag{6}$$

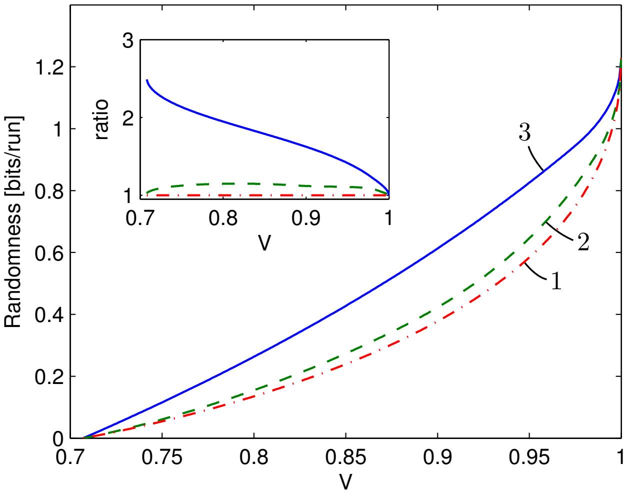


Figure 1 Rate of private randomness generation certified by the measure 1, 2 and 3 for a singlet state mixed with white noise. The inset presents the ratio of the curves to the lowest one.

where A_x, B_y are Alice's and Bob's observables, are the same for all V:

$$A_0 = \sigma_z, \quad A_1 = \sigma_x, \quad B_y = \frac{\sigma_z + (-1)^y \sigma_x}{2}. \tag{7}$$

We thus computed for this state and settings the three different rates of randomness mentioned above. The result is presented in Figure 1.

The randomness rate obtained in case 2 (middle curve) can be certified with the help of the following Bell expression:

$$\alpha\langle A_0 B_0\rangle + \langle A_0 B_1\rangle + \langle A_1 B_0\rangle - \beta\langle A_1 B_1\rangle, \tag{8}$$

where the values of α and β depend on V (see [2] for a description of this dependence). The inset in Figure 1 shows that the advantage provided by using this optimized Bell expression is however quite limited.

The largest amount of randomness is obtained in case 3 when considering the outcomes observed when all settings are used with the same probability (i. e. $p(x, y) = 1/4$). As mentioned in [1], the improvement, of the order of a factor of 2, is certified with the usual CHSH inequality.

2.2 Dephasing noise

Second, we consider measurement of the state

$$\rho(p) = p|\phi^+\rangle\langle\phi^+| + (1-p)(|00\rangle\langle 00| + |11\rangle\langle 11|). \tag{9}$$

The optimal violation of the CHSH inequality by this state, $S = 2\sqrt{1+p^2}$, is provided by using the following settings [19]:

$$\begin{aligned} &A_0 = \sigma_z, \quad A_1 = \sigma_x, \\ &B_y = \cos\chi\,\sigma_z + (-1)^y \sin\chi\,\sigma_x, \end{aligned} \tag{10}$$

with $\chi = \arctan(p)$.

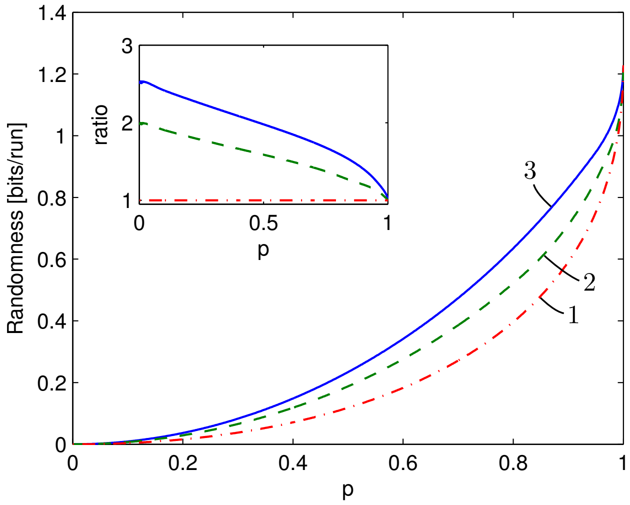


Figure 2 Rate of private randomness generation certified by the measure 1, 2 and 3 for a singlet state mixed with dephasing noise. The inset presents the ratio of the curves to the lowest one.

The three randomness rates obtained with these state and settings are presented in Figure 2. Similarly to the previous case, strictly more randomness can be certified in case 3 than in case 2, and in case 2 than in case 1. The inequality that certifies the largest amount of randomness is again CHSH in case 3, and a different inequality in case 2. One can check that this inequality, however, beyond being a correlation inequality presents no special symmetry. In particular, it is not of the form (8). Nevertheless, we note that when the randomness is extracted from a single set of settings, using an optimized inequality provides a larger advantage for this dephasing noise than it did in the case of mixture with white noise (as shown in the inset of Figure 2).

3 Conclusion

We have presented an application of the techniques presented in [1, 2] to the case where the measured state is a singlet mixed with either white noise or dephasing noise. While a significant advantage in terms of randomness rate can be obtained in both cases when randomness is extracted uniformly from all settings, the advantage for extraction from a fixed choice of settings is much more significant in the case of dephasing noise.

In a practical experiment, characteristics of both white and dephasing noise are expected to appear [20], as well as various other kind of noises and imperfections [16]. The present analysis is not meant to exhaust all the parameter space of a realistic experiment; but it should be clear that the techniques used here can be extended to describe experiments with all their features.

We have focused here on the asymptotic rate of randomness generation. It would be interesting to extend our analysis to take into account finite statistics, maybe in a way similar to [14] or [21]. This would allow one to quantify how many random bits can be extracted from a Bell experiment which involves only a finite number of rounds.

References

1 J.-D. Bancal, L. Sheridan, V. Scarani, New J. Phys. **16**, 033011 (2014).

2 O. Nieto-Silleras, S. Pironio, J. Silman, New J. Phys. **16**, 013035 (2014).

3 N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).

4 R. Colbeck, Ph. D. thesis, University of Cambridge, arXiv:0911.3814 (2006).

5 S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, C. Monroe, Nature **464**, 1021 (2010).

6 U. V. Vazirani, T. Vidick, arXiv:1111.6054

7 C. A. Miller, Y. Shi, arXiv:1402.0489

8 M. Coudron, H. Yuen, arXiv:1310.6755

9 R. Colbeck, R. Renner, Nature Physics **8**, 450–454 (2012)

10 J. Bouda, M. Pawlowski, M. Pivoluska, M. Plesch, arXiv:1402.0974

11 K.-M. Chung, Y. Shi, X. Wu, arXiv:1402.4797

12 R. Gallego, L. Masanes, G. de la Torre, C. Dhara, L. Aolita, A. Acin, Nature Communications **4**, 2654 (2013)

13 T. P. Le, L. Sheridan, V. Scarani, Phys. Rev. A **87**, 062121 (2013).

14 S. Pironio and S. Massar, Phys. Rev. A **87**, 012336 (2013).

15 Y. Z. Law, T.P. Le, J-D. Bancal and V. Scarani, arXiv:1401.4243.

16 V. Caprara-Vivoli, P. Sekatski, J.-D. Bancal, C.C.W. Lim, B.G. Christensen, A. Martin, R. T. Thew, H. Zbinden, N. Gisin, N. Sangouard, arXiv:1405.1939

17 M. Navascues, S. Pironio, A. Acín, Phys. Rev. Lett. **98**, 010401 (2007).

18 T. Moroder, J-D. Bancal, Y-C. Liang, M. Hofmann, O. Gühne, Phys. Rev. Lett. **111**, 030501 (2013).

19 R. Horodecki, M. Horodecki, P. Horodecki, Phys. Lett. A **200**, 340 (1995)

20 J. Kofler, S. Ramelow, M. Giustina, A. Zeilinger, arXiv:1307.6475.

21 Y. Zhang, S. Glancy, E. Knill, Phys. Rev. A **84**, 062118 (2011).

Exact Classical Simulation of the GHZ Distribution

Gilles Brassard[1], Luc Devroye[2], and Claude Gravel[3]

1 **Université de Montréal, CIFAR and ETH-ITS,** brassard@iro.umontreal.ca
2 **McGill University,** lucdevroye@gmail.com
3 **Université de Montréal,** claude.gravel@bell.net

Abstract

John Bell has shown that the correlations entailed by quantum mechanics cannot be reproduced by a classical process involving non-communicating parties. But can they be simulated with the help of bounded communication? This problem has been studied for more than twenty years and it is now well understood in the case of bipartite entanglement. However, the issue was still widely open for multipartite entanglement, even for the simplest case, which is the tripartite Greenberger–Horne–Zeilinger (GHZ) state. We give an exact simulation of arbitrary independent von Neumann measurements on general n-partite GHZ states. Our protocol requires $O(n^2)$ bits of expected communication between the parties, and $O(n \log n)$ expected time is sufficient to carry it out in parallel. Furthermore, we need only an expectation of $O(n)$ independent unbiased random bits, with no need for the generation of continuous real random variables nor prior shared random variables. In the case of equatorial measurements, we improve earlier results with a protocol that needs only $O(n \log n)$ bits of communication and $O(\log^2 n)$ parallel time. At the cost of a slight increase in the number of bits communicated, these tasks can be accomplished with a constant expected number of rounds.

1998 ACM Subject Classification J.2 Physical Sciences and Engineering: Physics

Keywords and phrases Entanglement simulation, Greenberger–Horne–Zeilinger (GHZ) state, Multiparty entanglement, von Neumann's rejection algorithm, Knuth-Yao's sampling algorithm

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.7

1 Introduction

The issue of non-locality in quantum physics was raised in 1935 by Einstein, Podolsky and Rosen when they introduced the notion of entanglement [10]. Thirty years later, Bell proved that the correlations entailed by entanglement cannot be reproduced by classical local hidden variable theories between noncommunicating parties [2]. This momentous discovery led to the natural question of *quantifying* quantum non-locality.

A natural quantitative approach to the non-locality inherent in a given entangled quantum state is to study the amount of resources that would be required in a purely classical theory to reproduce exactly the probabilities corresponding to measuring this state. More formally, we consider the problem of *sampling* the joint discrete probability distribution of the outcomes obtained by people sharing this quantum state, on which each party applies locally some measurement on his share. Each party is given a description of his own measurement but not informed of the measurements assigned to the other parties. This task would be easy (for a theoretician!) if the parties were indeed given their share of the quantum state, but they are not. Instead, they must *simulate* the outcome of these measurements without any quantum resources, using as little *classical communication* as possible.

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 7–23
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

This conundrum was introduced by Maudlin in 1992 in the simplest case of linear polarization measurements at arbitrary angles on the two photons that form a Bell state [17]. Similar concepts were reinvented independently years later by other researchers [5, 20]. This led to a series of results, culminating with the protocol of Toner and Bacon to simulate arbitrary von Neumann measurements on a Bell state with a single bit of communication in the worst case [21]. Later, Regev and Toner extended this result by giving a simulation of the correlations entailed by arbitrary binary von Neumann measurements on arbitrary bipartite states of any dimension using two bits of communication, also in the worst case [19]. Inspired by Ref. [20], Cerf, Gisin and Massar showed that the effect of an arbitrary pair of positive-operator-valued measurements (POVMs) on a Bell state can also be simulated with a bounded amount of expected communication [8]. A more detailed early history of the simulation of quantum entanglement can be found in Ref. [4, Sect. 6].

All this prior work is concerned strictly with the simulation of *bipartite* entanglement. Much less is known when it comes to simulating multipartite entanglement with classical communication, a topic that is still teeming with major open problems. Consider the simplest case, which is the simulation of independent arbitrary von Neumann measurements on the tripartite GHZ state, named after Greenberger, Horne and Zeilinger [14], which we shall denote $|\Psi_3\rangle = \frac{1}{\sqrt{2}}|000\rangle + \frac{1}{\sqrt{2}}|111\rangle$, or more generally on its n-partite generalization $|\Psi_n\rangle = \frac{1}{\sqrt{2}}|0^n\rangle + \frac{1}{\sqrt{2}}|1^n\rangle$.

The easiest situation arises in the special case of *equatorial* measurements (defined in Section 2) on the GHZ state because all the marginal probability distributions obtained by tracing out one or more of the parties are uniform. Hence, it suffices in this case to simulate the n-partite correlation. Once this has been achieved, all the marginals can easily be made uniform [11]. Making the best of this observation, Bancal, Branciard and Gisin have given a protocol to simulate equatorial measurements on the tripartite and fourpartite GHZ states at an expected cost of 10 and 20 bits of communication, respectively [1]. Later on, Branciard and Gisin improved this in the tripartite case with a protocol using 3 bits of communication in the worst case [3]. The simulation of equatorial measurements on $|\Psi_n\rangle$ for $n \geq 5$ was handled subsequently by Brassard and Kaplan in a paper published in the 2012 edition of this Conference on Theory of Quantum Computation, Communication and Cryptography, with an expected cost of $O(n^2)$ bits of communication [6]. This was the best result obtained until now on this line of work.

Despite substantial effort, the case of *arbitrary* von Neumann measurements, even on the original tripartite GHZ state $|\Psi_3\rangle$, was still wide open. Here, we solve this problem in the general case of the simulation of the n-partite GHZ state $|\Psi_n\rangle$, for any n, under the *random bit model* introduced in 1976 by Knuth and Yao [16], in which the only source of randomness comes from the availability of independently distributed unbiased random bits. Furthermore, we have no needs for prior shared random variables between the parties. An expected number of $6n + 17$ perfect random bits suffices to carry out our simulation. The expected communication cost is $O(n^2)$ bits, but only $O(n \log n)$ *time* if we count one step for sending bits in parallel according to a realistic scenario in which no party has to send or receive more than one bit in any given step. Furthermore, in the case of equatorial measurements, we improve the earlier best result [6] with an expected communication cost of only $O(n \log n)$ bits and $O(\log^2 n)$ parallel time. At the cost of a slight increase in the number of bits communicated and the number of required random bits, these tasks can be accomplished with a constant expected number of rounds.

More formally, the quantum task that we want to simulate is as follows. Each party i holds one qubit from state $|\Psi_n\rangle = \frac{1}{\sqrt{2}}|0^n\rangle + \frac{1}{\sqrt{2}}|1^n\rangle$ and is given the description of a von Neumann measurement M_i. By local operations, they collectively perform $\otimes_{i=1}^{n} M_i$ on $|\Psi_n\rangle$, thus obtaining one outcome each, say $b_i \in \{-1,+1\}$, which is their output. The joint probability distribution $p(b)$ of the b_i's is defined by the joint set of measurements (see Section 2). Our purpose is to sample *exactly* this joint probability distribution by a purely classical process that involves no prior shared random variables and as little communication as possible. Our complete solution builds on four ingredients: (1) Gravel's decomposition of $p(b)$ as a convex combination of two sub-distributions [12, 13]; (2) Knuth and Yao's algorithm to sample exactly probability distributions assuming only a source of unbiased identically independently distributed (i.i.d.) bits [16]; (3) the *universal method of inversion* [9, for instance]; and (4) our own distributed version of the classic *von Neumann's rejection algorithm* [18].

We define precisely our problem in Section 2 and we formulate our convex decomposition of the GHZ distribution, which is the key to its simulation. Then, we explain how to sample according to a Bernoulli distribution even when only approximations to the distribution's parameter are available. We also explain how the classic von Neumann rejection algorithm can be used to sample in the sub-distributions defined by our convex decomposition. However, little attention is paid in Section 2 to the fact that the various parameters that define the joint distribution are not available in a single place. Section 3 is concerned with the communication complexity issues. It culminates with a complete protocol to solve our problem, as well as its complete analysis. This is followed by variations on the theme, in which we consider a parallel model of communication, an expected bounded-round solution, and improvements on the prior art for the simulation of equatorial measurements. We conclude with a discussion and open problems in Section 4.

2 Sampling exactly the GHZ distribution in the random bit model

Any von Neumann measurement on a single qubit can be conveniently represented by a point on the surface of a three-dimensional sphere, known as the Bloch sphere, whose spherical coordinates can be specified by an *azimuthal* angle $\theta \in [0, 2\pi)$ and an *elevation* angle $\varphi \in [-\pi/2, \pi/2]$. These parameters defines a Hermitian idempotent operator

$$M = x\,\sigma_1 + y\,\sigma_2 + z\,\sigma_3 = \begin{pmatrix} \sin\varphi & e^{-\imath\theta}\cos\varphi \\ e^{\imath\theta}\cos\varphi & -\sin\varphi \end{pmatrix},$$

where $x = \cos\theta\cos\varphi$, $y = \sin\theta\cos\varphi_j$, $z = \sin\varphi$, and σ_1, σ_2 and σ_3 are the Pauli operators. In turn, this operator defines a measurement in the usual way, which we shall also call M for convenience, whose outcome is one of its eigenvalues $+1$ or -1. The azimuthal angle θ represents the equatorial part of the measurement and the elevation angle φ represents its real part. A von Neumann measurement is said to be *equatorial* when its elevation angle $\varphi = 0$ vanishes and it is said to be *in the computational basis* when $\varphi = \pm\pi/2$.

Consider a set of n von Neumann single-qubit measurements M_j, represented by their parameters (θ_j, φ_j), $1 \le j \le n$. This set of operators defines a joint measurement $M = \otimes_{j=1}^{n} M_j$. In turn, this measurement defines a probability distribution p, which we shall call the *GHZ distribution*, on the set $\{-1,+1\}^n$. This distribution corresponds to the probability of all possible outcomes when the n-partite GHZ state $|\Psi_n\rangle = \frac{1}{\sqrt{2}}|0^n\rangle + \frac{1}{\sqrt{2}}|1^n\rangle$ is measured according to M.

It is shown in [12, 13], albeit in the usual computer science language in which von Neumann measurements are presented as a unitary transformation followed by a measurement in the computational basis, that the probability $p(b)$ of obtaining $b = (b_1, \ldots, b_n)$ in $\{-1,+1\}^n$ can be decomposed as

$$p(b) = \cos^2\left(\tfrac{\theta}{2}\right) p_1(b) + \sin^2\left(\tfrac{\theta}{2}\right) p_2(b)\,, \text{ where } \theta = \textstyle\sum_{j=1}^n \theta_j \text{ and} \tag{1}$$

$$p_1(b) = \frac{1}{2}\big(\mathrm{a}_1(b) + \mathrm{a}_2(b)\big)^2, \qquad p_2(b) = \frac{1}{2}\big(\mathrm{a}_1(b) - \mathrm{a}_2(b)\big)^2, \tag{2}$$

$$\mathrm{a}_1(b) = \prod_{j=1}^n \cos\left(\tfrac{1}{2}\left(\varphi_j - \tfrac{\pi}{2} b_j\right)\right), \qquad \mathrm{a}_2(b) = \prod_{j=1}^n -\sin\left(\tfrac{1}{2}\left(\varphi_j - \tfrac{\pi}{2} b_j\right)\right). \tag{3}$$

Hence, we see that distribution $p(b)$ is a convex combination of sub-distributions $p_1(b)$ and $p_2(b)$, in which the coefficients $\cos^2(\theta/2)$ and $\sin^2(\theta/2)$ depend only on the equatorial part of the measurements, whereas the sub-distributions depend only on their real part. Furthermore, the squares of a_1 and a_2 are themselves discrete probability distributions.

Sampling p is therefore a matter of sampling a Bernoulli distribution with defining parameter $\cos^2(\theta/2)$ before sampling either p_1 or p_2, whichever is the case. Notice that sampling p_2 reduces to sampling p_1 if, say, we replace φ_1 by $\varphi_1 + 2\pi$. As we shall see, full knowledge of the parameters is not required to sample p *exactly*. We shall see in subsection 2.1 how to sample a Bernoulli distribution with an arbitrary $p \in [0,1]$ as parameter (not the same p as our probability distribution for GHZ) using a sequence of approximants converging to p and using an expected number of only five unbiased identically independently distributed (i.i.d.) random bits. Subsequently, we shall see in subsection 2.2 how to sample p_1 by modifying von Neumann's rejection algorithm in a way that it uses sequences of approximants and unbiased i.i.d. random bits. For simulating exactly the GHZ distribution, an expected number of $6n + 17$ perfect random bits is sufficient.

2.1 Sampling a Bernoulli distribution

Assume that only a random bit generator is available to sample a given probability distribution and that the parameters that specify this distribution are only accessible as follows: we can ask for any number of bits of each parameter, but will be charged one unit of cost per bit that is revealed. We shall also be charged for each random bit requested from the generator

To warm up to this conundrum, consider the problem of generating a Bernoulli random variable Y with parameter $p \in [0,1]$. If $U = 0.U_1U_2\ldots$ is the binary expansion of a uniform $[0,1)$ random variable, i.e. $U_1, U_2, \ldots$ is our source of unbiased independent random bits, and if $p = 0.p_1p_2\ldots$ is the binary expansion of p (in case $p = 1$ we can proceed as if it were $0.p_1p_2\ldots$ with each $p_i = 1$), we compare bits U_i and p_i for $i = 1, 2, \ldots$ until for the first time $U_i \neq p_i$. Then, if $U_i = 0 < p_i = 1$, we return $Y = 1$, and if $U_i = 1 > p_i = 0$, we return $Y = 0$. It is clear that $Y = 1$ if and only if $U < p$. Therefore, Y is Bernoulli(p). The expected number of bits required from p is precisely 2. The expected number of bits needed from our random bit source is also 2.

Now, suppose that the parameter p defining our Bernoulli distribution is given by $p = \cos^2(\theta/2)$, as in the case of our decomposition of the GHZ distribution. None of the parties can know θ precisely since it is distributed as a sum of θ_i's, each of which is known only by one individual party. If we could obtain as many physical bits of p as needed (although the expected number of required bits is as little as 2), we could use the idea given above in order to sample according to this Bernoulli distribution. However, it is not possible in general to know even the first bit of p given any fixed number of bits of the θ_i's. (For instance, if θ is arbitrarily close to $\pi/2$, we need arbitrarily many bits of precision about it before we can tell if the first bit in the binary expansion of $\cos^2(\theta/2)$ is 0 or 1). Nevertheless, we can use *approximations* of p, rather than *truncations*, which in turn can come from approximations of the θ_i's.

▶ **Definition 1.** A *k-bit approximation* of a quantity v is any $\hat{v}$ such that $|v - \hat{v}| \le 2^{-k}$. A special case of k-bit approximation is the *k-bit truncation* $\hat{v} = \lfloor v2^k \rfloor / 2^k$. For convenience, we sometimes use the shorthands *k-approximation* and *k-truncation.* Note that the value of k corresponds to the number of bits in the fractional part, without limitation on the size of the integer part.

We postpone to Section 3.1 the detail of how these approximations can be obtained in a distributed setting. For the moment, assume that, for any k, we can obtain $p(k)$ so that $|p(k)-p| \le 1/2^k$. Then, setting $U(k) = 0.U_1 \ldots U_k$, we have that $U \le p$ if $U(k) \le p(k) - 2/2^k$ whereas $U \ge p$ if $U(k) \ge p(k) + 1/2^k$. Thus, one can check if $U < p$ by generating only as many bits of U and increasingly good approximations of p as needed. These ideas are formalized in Algorithm 1. It is elementary to verify that the Y generated by this algorithm is Bernoulli(p) because $\mathbf{P}\{U < p\} = p$ if U is a continuous uniform random variable on $(0, 1)$.

Algorithm 1 Sampling a Bernoulli random variable with approximate defining parameter

```
Set $k \leftarrow 1$
Set $U(0) \leftarrow 0$
repeat forever
    Generate an i.i.d. unbiased bit $U_k$
    Compute $U(k) \leftarrow U(k-1) + U_k/2^k$ {hence $U(k) = 0.U_1 \ldots U_k$}
    Obtain $p(k)$ so that $|p(k) - p| \le 1/2^k$
    if $U(k) \le p(k) - 2/2^k$ then
        return $Y = 1$
    else if $U(k) \ge p(k) + 1/2^k$ then
        return $Y = 0$
    else
        $k \leftarrow k + 1$
    end if
end repeat
```

The number of iterations before Algorithm 1 returns a value, which is also its required number of independent unbiased random bits, is a random variable, say K. We have seen above that $\mathbf{E}\{K\}$, the expected value of K, would be exactly 2 if we could generate arbitrarily precise truncations of p. But since we can only obtain arbitrarily precise approximations instead, which is why we needed Algorithm 1 in the first place, we shall have to pay the price of a small increase in $\mathbf{E}\{K\}$.

$$\mathbf{P}\{K > k\} \le \mathbf{P}\left\{|U(k) - p(k)| \le \frac{2}{2^k}\right\} \le \mathbf{P}\left\{|U - p| \le \frac{4}{2^k}\right\} \le \frac{8}{2^k}.$$

Therefore,

$$\mathbf{E}\{K\} = \sum_{k=0}^{\infty} \mathbf{P}\{K > k\} \le \sum_{k=0}^{\infty} \min\left(1, \frac{8}{2^k}\right) = 5.$$

2.2 Sampling p_1 (or p_2) in the random bit model

As mentioned already, it suffices to concentrate on p_1 since one can sample p_2 in exactly the same way provided one of the angles φ_i is replaced by $\varphi_i + 2\pi$: this introduces the required minus sign in front of a_2 to transform p_1 into p_2. Let us define

$$\alpha_j = \cos\bigl(\tfrac{1}{2}(\varphi_j - \tfrac{\pi}{2})\bigr) = \sin\bigl(\tfrac{1}{2}(\varphi_j + \tfrac{\pi}{2})\bigr) \quad \text{and} \quad \beta_j = \cos\bigl(\tfrac{1}{2}(\varphi_j + \tfrac{\pi}{2})\bigr) = -\sin\bigl(\tfrac{1}{2}(\varphi_j - \tfrac{\pi}{2})\bigr)\,. \quad (4)$$

Consider n Rademacher[1] random variables B_j that take value -1 with probability β_j^2 and $+1$ with complementary probability α_j^2. The random vector with independent components given by $(B_1, \ldots, B_n)$ is distributed according to

$$q_1(b) \stackrel{\text{def}}{=} \prod_{j \in F_b} \beta_j^2 \prod_{j \in G_b} \alpha_j^2\,,$$

where $F_b = \{j \mid b_j = -1\}$ and $G_b = \{j \mid b_j = +1\}$ for all $b = (b_1, \ldots, b_n) \in \{-1, +1\}^n$. It is easy to verify that $q_1(b) = \mathrm{a}_1^2(b)$ for all b, where a_1 is given in Equation (3). Similarly, the random vector with independent components given by $(-B_1, \ldots, -B_n)$ is distributed according to

$$q_2(b) \stackrel{\text{def}}{=} \prod_{j \in F_b} \alpha_j^2 \prod_{j \in G_b} \beta_j^2 = \mathrm{a}_2^2(b)\,.$$

The key observation is that both q_1 and q_2 can be sampled without any needs for communication because each party j knows his own parameters α_j^2 and β_j^2, which is sufficient to draw independently according to local Rademacher random variable B_j or $-B_j$. Moreover, a single unbiased independent random bit s drawn by a designated party suffices to sample collectively from distribution $q = \frac{q_1 + q_2}{2}$, provided this bit is transmitted to all parties: everybody samples according to q_1 if $s = 0$ or to q_2 if $s = 1$. Now, It follows from Equation (2) that $p_1(b) + p_2(b) = \mathrm{a}_1^2(b) + \mathrm{a}_2^2(b) = q_1(b) + q_2(b)$ for all $b \in \{-1, +1\}^n$, and therefore $p_1(b) \le q_1(b) + q_2(b) = 2q(b)$.

The relevance of all these observations is that we can apply von Neumann's rejection algorithm [18] to sample $p_1(b)$ since it is bounded by a small constant (2) times an easy-to-draw probability distribution (q). For the moment, we assume once again the availability of a continuous uniform random generator, which we shall later replace by a source of unbiased independent random bits. We also assume for the moment that we can compute the α_i's, $p_1(b)$, $q_1(b)$ and $q_2(b)$ exactly. This gives rise to Algorithm 2.

Algorithm 2 Sampling p_1 using von Neumann's rejection algorithm

```
repeat
    Generate $U$ uniformly on $[0, 1)$
    Generate independent Rademacher random variables $B_1, \ldots, B_n$
    with parameters $\alpha_1^2, \ldots, \alpha_n^2$
    Generate an unbiased independent random bit $S$
    if $S = 1$ then
        set $B \leftarrow (B_1, \ldots, B_n)$
    else
        set $B \leftarrow (-B_1, \ldots, -B_n)$
    end if
until $(q_1(B) + q_2(B))\, U \le p_1(B)$
```

By the general principle of von Neumann's rejection algorithm, probability distribution p_1 is successfully sampled after an expected number of 2 iterations round the loop because

[1] A Rademacher random variable is like a Bernoulli, except that it takes value ±1 rather than 0 or 1.

$p_1(b) \leq 2q(b)$ for all $b \in \{-1,+1\}^n$. Within one iteration, 2 expected independent unbiased random bits suffice to generate each of the n Rademacher random variables by a process similar to what is explained in the second paragraph of Section 2.1. Hence an expected total of $2n+1$ random bits are needed each time round the loop for an expected grand total of $4n+2$ bits to sample p_1. But of course, this does not take account of the (apparent) need to generate continuous uniform $[0,1)$ random variable U. It follows that the expected total amount of work required by Algorithm 2 is $O(n)$, provided we count infinite real arithmetic at unit cost. Furthermore, the time taken by this algorithm, divided by n, is stochastically smaller than a geometric random variable with constant mean, so its tail is exponentially decreasing.

Now, we modify and adapt this algorithm to eliminate the need for the continuous uniform U (and hence its generation), which is not allowed in the random *bit* model. Furthermore, we eliminate the need for infinite real arithmetic and for the exact values of $q_1(B)$, $q_2(B)$ and $p_1(B)$, which would be impossible to obtain in our distributed setting since the parameters needed to compute these values are scattered among all parties, and replace them with approximations—we postpone to Section 3.2 the issue of how these approximations can be computed. (On the other hand, arbitrarily precise values of the α_i's *are* available to generate independent Rademacher random variables with these parameters because each party will be individually responsible to generate his own Rademacher.)

In each iteration of Algorithm 2, we generated a pair (U,B). However, we did not really need U: we merely needed to generate a Bernoulli random variable Y for which

$$\mathbf{P}\{Y=1\} = \mathbf{P}\left\{(q_1(B)+q_2(B))\,U \leq p_1(B)\right\}.$$

For this, we adapt the method developed for Algorithm 1. Again, we denote by $U(k)$ the k-bit truncation of U, so that $U(k) < U < U(k)+2^{-k}$, except with probability 0. Furthermore, we use L_k (L for *left*) and R_k (R for *right*) to denote k-bit approximations of $q_1(B)+q_2(B)$ and $p_1(B)$, respectively, so that $|L_k - \big(q_1(B)+q_2(B)\big)| \leq 2^{-k}$ and $|R_k - p_1(B)| \leq 2^{-k}$. Then using ε_k to denote arbitrary real numbers in the interval $(-1,1)$,

$$\begin{aligned} |U(k)L_k - U(q_1(B)+q_2(B))| &= \left|U(k)L_k - U\Big(L_k + \frac{\varepsilon_k}{2^k}\Big)\right| \\ &= \left|(U(k)-U)L_k - \frac{U\varepsilon_k}{2^k}\right| \;\leq\; \frac{L_k}{2^k} + \frac{1}{2^k} \;\leq\; \frac{3}{2^k}\,. \end{aligned}$$

Similarly, $|R_k - p_1(B)| \leq \dfrac{1}{2^k}$.

Thus, we know that $Y=1$ whenever $U(k)L_k + 3/2^k < R_k - 1/2^k$, whereas $Y=0$ whenever $U(k)L_k - 3/2^k > R_k + 1/2^k$. Otherwise, we are in the uncertainty zone and we need more bits of U, $q_1(B)+q_2(B)$ and $p_1(B)$ before we can decide on the value of Y. This is formalized in Algorithm 3 (on next page).

It follows from the above discussion that this algorithm can be used to sample random variable Y, which is used as terminating condition in Algorithm 2, in order to eliminate the need for the generation of a continuous uniform random variable $U \in [0,1)$ and for the precise values of $q_1(B)$, $q_2(B)$ and $p_1(B)$. Since $L_k \to q_1(B)+q_2(B)$ and $R_k \to p_1(B)$ as $k \to \infty$, Algorithm 3 halts with probability 1. Let K be a random variable corresponding to the value of k upon exiting from the **repeat forever** loop in the algorithm, which is the number of times round the loop and hence the number of bits needed from U and the precision in $q_1(B)+q_2(B)$ and $p_1(B)$ required in order to sample correctly Bernoulli random variable Y. Next, we calculate an upper-bound on $\mathbf{E}\{K\}$, the expected value of K.

Algorithm 3 Generator for the stopping condition in Algorithm 2

```
Note: $B \in \{-1,+1\}^n$ is given to the algorithm, generated according to $\frac{q_1+q_2}{2}$
Set $k \leftarrow 1$
Set $U(0) \leftarrow 0$
repeat forever
    Generate an i.i.d. unbiased bit $U_k$
    Compute $U(k) \leftarrow U(k-1) + U_k/2^k$ {hence $U(k) = 0.U_1 \ldots U_k$}
    Compute $L_k$ and $R_k$ from $B$
    if $U(k)\,L_k - R_k < -\frac{4}{2^k}$ then
        return $Y = 1$
    else if $U(k)\,L_k - R_k > \frac{4}{2^k}$ then
        return $Y = 0$
    else
        $k \leftarrow k + 1$
    end if
end repeat
```

If the algorithm has not yet halted after having processed $U(k)$, L_k and R_k, then we know that

$$\begin{aligned}
&|U(q_1(B)+q_2(B)) - p_1(B)| \\
&\quad = \left|\big(U(q_1(B)+q_2(B)) - U(k)L_k\big) + \big(R_k - p_1(B)\big) + \big(-R_k + U(k)L_k\big)\right| \\
&\quad \leq |U(q_1(B)+q_2(B)) - U(k)L_k| + |R_k - p_1(B)| + |R_k - U(k)L_k| \\
&\quad \leq \frac{3}{2^k} + \frac{1}{2^k} + \frac{4}{2^k} \quad = \quad \frac{8}{2^k}\,.
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbf{P}\{K > k \mid B\} &\leq \mathbf{P}\{|U(q_1(B)+q_2(B)) - p_1(B)| \leq 8/2^k \mid B\} \\
&= \mathbf{P}\left\{U \in \left(\frac{p_1(B)}{2q(B)} - \frac{1}{2}\frac{8}{2^k}\frac{1}{q(B)}\ ,\ \frac{p_1(B)}{2q(B)} + \frac{1}{2}\frac{8}{2^k}\frac{1}{q(B)}\right)\right\} \\
&\leq \frac{8}{2^k}\frac{1}{q(B)}\,.
\end{aligned}$$

Thus, using k_0 to denote $\left\lceil 3 + \log_2\left(\frac{1}{q(B)}\right)\right\rceil$,

$$\begin{aligned}
\mathbf{E}\{K \mid B\} &= \sum_{k=0}^{\infty} \mathbf{P}\{K > k \mid B\} \\
&\leq \sum_{k=0}^{\infty} \min\left(1, \frac{8}{2^k q(B)}\right) \\
&\leq \sum_{k<k_0} 1 + \sum_{k \geq k_0} \frac{8}{2^k q(B)} \\
&\leq 5 + \log_2\left(\frac{1}{q(B)}\right) \qquad \text{(this step requires a messy calculation).}
\end{aligned}$$

Now, we uncondition in order to conclude:

$$\begin{aligned}\mathbf{E}\{K\} &\le 5 + \sum_{b\in\{-1,+1\}^n} q(b)\log_2\left(\frac{1}{q(b)}\right)\\ &= H(q)+5 && (5)\\ &\le n+5\,, && (6)\end{aligned}$$

where $H(q)$ denote the entropy of distribution $q = \frac{q_1+q_2}{2}$.

3 Communication complexity of sampling

In this section, we consider the case in which the sampler of the previous section no longer has full knowledge of the GHZ distribution to be simulated. The sampler, whom we call *the leader* in a distributed setting, has to communicate through classical channels in order to obtain partial knowledge of the parameters belonging to the other parties. Partial knowledge results in approximation of the parameters involved in sampling the GHZ distribution, but, as we saw in the previous section, we know how to sample *exactly* in the random bit model using such approximations.

3.1 Sampling a Bernoulli distribution whose parameter is distributed

In order to sample the GHZ distribution, we know from Section 2 that we must first sample the Bernoulli distribution with parameter $\cos^2(\theta/2)$, where $\theta = \sum_{j=1}^n \theta_j$. Let us say that the leader is party number 1. Since he knows only θ_1, he must communicate with the other parties to obtain partial knowledge about θ_i for $i \ge 2$. The problem of sampling a Bernoulli distribution with probability $\cos^2(\theta/2)$ reduces to learning the sum θ with sufficient precision in order to use Algorithm 1.

The problem of computing a k-bit approximation of $\cos^2(\theta/2) = \cos^2\bigl(\sum_{i=1}^n \theta_i/2\bigr)$ is relatively easy. Define $\vartheta = \theta/2$ and $\vartheta_i = \theta_i/2$ for each i. If the leader obtains an ℓ-bit approximation $\hat{\vartheta}_i$ of each ϑ_i, $i \ge 2$, and if we define $\hat{\vartheta} = \sum_{i=1}^n \hat{\vartheta}_i$, we need to find the value of ℓ for which $\cos^2(\hat{\vartheta})$ is a k-bit approximation of $\cos^2(\vartheta)$. It is an elementary exercise in Taylor series expansion to verify that $|\cos^2(\vartheta) - \cos^2(\hat{\vartheta})| \le n/2^\ell$. Hence, it suffices to choose $\ell = k + \lceil \log_2 n \rceil$ in order to conclude as required that $|\cos^2(\vartheta) - \cos^2(\hat{\vartheta})| \le 2^{-k}$. Taking into account the integer part of each ϑ_i, which must also be communicated, and remembering that $0 \le \vartheta_i \le 2\pi$ since it is an angle [2], the required number of communicated bits in the sequential model is therefore $(n-1)(\ell+3) = (n-1)\bigl(3+k+\lceil \log_2 n\rceil\bigr)$, which is $O(kn + n\log n)$. In our case, the expected value of k is bounded by 5 (see the analysis of the Bernoulli sampling Section 2.1), so that this operation requires an expected communication of $O(n \log n)$ bits.

3.2 Approximating a product of bounded numbers

Once the leader has produced a bit Z with probability $\cos^2(\theta/2)$, he samples either p_1 or p_2, depending on whether he got $Z=0$ or $Z=1$. The problem of sampling p_2 reduces to sampling p_1 if the leader replaces his own φ_1 with $\varphi_1 + 2\pi$; thus we concentrate on sampling p_1. Of course, the leader does not know φ_i for $i \ge 2$. This problem reduces

[2] Actually, $0 \le \vartheta_i \le \pi$ since ϑ_i is a *half* angle and one fewer bit is needed to communicate its integer part, but we prefer to consider here the more general case of approximating the cosine square of a sum of arbitrary angles.

to learning with sufficient precision the products $\mathrm{a}_1(B) = \prod_{j=1}^{n} \cos\left(\frac{1}{2}\left(\varphi_j - \frac{\pi}{2}B_j\right)\right)$ and $\mathrm{a}_2(B) = \prod_{j=1}^{n} - \sin\left(\frac{1}{2}\left(\varphi_j - \frac{\pi}{2}B_j\right)\right)$, given that the B_j's are independent Rademacher distributions with parameters α_j^2, $1 \leq i \leq n$ defined in Equation (4). Once these products are known with $k+2$ bits of precision, the left and right k-bit approximations L_k and R_k are easily computed, which allows us to run the modified von Neumann's rejection algorithm from Section 2.2.

In this section, we explain how to compute a k-bit approximation to $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$ at an expected communication cost of $O(kn + n\log n)$ bits. For our specific application of simulating the GHZ distribution, we proved at the end of Section 2.2 (Equation 6) that the expected value of k is bounded by $n+5$. It follows that an expected cost of $O(n^2)$ bits suffices to carry out the simulation.

Given $B = (B_1, \ldots, B_n)$ with the B_i's distributed according to non-identical independent Rademachers with parameter $\cos^2\left(\frac{1}{2}\left(\varphi_i - \frac{\pi}{2}\right)\right)$ or $\cos^2\left(\frac{1}{2}\left(\varphi_i + \frac{\pi}{2}\right)\right)$, we need to compute k-bit approximations of $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$. We use c_j and s_j to denote $\cos\left(\frac{1}{2}\left(\varphi_j - \frac{\pi}{2}B_j\right)\right)$ and $-\sin\left(\frac{1}{2}\left(\varphi_j - \frac{\pi}{2}B_j\right)\right)$, respectively, as well as $\hat{c}_j$ and $\hat{s}_j$ to denote their respective ℓ-truncations. We need to determine ℓ such that the products $\prod_{j=1}^{n} \hat{c}_j$ and $\prod_{j=1}^{n} \hat{s}_j$ are k-approximations of $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$, respectively. *Notice that each party knows exactly his own c_j and s_j, and hence $\hat{c}_j$ and $\hat{s}_j$ can be transmitted directly to the leader, rather than approximations of the φ_i's.* For each c_j, there exists $\varepsilon_j \in [-1, 1]$ such that $c_j = \hat{c}_j + \frac{\varepsilon_j}{2^\ell}$; thus, using I to denote $\{1, 2, \ldots, n\}$, we have

$$\prod_{j=1}^{n} c_j = \sum_{A \in \mathcal{P}(I)} \prod_{j \in A} \hat{c}_j \prod_{j \notin A} \frac{\varepsilon_j}{2^\ell} = \prod_{j=1}^{n} \hat{c}_j + \sum_{A \in \mathcal{P}(I) \setminus I} \prod_{j \in A} \hat{c}_j \prod_{j \notin A} \frac{\varepsilon_j}{2^\ell}$$

and hence we can bound the error as follows:

$$\left| \prod_{j=1}^{n} c_j - \prod_{j=1}^{n} \hat{c}_j \right| \leq \sum_{j=1}^{n} \left(\binom{n}{j} \frac{1}{2^{j\ell}} \right) - 1 = \left(1 + \frac{1}{2^\ell}\right)^n - 1 .$$

Setting $\ell = \left\lceil -\log_2 \left(\left(1 + 2^{-k}\right)^{1/n} - 1 \right) \right\rceil \leq k + \lceil \log_2 n \rceil + 2$, we have

$$\left| \prod_{j=1}^{n} c_j - \prod_{j=1}^{n} \hat{c}_j \right| \leq \frac{1}{2^k} .$$

Taking account of the need to transmit the ℓ-truncations to both c_j and s_j, which consists of the sign of these numbers in addition to the first ℓ bits of their binary expansion, the expected communication cost is $2(n-1)(\ell+1)$ bits, which indeed is $O(kn + n\log n)$.

3.3 Protocol for sampling the GHZ distribution

We are finally ready to glue all the pieces together into Algorithm 4 (on next page), which samples exactly the GHZ distribution under arbitrary von Neumann measurements, thus solving our conundrum. Its correctness is proven below, and it is shown that the expected amount of randomness used in this process is upper-bounded by $6n + 17$ bits and an expected $O(n^2)$ bits of communication suffice to complete the task. Variations are discussed subsequently.

Correctness of the protocol: The part occurring before the first "repeat" (line 5) samples a Bernoulli with parameter $\cos^2\left(\sum_{i=1}^{n} \theta_i/2\right)$, which allows the leader to decide whether to

Algorithm 4 Complete protocol for sampling the GHZ distribution in the sequential model

```
The leader, who is party number 1, communicates with the other parties in order to obtain increasingly precise approximations of $\theta = \sum_{i=1}^{n} \theta_i$ until he can sample random bit $Z$ according to *exact* Bernoulli random distribution with parameter $\cos^2(\theta/2)$
if $Z = 1$ then
    The leader adds $2\pi$ to his own $\varphi$-parameter i.e. $\varphi_1 \leftarrow \varphi_1 + 2\pi$
    {to sample $p_2$ rather than $p_1$}
end if
{Now entering the modified von Neumann's "distributed" sampler for $p_1$}
repeat
    The leader generates a fair random bit $S$ and broadcasts it to the other parties
    {The bit $S$ determines whether to sample $q_1$ or $q_2$}
    Locally, each party $j$ generates a random $B_j \in \{-1, +1\}$ according to an independent Rademacher distribution so that $B_j = +1$ with probability $\cos^2\big(\frac{1}{2}\big(\varphi_j - \frac{\pi}{2}\big)\big)$
    {Random variable $B = (B_1, \ldots, B_n)$ is now sampled according to $q_1$}
    if $S = 1$ then
        Each party does $B_j \leftarrow -B_j$
        {In this case, random variable $B = (B_1, \ldots, B_n)$ is now sampled according to $q_2$}
    end if
    {Random variable $B = (B_1, \ldots, B_n)$ is sampled according to $q = \frac{q_1+q_2}{2}$}
    {The leader starts talking with the other parties to decide whether to accept $B$}
    Each party computes $c_j = \cos\big(\frac{1}{2}\big(\varphi_j - \frac{\pi}{2} B_j\big)\big)$ and $s_j = -\sin\big(\frac{1}{2}\big(\varphi_j - \frac{\pi}{2} B_j\big)\big)$
    The leader sets $k \leftarrow 1$
    The leader sets $U(0) \leftarrow 0$
    repeat forever
        The leader generates an i.i.d. unbiased bit $U_k$
        The leader computes $U(k) \leftarrow U(k-1) + U_k/2^k$ {hence $U(k) = 0.U_1 \ldots U_k$}
        The leader requests $(k + 3 + \lceil \log_2 n \rceil)$-approx. of $c_j$ and $s_j$ from each party $j \geq 2$
        The leader uses this information to compute $(k+2)$-approximations of $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$, which are used to compute $k$-bit approximations $L_k$ of $\mathrm{a}_1^2(B) + \mathrm{a}_2^2(B)$ and $R_k$ of $p_1(B)$
        if $U(k)L_k - R_k < -\frac{4}{2^k}$ then
            Set $Y \leftarrow 1$ and break from the repeat forever loop. {Vector $B$ is accepted}
        else if $U(k)L_k - R_k > \frac{4}{2^k}$ then
            Set $Y \leftarrow 0$ and break from the repeat forever loop. {Vector $B$ is rejected}
        else
            Set $k \leftarrow k + 1$ and continue the repeat forever loop
            {The leader does not yet have enough information to decide whether to accept or reject $B$. Therefore, he needs to compute the next bit of $\mathrm{a}_1(B)$ and $\mathrm{a}_2(B)$. For this, he needs more information from all the other parties.}
        end if
    end repeat
until $Y = 1$ {accepting}
The leader informs all the other parties that the simulation is complete and, therefore, that the time has come for each party $j$ (including the leader himself) to output his current value of $B_j$
```

sample B according to p_1 (by leaving his φ_1 unchanged) or according to p_2 (by adding 2π to his φ_1). Notice that the leader does not have to inform the other parties of this decision since they do not need to know if the sampling will be done according to p_1 or p_2. In Section 3.1, we showed how to sample exactly a Bernoulli with parameter $\cos^2\bigl(\sum_{i=1}^{n}\theta_i/2\bigr)$ when the θ_i's are not known to the leader for $i \geq 2$.

The part within the outer "repeat" loop (lines 5 to 27) is essentially von Neumann's rejection algorithm, which has been adapted and modified to work in a distributed scenario. The leader must first know which of q_1 or q_2 to sample. For this purpose, he generates an unbiased random bit S and broadcasts it to the other parties. Sampling either q_1 or q_2 can now be done locally and independently by each party j, yielding a tentative $B_j \in \{-1, +1\}$. The parties will output these B_j's only at the end, provided this round is not rejected. Now, each party uses his B_j to compute locally $c_j = \cos\bigl(\frac{1}{2}\bigl(\varphi_j - \frac{\pi}{2}B_j\bigr)\bigr)$ and $s_j = -\sin\bigl(\frac{1}{2}\bigl(\varphi_j - \frac{\pi}{2}B_j\bigr)\bigr)$, which will be sent bit by bit to the leader upon request, thus allowing him to compute increasingly precise approximations L_k and R_k of $q_1(B) + q_2(B)$ and $p_1(B)$, respectively. These values are used to determine whether a decision can be made to accept or reject this particular B, or whether more information is needed to make this decision. As shown at the end of Section 2.2 (Equation 6), the expected number of bits needed in L_k and R_k before we can break out of the "repeat forever" loop is $k \leq n + 5$. At that point, flag Y tells the leader whether or not this was a successful run of von Neumann's rejection algorithm. If $Y = 0$, the entire process has to be restarted from scratch, except for the initial Bernoulli sampling, at line 6. On the other hand, once the leader gets $Y = 1$, he can finally tell the other parties that they can output their B_j's because, according to von Neumann's rejection algorithm, this signals that the vector $(B_1, \ldots, B_n)$ is distributed according to p_1 (or p_2, depending on the initial Bernoulli). Also according to von Neumann's rejection algorithm, we have an expectation of $C = 2$ rounds of the outer "repeat" loop before we can thus conclude successfully.

Expected communication cost and number of random coins: The expected amount of randomness used in this process is upper-bounded by $6n + 17$ bits. This is calculated as follows: the expected number of bits for sampling Bernoulli Z is bounded by 5. This is followed by an expectation of $C = 2$ rounds of von Neumann's rejection algorithm (the outer "repeat" loop). In each of these rounds, we need 1 bit for S and expect 2 bits for each of the B_j's (hence $2n$ in total), before entering the "repeat forever" loop. The expected number of times round this loop is bounded by $n + 5$, and one more random bit U_k is needed each time. Putting it all together, the expected number of random bits is bounded by $5 + 2(1 + 2n + (n + 5)) = 6n + 17$.

The expected amount of communication is dominated by the leader's need to obtain increasingly accurate approximations of c_j and s_j from all other parties at line 17 in order to compute increasingly accurate approximations of L_k and R_k, which he needs in order to decide whether or not to break from the "repeat forever" loop and, in such case, whether or not to accept B as final output. On the k-th time round the loop, the leader needs $k + 3 + \lceil \log_2 n \rceil$ bits of precision plus one bit of sign about each c_j and s_j, $j \geq 2$ (in addition to having full knowledge about his own c_1 and s_1, of course). This would be very expensive if all those bits had to be resent each time round the loop, with increasing values of k. Fortunately, this process works well if the parties send *truncations* of these values to the leader, because each truncation simply adds one bit of precision to the previous one. Hence, it suffices for the leader to request $2(5 + \lceil \log_2 n \rceil)$ bits from each other party at the onset, when $k = 1$, and only two additional bits per party are needed afterwards for each subsequent trip round the loop

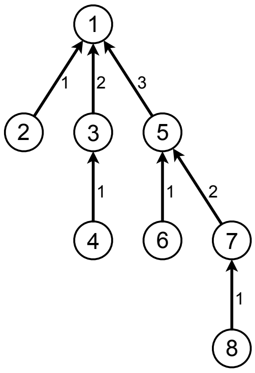


Figure 1 Binomial tree structure defining the parallel model.

(one for c_j and one for s_j). All counted, a total of $2(n-1)(k+5+\lceil \log_2 n \rceil)$ bits will have been requested from all other parties by the time we have gone through the "repeat forever" loop k times. Since the expected value of k upon exiting this loop is bounded by $n+5$, the expected number of bits that have to be communicated to the leader to complete von Neumann's rejection algorithm (lines 5 to 27) is bounded by $2(n-1)((n+5)+5+\lceil \log_2 n \rceil)$. This is $O(n^2)$ expected bits of communication. The additional amount of communication required to sample Bernoulli Z at step 1 (which is $(n-1)(5+\log_2 n)$ bits) and for the leader to broadcast to all parties the value of S, as well as synchronization bits by which he needs to inform the other parties of success or failure each time round the loop is negligible. All counted, Algorithm 4 needs $O(n)$ bits of randomness and $O(n^2)$ bits of communication in order to sample exactly the GHZ distribution under arbitrary von Neumann measurements.

Using Equation (5) rather than Equation (6), we shall show in the final journal version of this work that Algorithm 4 needs only $O(n \log n)$ bits of communication in order to sample exactly the GHZ distribution under computational-basis von Neumann measurements. Of course, $O(n)$ bits of communication would suffice, even in the worst case, if we knew ahead of time that all measurements are in the computational basis, but our protocol works seamlessly with $O(n \log n)$ expected bits of communication even if the measurements are not *exactly* in the computational basis, and if up to $O(\log n)$ of the measurements are arbitrary.

3.4 Variations on the theme

We can modify Algorithm 4 in a variety of ways to improve different parameters at the expense of others. Here, we mention briefly three of these variations: the parallel model, bounding the number of rounds, and the simulation of equatorial measurements.

The parallel model: Until now, we have considered only a *sequential model* of communication, in which the leader has a direct channel with everyone else. In this model, communication takes place sequentially because the leader cannot listen to everyone at the same time. However, it is legitimate to consider a *parallel model*, in which arbitrary many pairs of parties can communicate simultaneously. In this model, any number of bits can be sent and received in the same time step, provided no party has to send or receive more than one bit at any given time. If we make the parties communicate with one another following the binomial tree structure shown in Fig. 1, with the leader at the root, we shall show in the final journal version of this work that the exact simulation of the GHZ distribution under arbitrary independent von Neumann measurements can be accomplished within $O(n\log n)$ expected parallel time. The expected total number of bits communicated with this approach is slightly greater than with Algorithm 4, but it remains $O(n^2)$.

Reducing the number of rounds: Algorithm 4 is efficient in terms of the number of bits of randomness as well as the number of bits of communication, but it requires an expected $O(n)$ rounds, in which the leader and all other parties take turn at sending messages. This could be prohibitive if they are far apart and their purpose is to try to convince examiners that they are actually using true entanglement and quantum processes to produce their joint outputs, because it would prevent them from responding quickly enough to be credible. We leave it for the reader as an exercise to verify that if we change line 24 of Algorithm 4 from "$k \leftarrow k+1$" to "$k \leftarrow 2k$", the expected number of rounds is decreased from $O(n)$ to $O(\log n)$. If in addition we start with "$k \leftarrow n$" instead of "$k \leftarrow 1$" at line 12, the expected number of rounds becomes a constant. (Alternatively, we could start with "$k \leftarrow n$" at line 12 and step with "$k \leftarrow k+n$" at line 24.)

Equatorial measurements: Recall that equatorial measurements are those for which $\varphi_j = 0$ for each party j. In this case, the leader can sample according to p_1 or p_2, without any help or communication from the other parties, since he has complete knowledge of their vanished elevation angles. Therefore, he can run steps 5 to 27 of Algorithm 4 all by himself! However, he needs to communicate in step 1 of Algorithm 4 in order to know from which of p_1 or p_2 to sample. The only remaining need for communication occurs in step 28, which has to be modified from "The leader informs all the other parties that the simulation is complete" to "The leader informs all the other parties of which value of $B_j \in \{-1,+1\}$ he has chosen for them".

Only step 1 requires significant communication since the new step 28 needs only the transmission of $n-1$ bits. We have already seen at the end of Section 3.1 that step 1, which is a distributed version of Algorithm 1, requires an expected communication of $O(n\log n)$ bits in the sequential model. This is therefore the complexity of our simulation, which is an improvement over the previously best technique known to simulate the GHZ distribution under arbitrary equatorial von Neumann measurements [6], which required an expectation of $O(n^2)$ bits of communication.

A more elegant protocol can be obtained if we adapt Equations (1), (2) and (3), which were given at the beginning of Section 2 to define the GHZ probability distribution $p(b)$ for $b \in \{-1,+1\}^n$, to the special case of equatorial measurements. Because all the elevation angles φ_j vanish, these formulas reduce to

$$p(b) = \begin{cases} 2^{1-n}\cos^2\left(\frac{\theta}{2}\right) & \text{if } b \in X \\ 2^{1-n}\sin^2\left(\frac{\theta}{2}\right) & \text{if } b \notin X \end{cases} \qquad \text{where } X = \left\{ b \in \{-1,+1\}^n \;\middle|\; \prod_{j=1}^{n} b_j = +1 \right\}.$$

Now, each party j other than the leader can simply choose an independent unbiased Rademacher $b_j \in \{-1,+1\}$ as final output, without any consideration of his own input θ_j nor communication with anyone else, and inform the leader of this choice. It simply remains for the leader to choose his own b_1 in order to make $\prod_{j=1}^{n} b_j$ equal to $+1$ with probability $\cos^2(\theta/2)$ or -1 with probability $\sin^2(\theta/2)$. For this, we still need step 1 from Algorithm 4, which requires an expected communication of $O(n\log n)$ bits. We shall show in the final journal version of this work that this process can be achieved with only $O(\log^2 n)$ expected time steps in the parallel model of communication.

4 Discussion and open problems

We have addressed the problem of simulating the effect of arbitrary independent von Neumann measurements on the qubits forming the general GHZ state $\frac{1}{\sqrt{2}}|0^n\rangle + \frac{1}{\sqrt{2}}|1^n\rangle$ distributed among n parties. Rather than doing the actual quantum measurements, the parties must sample the exact GHZ probability distribution by purely classical means, which necessarily requires communication in view of Bell's theorem. Our main objective was to find a protocol that solves this conundrum with a finite amount of expected communication, which had only been known previously to be possible when the von Neumann measurements are restricted to being equatorial (a severe limitation indeed). Our solution needs only $O(n^2)$ bits of communication, which can be dispatched in $O(n \log n)$ time if bits can be sent in parallel according to a realistic scenario in which nobody has to send or receive more than one bit in any given step. We also improved on the former art in the case of equatorial measurements, with $O(n \log n)$ bits of communication and $O(\log^2 n)$ parallel time.

Knuth and Yao [16] initiated the study of the complexity of generating random integers (or bit strings) with a given probability distribution $p(b)$, assuming only the availability of a source of unbiased identically independently distributed random bits. They showed that any sampling algorithm must use an expected number of bits at least equal to the entropy $\sum_b p(b) \log_2(1/p(b))$ of the distribution, and that the best algorithm does not need more than two additional bits. For further results on the bit model in random variate generation, see Ref. [9, Chap. XIV].

The GHZ distribution has an entropy no larger than n, and therefore Knuth and Yao have shown that it could be sampled with no more than $n + 2$ expected random bits if all the parameters were concentrated in a single place [16]. Even though we have studied the problem of sampling this distribution in a setting in which the defining parameters (here the description of the von Neumann measurements) are distributed among n parties, and despite the fact that our main purpose was to minimize communication between these parties, we were able to succeed with $6n + 17$ expected random bits, which is just above six times the bound of Knuth and Yao. The amount of randomness required by our protocols does not depend significantly on the actual measurements they have to simulate. However, some sets of measurements entail a probability distribution $p(B)$ whose entropy $H(p)$ is much smaller than n. In the extreme case of having all measurements in the computational basis, $H(p)$ is a single bit! Can there be protocols that succeed with as few as $H(p) + 2$ expected random bits, thus meeting the bound of Knuth and Yao, or failing this as few as $O(H(p))$ expected random bits, no matter how small $H(p)$ is for the given set of von Neumann measurements? Notice that all the protocols presented here require $\Omega(n)$ random bits since they ask each party to sample independently at least once a Rademacher random variable, a hurdle that can only be alleviated in the case of measurements in the computational basis.

Are our protocols optimal in terms of the required amount of communication? Could we simulate arbitrary von Neumann measurements as efficiently as the case of equatorial measurements, i.e. with $O(n \log n)$ bits of communication? We leave this as open question, but point out that Broadbent, Chouha and Tapp have proved an $\Omega(n \log n)$ lower bound on the *worst case* communication complexity of simulating measurements on n-partite GHZ states [7], a result that holds even for equatorial measurements, and even under the promise that $\cos \sum_{i=1}^{n} \theta_i = \pm 1$ [15].

As a recent development, which we shall formalize in the final journal version of this work, we have discovered how to simulate more general multipartite states than the GHZ state. For instance, we know how to simulate the so-called W state $\frac{1}{\sqrt{3}}|100\rangle + \frac{1}{\sqrt{3}}|010\rangle + \frac{1}{\sqrt{3}}|001\rangle$

and more generally

$$W_n = \frac{1}{\sqrt{n}} \left(|10^{n-1}\rangle + |010^{n-2}\rangle + |0010^{n-3}\rangle + \cdots + |0^{n-1}1\rangle \right)$$

with $O(n^3)$ expected bits of communication and the need of only $O(n^2)$ expected unbiased independent random bits. However, we leave for further research the problem of simulating arbitrary positive-operator-valued measurements (POVMs) on the single-qubit shares of GHZ states (or on more general multipartite states), as well as the problem of simulating multipartite entanglement other than equatorial von Neumann measurements on the tripartite GHZ state [3] with *worst-case* bounded classical communication.

Acknowledgements. We wish to thank Marc Kaplan and Nicolas Gisin for stimulating discussions about the simulation of entanglement. Furthermore, Marc has carefully read Ref. [12], in which the decomposition of the GHZ distribution as a convex combination of two sub-distributions was first accomplished, and he has pointed out that the lower bound from Ref. [7] applies even in the case of equatorial measurements. Alain Tapp pointed out that the entropy of the GHZ distribution can be as small as one bit in the case of measurements in the computational basis. We are also grateful for the useful feedback provided by the TQC reviewers.

G. B. is supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chair program, the Canadian Institute for Advanced Research (CIFAR), and the Institute for Theoretical Studies (ITS) at the ETH Zürich. L. D. is supported in part by NSERC and Fonds de recherche du Québec – Nature et technologies (FRQNT).

References

1 J.-D. Bancal, C. Branciard and N. Gisin, "Simulation of equatorial von Neumann measurements on GHZ states using nonlocal resources", *Advances in Mathematical Physics* **2010**:293245, 2010.

2 J. S. Bell, "On the Einstein-Podolsky-Rosen paradox", *Physics* **1**:195–200, 1964.

3 C. Branciard and N. Gisin, "Quantifying the nonlocality of Greenberger-Horne-Zeilinger quantum correlations by a bounded communication simulation protocol", *Physical Review Letters* **107**:020401, 2011.

4 G. Brassard. "Quantum communication complexity", *Foundations of Physics* **33**(11): 1593–1616, 2003.

5 G. Brassard, R. Cleve and A. Tapp, "Cost of exactly simulating quantum entanglement with classical communication", *Physical Review Letters* **83**:1874–1877, 1999.

6 G. Brassard and M. Kaplan, "Simulating equatorial measurements on GHZ states with finite expected communication cost", *Proceedings of 7th Conference on Theory of Quantum Computation, Communication, and Cryptography (TQC)*, pp. 65–73, 2012.

7 A. Broadbent, P. R. Chouha and A. Tapp, "The GHZ state in secret sharing and entanglement simulation", *Proceedings of the Third International Conference on Quantum, Nano and Micro Technologies*, pp. 59–62, 2009.

8 N. Cerf, N. Gisin and S. Massar, "Classical teleportation of a quantum bit", *Physical Review Letters* **84**(11):2521–2524, 2000.

9 L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

10 A. Einstein, B. Podolsky and N. Rosen, "Can quantum-mechanical description of physical reality be considered complete?", *Physical Review* **47**:777–780, 1935.

11 N. Gisin, personal communication, 2010.

12 C. Gravel, “Structure de la distribution de probabilité de l’état GHZ sous l’action de mesures de von Neumann locales”, M.Sc. thesis, Department of Computer Science and Operations Research, Université de Montréal: `https://papyrus.bib.umontreal.ca/jspui/handle/1866/5511`, 2011.

13 C. Gravel, “Structure of the probability distribution for the GHZ quantum state under local von Neumann measurements”, *Quantum Physics Letters* **1**(3):87–96, 2012.

14 D. M. Greenberger, M. A. Horne and A. Zeilinger, “Going beyond Bell’s theorem”, in *Bell’s Theorem, Quantum Theory and Conceptions of the Universe*, edited by M. Kafatos (Kluwer Academic, Dordrecht), pp. 69–72, 1989.

15 M. Kaplan, personal communication, 2013.

16 D. E. Knuth and A. C. Yao, “The complexity of nonuniform random number generation”, in: *Algorithms and Complexity*, edited by J. E. Traub, pp. 357–428, Academic Press, New York, 1976.

17 T. Maudlin, “Bell’s inequality, information transmission, and prism models”, *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pp. 404–417, 1992.

18 J. von Neumann, “Various techniques used in connection with random digits. Monte Carlo methods”, *National Bureau Standards* **12**:36–38, 1951. Reprinted in *Collected Works*, **5**:768–770, Pergamon Press, 1963.

19 O. Regev and B. Toner, “Simulating quantum correlations with finite communication”, *SIAM Journal on Computing* **39**(4):1562–1580, 2009.

20 M. Steiner, “Towards quantifying non-local information transfer: finite-bit non-locality”, *Physics Letters A* **270**:239–244, 2000.

21 B. Toner and D. Bacon, “Communication cost of simulating Bell correlations”, *Physical Review Letters* **91**:187904, 2003.

On the Parallel Repetition of Multi-Player Games: The No-Signaling Case

Harry Buhrman[1,2], Serge Fehr[1], and Christian Schaffner[2,1]

1 **Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands**
{h.buhrman,s.fehr}@cwi.nl

2 **Institute for Logic, Language and Computation (ILLC), University of Amsterdam, The Netherlands**
c.schaffner@uva.nl

Abstract

We consider the natural extension of two-player nonlocal games to an arbitrary number of players. An important question for such nonlocal games is their behavior under parallel repetition. For *two-player* nonlocal games, it is known that both the *classical* and the *non-signaling* value of any game converges to zero exponentially fast under parallel repetition, given that the game is non-trivial to start with (i.e., has classical/non-signaling value < 1). Very recent results [7, 5, 10] show similar behavior of the *quantum* value of a two-player game under parallel repetition. For nonlocal games with three or more players, very little is known up to present on their behavior under parallel repetition; this is true for the classical, the quantum and the non-signaling value.

In this work, we show a parallel repetition theorem for the *non-signaling* value of a large class of multi-player games, for an arbitrary number of players. Our result applies to all multi-player games for which all possible combinations of questions have positive probability; this class in particular includes all *free* games, in which the questions to the players are chosen independently. Specifically, we prove that if the original game $\mathcal{G}$ has a non-signaling value $v_{\text{ns}}(\mathcal{G}) < 1$, then the non-signaling value of the n-fold parallel repetition is exponentially small in n. Stronger than that, we prove that the probability of winning more than $(v_{\text{ns}}(\mathcal{G}) + \delta) \cdot n$ parallel repetitions is exponentially small in n (for any $\delta > 0$).

Our parallel repetition theorem for multi-player games is weaker than the known parallel repetition results for two-player games in that the rate at which the non-signaling value of the game decreases not only depends on the non-signaling value of the original game (and the number of possible responses), but on the complete description of the game. Nevertheless, we feel that our result is a first step towards a better understanding of the parallel repetition of nonlocal games with more than two players.

1998 ACM Subject Classification E.4 Coding and Information Theory

Keywords and phrases Parallel repetition, non-signaling value, multi-player non-local games

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.24

1 Introduction

Background

In an m-player nonlocal game $\mathcal{G}$, m players receive respective questions $x_1, \ldots, x_m$, chosen according to some joint probability distribution, and the task of the m players is to provide "good" answers $a_1, \ldots, a_m$, *without communicating* with each other. The players are said to *win* the game if the given answers jointly satisfy some specific property with respect to the given questions. The *value* of a given game is defined to be the maximal winning probability

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 24–35
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of the players. One distinguishes between the classical, the quantum, and the non-signaling value, depending on whether the players are restricted to be classical, may share entanglement and do quantum measurements, or are allowed to make use of any hypothetical strategy that does not violate non-signaling.

An important question for nonlocal games is their behavior under parallel repetition. This question is somewhat understood in the case of *two* players, where $m = 2$. Indeed, Raz showed in his celebrated parallel repetition theorem [14] that if the classical value of a two-player game $\mathcal{G}$ is $v_{\mathrm{c}}(\mathcal{G}) < 1$ then the classical value $v_{\mathrm{c}}(\mathcal{G}^n)$ of the n-fold parallel repetition of $\mathcal{G}$ satisfies $v_{\mathrm{c}}(\mathcal{G}^n) \leq \bar{v}_{\mathrm{c}}(\mathcal{G})^{n/\log(s)}$, where s denotes the number of possible pairs of answers a_1 and a_2, and $\bar{v}_{\mathrm{c}}(\mathcal{G}) < 1$ only depends on $v_{\mathrm{c}}(\mathcal{G})$. Raz's result was improved and simplified by Holenstein [9], who gave an explicit and tighter dependency between $\bar{v}_{\mathrm{c}}(\mathcal{G})$ and $v_{\mathrm{c}}(\mathcal{G})$, namely $\bar{v}_{\mathrm{c}}(\mathcal{G}) = 1 - \frac{1}{6000}(1 - v_{\mathrm{c}}(\mathcal{G}))^3$. Holenstein also showed that a similar result holds for the non-signaling value of any two-player game: $v_{\mathrm{ns}}(\mathcal{G}^n) \leq \bar{v}_{\mathrm{ns}}(\mathcal{G})^n$ for $\bar{v}_{\mathrm{ns}}(\mathcal{G}) = 1 - \frac{1}{6400}(1 - v_{\mathrm{ns}}(\mathcal{G}))^2$. Parallel repetition results for the quantum value of two-player games were first derived for certain special classes of games, like XOR-games [6] or unique games [11], or for a non-standard parallel repetition where the different repetitions of the original game are intertwined with modified versions of the original game [12]. Recently, several results about the parallel repetition of more general quantum games have been obtained [7, 5, 10].

There are further improvements to the above results on two-player games. For instance, Rao [13] showed a *concentration* result for the classical value of any two-player game, saying that the probability to win more than $(v_{\mathrm{ns}}(\mathcal{G}) + \delta) \cdot n$ out of the n repetitions is exponentially small (for any $\delta > 0$).[1] Furthermore, he improved the bound on the classical value under parallel repetition for *projection* games. A similar improvement on the bound on the classical value under parallel repetition was given by Barak *et al.* [1] for *free* games, together with a further improvement, namely a *strong* parallel repetition theorem (meaning that meaning that $v_{\mathrm{c}}(\mathcal{G}^n) \leq v_{\mathrm{c}}(\mathcal{G})^{\Omega(n)}$), for *free projection* games.

When considering multi-player nonlocal games with strictly more than 2 players, to the best of our knowledge, very little is known about their behavior under parallel repetition, except for trivial cases. This applies to the classical, the quantum, and the non-signaling value. In [15], Rosen proved a parallel-repetition result for more than 2 players. While her proof strategy is very similar to ours (closely following [9]), a somewhat unnatural definition of multi-player non-signaling correlations is used where no $m - 1$ provers together can signal to the remaining prover. In our (standard) model, one also demands that any subset (of arbitrary size) of provers can not signal to the remaining provers.

Another result about multi-player games is by Briët *et al.* [2] about the related question of XOR repetition. They show the existence of a 3-player XOR game whose classical value of the XOR repetition is bounded from below by a constant (independent of the number of repetitions). Hence, XOR repetition does not hold for this game (but parallel repetition might still hold). Our result does not imply anything about those games, because the non-signaling value of XOR games is always 1.

Possible applications of our result could be of cryptographic nature where the hardness of a basic task is amplified by parallel repetition. A likely scenario for applying our results (and our original motivation to study the problem) is position-based quantum cryptography [3, 4], in the spirit of a recent result on parallel repetition of a particular game [17]. However, as

[1] Rao claims the concentration result only for the classical value, but the same techniques also apply to the non-signaling value.

our result only applies to a restricted class of games, we were not able yet to apply it to in this cryptographic context.

Our Results

We show a parallel repetition and a concentration theorem for the non-signaling value of m-player games for any m, for a large class of games. The class of games to which our result applies consists of all multi-player games with *complete support*, meaning that all possible combinations of questions $x_1, \ldots, x_m$ must have positive probability of being asked. This class of games in particular includes all *free* games, in which the questions to the different players are chosen independently. For any m-player game $\mathcal{G}$ with complete support, we show that if $v_{\mathrm{ns}}(\mathcal{G}) < 1$ then there exists $\bar{v}_{\mathrm{ns}}(\mathcal{G}) < 1$ so that $v_{\mathrm{ns}}(\mathcal{G}^n) \leq \bar{v}_{\mathrm{ns}}(\mathcal{G})^n$, and the probability of winning more than $(v_{\mathrm{ns}}(\mathcal{G}) + \delta) \cdot n$ out of the n repetitions with an arbitrary non-signaling strategy is exponentially small (for any $\delta > 0$).

We point out that our parallel repetition result for multi-player games (with complete support) is of a weaker nature than the parallel repetition results for two-player games discussed above, in that in our result the constant $\bar{v}_{\mathrm{ns}}(\mathcal{G})$ depends on the complete description of the game $\mathcal{G}$, and not just on its non-signaling value $v_{\mathrm{ns}}(\mathcal{G})$. Still, our result is the first that shows a parallel repetition result for a large class of m-player games with $m > 2$ for one of the three values (the classical, quantum or non-signaling) of interest.

For proving our results, we borrow and extend tools from [9] and [13], and combine them with some new technique. The new technique involves considering strategies that are *almost* non-signaling, meaning that the non-signaling properties only hold up to some small error. We then show (Proposition 18) and use in our proof that the non-signaling value of a game is *robust* under extending the quantification over all non-signaling strategies to all almost non-signaling strategies.

2 Preliminaries

2.1 Basic Notation

For any m-partite set $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$, any m-tuple $x = (x_1, \ldots, x_m) \in \mathcal{X}$, and any index set $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$, we write $\mathcal{X}_I$ to denote the k-partite set $\mathcal{X} = \mathcal{X}_{i_1} \times \cdots \times \mathcal{X}_{i_k}$, and we write x_I to denote the k-tuple $x = (x_{i_1}, \ldots, x_{i_k}) \in \mathcal{X}_I$. To denote elements from the n-fold Cartesian product of an m-partite set $\mathcal{X}$ as above, we write $\boldsymbol{x} = (x^1, \ldots, x^n) \in \mathcal{X} \times \cdots \times \mathcal{X}$ with $x^i = (x^i_1, \ldots, x^i_m) \in \mathcal{X}$. For $i \in \{1, \ldots, m\}$, we then write $\boldsymbol{x}_i$ for $\boldsymbol{x}_i = (x^1_i, \ldots, x^n_i)$, and for $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$, x^ℓ_I is naturally understood as $x^\ell_I = (x^\ell_{i_1}, \ldots, x^\ell_{i_k})$ and $\boldsymbol{x}_I$ as $\boldsymbol{x}_I = (\boldsymbol{x}_{i_1}, \ldots, \boldsymbol{x}_{i_k})$. Corresponding notation is used for random variables X over $\mathcal{X}$ and $\boldsymbol{X}$ over $\mathcal{X} \times \cdots \times \mathcal{X}$.

2.2 Probabilities and Random Variables

We consider finite probability spaces, given by a non-empty finite sample space Ω and a probability function $P : \Omega \to [0, 1]$. A random variable is a function $X : \Omega \to \mathcal{X}$ from Ω into some finite set $\mathcal{X}$. The distribution of X, denoted as P_X, is given by $P_X(x) = P[\,X\!=\!x\,] = P[\,\{\omega \in \Omega \,|\, X(\omega)\!=\!x\}\,]$. The joint distribution of a pair of random variables X and Y is denoted by P_{XY}, i.e., $P_{XY}(x, y) = P[\,X\!=\!x \wedge Y\!=\!y\,]$, and the conditional distribution of X given Y is denoted by $P_{X|Y}$ and defined as $P_{X|Y}(x|y) = P_{XY}(x, y)/P_Y(y)$ for all x and y with $P_Y(y) > 0$. An event $\mathcal{E}$ is a subset of Ω, and the conditional distribution of a random variable X given $\mathcal{E}$ is denoted as $P_{X|\mathcal{E}}$ and given by $P_{X|\mathcal{E}}(x) = P[\,X\!=\!x \wedge \mathcal{E}\,]/P[\,\mathcal{E}\,]$.

The variational (or statistical) distance between two probability distributions P_X and Q_X for the same random variable $X : \Omega \to \mathcal{X}$ over two probability spaces (Ω, P) and (Ω, Q) (with the same Ω), is defined as

$$\|P_X - Q_X\| := \frac{1}{2} \sum_{x \in \mathcal{X}} |P_X(x) - Q_X(x)|$$

If P_X and Q_X are ε-close in variational distance, we also write $P_X \approx_\varepsilon Q_X$.

Usually, we leave the probability space (Ω, P) etc. implicit, and understand random variables $X, Y, \ldots$ to be defined by their joint distribution $P_{XY\ldots}$, or by some "experiment" that uniquely determines their joint distribution.

2.3 Some Useful Facts

The following lemma states that the variational distance cannot increase when less information is taken into account.

▶ **Lemma 1.** *Let P_{XY} and Q_{XY} be joint distributions for random variables X and Y with respective ranges $\mathcal{X}$ and $\mathcal{Y}$, and let P_X and Q_X be the corresponding marginals. Then,*

$$\|P_X - Q_X\| \le \|P_{XY} - Q_{XY}\| .$$

Proof.

$$\begin{aligned}\|P_X - Q_X\| &= \frac{1}{2} \sum_{x \in \mathcal{X}} |P_X(x) - Q_X(x)| = \frac{1}{2} \sum_{x \in \mathcal{X}} \left| \sum_{y \in \mathcal{Y}} \big(P_{XY}(x,y) - Q_{XY}(x,y)\big) \right| \\ &\le \frac{1}{2} \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} |P_{XY}(x,y) - Q_{XY}(x,y)| = \|P_{XY} - Q_{XY}\| .\end{aligned}$$

◀

The next lemma is due to Holenstein [9] (a simplified version of his Corollary 6).

▶ **Lemma 2.** *Let T and $U^1, \ldots, U^L$ be random variables with distribution $P_{TU^1\cdots U^L} = P_T \cdot P_{U^1|T} \cdots P_{U^L|T}$ (i.e. the U^ℓ's are conditionally independent given T), and let $\mathcal{E}$ be an event. Then*

$$\sum_{\ell=1}^{L} \left\| P_{TU^\ell|\mathcal{E}} - P_{T|\mathcal{E}} \cdot P_{U^\ell|T} \right\| \le \sqrt{L \log\!\left(\frac{1}{P[\mathcal{E}]}\right)} .$$

The following is Hoeffding Inequality's for sampling without replacement [8].

▶ **Theorem 3** (Hoeffding Inequality for sampling without replacement)**.** *Let $w \in \{0,1\}^n$ be an n-bit string with $\frac{1}{n}\sum_{\ell=1}^{n} w_i = \overline{w}$. Let the random variables $D_1, D_2, \ldots, D_K$ be obtained by sampling K random entries from w* without replacement*. Then, for any $\varepsilon > 0$, the random variable $\overline{D} := \frac{1}{K}\sum_k D_k$ satisfies*

$$P\big[\, \overline{D} \le \overline{w} - \varepsilon \,\big] \le \exp\!\big(-2\varepsilon^2 K\big) .$$

Finally, we will make use of the Azuma-Hoeffding Inequality, stated below. We first define the notion of a supermartingale.

▶ **Definition 4** (Supermartingale)**.** A sequence of real valued random variables $M_0, M_1, \ldots, M_K$ is called a *supermartingale* if $\mathbb{E}[M_k | M_0 \cdots M_{k-1}] \le M_{k-1}$ (with probability 1) for every $k \ge 1$.

▶ **Theorem 5** (Azuma-Hoeffding Inequality)**.** *If $M_0, M_1, \ldots, M_K$ is a supermartingale with $M_k \le M_{k-1} + 1$, then*

$$P\big[\, M_K > M_0 + \varepsilon K \,\big] \le \exp\!\big(-\varepsilon^2 K/2\big) .$$

2.4 Nonlocal Games

▶ **Definition 6.** An *m-player nonlocal game*, or simply *(m-player) game* $\mathcal{G}$ consists of two m-partite sets $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ and $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_m$, a probability distribution $\pi : \mathcal{X} \to [0,1]$ on $\mathcal{X}$, i.e., $\sum_x \pi(x) = 1$, and a verification predicate $\mathsf{V} : \mathcal{X} \times \mathcal{A} \to \{0,1\}$.

▶ **Definition 7.** A *strategy* for an m-player game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ is a conditional probability distribution $q(\cdot|\cdot) : \mathcal{A} \times \mathcal{X} \to [0,1]$, i.e., $\sum_a q(a|x) = 1$ for all $x \in \mathcal{X}$.

▶ **Definition 8.** For any m-player game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ and any strategy q for $\mathcal{G}$, the *value* of the game with respect to q is given by

$$v[q](\mathcal{G}) := \sum_{\substack{x \in X \\ a \in A}} \pi(x)\, q(a|x)\, \mathsf{V}(x,a)\,.$$

Any m-player game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ and any strategy q for $\mathcal{G}$ together naturally define a probability space with random variables $X = (X_1, \ldots, X_m)$ and $A = (A_1, \ldots, A_m)$ with joint probability distribution P_{XA} given by $P_{XA}(x,a) = \pi(x)q(a|x)$. The random variable X describes the choice of the input $x \in \mathcal{X}$ according to π, and the random variable A then describes the reply $a \in \mathcal{A}$ chosen according to the distribution $q(\cdot|x)$. It obviously holds that $P_X = \pi$, and $P_{A|X}(\cdot|x) = q(\cdot|x)$ for any $x \in \mathcal{X}$ with $P_X(x) > 0$. A subtlety is that for $x \in \mathcal{X}$ with $P_X(x) = 0$, the distribution $P_{A|X}(\cdot|x)$ is strictly speaking not defined whereas $q(\cdot|x)$ is. The value of the game with respect to strategy q can be written in terms of these random variables as $v[q](\mathcal{G}) = P[\,\mathsf{V}(X,A) = 1\,]$. In the following we define the classical, quantum and non-signaling values of m-player games. Only the last one will be used in the rest of the paper, but we provide all of them for the sake of completeness.

▶ **Definition 9.** A *strategy* q for an m-player game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ is *classical* (or *local*) if there exists a probability distribution p on a set $\mathcal{W}$ and conditional probability distributions $q_1, \ldots, q_m$ such that

$$q(a_1, \ldots, a_m | x_1, \ldots, x_m) = \sum_{w \in \mathcal{W}} p(w) \prod_{i=1}^{m} q_i(a_i | x_i, w)\,.$$

The *classical value* of a game $\mathcal{G}$ is defined as $v_{\mathrm{c}}(\mathcal{G}) := \sup_q v[q](\mathcal{G})$, where the supremum is over all classical strategies q for $\mathcal{G}$.

▶ **Definition 10.** A *strategy* q for an m-player game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ is *quantum* if there exists an m-partite quantum state $|\psi\rangle \in \mathcal{H}_{\mathsf{A}_1} \otimes \cdots \otimes \mathcal{H}_{\mathsf{A}_m}$ and for every $x = (x_1, \ldots, x_m) \in \mathcal{X}$ there exist POVMs $\mathbf{E}^1_{x_1} = \{E^1_{x_1,a_1}\}_{a_1 \in \mathcal{A}_1}, \ldots, \mathbf{E}^m_{x_m} = \{E^m_{x_m,a_m}\}_{a_m \in \mathcal{A}_m}$ such that for all $a = (a_1, \ldots, a_m) \in \mathcal{A}$ and $x = (x_1, \ldots, x_m) \in \mathcal{X}$:

$$q(a|x) = \langle \psi | E^1_{x_1,a_1} \otimes \cdots \otimes E^m_{x_m,a_m} | \psi \rangle$$

The *quantum value* of a game $\mathcal{G}$ is defined as $v_{\mathrm{qu}}(\mathcal{G}) := \sup_q v[q](\mathcal{G})$, where the supremum is over all quantum strategies q for $\mathcal{G}$.

▶ **Definition 11.** A *strategy* q for an m-player game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ is *non-signaling* if for any index subset $I \subset \{1, \ldots, m\}$ and its complement $J = \{1, \ldots, m\} \setminus I$, it holds that

$$\sum_{a_J \in \mathcal{A}_J} q(a_I, a_J | x_I, x_J) = \sum_{a_J \in \mathcal{A}_J} q(a_I, a_J | x_I, x'_J) \quad \text{for all } a_I \in \mathcal{A}_I,\ x_I \in \mathcal{X}_I \text{ and } x_J, x'_J \in \mathcal{X}_J\,.$$

The *non-signaling value* of a game $\mathcal{G}$ is defined as $v_{\mathrm{ns}}(\mathcal{G}) := \sup_q v[q](\mathcal{G})$, where the supremum is over all non-signaling strategies q for $\mathcal{G}$.

The following relaxed notion of non-signaling is crucial for the understanding of our parallel-repetition proof.

▶ **Definition 12.** A *strategy* q for an m-player game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ is ε*-almost non-signaling* if for any index subset $I \subset \{1, \ldots, m\}$ and its complement $J = \{1, \ldots, m\} \setminus I$, it holds that

$$\left| \sum_{a_J \in \mathcal{A}_J} q(a_I, a_J | x_I, x_J) - \sum_{a_J \in \mathcal{A}_J} q(a_I, a_J | x_I, x'_J) \right| \leq \varepsilon \quad \text{for all } a_I \in \mathcal{A}_I,\, x_I \in \mathcal{X}_I \text{ and } x_J, x'_J \in \mathcal{X}_J\,.$$

3 A Multi-Player Parallel Repetition Theorem

3.1 The Parallel Repetition of Nonlocal Games

Given a game $\mathcal{G}$, the n-fold parallel repetition $\mathcal{G}^n$ is the game where the referees samples n independent inputs $\boldsymbol{x} = (x^1, \ldots, x^n) \in \mathcal{X} \times \cdots \times \mathcal{X}$ and $\mathcal{G}^n$ is won if and only if all its sub-games are won. For the sake of notational convenience, we also introduce the following way of denoting the fact that t of the n parallel repetitions are won.

▶ **Definition 13** (t-out-of-n Parallel Repetition)**.** For any $n \in \mathbb{N}$ and $t \in \mathbb{R}$, the *t-out-of-n parallel repetition* of a game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ is given by the game $\mathcal{G}^{t/n} = (\mathcal{X}^n, \mathcal{A}^n, \pi^n, \mathsf{V}^{t/n})$ where $\mathcal{X}^n = \mathcal{X} \times \cdots \times \mathcal{X}$ and $\mathcal{A}^n = \mathcal{A} \times \cdots \times \mathcal{A}$, and for all $\boldsymbol{x} = (x^1, \ldots, x^n) \in \mathcal{X}^n$ and $\boldsymbol{a} = (a^1, \ldots, a^n) \in \mathcal{A}^n$

$$\pi^n(\boldsymbol{x}) := \prod_{\ell=1}^{n} \pi(x^\ell) \qquad \text{and} \qquad \mathsf{V}^{t/n}(\boldsymbol{x}, \boldsymbol{a}) := \begin{cases} 1 & \text{if } \sum_{\ell=1}^{n} \mathsf{V}(x^\ell, a^\ell) \geq t \\ 0 & \text{else} \end{cases} \,.$$

The (standard) *n-fold parallel repetition* of a game $\mathcal{G}$ is given by the game $\mathcal{G}^n := \mathcal{G}^{n/n}$.

Similar to the observation after Definition 8, for any game $\mathcal{G}$ and for any strategy[2] $q^{(n)}$ for the t-out-of-n (or the n-fold) parallel repetition, random variables $\boldsymbol{X} = (X^1, \ldots, X^n)$ and $\boldsymbol{A} = (A^1, \ldots, A^n)$, together with their joint distribution $P_{\boldsymbol{XA}}$, are naturally determined.

Note that for any $\ell \in \{1, \ldots, n\}$, X^ℓ is of the form $X^\ell = (X^\ell_1, \ldots, X^\ell_m)$, where X^ℓ_i represents the question to the i-th player in the ℓ-th repetition of $\mathcal{G}$ (and is distributed over $\mathcal{X}_i$). Therefore, for any $i \in \{1, \ldots, m\}$, we write $\boldsymbol{X}_i$ for $\boldsymbol{X}_i = (X^1_i, \ldots, X^n_i)$, and for any $I = \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, m\}$, X^ℓ_I should be understood as $X^\ell_I = (X^\ell_{i_1}, \ldots, X^\ell_{i_k})$ and $\boldsymbol{X}_I$ as $\boldsymbol{X}_I = (\boldsymbol{X}_{i_1}, \ldots, \boldsymbol{X}_{i_k})$. The corresponding holds for $\boldsymbol{A}$.

To simplify notation, for the n-fold repetition of a given game $\mathcal{G}$ with a given strategy $q^{(n)}$, we define W_ℓ to be the random variable $W_\ell := \mathsf{V}(X^\ell, A^\ell)$ that indicates if the ℓ-th repetition of $\mathcal{G}$ is won, and we define $\overline{W} := \frac{1}{n} \sum_{\ell=1}^{n} W_\ell$ to be the fraction of repetitions that are won. Obviously, $v[q^{(n)}](\mathcal{G}^{t/n}) = P[\overline{W} \geq t/n]$.

3.2 Concentration and Parallel Repetition Theorems

Our concentration and parallel repetition theorems below hold for all multi-player nonlocal games $\mathcal{G}$ up to the following restriction on the distribution π.

▶ **Definition 14.** We say that an m-player game $\mathcal{G} = (\mathcal{X}, \mathcal{A}, \pi, \mathsf{V})$ has *complete support* if $\pi(x) > 0$ for all $x \in \mathcal{X}$, i.e., every $x \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_m$ is a "valid input" to the game.

[2] We write $q^{(n)}$ (rather than e.g. q^n) to emphasize that it is a strategy for an n-fold repetition of $\mathcal{G}$, but it is *not* (necessarily) the n-fold independent execution of a strategy q for $\mathcal{G}$.

An important class of games that satisfy the complete-support property are the so-called *free* games, as studied for instance in [1]. In a free game, π is required to be a *product distribution*, i.e., $\pi(x) = \pi_1(x_1)\cdots\pi_m(x_m)$ for all $x = (x_1, \ldots, x_m) \in \mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_m$. Such a game has obviously full support.[3]

▶ **Theorem 15** (Concentration Theorem). *Let $\mathcal{G}$ be an arbitrary m-player game with complete support. Then there exists a constant $\mu > 0$, depending on $\mathcal{G}$, such that for any $\delta > 0$, any $n \in \mathbb{N}$, and for $t = (v_{\text{ns}}(\mathcal{G})+\delta)n$:*

$$v_{\text{ns}}(\mathcal{G}^{t/n}) \leq 8\exp\bigl(-\delta^4\mu n\bigr) .$$

As an immediate consequence, we get the following parallel-repetition theorem.

▶ **Theorem 16** (Parallel-Repetition Theorem). *Let $\mathcal{G}$ be an arbitrary m-player game with complete support and non-signaling value $v_{\text{ns}}(\mathcal{G}) < 1$. Then there exists $\nu < 1$, depending on $\mathcal{G}$, such that $v_{\text{ns}}(\mathcal{G}^n) < 8\nu^n$ for any $n \in \mathbb{N}$.*

We point out that the constants μ (in Theorem 15) and ν (in Theorem 16) not only depend on the non-signaling value $v_{\text{ns}}(\mathcal{G})$ of $\mathcal{G}$, but on the game $\mathcal{G}$ itself. The restriction to games with complete support stems from the fact that μ becomes 0 when the smallest probability in the distribution π goes to 0, rendering the bound useless.

3.3 The Proof

A central idea of our proof is the *robustness* of the non-signaling value of a game. We will use the following result from [16, Section 10.4] about the sensitivity analysis of linear programs.

▶ **Lemma 17.** *Let A be an $m \times n$-matrix, and let A be such that for each nonsingular submatrix B of A, all entries of B^{-1} are at most Δ in absolute value. Let c be a row n-vector, and let b' and b'' be column m-vectors such that both $\max_x\{cx \mid Ax \leq b'\}$ and $\max_x\{cx \mid Ax \leq b''\}$ are finite. Then* [4]

$$\left| \max_{x\in\mathbb{R}^n}\{cx \mid Ax \leq b''\} - \max_{x\in\mathbb{R}^n}\{cx \mid Ax \leq b'\} \right| \leq n\Delta\|c\|_1 \cdot \|b'' - b'\|_\infty .$$

▶ **Proposition 18** (Robustness of $v_{\text{ns}}(\mathcal{G})$). *Let $\mathcal{G}$ be an m-player game with non-signaling value $v_{\text{ns}}(\mathcal{G})$. Then, there exists a constant $c(\mathcal{G})$ such that for any $\varepsilon \geq 0$ and for any strategy q for $\mathcal{G}$ that is ε-almost non-signaling, the value of $\mathcal{G}$ with respect to q is bounded by $v[q](\mathcal{G}) \leq v_{\text{ns}}(\mathcal{G}) + c(\mathcal{G}) \cdot \varepsilon$.*

[3] After possibly having restricted the sets $\mathcal{X}_1, \ldots, \mathcal{X}_m$ appropriately.

[4] For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the norms are defined as $\|x\|_1 = \sum_i |x_i|$ and $\|x\|_\infty = \max_i |x_i|$.

Proof. The non-signaling value $v_{\mathrm{ns}}(\mathcal{G})$ is the optimal value of the following linear program:

$$\text{maximize} \quad \sum_{\substack{x\in\mathcal{X}\\ a\in\mathcal{A}}} \pi(x)\,\mathsf{V}(x,a)\,q(a|x)$$

$$\text{subject to} \tag{1}$$

$$q(a|x) \geq 0 \quad \text{for all } a\in\mathcal{A},\ x\in\mathcal{X}, \tag{2}$$

$$\sum_{a\in\mathcal{A}} q(a|x) = 1 \quad \text{for all } x\in\mathcal{X}, \tag{3}$$

$$\sum_{a_J\in\mathcal{A}_J} q(a_I,a_J|x_I,x_J) - q(a_I,a_J|x_I,x'_J) = 0 \text{ for all } I\subset\{1,\ldots,m\},\ J=\{1,\ldots,m\}\setminus I \\ \text{and for all } a_I\in\mathcal{A}_I,\ x_I\in\mathcal{X}_I \text{ and } x_J,x'_J\in\mathcal{X}_J\,. \tag{4}$$

Lemma 17 gives a bound on how much the optimal value of this linear program can vary if we optimize over ε-almost non-signaling strategies instead of a fully non-signaling strategies. Formally, we can express the linear program above in the "standard form" $\max\{cx \mid Ax \leq b'\}$ by expanding the equality constraints (3) and (4) as $\leq$ and $\geq$ inequality constraints. According to Definition 12, ε-almost non-signaling strategies fulfill the constraints (4) only up to an error of at most 2ε. Hence, relaxing the constraints from non-signaling to ε-almost non-signaling amounts to change the b'-coordinates corresponding to the non-signaling constraints (4) from 0 to 2ε. Hence, the parameters of Lemma 17 are $\|b''-b'\|_\infty = 2\varepsilon$, $n=|\mathcal{X}|\cdot|\mathcal{A}|$, $\|c\|_1 = \sum_{\substack{x\in\mathcal{X}\\ a\in\mathcal{A}}} |\pi(x)\,\mathsf{V}(x,a)| \leq |\mathcal{A}|$ and Δ is a finite constant that depends on the number of players m and the number of answers $|\mathcal{A}|$ and questions $|\mathcal{X}|$.[5] Finally, we note that we can apply the lemma, because the objective function is at most one (and thus finite) irrespective of which strategies we are considering. Setting $c(\mathcal{G}) := 2|\mathcal{X}||\mathcal{A}|^2\Delta$ yields the claim. ◀

▶ **Lemma 19** (Main Lemma). *Let $\mathcal{G}$ be a game with complete support. Consider an n-fold repetition $\mathcal{G}^n$ of $\mathcal{G}$ with an arbitrary non-signaling strategy $q^{(n)}$ for $\mathcal{G}^n$. Let $\mathcal{E}$ be an arbitrary event (in the underlying probability space). Then for any subset $S=\{v_1,\ldots,v_k\}\subset\{1,\ldots,n\}$, the probability $P[\,W_V\!=\!1\,|\,\mathcal{E}\,]$ for a randomly chosen V in $\{1,\ldots,n\}\setminus S$ is bounded by*

$$P\big[\,W_V\!=\!1\,\big|\,\mathcal{E}\,\big] \leq v_{\mathrm{ns}}(\mathcal{G}) + c'(\mathcal{G})\cdot\sqrt{\tfrac{1}{n-k}\log\!\big(\tfrac{1}{P[\,\mathcal{E}\,]}\big)}$$

where $c'(\mathcal{G}) = 3\cdot 2^m c(\mathcal{G})/\min_x \pi(x)$ is some constant that only depends on $\mathcal{G}$.

The following is an immediate consequence.

▶ **Corollary 20.** *Let $\mathcal{G}$ be a game with complete support. Consider an execution of the n-fold repetition $\mathcal{G}^n$ with an arbitrary non-signaling strategy for $\mathcal{G}^n$. For any $\ell\in\{1,\ldots,n\}$, let $\mathcal{E}_\ell$ be the event that the ℓ-th repetition is accepted, i.e. $W_\ell = 1$. Then for any subset $S=\{v_1,\ldots,v_k\}\subset\{1,\ldots,n\}$, there exists $v_{k+1}\in\{1,\ldots,n\}\setminus S$ such that*

$$P\big[\,\mathcal{E}_{v_{k+1}}\,\big|\,\mathcal{E}_{v_1}\wedge\ldots\wedge\mathcal{E}_{v_k}\,\big] \leq v_{\mathrm{ns}}(\mathcal{G}) + c'(\mathcal{G})\cdot\sqrt{\tfrac{1}{n-k}\log\!\big(\tfrac{1}{P[\,\mathcal{E}_{v_1}\wedge\ldots\wedge\mathcal{E}_{v_k}\,]}\big)}$$

where $c'(\mathcal{G})$ is some constant that only depends on $\mathcal{G}$.

[5] In our case, the relevant constraint matrix A has $n = |\mathcal{X}|\cdot|\mathcal{A}|$ columns and at most $2\big((|\mathcal{A}|\cdot|\mathcal{X}|+|\mathcal{X}|^2)^m+|\mathcal{X}|\big)$ rows. Let $\Delta := \max\big\{\big|(B^{-1})_{ij}\big| \mid \text{B a nonsingular submatrix of } A\big\}$, which depends only $m, |\mathcal{A}|, |\mathcal{X}|$.

Proof (of Lemma 19). Let $\pi_\circ > 0$ be such that $\pi(x) \geq \pi_\circ$ for all $x \in \mathcal{X}$; by assumption on $\mathcal{G}$, such a $\pi_\circ$ exists. By re-ordering the (strategies of the) n executions, we may assume without loss of generality that $S = \{n-k+1, \dots, n\}$, and we now need to argue about the probability over a random V in $\{1, \dots, n-k\}$. To simplify notation, let us define

$$\varepsilon := \sqrt{\tfrac{1}{n-k} \log\big(\tfrac{1}{P[\mathcal{E}]}\big)}\,.$$

Fix a subset $I \subseteq \{1, \dots, m\}$ and let $J = \{1, \dots, m\} \setminus I$ be the complement of I. Consider the distribution

$$P_{\boldsymbol{X}_I \boldsymbol{X}_J \boldsymbol{A}_I} = P_{\boldsymbol{X}_I \boldsymbol{A}_I} \cdot P_{\boldsymbol{X}_J | \boldsymbol{X}_I \boldsymbol{A}_I} = P_{\boldsymbol{X}_I \boldsymbol{A}_I} \cdot P_{\boldsymbol{X}_J | \boldsymbol{X}_I} = P_{\boldsymbol{X}_I \boldsymbol{A}_I} \cdot \prod_{\ell=1}^{n} P_{X_J^\ell | \boldsymbol{X}_I} = P_{\boldsymbol{X}_I \boldsymbol{A}_I} \cdot \prod_{\ell=1}^{n} P_{X_J^\ell | \boldsymbol{X}_I \boldsymbol{A}_I}$$

where the second equality is due to non-signaling, the third due to the independence of every pair (X_I^ℓ, X_J^ℓ), and the third again due to non-signaling. We can thus apply Lemma 2 (with $T = (\boldsymbol{X}_I, \boldsymbol{A}_I)$ and $U^\ell = X_J^\ell$) and obtain

$$\begin{aligned}(n-k) \cdot \varepsilon = \sqrt{(n-k) \log\big(\tfrac{1}{P[\mathcal{E}]}\big)} &\geq \sum_{\ell=1}^{n-k} \big\| P_{\boldsymbol{X}_I X_J^\ell \boldsymbol{A}_I | \mathcal{E}} - P_{\boldsymbol{X}_I \boldsymbol{A}_I | \mathcal{E}} \cdot P_{X_J^\ell | \boldsymbol{X}_I \boldsymbol{A}_I} \big\| \\ &\geq \sum_{\ell=1}^{n-k} \big\| P_{X_I^\ell X_J^\ell A_I^\ell | \mathcal{E}} - P_{X_I^\ell A_I^\ell | \mathcal{E}} \cdot P_{X_J^\ell | X_I^\ell A_I^\ell} \big\| = \sum_{\ell=1}^{n-k} \big\| P_{X_I^\ell X_J^\ell A_I^\ell | \mathcal{E}} - P_{X_I^\ell A_I^\ell | \mathcal{E}} \cdot P_{X_J^\ell | X_I^\ell} \big\| \\ &= \sum_{\ell=1}^{n-k} \big\| P_{X_I^\ell X_J^\ell | \mathcal{E}} \cdot P_{A_I^\ell | X_I^\ell X_J^\ell \mathcal{E}} - P_{X_I^\ell | \mathcal{E}} \cdot P_{A_I^\ell | X_I^\ell \mathcal{E}} \cdot P_{X_J^\ell | X_I^\ell} \big\| \,.\end{aligned}$$

The first inequality holds by Lemma 2. The second inequality follows from Lemma 1 which states that the distance of the random variables $X_I^\ell, X_J^\ell, A_I^\ell$ cannot be larger than the distance of all random variables $\boldsymbol{X}_I, X_J^\ell, \boldsymbol{A}_I$. The subsequent equality holds due to the non-signaling condition between subsets I and J, and the last equality is a simple re-writing of some probabilities.

By means of Lemma 2 (setting T to be a constant), we can also conclude that $\sum_\ell \| P_{X_I^\ell X_J^\ell | \mathcal{E}} - P_{X_I^\ell X_J^\ell} \|$, and thus in particular $\sum_\ell \| P_{X_I^\ell | \mathcal{E}} - P_{X_I^\ell} \|$, is upper bounded by $(n-k)\varepsilon$. Therefore, noting that $P_{X_I^\ell X_J^\ell} = P_{X_I X_J}$, we can conclude that

$$\sum_{\ell=1}^{n-k} \big\| P_{X_I X_J} \cdot P_{A_I^\ell | X_I^\ell X_J^\ell \mathcal{E}} - P_{X_I X_J} \cdot P_{A_I^\ell | X_I^\ell \mathcal{E}} \big\| \leq 3(n-k)\varepsilon \,.$$

By summing over all subsets $I \subseteq \{1, \dots, m\}$ (and letting J be its complement), changing the order of the summation, and defining

$$\varepsilon_\ell := \sum_I \big\| P_{X_I X_J} \cdot P_{A_I^\ell | X_I^\ell X_J^\ell \mathcal{E}} - P_{X_I X_J} \cdot P_{A_I^\ell | X_I^\ell \mathcal{E}} \big\|$$

we get

$$\sum_{\ell=1}^{n-k} \varepsilon_\ell \leq 3 \cdot 2^m (n-k) \varepsilon \,.$$

Note that by definition of ε_ℓ, for any choice of I and $J = \{1, \dots, m\} \setminus I$, it holds that

$$\big\| P_{X_I X_J} \cdot P_{A_I^\ell | X_I^\ell X_J^\ell \mathcal{E}} - P_{X_I X_J} \cdot P_{A_I^\ell | X_I^\ell \mathcal{E}} \big\| \leq \varepsilon_\ell \,,$$

and hence, by the lower bound $\pi_\circ$ on $P_{X_I X_J}$, that

$$\big\|P_{A_I^\ell|X_I^\ell X_J^\ell \mathcal{E}}(\cdot|x_I,x_J) - P_{A_I^\ell|X_I^\ell \mathcal{E}}(\cdot|x_I)\big\| \leq \frac{\varepsilon_\ell}{\pi_\circ}$$

for any x_I and x_J. For any $\ell \in \{1,\ldots,n-k\}$, consider the strategy $\tilde{q}_\ell$ for (one execution of) $\mathcal{G}$, defined by $\tilde{q}_\ell(a|x) = P_{A^\ell|X^\ell\mathcal{E}}(a|x)$. By the above, $\tilde{q}_\ell$ is $(\varepsilon_\ell/\pi_\circ)$-almost non-signaling.

Furthermore, by the definition of $\tilde{q}_\ell$, the probability $P[\,\mathcal{E}_\ell\,|\,\mathcal{E}\,]$ that the ℓ-th repetition of the n-fold repetition of $\mathcal{G}$ is accepted equals the probability $v[\tilde{q}_\ell](\mathcal{G})$ that a *single* execution of $\mathcal{G}$ is accepted when strategy $\tilde{q}_\ell$ is played. Since $\tilde{q}_\ell$ is $(\varepsilon_\ell/\pi_\circ)$-almost non-signaling, it follows from Proposition 18 that this probability is at most $v_{\mathrm{ns}}(\mathcal{G}) + c(\mathcal{G})\cdot\varepsilon_\ell/\pi_\circ$. The claimed bound on $P[\,\mathcal{E}_V\,|\,\mathcal{E}\,]$ for a randomly chosen V in $\{1,\ldots,n-k\}$ now follows from the bound on $\sum_\ell \varepsilon_\ell$, where $c'(\mathcal{G})$ is given by $3\cdot 2^m c(\mathcal{G})/\pi_\circ$. ◀

We are now ready to prove our main concentration bound.

Proof (of Theorem 15). Let K be some integer parameter, to be defined later. Let $V_1,\ldots,V_K$ be a random subset of distinct integers from $\{1,\ldots,n\}$, and let D_k be the random variable $D_k = W_{V_k} = \mathsf{V}(X^{V_k},A^{V_k})$ for any $k \in \{1,\ldots,K\}$. Understanding $V_1,\ldots,V_K$ as a "sample subset" of the n parallel repetitions of $\mathcal{G}$, D_k indicates whether the k-th game in the sample is won. A pair $(d_1,\ldots,d_k) \in \{0,1\}^k$ and $(v_1,\ldots,v_k) \in \{1,\ldots,n\}$ of k-tuples is called *typical* if $P_{D_1\cdots D_k|V_1\cdots V_k}(d_1,\ldots,d_k|v_1,\ldots,v_k) \geq 2^{-2K}$. Let $\mathcal{T}_k$ be the event that $(D_1\cdots D_k)$ and $(V_1\cdots V_k)$ form a typical pair. Note that the corresponding complementary events satisfy $\bar{\mathcal{T}}_k \Rightarrow \bar{\mathcal{T}}_{k+1}$ as well as

$$P[\,\bar{\mathcal{T}}_k\,] = \sum_{\substack{\text{atypical pairs}\\ (d_1\ldots d_k),(v_1\ldots v_k)}} P_{V_1\cdots V_k}(v_1,\ldots,v_k)\, P_{D_1\cdots D_k|V_1\cdots V_k}(d_1\cdots d_k|v_1\cdots v_k) < 2^{-K}\,.$$

Let $\gamma := 1 - v_{\mathrm{ns}}(\mathcal{G}) - \varepsilon$ where $\varepsilon := \delta/3$. Note that we obviously may assume that $\delta \leq 1 - v_{\mathrm{ns}}(\mathcal{G})$ so that $\gamma > 0$. We now define a sequence of random variables $M_0,\ldots,M_K$ as follows. Random variable M_0 takes the value 0 with certainty, and M_{k+1} is inductively defined as

$$M_{k+1} := \begin{cases} M_k + \gamma & \text{if } D_{k+1}\!=\!1 \text{ and } \mathcal{T}_k \\ M_k - (1-\gamma) & \text{otherwise}\,. \end{cases}$$

We want to show that $M_0,\ldots,M_K$ forms a supermartingale. We fix $k \in \{0,\ldots,K-1\}$ and we fix values $(v_1,\ldots,v_k)$ for the random variables $V_1,\ldots,V_k$. Up to the end of this paragraph, all probabilities etc. are to be understood conditioned on these values. We define $\mathcal{E}$ to be the event that $D_1,\ldots,D_k$ take on some arbitrary but fixed values $(d_1,\ldots,d_k)$. If the pair $(d_1,\ldots,d_k)$ and $(v_1,\ldots,v_k)$ is atypical, then conditioned on $\mathcal{E}$ we have $M_{k+1} = M_k + \gamma - 1 < M_k$ and thus $\mathbb{E}[M_{k+1}|M_0\cdots M_k] < \mathbb{E}[M_k|M_0\cdots M_k] = M_k$. In the other case, if the pair $(d_1,\ldots,d_k)$ and $(v_1,\ldots,v_k)$ is typical then $P[\,\mathcal{E}\,] \geq 2^{-2K}$. Furthermore, Lemma 19 implies that $P_{D_{k+1}|\mathcal{E}}(1) = P[\,\mathcal{E}_{V_{k+1}}\,|\,\mathcal{E}\,] \leq v_{\mathrm{ns}}(\mathcal{G}) + c'(\mathcal{G})\sqrt{\log(1/P[\,\mathcal{E}\,])/(n-k)} \leq v_{\mathrm{ns}}(\mathcal{G}) + c'(\mathcal{G})\sqrt{2K/(n-K)}$. We want this last term to be upper bounded by $v_{\mathrm{ns}}(\mathcal{G}) + \varepsilon = 1-\gamma$, which we achieve by choosing K as $K := \lfloor \alpha n \rfloor$ where $\alpha := \min\{\varepsilon^2/(3c'(\mathcal{G})^2), 1/3\}$, as can easily be verified. It follows that $\mathbb{E}[M_{k+1}|M_0\cdots M_k] \leq (1-\gamma)(M_k+\gamma) + \gamma(M_k - (1-\gamma)) = M_k$ (when conditioning on $\mathcal{E}$). Since the argument that the $M_0,\ldots,M_K$ form a supermartingale holds independent of the choice of $(d_1,\ldots,d_k)$ and of the choice of $(v_1,\ldots,v_k)$, $M_0,\ldots,M_K$ indeed forms a supermartingale in the original probability space (without conditioning on the values for

$V_1, \ldots, V_k)$. Therefore,

$$\begin{aligned}
&P\Big[\sum_{k=1}^{K} D_k \geq (v_{\mathrm{ns}}(\mathcal{G}){+}2\varepsilon)K\Big] \\
&\qquad \leq P\big[\,\bar{\mathcal{T}}_K\,\big] + P\big[\,M_K \geq (v_{\mathrm{ns}}(\mathcal{G}){+}2\varepsilon)K\gamma - (1{-}v_{\mathrm{ns}}(\mathcal{G}){-}2\varepsilon)K(1{-}\gamma)\,\big] \\
&\qquad \leq 2^{-K} + P\big[\,M_K \geq (\gamma - 1 + v_{\mathrm{ns}}(\mathcal{G}) + 2\varepsilon)K\,\big] \;=\; 2^{-K} + P[\,M_K \geq \varepsilon K\,] \\
&\qquad \leq 2^{-K} + \exp(-\varepsilon^2 K/2) \;<\; 2\exp(-\varepsilon^2 K/2)\,.
\end{aligned}$$

The first inequality holds by definition of M_K, and the second by a simple manipulation of the terms. The equality holds by definition of γ, and the subsequent inequality by the Azuma-Hoeffding Inequality. Finally, the last inequality holds since $\varepsilon < 1$ and $\exp(\frac{1}{2}) < 2$.

On the other hand, setting $\overline{D} := \frac{1}{K}\sum_{k=1}^{K} D_k$, we can also write

$$P\big[\,\overline{D} \geq v_{\mathrm{ns}}(\mathcal{G}) + 2\varepsilon\,\big] \geq P\big[\,\overline{W} > v_{\mathrm{ns}}(\mathcal{G}) + \delta\,\big] \cdot P\big[\,\overline{D} \geq v_{\mathrm{ns}}(\mathcal{G}) + 2\varepsilon \,\big|\, \overline{W} > v_{\mathrm{ns}}(\mathcal{G}) + \delta\,\big]$$

where by the Hoeffding Inequality (and using that $\varepsilon = \delta/3$)

$$P\big[\,\overline{D} \geq v_{\mathrm{ns}}(\mathcal{G}){+}2\varepsilon \,\big|\, \bar{W} > v_{\mathrm{ns}}(\mathcal{G}) + \delta\,\big] \geq 1 - \exp(-2\varepsilon^2 K)\,.$$

Therefore,

$$P\big[\,\overline{W} > v_{\mathrm{ns}}(\mathcal{G}) + \delta\,\big] \leq \frac{2\exp(-\varepsilon^2 K/2)}{1 - \exp(-2\varepsilon^2 K)}\,.$$

In case that $\exp(-2\varepsilon^2 K) < \frac{1}{4}$, we obtain the bound

$$P\big[\,\overline{W} > v_{\mathrm{ns}}(\mathcal{G}) + \delta\,\big] \leq \frac{8}{3}\exp(-\varepsilon^2 K/2)\,. \tag{5}$$

Note that in the other case, if $\exp(-2\varepsilon^2 K) \geq \frac{1}{4}$, then $2\exp(-\varepsilon^2 K/2) \geq 1$ and the bound (5) holds trivially.

Setting $\mu := 1/(2 \cdot 3^5 \cdot c'(\mathcal{G})^2)$, and recalling that $\varepsilon = \delta/3$ and $K := \lfloor \alpha n \rfloor$ with α chosen as $\alpha := \min\{\varepsilon^2/(3c'(\mathcal{G})^2), 1/3\}$, leads to the claim. ◀

4 Conclusion and Open Questions

This article initiates the investigation of the behavior of multi-player nonlocal games under parallel repetition. For the case of the non-signaling value, we provide a concentration bound for games with complete support. Our results might serve as a stepping stone for the investigation of the quantum and classical values. Other interesting questions include improving the rate of repetition (e.g. by making it independent of the minimal probability that any question is asked) or finding cryptographic applications, for instance in position-based cryptography.

Acknowledgments. We would like to thank Tobias Müller for insightful discussions about the sensitivity of linear programs. HB is supported by a 7th framework EU SIQS grant. CS is supported by an NWO VENI grant.

References

1 Boaz Barak, Anup Rao, Ran Raz, Ricky Rosen, and Ronen Shaltiel. Strong parallel repetition theorem for free projection games. In Irit Dinur, Klaus Jansen, Joseph Naor, and

José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 12th International Workshop, APPROX 2009, and 13th International Workshop, RANDOM 2009, Berkeley, CA, USA, August 21–23, 2009. Proceedings*, volume 5687 of *Lecture Notes in Computer Science*, pages 352–365. Springer, 2009.

2 Jop Briët, Harry Buhrman, Troy Lee, and Thomas Vidick. Multipartite entanglement in xor games. *Quantum Information & Computation*, 13(3-4):334–360, March 2013.

3 Harry Buhrman, Nishanth Chandran, Serge Fehr, Ran Gelles, Vipul Goyal, Rafail Ostrovsky, and Christian Schaffner. Position-based quantum cryptography: Impossibility and constructions. In Phillip Rogaway, editor, *Advances in Cryptology – CRYPTO 2011*, volume 6841 of *Lecture Notes in Computer Science*, pages 429–446. Springer Berlin / Heidelberg, 2011.

4 Harry Buhrman, Serge Fehr, Christian Schaffner, and Florian Speelman. The garden-hose model. In *Innovations in Theoretical Computer Science, ITCS'13, Berkeley, CA, USA, January 9–12, 2013*, pages 145–158. ACM, 2013.

5 A. Chailloux and G. Scarpa. Parallel Repetition of Entangled Games with Exponential Decay via the Superposed Information Cost. arxiv:1310.7787, 2013.

6 Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect parallel repetition theorem for quantum xor proof systems. *Computational Complexity*, 17(2):282–299, 2008.

7 I. Dinur, D. Steurer, and T. Vidick. A parallel repetition theorem for entangled projection games. arxiv:1310.4113, 2013.

8 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

9 Thomas Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(1):141–172, 2009.

10 R. Jain, A. Pereszlényi, and P. Yao. A parallel repetition theorem for entangled two-player one-round games under product distributions. arxiv:1311.6309, 2013.

11 Julia Kempe, Oded Regev, and Ben Toner. Unique games with entangled provers are easy. *SIAM J. Comput.*, 39(7):3207–3229, July 2010.

12 Julia Kempe and Thomas Vidick. Parallel repetition of entangled games. In *Proceedings of the 43rd annual ACM symposium on Theory of computing*, STOC'11, pages 353–362, New York, NY, USA, 2011. ACM.

13 Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM J. Comput.*, 40(6):1871–1891, 2011.

14 Ran Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, June 1998.

15 Ricky Rosen. A k-provers parallel repetition theorem for a version of no-signaling model. *Discrete Math., Alg. and Appl.*, 2(4):457–468, 2010.

16 A. Schrijver. *Theory of Linear and Integer Programming*. Wiley Series in Discrete Mathematics & Optimization. John Wiley & Sons, 1998.

17 Marco Tomamichel, Serge Fehr, Jędrzej Kaniewski, and Stephanie Wehner. A monogamy-of-entanglement game with applications to device-independent quantum cryptography. *New Journal of Physics*, 15(10):103002, 2013.

Quantum Communication Complexity with Coherent States and Linear Optics

Juan Miguel Arrazola[1] and Norbert Lütkenhaus[1,2]

1 **Institute for Quantum Computing, University of Waterloo**
200 University Avenue West, N2L 3G1 Waterloo, Ontario, Canada
2 **Department of Physics and Astronomy, University of Waterloo**
200 University Avenue West, N2L 3G1 Waterloo, Ontario, Canada

Abstract

We introduce a general mapping for encoding quantum communication protocols involving pure states of multiple qubits, unitary transformations, and projective measurements into another set of protocols that employ coherent states of light in a superposition of optical modes, linear optics transformations and measurements with single-photon threshold detectors. This provides a general framework for transforming a wide class of protocols in quantum communication into a form in which they can be implemented with current technology. In particular, we apply the mapping to quantum communication complexity, providing general conditions under which quantum protocols can be implemented with coherent states and linear optics while retaining exponential separations in communication complexity compared to the classical case. Finally, we make use of our results to construct a protocol for the Hidden Matching problem that retains the known exponential gap between quantum and classical one-way communication complexity.

1998 ACM Subject Classification J.2 Physics, E.4 Formal Models of Communication

Keywords and phrases Quantum Communication Complexity, Quantum Optics

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.36

1 Introduction

What information-processing tasks are unachievable in a classical world but become possible when exploiting the intrinsic quantum mechanical properties of physical systems? This question has been a driving force of numerous research endeavours over the last two decades and remarkable progress has been made in our understanding of the advantages that quantum mechanics can provide, as well is in developing the experimental platforms that will allow them to be realized in practice [18, 7, 13, 22]. An example pertains to the field of quantum communication [14], where quantum systems can be used, for instance, to distribute secret keys [4, 5] or reduce the amount of communication required for joint computations [8, 9, 19, 2].

In terms of experimental implementations, only quantum key distribution (QKD) has been routinely demonstrated and deployed over increasingly complex networks and large distances [24, 21]. This is possible largely due to the fact that, fundamentally, QKD can be carried out with sequences of independent signals and measurements [22]. QKD and other cryptographic applications are easier to implement, as imperfections in implementations only need to be overcome to the point of being able to achieve their qualitative goal.

Other tasks, such as those in quantum communication complexity, face the additional challenge of demonstrating, in practice, their quantitative improvements over classical alternatives. Moreover, many of these tasks require sophisticated quantum states to be transmitted and measured. As such, there is a large set of quantum communication protocols

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 36–47
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

whose potential advantages currently escape the grasp of available technology. Thus, only a few proof-of-principle implementations have been reported [28, 26, 16].

Confronted with these challenges we face two alternatives: We can either strive to improve current technology or we can flip the issue around and ask: Can protocols in quantum communication be adapted to a form that makes them ready to be deployed with available techniques? To adopt the latter strategy is to push for a theoretical reformulation that converts previously intractable protocols into a form that, while conserving their relevant features, eliminates the obstacles affecting their implementation. This is precisely the road that has already been successfully followed for QKD.

In this work, we describe in detail an abstract mapping that converts quantum communication protocols that use pure states of multiple qubits, unitary operations, and projective measurements into another class of protocols that use only a sequence of coherent states, linear optics operations, and measurements with single-photon threshold detectors. The new class of protocols requires a number of optical modes equal to the dimension of the original states, but the number of photons can be chosen freely and is typically very small. This results in the signal states occupying a small Hilbert space, so that they can only be used to transmit the equivalent of a number of qubits that is only logarithmic in the number of modes used. We proceed by examining how the mapping may be generally applied in the context of quantum communication complexity and conclude by illustrating a coherent-state protocol for the Hidden Matching problem.

2 Coherent-state Protocols

We consider a wide class of quantum communication protocols that require only three basic operations: the preparation of pure states of a fixed dimension, unitary transformations on these states, and projective measurements on a canonical basis. This set of protocols is not completely general since we are not accounting for the possibility of shared entanglement or non-unitary evolution, although these extensions could potentially be considered. The simplest form of a protocol in this class is one in which Alice prepares a state $|\psi\rangle$ and sends it to Bob, who then applies a unitary transformation U_B to that state, followed by a projective measurement on the canonical basis. More complex protocols can be constructed by increasing the number of these basic operations as well as the number of parties. Even though these protocols generally involve states of some arbitrary dimension d, we can always think of them as corresponding to a system of $\mathcal{O}(\log_2 d)$ qubits. Hence, we refer to them as *qubit protocols*.

An exact implementation of such protocols can be achieved without the use of actual physical qubits by instead considering a single photon in a superposition of optical modes. Any pure state $|\psi\rangle = \sum_{k=1}^{d} \lambda_k |k\rangle$, with $\sum_{k=1}^{d} |\lambda_k|^2 = 1$, can be equally thought of as the state of a single photon in a superposition of d modes

$$a_\psi^\dagger |0\rangle = \sum_{k=1}^{d} \lambda_k |1\rangle_k \,, \tag{1}$$

where $a_\psi^\dagger = \sum_{k=1}^{d} \lambda_k b_k^\dagger$ for a collection of creation operators $\{b_1, b_2, \ldots, b_d\}$ corresponding to d optical modes, and where $|1\rangle_k$ is the state of a single photon in the k-th mode.

In this picture, unitary operations correspond exactly to linear optics transformations, and measurements in the canonical basis are equivalent to a photon counting measurement in each of the modes. From a practical perspective, the issue with implementing qubit protocols in terms of a single photon in a superposition of modes is that the experimental preparation of these states also presents daunting challenges. Instead, we are interested in an adaptation of this formulation of qubit protocols into another that is more readily implementable in practice.

With this purpose in mind, we outline a method for converting qubit protocols into another class of protocols that, although seemingly disparate, actually retain the essential properties of the original ones. We call these *coherent-state protocols* since they can be implemented by using only coherent states of light and linear optics operations. The recipe for constructing coherent-state protocols is specified by the following rules:

Coherent-state Mapping

1. The original Hilbert space $\mathcal{H}$ of dimension d with canonical basis $\{|1\rangle, |2\rangle, \ldots, |d\rangle\}$ is mapped to a set of d orthogonal optical modes with corresponding annihilation operators $\{b_1, b_2, \ldots, b_d\}$:

$$|k\rangle \longrightarrow b_k. \tag{2}$$

2. A state $|\psi\rangle = \sum_{k=1}^{d} \lambda_k |k\rangle$ is mapped to a coherent state with parameter α in the mode $a_\psi = \sum_{k=1}^{d} \lambda_k b_k$:

$$\begin{aligned} |\psi\rangle \longrightarrow |\alpha, \psi\rangle &:= D_{a_\psi}(\alpha)|0\rangle \\ &= \bigotimes_{k=1}^{d} |\alpha\,\lambda_k\rangle_k\,, \end{aligned} \tag{3}$$

where $|\alpha\,\lambda_k\rangle_k$ is a coherent state with parameter $\alpha\,\lambda_k$ in the k-th mode. The value of α can be chosen freely but remains fixed.

3. A unitary operation U acting on a state in $\mathcal{H}$ is mapped into linear optics transformation corresponding to the same unitary operator U acting on the modes $\{b_1, b_2, \ldots, b_d\}$. Thus, the transformation of a state is linked to a transformation of the modes as:

$$|\psi'\rangle = U|\psi\rangle \longrightarrow \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_d \end{pmatrix} = U \begin{pmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_d \end{pmatrix}. \tag{4}$$

4. A projective measurement in the canonical basis $\{|1\rangle, |2\rangle, \ldots, |d\rangle\}$ is mapped into a two-outcome measurement in each of the modes:

$$\{|k\rangle\langle k|\} \longrightarrow \{F^k_{\text{no-click}}, F^k_{\text{click}}\}, \tag{5}$$

where $F^k_{\text{no-click}} = |0\rangle\langle 0|$ is a projection onto the vacuum, $F^k_{\text{no-click}} = \sum_{n=1}^{\infty} |n\rangle_k\langle n|_k$ and $|n\rangle_k$ is a state with n photons in the k-th mode. As such, an outcome in a coherent-state protocol corresponds to a pattern of clicks in the modes. For coherent-state protocols, the observation of N clicks correspond to a particular pattern of N outcomes of a qubit protocol. Thus, an immediate interpretation of the outcomes is not provided by the mapping, but instead must be chosen according to the particular context.

Since any qubit protocol can be constructed from the basic operations of state preparation, unitary transformations, and projective measurements, the above instructions are sufficient to construct the coherent-state version of any qubit protocol up to an interpretation of the

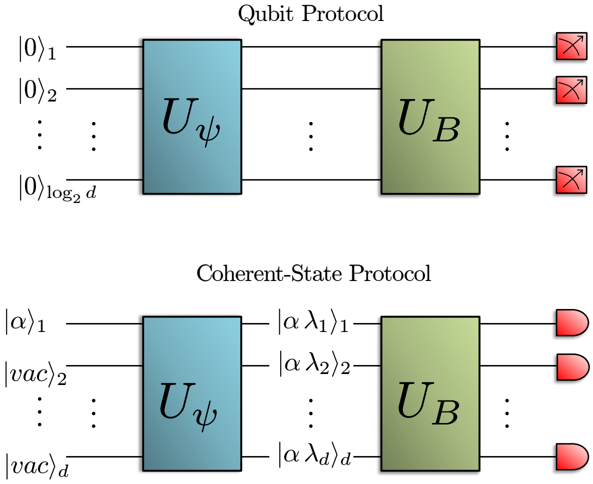


Figure 1 (Color online) In a simple qubit protocol, Alice prepares a state $|\psi\rangle = \sum_{k=1}^{d} \lambda_k |k\rangle$ of $\log_2 d$ qubits by applying a unitary transformation U_ψ on an inital state $|\bar{0}\rangle := |0\rangle^{\otimes \log_2 d}$. She sends the state to Bob, who applies a unitary transformation U_B and measures the resulting state in the computational basis. In the equivalent coherent-state protocol, the initial state corresponds to a coherent state in a single mode and the vacuum on the others. The state $|\alpha, \psi\rangle = \bigotimes_{k=1}^{d} |\alpha\,\lambda_k\rangle_k$ is prepared by applying the transformation U_ψ to the optical modes. This state is sent to Bob, who applies the transformation U_B and consequently measures each mode for the presence of photons with threshold single-photon detectors.

measurement outcomes. As an illustration, a simple qubit protocol and its coherent-state counterpart are depicted in Fig. 1.

An immediate appealing property of coherent-state protocols is that their implementation faces much lesser obstacles than their qubit counterparts. Indeed, the fundamental challenge of a quantum-optical implementation of qubit protocols lies in the difficulty of generating entangled states of many qubits and performing global unitary transformations on them. On the other hand, coherent-state protocols face significantly less daunting obstacles. The experimental generation of coherent states is a commonplace task and the construction of linear-optical circuits can, in principle, be realized with simple devices such as beam splitters and phase-shifters [20], though experimental challenges may remain depending on the required unitary operation. Moreover, the platforms for linear optics experiments continue to improve at a fast rate, most notably with the development of integrated optics [25] and time-bin encodings [17, 10].

As we have mentioned already, an advantage of coherent-state protocols is that they employ a coherent state in a superposition of modes, which is equivalent to a tensor product of individual coherent states across the various modes. However, qubit protocols usually require high amounts of entanglement. This seems to indicate that the 'quantumness' of the original qubit protocol has been lost through the mapping. Nevertheless, it is important to realize that this is not the case, as coherent-state protocols showcase a truly quantum property: non-orthogonality. Given two states $|\alpha, \psi\rangle = \bigotimes_{k=1}^{d} |\alpha\,\lambda_k\rangle_k$ and $|\alpha, \varphi\rangle = \bigotimes_{k=1}^{d} |\alpha\,\nu_k\rangle_k$, with $d \gg \alpha$, the individual coherent states in each mode will typically be highly non-orthogonal, i.e. $\langle \alpha\,\nu_k | \alpha\,\lambda_k \rangle \approx 1$. Moreover, the presence of single-photon detectors also permit truly quantum phenomena, such as unambiguous state discrimination of non-orthogonal states. In fact, it can be useful to intuitively think of the coherent-state mapping as an exchange between entanglement and non-orthogonality, since an implementation of qubit protocols with actual physical qubits usually requires entanglement amongst the qubits.

In coherent-state protocols, the average photon number, $|\alpha|^2$, is a parameter that can be chosen independently of the dimension of the states of the original qubit protocol. This is to be put in contrast with any quantum-optical realization of a qubit protocol, which inevitably requires a number of photons that scales with the dimension of the states. Hence, coherent-state protocols offer an intrinsic saving in the number of photons required for their implementation. The drawback, of course, is that the number of optical modes is exponentially larger than the number of qubits in the original protocol. This means that the mapping is only suitable for its application to protocols that originally require only a small number of qubits. From a theoretical perspective, the relationship between these two types of protocols may provide an insight into the trade-offs between different resources in quantum communication, as well as into the interplay between entanglement and non-orthogonality in quantum mechanics.

Now that we have outlined the coherent-state mapping, we continue by describing how these techniques can be applied in the construction of protocols in quantum communication complexity.

3 Quantum Communication Complexity

Communication complexity is the study of the amount of communication that is required to perform distributed information-processing tasks. This corresponds to the scenario in which two parties, Alice and Bob, respectively receive inputs $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$ and their goal is to collaboratively compute the value of a Boolean function $f(x,y)$ with as little communication as possible [27]. Although they can always do this by communicating their entire input, the pertaining question in communication complexity is: What is the minimum amount of communication that is really needed? Likewise, quantum communication complexity studies the case where the parties are allowed to employ quantum resources such as quantum channels and shared entanglement [6, 7].

Remarkably, it has been proven that there exist various problems for which the use of quantum resources offer exponential savings in communication compared to their classical counterparts [9, 19, 2, 12, 8]. As discussed previously, coherent-state protocols require a number of modes that is exponentially larger than the number of qubits of the original protocol. Thus, from a practical perspective, the exponential savings that are possible for certain tasks in quantum communication complexity conveniently balance the exponential increase in the number of modes, making them a natural candidate for the application of the coherent-state mapping.

We are first interested in quantifying the amount of communication that takes place in a quantum communication complexity protocol. Informally, this is done by counting the number of qubits that are employed. But what happens if a protocol uses physical systems that are manifestly *not* qubits? In that case, we quantify the amount of communication in terms of the smallest number of qubits that would be required, in principle, to replicate the performance of the protocol. More precisely, if a quantum communication protocol uses states in a Hilbert space of dimension d, this space can be associated to a system of $\mathcal{O}(\log_2 d)$ qubits. Therefore, the amount of communication C in a quantum protocol is generally given by

$$C = \log_2[\dim(\mathcal{H})] \tag{6}$$

where $\mathcal{H}$ is the smallest Hilbert space containing all states of the protocol. Moreover, Holevo's theorem [15] guarantees that no more than $\log_2 d$ classical bits of information could be transmitted, on average, by a quantum protocol that uses state in a Hilbert space of dimension d.

Quantifying communication in qubit protocols is straightforward. For coherent-state protocols, even though the *actual* Hilbert space associated to all possible signal states is large (distinct coherent states are linearly independent), they effectively occupy a small Hilbert space, as is expressed in the following theorem:

▶ **Theorem 1.** *[1] For any state $|\psi\rangle$ in a Hilbert space of dimension d, let $|\alpha,\psi\rangle$ be the state associated to it through the coherent-state mapping. Then, for any $\epsilon > 0$, there exists a Hilbert space $\mathcal{H}_\alpha$ of dimension d_α such that*

$$\log_2 d_\alpha = \mathcal{O}(\log_2 d),$$
$$\langle\alpha,\psi|P_{\mathcal{H}_\alpha}|\alpha,\psi\rangle \geq 1-\epsilon,$$

and where $P_{\mathcal{H}_\alpha}$ is the projector onto $\mathcal{H}_\alpha$.

Proof. For a given $\Delta > 0$, let $\mathcal{H}_\alpha$ be the subspace spanned by the set of Fock states $\{|n_1\rangle \otimes |n_2\rangle \otimes \ldots \otimes |n_d\rangle\}$ over d modes whose total photon number $n = \sum_{k=1}^d n_k$ satisfies $|n - |\alpha|^2| \leq \Delta$. The dimension of $\mathcal{H}_\alpha$ is equal to the the number of distinct ways in which n photons can be distributed into the d different modes. Since the photons are indistinguishable, this quantity is given by the binomial factor $\binom{n+d-1}{d-1}$ [23]. In the case of $\mathcal{H}_\alpha$, there are 2Δ different possible values of n, the largest being $n = |\alpha|^2 + \Delta$. Thus, the dimension d_α of this subspace satisfies

$$\begin{aligned}\log_2 d_\alpha \leq \log_2 &\left[2\Delta\binom{|\alpha|^2+\Delta+d-1}{d-1}\right]\\ \leq(|\alpha|^2+\Delta)\log_2 &\left[(|\alpha|^2+\Delta+d-1)\right]+\log_2(2\Delta),\end{aligned} \tag{7}$$

which is $\mathcal{O}(\log_2 d)$ for any fixed α and Δ.

Now notice that the number $\langle\alpha,\psi|P_{\mathcal{H}_\alpha}|\alpha,\psi\rangle$ is equal to the probability of performing a photon number measurement on $|\alpha,\psi\rangle$ and obtaining a value satisfying $|n-|\alpha|^2| \leq \Delta$. Since any coherent state $|\alpha,\psi\rangle$ has a Poissonian photon number distribution with mean $|\alpha|^2$, we can use the properties of this distribution to calculate the probability that the measured number of photons lies within the desired range. This probability satisfies [11]

$$P(|n-|\alpha|^2| \geq \Delta) \leq 2e^{-|\alpha|^2}\left(\frac{e|\alpha|^2}{|\alpha|^2+\Delta}\right)^{|\alpha|^2+\Delta} \tag{8}$$

which can be made equal to any $\epsilon > 0$ by choosing Δ accordingly while keeping α fixed. ◀

Therefore, the fact that the mean photon number $|\alpha|^2$ is fixed in coherent-state protocols leads to the states involved effectively occupying a Hilbert space of dimension that is comparable to that of the original one. This implies that the asymptotic behaviour of the amount of communication is the same for both classes of protocols. Moreover, the effectively unused sections of the entire Hilbert space can still be used, in principle, for other purposes such as the transmission of additional information.

We now focus on the bounded-error model in which Alice and Bob have randomness at their disposal and need only determine the value of the function $f(x,y)$ with probability greater or equal to $1-\epsilon$ for all possible values of x and y. They can send quantum states to each other, apply unitary transformations on these states, and make measurements in the same way as the quantum communication protocols discussed before. Since they are only interested in learning the value of the function, their final measurement can always be thought of as a projective measurement onto two orthogonal subspaces H_0 and H_1, corresponding to $f(x, y) = 0$ and $f(x, y) = 1$ respectively.

In a coherent-state version of this model, the crucial difference lies in the measurement stage, where the subspaces H_0 and H_1 are mapped onto sets of modes S_0 and S_1. Unlike the qubit protocol, there can be clicks happening in both sets of modes and as a consequence, checking for the presence of clicks does not suffice to determine the value of the function. Instead, in order to decide between both possible values of $f(x, y)$, we opt for the strategy of counting the number of clicks that occur in each set of modes and selecting the one with the largest number of clicks.

Let C_b be the random variable corresponding to the number of clicks observed in the set of modes S_b, with $b = 0, 1$. The distribution of C_b is known as a Poisson-binomial distribution and its expectation value is given by

$$\mathbb{E}(C_b) = \sum_{k\in S_b} p_{\alpha,k} := \mu_b, \tag{9}$$

where $p_{\alpha,k} = 1 - \exp(-|\alpha\, \lambda_k|^2)$ is the probability of obtaining a click in the k-th mode.

This distribution can be difficult to work with in its exact form, so it is usual to approximate it by a Poisson distribution with the same mean. This approximation can be made precise through the following result:

▶ **Theorem 2.** *[3] Let C_b be a Poisson-binomial random variable with mean μ_b. Similarly, let L_b be a Poisson random variable with the same mean μ_b. Then, for any set A, it holds that*

$$|\Pr(C_b \in A) - \Pr(L_b \in A)| \leq \min(1, \mu_b^{-1})\tau_b, \tag{10}$$

where $\tau_b := \sum_{k\in S_b}(p_{\alpha,k})^2$ and $p_{\alpha,k}$ is the probability of obtaining a click on the k-th mode.

We can use this fact to show that, under certain conditions, a coherent-state version of a bounded-error qubit protocol also gives the correct value of the function with bounded error.

▶ **Theorem 3.** *Let a qubit protocol for communication complexity have a probability of success $P \geq 1 - \epsilon$. Then the corresponding coherent-state protocol has a probability of success $P_\alpha > 1 - \epsilon$ if there exists a $\mu = |\alpha|^2$ such that*

$$2e^{-P\mu}(2eP\mu)^{\mu/2} + \max_{\mu_0,\mu_1}\{\min(1, \mu_b^{-1})\}\tau \leq \epsilon \tag{11}$$

where μ_b is the expected number of clicks in the set of modes S_b and $\tau = \sum_k (p_{\alpha,k})^2$.

Proof. Without loss of generality, we take $f(x, y) = 0$ to correspond to the correct value of the function. We can bound the success probability as

$$\begin{aligned} P_\alpha &= \Pr(C_0 > C_1) \\ &\geq \Pr(C_0 > \tfrac{\mu}{2}) \Pr(C_1 < \tfrac{\mu}{2}) \\ &= (1 - \Pr(C_0 < \tfrac{\mu}{2}))(1 - \Pr(C_1 > \tfrac{\mu}{2})). \end{aligned}$$

From Theorem 2 we can also write

$$\begin{aligned} \Pr(C_0 < \tfrac{\mu}{2}) &\leq \Pr(L_0 < \tfrac{\mu}{2}) + \min(1, \mu_0^{-1})\tau_0 \\ &\leq e^{-\mu_0}\left(\frac{2e\mu_0}{\mu}\right)^{\mu/2} + \min(1, \mu_0^{-1})\tau_0, \end{aligned}$$

where we have bounded the Poisson distribution as in Eq. (8). Similarly we have

$$\Pr(C_1 > \tfrac{\mu}{2}) \leq e^{-\mu_1}\left(\frac{2e\mu_1}{\mu}\right)^{\mu/2} + \min(1,\mu_1^{-1})\tau_1.$$

Putting these together we get

$$\begin{aligned}P_\alpha \geq &\left(1-e^{-\mu_0}\left(\frac{2e\mu_0}{\mu}\right)^{\mu/2} - \min(1,\mu_0^{-1})\tau_0\right)\left(1-e^{-\mu_1}\left(\frac{2e\mu_1}{\mu}\right)^{\mu/2} - \min(1,\mu_1^{-1})\tau_1\right)\\ >&1-e^{-\mu_0}\left(\frac{2e\mu_0}{\mu}\right)^{\mu/2} - e^{-\mu_1}\left(\frac{2e\mu_1}{\mu}\right)^{\mu/2} - \min(1,\mu_0^{-1})\tau_0 - \min(1,\mu_1^{-1})\tau_1\\ \geq&1-e^{-P\mu}(2eP\mu)^{\mu/2} - e^{-(1-P)\mu}(2e(1-P)\mu)^{\mu/2} - \max_{\mu_0,\mu_1}\{\min(1,\mu_b^{-1})\}\tau,\end{aligned}$$

where $\tau = \tau_0 + \tau_1 = \sum_k (p_{\alpha,k})^2$ and we have used the fact that

$$P\mu = \sum_{k\in S_0} |\alpha|^2 p_k > \sum_{k\in S_0}(1-e^{-|\alpha|^2 p_k}) = \mu_0 \tag{12}$$

and similarly $(1-P)\mu > \mu_1$. Whenever $P > 1/2$, it holds that $e^{-P\mu}(2eP\mu) > e^{-(1-P)\mu}(2e(1-P)\mu)$ so we can finally write

$$P_\alpha > 1 - 2e^{-P\mu}(2eP\mu)^{\mu/2} - \max_{\mu_0,\mu_1}\{\min(1,\mu_b^{-1})\}\tau. \tag{13}$$

From this expression it is clear that whenever condition (11) holds, $P_\alpha > 1-\epsilon$ as desired. ◀

Notice that the quantity $2e^{-P\mu}(2eP\mu)^{\mu/2}$ can be made arbitrarily small for any $P > 1-\epsilon$ by choosing a large enough value of $\mu = |\alpha|^2$. However, large values of μ result in higher values of the individual click probabilities $\{p_{k,\alpha}\}$ and consequently larger values of $\tau = \sum_k (p_{\alpha,k})^2$, making it harder for the quantity $\max_{\mu_0,\mu_1}\{\min(1,\mu_b^{-1})\}\tau$ to be small. Therefore, condition (11) may only be satisfied when the original probabilities $\{p_i\}$ are very small, as this results in a small τ even when μ is large. Of course, whenever the communicated states lie in a Hilbert space of large dimension, we expect the outcome probabilities to be small and, consequently, the coherent-state protocol to function adequately.

We would like to apply the coherent-state mapping to known protocols in quantum communication complexity. In fact, this has already been demonstrated in [1], where, essentially, a coherent-state mapping was used to construct a protocol for quantum fingerprinting. We now discuss how the mapping can be used to construct a protocol for the Hidden Matching Problem.

The Hidden Matching Problem. In this communication complexity problem, Alice receives an n-bit string $x \in \{0,1\}^n$ as input, with n an even number. Additionally, Bob receives a perfect matching $M = \{(i_1,j_1),(i_2,j_2),\ldots,(i_{n/2},j_{n/2})\}$ on the set of numbers $\{1,2,\ldots,n\}$, i.e. a partition into $n/2$ disjoint pairs. A perfect matching can be visualized as a graph with n vertices and $n/2$ edges, where each vertex is connected to only one other vertex. Only one-way communication from Alice to Bob is permitted and their goal is to output an element of the matching (i,j) and a bit value b such that $b = x_i \oplus x_j$, where x_i is the i-th bit of the string x.

It has been shown that any classical protocol requires $\Omega(\sqrt{n})$ bits of communication, even in the presence of errors [2]. On the other hand, there exists an efficient quantum

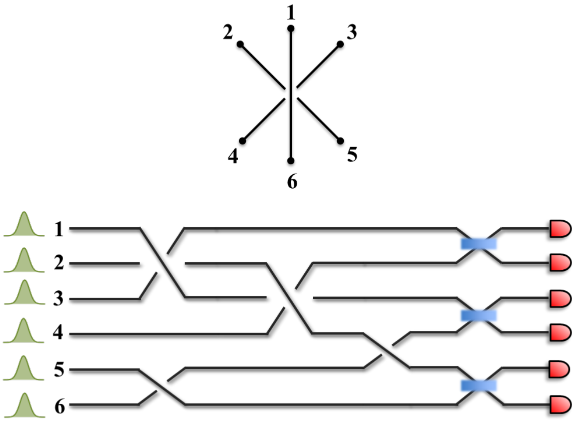


Figure 2 (Color online) An example of an implementation of a coherent-state protocol for the Hidden Matching problem. Alice receives a string of six bits and Bob receives the matching $(1,6),(2,5),(3,4)$, as represented in the graph. Alice encodes her input values in the phases of six coherent states in different modes and sends them to Bob. His measurement consists of a circuit in which the modes are permutated in accordance with the matching and then interefere pairwise in three balanced beamsplitters. Bob can output a correct solution to the problem based on the detectors that click.

protocol that uses only $\mathcal{O}(\log_2 n)$ qubits of communication and outputs the correct answer with certainty. In this protocol, Alice prepares the state

$$|x\rangle = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(-1)^{x_i}|i\rangle \tag{14}$$

and sends it to Bob, who measures it in the basis

$$\{\tfrac{1}{\sqrt{2}}(|i\rangle \pm |j\rangle)\}, \tag{15}$$

with $(i,j) \in M$. The outcome $\frac{1}{\sqrt{2}}(|i\rangle+|j\rangle)$ only occurs if $x_i \oplus x_j = 0$ and similarly, $\frac{1}{\sqrt{2}}(|i\rangle-|j\rangle)$ only occurs if $x_i \oplus x_j = 1$. This allows Bob to give a correct output after performing his measurement. Note that Bob's measurement basis is constructed from the canonical basis by applying a Hadamard transformation to the subspaces $\{|i\rangle, |j\rangle\}$, with $(i,j) \in M$.

To construct a coherent-state protocol for the Hidden Matching problem, we just have to apply the rules of the mapping. In this case, Alice prepares the state

$$|\alpha, x\rangle = \bigotimes_{i=1}^{n}\left|(-1)^{x_i}\tfrac{\alpha}{\sqrt{n}}\right\rangle \tag{16}$$

and sends it to Bob. The linear-optical equivalent of a Hadamard gate is a balanced beam-splitter, so Bob's measurement consists of interfering each of the pairs of modes $\{b_i, b_j\}$ (with $(i,j) \in M$) in a balanced beam-splitter and detecting photons in the outputs as illustrated in Fig. 2. If the incoming states to the input ports of the beam splitter are

$$\left|(-1)^{x_i}\tfrac{\alpha}{\sqrt{n}}\right\rangle \otimes \left|(-1)^{x_j}\tfrac{\alpha}{\sqrt{n}}\right\rangle, \tag{17}$$

the output states will be

$$\left|\left(1+(-1)^{x_i\oplus x_j}\right)\tfrac{\alpha}{\sqrt{n}}\right\rangle \otimes \left|\left(1-(-1)^{x_i\oplus x_j}\right)\tfrac{\alpha}{\sqrt{n}}\right\rangle. \tag{18}$$

Notice that for each possible value of $x_i \oplus x_j$, one of the output states will be a vacuum while the other is a coherent-state with non-zero amplitude. Therefore, we can associate a value of $x_i \oplus x_j$ to each of the output detectors so that whenever a click occurs, the correct value can be inferred with certainty. Even if there are many clicks, they will always correspond to a correct value. Thus, the only issue that can arise is that no-clicks occur and the probability that this happens is given by

$$P_{\text{no-click}} = e^{-|\alpha|^2}, \tag{19}$$

which can be made arbitrarily small by choosing α appropriately. Moreover, Theorem 1 guarantees that the amount of communication in the coherent-state protocol is $\mathcal{O}(\log_2 n)$ and an exponential separation in communication complexity is maintained.

4 Conclusions

We have outlined a general framework for encoding quantum communication protocols involving pure states, unitary transformations, and projective measurements, into another set of protocols that employs coherent states of light in a superposition of modes, linear optics transformations, and measurements with single-photon threshold detectors. Although seemingly disparate at first glance, qubit and coherent-state protocols share in fact many properties, including the amount of communication required and the outcome statistics. Moreover, since the mapping depends on a parameter α that can be freely chosen, coherent-state protocols offer increased tunability compared to qubit protocols.

This work thus provides a general method for mapping protocols in quantum communication into a form in which they can be implemented with current technology. It is of great interest to explore what other protocols in quantum communication, besides the ones we have outlined in this work, could be implemented by applying the coherent-state mapping to their qubit versions.

Fundamentally, coherent-state protocols require a fixed and small number of photons at the price of an exponentially large number of optical modes. For practical purposes, this implies that their application to protocols that originally require a large number of qubits will be difficult. Nevertheless, the fact that very few photons are needed not only implies an inherent savings in the energy cost of communication, but may also provide other practical advantages. For example, since optical multiplexing is limited by nonlinear effects arising from large amplitudes of the electromagnetic field, the fact that coherent-state protocols employ signals with very few photons implies that they can be easily assimilated into multiplexing schemes, or even provide a new way of multiplexing quantum messages, for example by utilizing the unused sections of the entire Hilbert space. Additionally, the low photon number may result in increased clock rates: Since only a few clicks are expected to occur even in cases when many modes are transmitted, the detector dead times and jitter times do not pose a barrier to the achievable rates.

From a theoretical perspective, the coherent-state mapping can be thought of as a tool for understanding fundamental aspects about quantum communication and information. In particular, the mapping provides us with a connection between two intrinsically quantum properties: entanglement and non-orthogonality. It may also serve as a theoretical test bed for proving results regarding qubit protocols, in the same way as many other dualities have been useful in both physics and mathematics.

Acknowledgements. J. M. Arrazola would like to thank A. Ignjatovic for her help in preparing this manuscript and designing the figures, and he is grateful for the support of the

Mike and Ophelia Lazaridis Fellowship. We acknowledge support from Industry Canada, the NSERC Strategic Project Grant (SPG) FREQUENCY and the NSERC Discovery Program.

References

1 Juan Miguel Arrazola and Norbert Lütkenhaus. Quantum fingerprinting with coherent states and a constant mean number of photons. *Phys. Rev. A*, 89:062305, Jun 2014.

2 Ziv Bar-Yossef, T. S. Jayram, and Iordanis Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *STOC*, pages 128–137, 2004.

3 Andrew D Barbour, Lars Holst, and Svante Janson. *Poisson approximation*. Clarendon press Oxford, 1992.

4 C. H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, 68(21):3121–3124, may 1992.

5 C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of IEEE International Conference on Computers, Systems, and Signal Processing, Bangalore, India*, pages 175–179, New York, dec 1984. IEEE.

6 Gilles Brassard. Quantum communication complexity. *Foundations of Physics*, 33(11):1593–1616, 2003.

7 Harry Buhrman, Richard Cleve, Serge Massar, and Ronald de Wolf. Nonlocality and communication complexity. *Rev. Mod. Phys.*, 82:665–698, Mar 2010.

8 Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87:167902, Sep 2001.

9 Harry Buhrman, Richard Cleve, and Avi Wigderson. Quantum vs. classical communication and computation. In *STOC*, pages 63–68, 1998.

10 John M. Donohue, Megan Agnew, Jonathan Lavoie, and Kevin J. Resch. Coherent ultrafast measurement of time-bin encoded photons. *Phys. Rev. Lett.*, 111:153602, Oct 2013.

11 Massimo Franceschetti, Olivier Dousse, David Tse, and Patrick Thiran. Closing the gap in the capacity of wireless networks via percolation theory. *Information Theory, IEEE Transactions on*, 53(3):1009–1018, 2007.

12 Dmitry Gavinsky. Classical interaction cannot replace a quantum message. In *STOC*, pages 95–102, 2008.

13 Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum metrology. *Physical review letters*, 96(1):010401, 2006.

14 Nicolas Gisin and Rob Thew. Quantum communication. *Nature Photonics*, 1(3):165–171, 2007.

15 Alexander Semenovich Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973.

16 Rolf T. Horn, S. A. Babichev, Karl-Peter Marzlin, A. I. Lvovsky, and Barry C. Sanders. Single-qubit optical quantum fingerprinting. *Phys. Rev. Lett.*, 95:150502, Oct 2005.

17 Peter C. Humphreys, Benjamin J. Metcalf, Justin B. Spring, Merritt Moore, Xian-Min Jin, Marco Barbieri, W. Steven Kolthammer, and Ian A. Walmsley. Linear optical quantum computing in a single spatial mode. *Phys. Rev. Lett.*, 111:150501, Oct 2013.

18 Thaddeus D Ladd, Fedor Jelezko, Raymond Laflamme, Yasunobu Nakamura, Christopher Monroe, and Jeremy L O'Brien. Quantum computers. *Nature*, 464(7285):45–53, 2010.

19 Ran Raz. Exponential separation of quantum and classical communication complexity. In *STOC*, pages 358–367, 1999.

20 M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.*, 73:58–61, 1994.

21 M Sasaki, M Fujiwara, H Ishizuka, W Klaus, K Wakui, M Takeoka, S Miki, T Yamashita, Z Wang, A Tanaka, et al. Field test of quantum key distribution in the tokyo qkd network. *Optics Express*, 19(11):10387–10409, 2011.

22 Valerio Scarani, Helle Bechmann-Pasquinucci, Nicolas J Cerf, Miloslav Dušek, Norbert Lütkenhaus, and Momtchil Peev. The security of practical quantum key distribution. *Reviews of modern physics*, 81(3):1301, 2009.

23 R. Sheldon. *A First Course In Probability, 6/E.* Pearson Education, 2002.

24 Damien Stucki, Nino Walenta, Fabien Vannel, Robert Thomas Thew, Nicolas Gisin, Hugo Zbinden, S Gray, CR Towery, and S Ten. High rate, long-distance quantum key distribution over 250 km of ultra low loss fibres. *New Journal of Physics*, 11(7):075003, 2009.

25 Sébastien Tanzilli, Anthony Martin, Florian Kaiser, Marc P De Micheli, Olivier Alibart, and Daniel B Ostrowsky. On the genesis and evolution of integrated quantum optics. *Laser & Photonics Reviews*, 6(1):115–143, 2012.

26 Pavel Trojek, Christian Schmid, Mohamed Bourennane, Časlav Brukner, Marek Żukowski, and Harald Weinfurter. Experimental quantum communication complexity. *Phys. Rev. A*, 72:050305, Nov 2005.

27 Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *STOC*, pages 209–213, 1979.

28 Jun Zhang, Xiao-Hui Bao, Teng-Yun Chen, Tao Yang, Adán Cabello, and Jian-Wei Pan. Experimental quantum "guess my number" protocol using multiphoton entanglement. *Phys. Rev. A*, 75:022302, Feb 2007.

Bounds on Entanglement Assisted Source-channel Coding Via the Lovász ϑ Number and Its Variants[*]

Toby Cubitt[1], Laura Mančinska[2], David Roberson[3], Simone Severini[4], Dan Stahlke[5], and Andreas Winter[6]

1 Universidad Complutense de Madrid and University of Cambridge
2 University of Waterloo and National University of Singapore
3 Nanyang Technological University
4 University College London
5 Carnegie Mellon University
6 ICREA and Universitat Autonòma de Barcelona

Abstract

We study zero-error entanglement assisted source-channel coding (communication in the presence of side information). Adapting a technique of Beigi, we show that such coding requires existence of a set of vectors satisfying orthogonality conditions related to suitably defined graphs G and H. Such vectors exist if and only if $\vartheta(\overline{G}) \leq \vartheta(\overline{H})$ where ϑ represents the Lovász number. We also obtain similar inequalities for the related Schrijver ϑ^- and Szegedy ϑ^+ numbers.

These inequalities reproduce several known bounds and also lead to new results. We provide a lower bound on the entanglement assisted cost rate. We show that the entanglement assisted independence number is bounded by the Schrijver number: $\alpha^*(G) \leq \vartheta^-(G)$. Therefore, we are able to disprove the conjecture that the one-shot entanglement-assisted zero-error capacity is equal to the integer part of the Lovász number. Beigi introduced a quantity β as an upper bound on α^* and posed the question of whether $\beta(G) = \lfloor \vartheta(G) \rfloor$. We answer this in the affirmative and show that a related quantity is equal to $\lceil \vartheta(G) \rceil$. We show that a quantity $\chi_{\text{vect}}(G)$ recently introduced in the context of Tsirelson's conjecture is equal to $\lceil \vartheta^+(\overline{G}) \rceil$.

1998 ACM Subject Classification E.4 Coding and Information Theory

Keywords and phrases source-channel coding, zero-error capacity, Lovász theta

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.48

1 Introduction

The *zero-error source-channel coding* problem is as follows. Suppose Alice wishes to send a message $x \in X$ to Bob through a noisy classical channel $\mathcal{N} : S \to V$ in such a way that Bob may deduce Alice's message with zero probability of error. Alice encodes her message via some function $f : X \to S$ before sending it through the channel. Bob is aided by some side

[*] TC is supported by the Royal Society. LM is supported by the Ministry of Education (MOE) and National Research Foundation Singapore, as well as MOE Tier 3 Grant "Random numbers from quantum processes" (MOE2012-T3-1-009). DR is supported by an NTU start-up grant awarded to D. V. Pasechnik. SS is supported by the Royal Society and the British Heart Foundation. DS is supported by the National Science Foundation through Grant PHY-1068331. AW is supported by the European Commission (STREPs "QCS" and "RAQUEL"), the European Research Council (Advanced Grant "IRQUAT") and the Philip Leverhulme Trust; furthermore by the Spanish MINECO, project FIS2008-01236, with the support of FEDER funds.

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 48–51
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

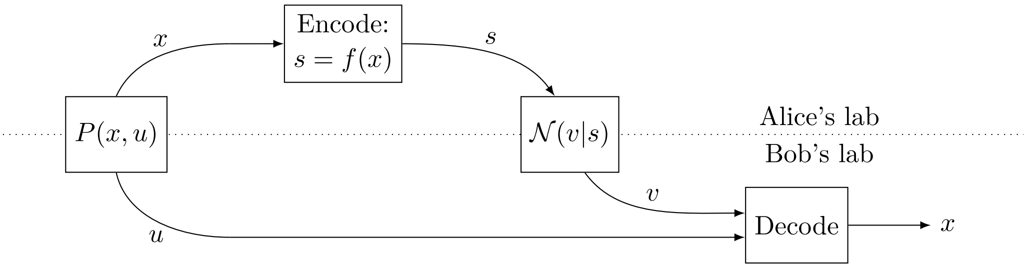


Figure 1 A zero-error source-channel coding scheme.

information $u \in U$ regarding Alice's message. Formally, we can imagine that the symbols x and u originate from a *dual source* with probability $P(x,u)$. See Fig. 1.

The success of this protocol can be analyzed using a pair of graphs: G with vertices from X and H with vertices from V, having edges

$$x \sim_G y \iff \exists u \in U \text{ such that } P(x,u)P(y,u) \neq 0 \tag{1}$$

$$s \sim_H t \iff \mathcal{N}(v|s)\mathcal{N}(v|t) = 0 \text{ for all } v \in V, \tag{2}$$

where $P(x,u)$ is the probability of input pair x, u and $\mathcal{N}(v|s)$ is the probability that the channel outputs v given input s. G is the *characteristic graph* of P and H is the complement of the *confusability graph* of $\mathcal{N}$. Intuitively, G represents the information that needs to be sent and H represents the information that survives the channel. Bob is able to decode x (with zero chance of error) if and only if Alice's encoding satisfies $x \sim_G y \implies f(x) \sim_H f(y)$ [11]. Such a function is called a *homomorphism* from G to H. If such a function exists then G is *homomorphic* to H, written $G \to H$.

Many graph quantities can be defined in terms of homomorphisms [8, 9], and the above protocol puts these in an operational context. If there is no side information then $G = K_n$, the complete graph on $n = |X|$ vertices. The largest n such that $K_n \to H$ is the clique number $\omega(H)$. Thus the largest number of error-free messages that can be sent through $\mathcal{N}$ is $\omega(H)$ (equivalently, $\alpha(\overline{H})$, the independence number of the complementary graph). If $\mathcal{N}$ is the perfect channel then $H = K_n$ with $n = |S|$. The smallest n such that $G \to K_n$ is the chromatic number $\chi(G)$. This is the size of the smallest channel that suffices to communicate inputs from a dual source with characteristic graph G.

Source-channel coding may also be considered in the case where Alice and Bob make use of an entanglement resource, Fig. 2 [3]. Now Alice's encoding operation consists of a POVM $\{M_s^x\}_{s\in S}$ depending on her input x and producing a value s to be sent to Bob through the channel. Bob can successfully decode if and only if

$$\rho_s^x \perp \rho_t^y \text{ for all } x \sim_G y \text{ and } s \not\sim_H t, \tag{3}$$

where ρ_s^x is Bob's share of the post-measurement entanglement resource after POVM outcome M_s^x. By analogy to the above, a successful protocol is called an *entanglement assisted homomorphism* from G to H. If such a thing exists, one writes $G \xrightarrow{*} H$. Also by way of analogy, the *entanglement assisted independence number* $\alpha^*(\overline{H})$ is the largest n such that $K_n \xrightarrow{*} H$ and the *entanglement assisted chromatic number* $\chi^*(G)$ is the smallest n such that $G \xrightarrow{*} K_n$. These have similar operational interpretations as $\alpha(\overline{H})$ and $\chi(G)$ discussed above.

We consider two relaxations of condition (3) for $G \xrightarrow{*} H$. The first we denote $G \xrightarrow{B} H$ since it reduces to a construction of Beigi [2] when $G = K_n$. We say $G \xrightarrow{B} H$ if there are vectors $|w\rangle$ and $|w_s^x\rangle$ such that

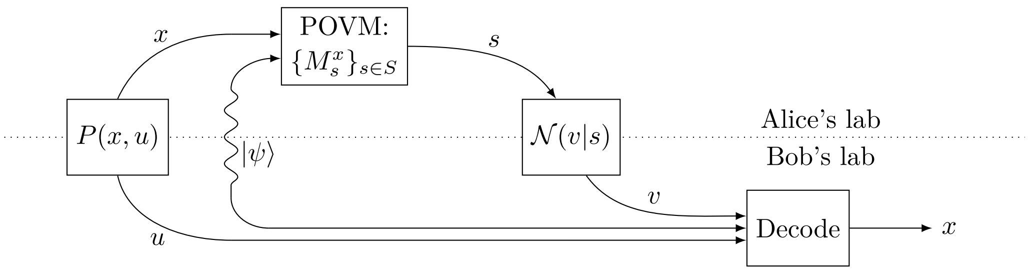


Figure 2 An entanglement assisted zero-error source-channel coding scheme.

1. $\langle w|w\rangle = 1$
2. $\sum_s |w_s^x\rangle = |w\rangle$
3. $\langle w_s^x|w_t^y\rangle = 0$ for all $x \sim_G y$, $s \not\sim_H t$
4. $\langle w_s^x|w_t^x\rangle = 0$ for all $s \neq t$.

Another relaxation $G \overset{+}{\to} H$ is defined similarly, except that the last condition is replaced by
4. $\langle w_s^x|w_t^y\rangle \geq 0$.

Since these are relaxed conditions, $G \overset{*}{\to} H$ implies $G \overset{B}{\to} H$ and $G \overset{+}{\to} H$. All of our results follow from two theorems. With $\bar{\vartheta}(G)$, $\bar{\vartheta}^-(G)$, and $\bar{\vartheta}^+(G)$ being the Lovász, Schrijver, and Szegedy numbers of the complementary graph $\overline{G}$, we have

▶ **Theorem 1.** *$G \overset{B}{\to} H$ if and only if $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$.*

▶ **Theorem 2.** *If $G \overset{+}{\to} H$ then $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$, $\bar{\vartheta}^-(G) \leq \bar{\vartheta}^-(H)$, and $\bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(H)$.*

A number of original results follow as immediate corollaries:

- Entanglement assisted zero-error source-channel coding ($G \overset{*}{\to} H$) requires $\bar{\vartheta}(G) \leq \bar{\vartheta}(H)$, $\bar{\vartheta}^-(G) \leq \bar{\vartheta}^-(H)$, and $\bar{\vartheta}^+(G) \leq \bar{\vartheta}^+(H)$.
- $\alpha^*(H) \leq \vartheta^-(H)$ (previously only $\alpha^*(H) \leq \vartheta(H)$ was known [2, 4]).
- The average number of channel uses required per input, in the asymptotic limit, is known as the *entanglement assisted cost rate* $\eta^*(G, \overline{H})$. Since $\bar{\vartheta}$ is multiplicative under appropriate graph products, $\eta^*(G, \overline{H}) \geq \log \bar{\vartheta}(G) / \log \bar{\vartheta}(H)$.
- Beigi defined $\beta(\overline{H})$ to be the largest n such that $K_n \overset{B}{\to} H$ (paraphrased into our terminology) and asked whether $\beta(\overline{H}) = \lfloor \bar{\vartheta}(H) \rfloor$. The answer is "yes" – this follows directly from Theorem 1.
- By considering instead $G \overset{B}{\to} K_n$ one can define a quantity similar to Beigi's, equal to $\lceil \bar{\vartheta}(H) \rceil$.

Also as immediate corollaries, we reproduce the following known results:

- $\chi^*(G) \geq \bar{\vartheta}^+(G)$ [3].
- There is a notion of a *quantum homomorphism* $G \overset{q}{\to} H$ defined in the context of a quantum pseudo-telepathy game [14, 13]. Since $G \overset{q}{\to} H \implies G \overset{*}{\to} H \implies G \overset{+}{\to} H$, the inequalities of Theorem 2 apply to $G \overset{q}{\to} H$ as well.

These various generalized homomorphisms can be arranged in a sequence of most to least strict:

$$G \to H \implies G \overset{q}{\to} H \implies G \overset{*}{\to} H \implies G \overset{+}{\to} H \implies G \overset{B}{\to} H. \tag{4}$$

It is known that the converse of the first implication does not hold [5, 14], and we show the converse of the last does not hold. The other two are open. The second converse holds if and only if entanglement assisted source-channel coding can always be accomplished using projective measurements and a maximally entangled state. The third converse holds if, loosely speaking, it is permissible to drop all mathematical structure from (3) except for the basic properties related to inner products $\langle \rho_s^x, \rho_t^y \rangle$.

It is not known whether there can be a gap between the asymptotic entanglement assisted zero-error capacity Θ^* and ϑ. To show such a gap requires a stronger bound on α^*. Since Beigi's β is now shown to be essentially no different from ϑ, this dashes the hope that β could be used to show such a gap. Our bound $\alpha^*(H) \leq \vartheta^-(H)$ would imply a gap, unless ϑ^- regularizes to ϑ in the asymptotic limit. Haemers provided a bound on Shannon capacity which is sometimes stronger than Lovász's bound [6, 7, 1, 12]; however, this bound does not apply to the entanglement assisted case [10].

References

1 Noga Alon. The shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998.

2 Salman Beigi. Entanglement-assisted zero-error capacity is upper-bounded by the Lovász ϑ function. *Physical Review A*, 82:010303, July 2010.

3 Jop Briët, Harry Buhrman, Monique Laurent, Teresa Piovesan, and Giannicola Scarpa. Zero-error source-channel coding with entanglement, 2013.

4 Runyao Duan, S. Severini, and A. Winter. Zero-error communication via quantum channels and a quantum Lovász ϑ-function. In *Proc. IEEE International Symposium on Information Theory (ISIT), 2011*, pages 64–68, August 2011.

5 Viktor Galliard and Stefan Wolf. Pseudo-telepathy, entanglement, and graph colorings. In *Proc. IEEE International Symposium on Information Theory (ISIT), 2002*, page 101, 2002.

6 Willem H. Haemers. An upper bound for the Shannon capacity of a graph. *Colloquia Mathematica Societatis Janos Bolyai*, 25:267–272, 1978.

7 Willem H. Haemers. On some problems of Lovász concerning the Shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25:231–232, 1979.

8 Geňa Hahn and Claude Tardif. Graph homomorphisms: structure and symmetry. In *Graph symmetry*, pages 107–166. Springer, 1997.

9 Pavol Hell and Jaroslav Nešetřil. *Graphs and Homomorphisms (Oxford Lecture Series in Mathematics and Its Applications)*. Oxford University Press, USA, 9 2004.

10 Debbie Leung, Laura Mančinska, William Matthews, Maris Ozols, and Aidan Roy. Entanglement can increase asymptotic rates of zero-error classical communication over classical channels. *Communications in Mathematical Physics*, 311(1):97–111, 2012.

11 Jayanth Nayak, Ertem Tuncel, and Kenneth Rose. Zero-Error Source-Channel Coding With Side Information. *IEEE Transactions on Information Theory*, 52(10):4626–4629, 2006.

12 René Peeters. Orthogonal representations over finite fields and the chromatic number of graphs. *Combinatorica*, 16(3):417–431, 1996.

13 David E. Roberson. *Variations on a Theme: Graph Homomorphisms*. PhD thesis, University of Waterloo, 2013.

14 David E. Roberson and Laura Mančinska. Graph Homomorphisms for Quantum Players, 2012.

Strong Converse for the Quantum Capacity of the Erasure Channel for Almost All Codes

Mark M. Wilde[1] and Andreas Winter[2]

1 Hearne Institute for Theoretical Physics
Department of Physics and Astronomy
Center for Computation and Technology
Louisiana State University
Baton Rouge, Louisiana 70808, USA
mwilde@lsu.edu

2 ICREA & Física Teórica
Informació i Fenomens Quántics
Universitat Autònoma de Barcelona
ES-08193 Bellaterra (Barcelona), Spain
andreas.winter@uab.cat

Abstract

A strong converse theorem for channel capacity establishes that the error probability in any communication scheme for a given channel necessarily tends to one if the rate of communication exceeds the channel's capacity. Establishing such a theorem for the quantum capacity of degradable channels has been an elusive task, with the strongest progress so far being a so-called "pretty strong converse." In this work, Morgan and Winter proved that the quantum error of any quantum communication scheme for a given degradable channel converges to a value larger than $1/\sqrt{2}$ in the limit of many channel uses if the quantum rate of communication exceeds the channel's quantum capacity. The present paper establishes a theorem that is a counterpart to this "pretty strong converse." We prove that the large fraction of codes having a rate exceeding the erasure channel's quantum capacity have a quantum error tending to one in the limit of many channel uses. Thus, our work adds to the body of evidence that a fully strong converse theorem should hold for the quantum capacity of the erasure channel. As a side result, we prove that the classical capacity of the quantum erasure channel obeys the strong converse property.

1998 ACM Subject Classification H.1.1 Systems and Information Theory, E.4 Coding and Information Theory, Error control codes

Keywords and phrases strong converse, quantum erasure channel, quantum capacity

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.52

1 Introduction

In his seminal paper on quantum error correction, Shor set out the task of determining the quantum capacity of a quantum channel [26], defined as the maximum rate at which it is possible to transmit qubits reliably over a noisy quantum communication channel. Subsequent to this, the coherent information was identified as being a relevant quantity for quantum capacity [23], a regularized upper bound on quantum capacity was established in terms of the coherent information [4, 5], and the coherent information lower bound on the quantum capacity was established by a sequence of works which are often said to bear

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 52–66
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

"increasing standards of rigor" [19, 27, 9]. [1] All of these works did not identify a tractable characterization of the quantum capacity in general, but Devetak and Shor subsequently proved that the coherent information is equal to the quantum capacity for a class of channels bearing the property of degradability [10]. Degradable channels are such that the receiver of the output of the channel can simulate the channel to the environment by applying a degrading map.

A particularly simple example of a degradable channel is the quantum erasure channel $\mathcal{N}_p$ [13], which has the following action on an input density operator ρ:

$$\mathcal{N}_p(\rho) \equiv (1-p)\rho + p|e\rangle\langle e|, \tag{1}$$

where $p \in [0,1]$ is the erasure probability and $|e\rangle$ is a state orthogonal to the input space (i.e., $\langle e|\rho|e\rangle = 0$ for all input ρ). One can readily show that the map to the environment is equivalent (up to isometry) to an erasure channel with the complementary probability:

$$\mathcal{N}_p(\rho) \equiv p\rho + (1-p)|e\rangle\langle e|. \tag{2}$$

The interpretation here is that if the receiver recovers the channel input, then the environment does not and instead receives the erasure flag, and vice versa.

The quantum capacity of the erasure channel was identified early on by employing a now well known "no-cloning" argument [7]. That is, when $p = 1/2$, the channels from input to the receiver and from input to the environment are the same, so that the quantum capacity of the original channel must vanish. If this were not the case, then it would be possible to send quantum data reliably to both the receiver and the environment of the channel, in violation of the no-cloning theorem. It is then possible to prove that the quantum capacity of the erasure channel in general is equal to $(1-2p)\log d$ for $p \geq 1/2$ and zero otherwise (in agreement with the aforementioned reasoning), where d is the dimension of the input space for the channel.

All of the above works established an understanding of quantum capacity in the following sense:

1. (Achievability) If the rate of quantum communication is below the quantum capacity, then there exists a scheme for quantum communication such that the fidelity approaches one in the limit of many channel uses.
2. (Weak Converse) If the rate of quantum communication is above the quantum capacity, then there cannot exist an error-free quantum communication scheme.

However, the theorem stated as such still leaves more to be desired. For example, it has been known for a long time that the classical capacity of a classical channel obeys the strong converse property [33, 1]: if the rate of communication exceeds capacity, then the error probability necessarily converges to one in the limit of many channel uses. Furthermore, many works have now established that the strong converse property holds for the classical capacity of several quantum channels [32, 22, 18, 31, 30, 3] and for the entanglement-assisted classical capacity of all quantum channels [6, 8, 14].

Thus, we are left with the strong converse question for the quantum capacity, with the goal being to sharpen our understanding of quantum capacity. In general, the quantum capacity of arbitrary channels can exhibit rather exotic behavior [28], so it seems reasonable to restrict attention for now to the class of degradable channels since they are more well behaved. In this spirit, a recent work has proved that the quantum capacity of all degradable

[1] However, see the later works in [16] and [15], which respectively set [19] and [27] on a firm foundation.

channels exhibits a property dubbed the "pretty strong converse" [20]. These authors have proven that the quantum error[2] of any quantum communication scheme for a degradable channel experiences a sudden jump from zero to at least $1/\sqrt{2}$ when the communicate rate crosses the quantum capacity threshold (this statement is in the limit of many channel uses). At the very least, we now know that the quantum capacity experiences this jump, but the work of [20] left open the question of whether the jump in quantum error is actually from zero to one in the limit of many channel uses.

In this paper, we prove a statement that is similar in spirit to the pretty strong converse: for almost all codes having a rate exceeding the quantum capacity of the erasure channel, the error necessarily converges to one in the limit of many channel uses. We should clarify that we do not prove a strong converse for all codes, but instead show that the strong converse property holds for almost all codes. We will be more precise in what follows with clarifying what we mean by "almost all codes," but suffice it for now to say if anyone devises a communication scheme for quantum communication over the erasure channel whose rate exceeds capacity, then the chances are very good that, regardless of the scheme, it will fail with probability converging to one in the limit of many channel uses.

In the absence of a proof that the strong converse holds, both the present paper and [20] are offering an increasing body of evidence that it should indeed hold for the class of quantum erasure channels. That is, both results allow us to conclude the following statement: all codes whose rate exceeds the quantum capacity of the erasure channel have a quantum error converging to $1/\sqrt{2}$ in the limit of many channel uses, and a large fraction of them in fact have quantum error converging to one.

This paper is organized as follows. The next section reviews the definition of an entanglement generation code. Section 3 then reviews the generalized divergence framework of Sharma and Warsi [25] for establishing bounds relating rate, error, and the channel of interest in any quantum communication protocol. Section 4 provides a proof for our main result: that the strong converse property holds for almost all codes used for quantum communication over the quantum erasure channel. We state some open directions in the conclusion. The appendix includes, as a side result, a proof that the strong converse holds for the classical capacity of the quantum erasure channel.

2 Entanglement generation codes

In this paper, we focus on entanglement generation codes, for which the goal is for the sender Alice to use the channel n times in order to share a state with the receiver Bob, such that this state is indistinguishable from a maximally entangled state. We focus on this task because the entanglement generation capacity of a quantum channel serves as an upper bound on its quantum capacity (this in turn is because a protocol for noiseless quantum communication can always be used to generate entanglement between sender and receiver). Thus, if one establishes an upper bound on the entanglement generation capacity, then this bound serves as an upper bound on the quantum capacity. However, we should emphasize again that our final statement is a bound that holds for almost all entanglement generation codes, so that we cannot conclude a full strong converse.

More formally, we now define an $(n, R, \varepsilon, \phi, D)$ entanglement generation code for a channel $\mathcal{N}$. Such a protocol begins with Alice preparing a state on $n+1$ systems, she sends n shares of the state through n instances of the channel, and then Bob decodes. That is,

[2] As quantified by the so-called "purified distance" (see Chapter 3 of [29], for example).

such a code begins with Alice preparing a state $|\phi\rangle_{AA_1\cdots A_n}$. The reduced state on system A has its rank equal to M, where $M = 2^{nR}$. Alice then transmits systems $A_1 \cdots A_n$ through n uses of the channel, leading to the state

$$\rho_{AB^n} \equiv \mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n}), \tag{3}$$

where $\mathcal{N}_{A^n\to B^n} \equiv \mathcal{N}^{\otimes n}$ and A^n is shorthand for $A_1 \cdots A_n$. Finally, Bob performs a decoding $D_{B^n\to\hat{B}}$, leading to the state

$$\omega_{A\hat{B}} \equiv D_{B^n\to\hat{B}}(\mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n})). \tag{4}$$

The fidelity of the code is given by

$$F \equiv \langle\Phi|_{A\hat{B}}\, \omega_{A\hat{B}}\, |\Phi\rangle_{A\hat{B}}, \tag{5}$$

where $|\Phi\rangle_{A\hat{B}}$ is the maximally entangled state

$$|\Phi\rangle_{A\hat{B}} \equiv \frac{1}{\sqrt{M}} \sum_{i=0}^{M-1} |i\rangle_A |i\rangle_{\hat{B}}, \tag{6}$$

so that the rate of entanglement generation is equal to $\frac{1}{n}\log_2 M$. An $(n, R, \varepsilon, \phi, D)$ code uses the state ϕ, the decoder D, the channel n times at rate R, and is such that the fidelity $F \geq 1-\varepsilon$. Note that without loss of generality, we can restrict our consideration to pure-state entanglement generation codes. For if the initial state is a mixed state $\rho_{AA_1\cdots A_n}$ and the following condition holds

$$\langle\Phi|_{A\hat{B}} D_{B^n\to\hat{B}}(\mathcal{N}_{A^n\to B^n}(\rho_{AA_1\cdots A_n}))|\Phi\rangle_{A\hat{B}} \geq 1-\varepsilon, \tag{7}$$

then there always exists at least one pure state in the spectral decomposition of $\rho_{AA_1\cdots A_n}$ which meets the same fidelity constraint given above.

3 Generalized divergence framework for quantum communication

We now recall the Sharma-Warsi framework for bounding fidelities in quantum communication [25]. We say that $\mathcal{D}(X||Y)$ is a *generalized divergence* if it satisfies the following monotonicity inequality for all quantum channels $\mathcal{M}$ and positive operators X and Y:

$$\mathcal{D}(X||Y) \geq \mathcal{D}(\mathcal{M}(X)||\mathcal{M}(Y)). \tag{8}$$

Let $I_{\mathcal{D}}(A\rangle B)_\rho$ denote the generalized coherent information of a bipartite state ρ_{AB}:

$$I_{\mathcal{D}}(A\rangle B)_\rho \equiv \min_{\sigma_B} \mathcal{D}(\rho_{AB}||I_A \otimes \sigma_B). \tag{9}$$

Let $I_{\mathcal{D}}(\mathcal{N})$ denote the generalized coherent information of a quantum channel $\mathcal{N}$:

$$I_{\mathcal{D}}(\mathcal{N}) \equiv \max_{\phi_{AA'}} I_{\mathcal{D}}(A\rangle B)_{\mathcal{N}_{A'\to B}(\phi_{AA'})} \tag{10}$$

$$= \max_{\phi_{AA'}} \min_{\sigma_B} \mathcal{D}(\mathcal{N}_{A'\to B}(\phi_{AA'})||I_A \otimes \sigma_B). \tag{11}$$

If the generalized divergence is equal to the von Neumann relative entropy, then the above expressions are equal to the usual coherent information of a quantum state and coherent information of a quantum channel, respectively.

We now establish a bound relating the rate and error of any entanglement generation code for a quantum channel $\mathcal{N}$ to the generalized coherent information of the tensor-power channel $\mathcal{N}^{\otimes n}$. For our purposes here, we begin by considering the generalized divergence between the state ρ_{AB^n} defined in (3) that is output from n uses of the channel and any other operator of the form $I_A \otimes \sigma_{B^n}$, where σ_{B^n} is a density operator on the systems B^n:

$$\mathcal{D}(\rho_{AB^n}||I_A \otimes \sigma_{B^n}). \tag{12}$$

By monotonicity under the application of the decoder $D_{B^n\to\hat{B}}$ to the system B^n, the following inequality holds

$$\mathcal{D}(\rho_{AB^n}||I_A \otimes \sigma_{B^n}) \geq \mathcal{D}\big(\omega_{A\hat{B}}||I_A \otimes D_{B^n\to\hat{B}}(\sigma_{B^n})\big). \tag{13}$$

Next, consider the following test (a completely positive trace-preserving map), which outputs a flag indicating whether a state is maximally entangled or not:

$$T_{A\hat{B}\to Z}(\cdot) \equiv \mathrm{Tr}\{\Phi_{A\hat{B}}(\cdot)\}|1\rangle\langle 1| + \mathrm{Tr}\{\big(I_{A\hat{B}} - \Phi_{A\hat{B}}\big)(\cdot)\}|0\rangle\langle 0|. \tag{14}$$

Intuitively, this test is simply asking, “Is the entanglement decoded or not?” Applying monotonicity of the generalized divergence under this test, we find that the following inequality holds

$$\mathcal{D}\big(\omega_{A\hat{B}}||I_A \otimes D_{B^n\to\hat{B}}(\sigma_{B^n})\big) \geq \mathcal{D}\big(T_{A\hat{B}\to Z}\big(\omega_{A\hat{B}}\big)||T_{A\hat{B}\to Z}\big(I_A \otimes D_{B^n\to\hat{B}}(\sigma_{B^n})\big)\big). \tag{15}$$

By defining

$$\rho_F \equiv F|1\rangle\langle 1| + (1-F)|0\rangle\langle 0|, \tag{16}$$

$$P_{\frac{1}{M}} \equiv \frac{1}{M}|1\rangle\langle 1| + \left(M - \frac{1}{M}\right)|0\rangle\langle 0|, \tag{17}$$

we see that

$$\mathcal{D}\big(T_{A\hat{B}\to Z}\big(\omega_{A\hat{B}}\big)||T_{A\hat{B}\to Z}\big(I_A \otimes D_{B^n\to\hat{B}}(\sigma_{B^n})\big)\big) = \mathcal{D}\Big(\rho_F||P_{\frac{1}{M}}\Big), \tag{18}$$

which follows from (5) and the fact that

$$\mathrm{Tr}\{\Phi_{A\hat{B}}\big(I_A \otimes D_{B^n\to\hat{B}}(\sigma_{B^n})\big)\} = \frac{1}{M}. \tag{19}$$

Thus, putting everything together, we obtain the following inequality

$$\mathcal{D}(\rho_{AB^n}||I_A \otimes \sigma_{B^n}) \geq \mathcal{D}\Big(\rho_F||P_{\frac{1}{M}}\Big). \tag{20}$$

This inequality holds for any choice of σ_{B^n}, so we can obtain the tightest upper bound on $\mathcal{D}(\rho_F||P_{\frac{1}{M}})$ for a particular entanglement generation code with initial state $\phi_{AA_1\cdots A_n}$ by taking a minimization over all such σ_{B^n}:

$$\min_{\sigma_{B^n}} \mathcal{D}(\rho_{AB^n}||I_A \otimes \sigma_{B^n}) \geq \mathcal{D}\Big(\rho_F||P_{\frac{1}{M}}\Big). \tag{21}$$

We can then remove the dependence of the bound on any particular entanglement generation code by taking a maximization over all initial states $\phi_{AA_1\cdots A_n}$:

$$\max_{\phi_{AA_1\cdots A_n}} \min_{\sigma_{B^n}} \mathcal{D}(\rho_{AB^n}||I_A \otimes \sigma_{B^n}) \geq \mathcal{D}\Big(\rho_F||P_{\frac{1}{M}}\Big). \tag{22}$$

By employing the definition in (10), we find that the bound is equivalent to

$$I_{\mathcal{D}}\big(\mathcal{N}^{\otimes n}\big) \geq \mathcal{D}\Big(\rho_F||P_{\frac{1}{M}}\Big). \tag{23}$$

3.1 Specializing to Rényi relative entropies

The above development applies for any divergence satisfying monotonicity, and the Rényi relative entropy is a particular example of a generalized divergence, defined as

$$D_\alpha(\rho||\sigma) \equiv \frac{1}{\alpha-1}\log_2 \ \text{Tr}\{\rho^\alpha \sigma^{1-\alpha}\}. \tag{24}$$

Monotonicity of $D_\alpha(\rho||\sigma)$ under quantum channels holds for all $\alpha \in [0,2]$ (see Appendix B of [29], for example). In the present paper, we are focused on $\alpha \in (1,2]$, especially when α is in a neighborhood near one in this interval. This is because the Rényi relative entropy converges to the von Neumann relative entropy as $\alpha \to 1$.

Now we can evaluate the bound in (21) for the case when the divergence is chosen to be the Rényi relative entropy:

$$\begin{aligned}
\min_{\sigma_{B^n}} \mathcal{D}(\rho_{AB^n}||I_A \otimes \sigma_{B^n}) &\geq D_\alpha\left(\rho_F||P_{\frac{1}{M}}\right) & (25)\\
&= \frac{1}{\alpha-1}\log_2\left[F^\alpha\left(\frac{1}{M}\right)^{1-\alpha} + (1-F)^\alpha\left(M-\frac{1}{M}\right)^{1-\alpha}\right] & (26)\\
&\geq \frac{1}{\alpha-1}\log_2\left[F^\alpha\left(\frac{1}{M}\right)^{1-\alpha}\right] & (27)\\
&= \frac{\alpha}{\alpha-1}\log_2[F] + \log_2 M & (28)\\
&= \frac{\alpha}{\alpha-1}\log_2[F] + nR & (29)
\end{aligned}$$

If we optimize over all entanglement generation codes, then we have the bound

$$\max_{\phi_{AA_1\cdots A_n}} \min_{\sigma_{B^n}} D_\alpha(\rho_{AB^n}||I_A\otimes\sigma_{B^n}) \geq \frac{\alpha}{\alpha-1}\log_2[F] + nR. \tag{30}$$

This is equivalent to

$$I_\alpha\left(\mathcal{N}^{\otimes n}\right) \geq \frac{\alpha}{\alpha-1}\log_2[F] + nR, \tag{31}$$

where we define the Rényi coherent information I_α of a quantum channel according to the recipe in (10). Rewriting this, the bound is equivalent to

$$F \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{n}I_\alpha\left(\mathcal{N}^{\otimes n}\right)\right)}. \tag{32}$$

▶ Remark. It is worth noting at this point that if it is possible to prove that $\frac{1}{n}I_\alpha(\mathcal{N}^{\otimes n})$ is an additive function of the channel $\mathcal{N}$, in the sense that

$$\frac{1}{n}I_\alpha\left(\mathcal{N}^{\otimes n}\right) = I_\alpha(\mathcal{N}) \tag{33}$$

for any finite n, then this would be sufficient to prove that the strong converse holds according to the argument of [22] (which has since been repeated in different contexts in both [25] and [14]). (In fact, any subadditivity relation of the following form would suffice: $I_\alpha(\mathcal{N}^{\otimes n}) \leq nI_\alpha(\mathcal{N})+o(n)$.) One could also consider using the recently developed sandwiched Rényi relative entropy [21, 31] in this context. So far, it is not clear to us whether either of the coherent information quantities derived from the traditional or sandwiched Rényi relative entropies are additive in the above sense for any degradable channel.

3.2 Application to the quantum erasure channel

We now specialize the above bounds to the case of the quantum erasure channel. Beginning from (25)-(29), we see that we can choose any state σ_{B^n} for establishing a bound relating rate and fidelity to an information quantity. So we choose $\sigma_{B^n} = [\mathcal{N}_p(\pi)]^{\otimes n} = ((1-p)\pi + p|e\rangle\langle e|)^{\otimes n}$, where $\pi = I/d$ is the maximally mixed qudit state on the input and $\mathcal{N}_p$ is the erasure channel defined in (1). This then leads to the following bound for any $(n, R, \varepsilon, \phi, D)$ entanglement generation code:

$$\frac{\alpha}{\alpha-1}\log_2[F(\phi)] + nR \leq \min_{\sigma_{B^n}} D_\alpha(\mathcal{N}_{A\to B^n}(\phi_{AA_1\cdots A_n})||I_A \otimes \sigma_{B^n}) \tag{34}$$

$$\leq D_\alpha\big(\mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n})||I_A \otimes [\mathcal{N}_p(\pi)]^{\otimes n}\big), \tag{35}$$

where $\mathcal{N}_{A^n\to B^n} = \mathcal{N}_p^{\otimes n}$ and $F(\phi)$ denotes the fidelity of an entanglement generation code with initial state ϕ.[3] Observe now that the output of n uses of the quantum erasure channel is rather special, in the sense that it can be written as a convex combination of 2^n density operators which are supported on orthogonal subspaces. We can index these by a binary string i (where ones in this string represent the systems that get erased and zeros represent systems that do not get erased), and we denote the density operators for $\mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n})$ by $\omega^i_{AB^n}$ and those for $[\mathcal{N}(\pi)]^{\otimes n}$ by $\tau^i_{B^n}$. Furthermore, let $\{i\}$ be the set of indices for the systems that get erased, so that we denote the systems that get erased by $A^{\{i\}}$ and those that do not by $A^{\{i\}^c}$. We then find that

$$D_\alpha\big(\mathcal{N}_{A^n\to B^n}(\phi_{AA_1\cdots A_n})||I_A \otimes [\mathcal{N}_p(\pi)]^{\otimes n}\big)$$

$$= \frac{1}{\alpha-1}\log\sum_{i\in\{0,1\}^n}(1-p)^{n-|i|}p^{|i|}\mathrm{Tr}\big\{\big[\omega^i_{AB^n}\big]^\alpha\big(I_A\otimes\big(\tau^i_{B^n}\big)^{1-\alpha}\big)\big\} \tag{36}$$

$$= \frac{1}{\alpha-1}\log\sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\mathrm{Tr}\{[\phi_{AA^{\{i\}^c}}]^\alpha\} \tag{37}$$

$$= \frac{1}{\alpha-1}\log\sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}, \tag{38}$$

where the last equality follows because the spectrum of $\phi_{AA^{\{i\}^c}}$ is equal to the spectrum of $\phi_{A^{\{i\}}}$ for a pure state. Rewriting (34)-(38), we obtain the following bound on the fidelity $F(\phi)$:

$$F(\phi) \leq \Big[2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\Big]\left[\sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}\right]^{\frac{1}{\alpha}}. \tag{39}$$

▶ Remark. By inspecting the above, we see that obtaining a general bound on the fidelity of an entanglement generation code for the quantum erasure channel is related to the quantum marginal problem [17], since the various terms $\mathrm{Tr}\{[\phi_{A^{\{i\}}}]^\alpha\}$ in the sum are the α-purities of all of the 2^n marginals of the quantum state $\phi_{AA_1\cdots A_n}$.

4 Strong converse for almost all codes

In the previous section, we established the bound (39) on the fidelity $F(\phi)$ of any $(n, R, \varepsilon, \phi, D)$ entanglement generation code. In this section, we prove our main result, i.e., that the large

[3] We could denote this fidelity as $F(\phi, D)$ because the fidelity of any code depends on the initial state ϕ and the decoder D, but the bound we find here is independent of the decoder D, so we suppress it from the notation.

fraction of capacity-exceeding entanglement generation codes satisfy the strong converse property. Before proving this result, we need to establish a measure on the set of all entanglement generation codes, in order to talk about the fraction of codes that satisfy the strong converse property. The most natural measure in this context is the unitarily invariant measure (Haar measure) on pure states, so that each possible initial state for an entanglement generation code is "receiving equal weight."

Now, suppose that we select the pure state ϕ_{AA^n} at random according to the Haar measure with $|A| = 2^{nR}$ and $|A_i| = d$ for all $i \in \{1, \ldots, n\}$. What makes the subsequent reasoning pertinent is the well-known fact that for $R < Q(\mathcal{N}_p) = (1-2p)\log d$, this choice results in a good code asymptotically with overwhelming probability. (Cf. for instance [15].)

We begin by analyzing the expectation of the fidelity $F(\phi)$:

$$\mathbb{E}_{\phi}\{F(\phi)\} \leq \mathbb{E}_{\phi}\left\{2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\left[\sum_{i\in\{0,1\}^n}\left[(1-p)d^{\alpha-1}\right]^{n-|i|}p^{|i|}\operatorname{Tr}\{[\phi_{A\{i\}}]^{\alpha}\}\right]^{\frac{1}{\alpha}}\right\} \tag{40}$$

$$\leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\left[\sum_{i\in\{0,1\}^n}\left[(1-p)d^{\alpha-1}\right]^{n-|i|}p^{|i|}\ \mathbb{E}_{\phi}\left\{\operatorname{Tr}\left\{\left[\phi_{A\{i\}}\right]^{\alpha}\right\}\right\}\right]^{\frac{1}{\alpha}}, \tag{41}$$

with the first inequality following from the development in the previous section and the second inequality following from concavity of $x^{\frac{1}{\alpha}}$ for $\alpha \in (1,2]$. So it remains to analyze the term $\mathbb{E}\{\operatorname{Tr}\{[\phi_{A\{i\}}]^{\alpha}\}\}$. Let $M_i^{\dagger}M_i = \phi_{A\{i\}}$ and consider that

$$\operatorname{Tr}\{[\phi_{A\{i\}}]^{\alpha}\} = \operatorname{Tr}\left\{(M_i^{\dagger}M_i)^{\alpha}\right\} = \operatorname{Tr}\left\{(M_i^{\dagger}M_i)^{\alpha-1}\left(M_i^{\dagger}M_i\right)\right\} \tag{42}$$

$$\leq \left(\|M_i\|_{\infty}^2\right)^{\alpha-1}\operatorname{Tr}\left\{\left(M_i^{\dagger}M_i\right)\right\} = \left(\|M_i\|_{\infty}^2\right)^{\alpha-1} \tag{43}$$

By employing the above inequalities and concavity of $x^{\alpha-1}$ for $\alpha \in (1,2]$, we find that

$$\mathbb{E}\{\operatorname{Tr}\{[\phi_{A\{i\}}]^{\alpha}\}\} \leq \left[\mathbb{E}\left\{\|M_i\|_{\infty}^2\right\}\right]^{\alpha-1}. \tag{44}$$

For a randomly chosen pure state ψ_{RS} on systems R and S and such that $\psi_R = M^{\dagger}M$, we have the estimate

$$\mathbb{E}\left\{\|M\|_{\infty}^2\right\} \leq C d_R^{-1}, \tag{45}$$

where $d_R = \dim(\mathcal{H}_R)$ and C is a universal constant independent of d_R [2]. This then implies the following bound for our setting:

$$\mathbb{E}\{\operatorname{Tr}\{[\phi_{A\{i\}}]^{\alpha}\}\} \leq \left(Cd^{-|i|}\right)^{\alpha-1} = C^{\alpha-1}d^{|i|(1-\alpha)}, \tag{46}$$

where we recall that d is the dimension of an individual input to the channel (so that the support of $\psi_{A\{i\}}$ has dimension $d^{|i|}$). Plugging back in to (41), we find the upper bound

$$\begin{aligned}
\mathbb{E}_{\phi}\{F(\phi)\} &\leq \Big[2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\Big]\left[\sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\mathbb{E}\{\mathrm{Tr}\{[\phi_{A\{i\}}]^{\alpha}\}\}\right]^{\frac{1}{\alpha}} && (47)\\
&\leq \Big[2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\Big]\left[\sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}C^{\alpha-1}d^{|i|(1-\alpha)}\right]^{\frac{1}{\alpha}} && (48)\\
&= 2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\ C^{\alpha-1}\left[\sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}\Big[pd^{(1-\alpha)}\Big]^{|i|}\right]^{\frac{1}{\alpha}} && (49)\\
&= 2^{-n\left(\frac{\alpha-1}{\alpha}\right)R}\ C^{\alpha-1}\ \big[(1-p)d^{\alpha-1}+d^{1-\alpha}p\big]^{\frac{n}{\alpha}} && (50)\\
&= 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]-\frac{\alpha}{n}\log C\right)}. && (51)
\end{aligned}$$

We now argue that if the rate R of quantum communication is strictly larger than the quantum capacity $(1-2p)\log d$ of the erasure channel, then we can pick α as a constant near one and n large enough such that

$$\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log\big[(1-p)d^{\alpha-1}+d^{1-\alpha}p\big]-\frac{\alpha}{n}\log C\right)>0. \tag{52}$$

So consider the term:

$$\frac{1}{\alpha-1}\log\big[(1-p)d^{\alpha-1}+d^{1-\alpha}p\big]. \tag{53}$$

Let us set $\alpha = 1+t$, so that the above is

$$\frac{1}{t}\log\big((1-p)d^{t}+d^{-t}p\big). \tag{54}$$

The limit of this quantity as $t\to 0$ ($\alpha\to 1$) is given by

$$\left.\frac{(1-p)d^{t}\log d-pd^{-t}\log d}{(1-p)d^{t}+d^{-t}p}\right|_{t=0}=(1-2p)\log d. \tag{55}$$

The other term $-\frac{\alpha}{n}\log C$ in the exponent becomes arbitrarily small as n becomes larger. Thus, it is always possible to pick a constant α and n large enough so that (52) is satisfied, and we recover a strong converse property for the expectation of the fidelity under randomly chosen entanglement generation codes.

Since the fidelity $F(\phi)$ is a non-negative random variable between zero and one, we can appeal to Markov's inequality to recover the following bound:

$$\begin{aligned}
&\Pr_{\phi}\left\{F(\phi)>2^{-\frac{1}{2}n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]-\frac{\alpha}{n}\log C\right)}\right\}\\
&\leq\frac{\mathbb{E}_{\phi}\{F(\phi)\}}{2^{-\frac{1}{2}n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log[(1-p)d^{\alpha-1}+d^{1-\alpha}p]-\frac{\alpha}{n}\log C\right)}}\\
&\qquad\qquad\leq 2^{-\frac{1}{2}n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1}+d^{1-\alpha}p\right]-\frac{\alpha}{n}\log C\right)}, \qquad (56)
\end{aligned}$$

where we used the bound in (51) for the second inequality. Thus, our conclusion is that if $R>(1-2p)\log d$, then we can choose α a constant and n large enough so that (52) holds, with the fraction of codes satisfying the strong converse property rapidly approaching one as the number of channel uses increases.

We can obtain an even sharper statement about the convergence by appealing to Levy's Lemma (see [12], for example):

▶ **Lemma 1** (Levy's Lemma). *Let $f : \mathbb{C}^d \to \mathbb{R}$ and $\eta > 0$ be such that for all pure states $|\varphi_1\rangle$ and $|\varphi_2\rangle$ in $\mathbb{C}^d$*

$$|f(|\varphi_1\rangle) - f(|\varphi_1\rangle)| \le \eta \| |\varphi_1\rangle - |\varphi_2\rangle \|_2.$$

Let $|\varphi\rangle$ be a random pure state in $\mathbb{C}^d$. Then for all $\delta \in [0, \eta]$, the following bound holds

$$\Pr\{|f(|\varphi\rangle) - \mathbb{E}\{f(|\varphi\rangle)\}| \ge \delta\} \le 4 \exp\left\{-\frac{d\delta^2}{c\eta}\right\},$$

where c is a positive constant.

We obtain a Lipschitz constant for the fidelity as a function of pure input states as follows:

$$\begin{aligned} |F(\varphi_1) - F(\varphi_2)| &\le |F(\varphi_1) - F(\varphi_2)| + |[1 - F(\varphi_1)] - [1 - F(\varphi_2)]| && (57)\\ &\le \|\varphi_1 - \varphi_2\|_1 && (58)\\ &\le 2 \| |\varphi_1\rangle - |\varphi_2\rangle \|_2. && (59) \end{aligned}$$

The first inequality is obvious, the second follows from monotonicity of trace distance under quantum operations (with these operations being a test for the maximally entangled state, the decoder, the channel and the encoder), and the third inequality is straightforward (see Lemma I.4 in [11], for example).

Since we have the bound

$$0 \le \mathbb{E}_\phi\{F(\phi)\} \le 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1} + d^{1-\alpha}p\right] - \frac{\alpha}{n}\log C\right)} \equiv g, \tag{60}$$

it follows from Levy's lemma that

$$\begin{aligned} \Pr\{F(\phi) \ge g + \delta\} &\le \Pr\{F(\phi) \ge \mathbb{E}_\phi\{F(\phi)\} + \delta\} && (61)\\ &\le 4\exp\left\{-\frac{2^{n[R+\log d]}\delta^2}{2c}\right\} && (62) \end{aligned}$$

We can take $\delta = g$, to find that

$$\begin{aligned} &\Pr\left\{F(\phi) \ge 2 \cdot 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1} + d^{1-\alpha}p\right] - \frac{\alpha}{n}\log C\right)}\right\}\\ &\qquad \le 4\exp\left\{-\frac{2^{n[R+\log d]}\left[2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R - \frac{1}{\alpha-1}\log\left[(1-p)d^{\alpha-1} + d^{1-\alpha}p\right] - \frac{\alpha}{n}\log C\right)}\right]^2}{2c}\right\}. \qquad (63) \end{aligned}$$

Now, without loss of generality, we can take $R \le \log d$ (otherwise the strong converse already holds for all codes), so that $R + \log d \ge 2\left(\frac{\alpha-1}{\alpha}\right)R$. Thus, we see that the fraction of codes with $R > (1 - 2p)\log d$ and obeying the strong converse approaches one doubly exponentially fast in the number of channel uses.

5 Conclusion

The main result of the present paper is a proof that the large fraction of codes with a quantum communication rate exceeding the quantum capacity of the erasure channel satisfy the strong converse. We view this result as adding to the evidence from [20] that a strong converse should hold for the quantum capacity of these channels. The main open question going forward from here is to prove that a fully strong converse holds for the quantum capacity of the erasure channel (i.e., that if the rate of any quantum communication scheme exceeds the quantum capacity of the erasure channel, then the quantum error necessarily converges to one).

The focus on the erasure channel of the present discussion may be justified by the simplicity of the channel (including its additivity). It also allowed us to give an illustration of the power of the Rényi divergence approach. At the same time, it seems to be true for all currently known random code ensembles achieving the coherent information for a channel $\mathcal{N}$ with Stinespring isometry $V : A' \hookrightarrow B \otimes E$ (with respect to a given input density ρ_A), that at rates above the same coherent information they have fidelity going to zero, with overwhelming probability. Of course this has to be verified for each ensemble separately, but rests on two properties that hold for most codes in the ensemble. Namely, with respect to the pure state $|\psi\rangle_{AB^nE^n} = (I \otimes V^{\otimes n})|\phi\rangle_{AA'^n}$:

1. **Typicality of B**. The channel output ψ_{B^n} is largely in the typical subspace of $\mathcal{N}(\rho_A)^{\otimes n}$ in the sense that $H^\delta_{\max}(B^n) \leq nS(\mathcal{N}(\rho_A)) + o(n)$.
2. **Saturation of E**. The complementary channel output ψ_{E^n} covers essentially uniformly the typical subspace of $\mathcal{N}^c(\rho_A)^{\otimes n}$ in the sense that $H^\delta_{\min}(E^n) \geq nS(\mathcal{N}^c(\rho_A)) - o(n)$.

[In fact, in practice the latter property tends to be true for most states in most code subspaces.] We refer to [29] (cf. [20]) for the definitions and necessary properties of (smooth) min- and max-entropies used in the following.

Now, if our code is supposed to generate entanglement at rate R with fidelity F, then by the decoupling principle,

$$H^{\sqrt{1-F^2}}_{\min}(A|E^n) \geq nR. \tag{64}$$

On the other hand, using relations between min- and max-entropies as well as chain rules,

$$\begin{aligned} H^{\sqrt{1-F^2}}_{\min}(A|E^n) &\lesssim H^\epsilon_{\max}(A|E^n) \\ &\lesssim H^\delta_{\max}(AE^n) - H^\delta_{\min}(E^n) \\ &= H^\delta_{\max}(B^n) - H^\delta_{\min}(E^n), \end{aligned} \tag{65}$$

where $\epsilon = \frac{1}{2}(1 - \sqrt{1-F^2})$ and $\delta = \frac{1}{4}\epsilon$, the inequalities are true up to terms of order $\log \frac{1}{\delta}$. By the typicality and saturation properties, (64) and (65) bound the rate as desired,

$$R \leq S(\mathcal{N}(\rho_A)) - S(\mathcal{N}^c(\rho_A)) + o(1) = I(A\rangle B) + o(1). \tag{66}$$

Acknowledgements. We are grateful to Naresh Sharma for many conversations from which the ideas in this paper arose. We thank the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge for organizing the semester "Mathematical Challenges in Quantum Information," at which we had an opportunity to discuss this research. MMW is grateful to the Department of Physics and Astronomy at Louisiana State University for startup funds that supported this research and acknowledges support from the DARPA Quiness Program through US Army Research Office award W31P4Q-12-1-0019. AW acknowledges financial support by the Spanish MINECO, project FIS2008-01236 with the support of FEDER funds, the EC STREP "RAQUEL", the ERC Advanced Grant "IRQUAT", and the Philip Leverhulme Trust.

References

1 Suguru Arimoto. On the converse to the coding theorem for discrete memoryless channels. *IEEE Transactions on Information Theory*, 19:357–359, May 1973.

2 Guillaume Aubrun, Stanislaw Szarek, and Elisabeth Werner. Non-additivity of Rényi entropy and Dvoretzky's theorem. October 2009. arXiv:0910.1189.

3 Bhaskar Roy Bardhan, Raul Garcia-Patron, Mark M. Wilde, and Andreas Winter. Strong converse for the classical capacity of all phase-insensitive bosonic Gaussian channels. January 2014. arXiv:1401.4161.

4 Howard Barnum, Emanuel Knill, and Michael A. Nielsen. On quantum fidelities and channel capacities. *IEEE Transactions on Information Theory*, 46(4):1317–1329, July 2000. arXiv:quant-ph/9809010.

5 Howard Barnum, M. A. Nielsen, and Benjamin Schumacher. Information transmission through a noisy quantum channel. *Physical Review A*, 57(6):4153–4175, June 1998. arXiv:quant-ph/9702049.

6 Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. Quantum reverse Shannon theorem. December 2012. arXiv:0912.5537.

7 Charles H. Bennett, David P. DiVincenzo, and John A. Smolin. Capacities of quantum erasure channels. *Physical Review Letters*, 78(16):3217–3220, April 1997. arXiv:quant-ph/9701015.

8 Mario Berta, Matthias Christandl, and Renato Renner. The quantum reverse Shannon theorem based on one-shot information theory. *Communications in Mathematical Physics*, 306(3):579–615, August 2011. arXiv:0912.3805.

9 Igor Devetak. The private classical capacity and quantum capacity of a quantum channel. *IEEE Transactions on Information Theory*, 51(1):44–55, January 2005. arXiv:quant-ph/0304127.

10 Igor Devetak and Peter W. Shor. The capacity of a quantum channel for simultaneous transmission of classical and quantum information. *Communications in Mathematical Physics*, 256(2):287–303, June 2005. arXiv:quant-ph/0311131.

11 Frederic Dupuis. *The decoupling approach to quantum information theory*. PhD thesis, Université de Montréal, November 2009. arXiv:1004.1641.

12 Omar Fawzi. *Uncertainty relations for multiple measurements with applications*. PhD thesis, McGill University, August 2012. arXiv:1208.5918.

13 Markus Grassl, Thomas Beth, and Thomas Pellizzari. Codes for the quantum erasure channel. *Physical Review A*, 56(1):33–38, July 1997. arXiv:quant-ph/9610042.

14 Manish K. Gupta and Mark M. Wilde. Multiplicativity of completely bounded p-norms implies a strong converse for entanglement-assisted capacity. October 2013. arXiv:1310.7028.

15 Patrick Hayden, Peter W. Shor, and Andreas Winter. Random quantum codes from Gaussian ensembles and an uncertainty relation. *Open Systems & Information Dynamics*, 15(1):71–89, March 2008. arXiv:0712.0975.

16 Rochus Klesse. A random coding based proof for the quantum coding theorem. *Open Systems & Information Dynamics*, 15(1):21–45, March 2008. arXiv:0712.2558.

17 Alexander Klyachko. Quantum marginal problem and representations of the symmetric group. September 2004. arXiv:quant-ph/0409113.

18 Robert Koenig and Stephanie Wehner. A strong converse for classical channel coding using entangled inputs. *Physical Review Letters*, 103:070504, August 2009. arXiv:0903.2838.

19 Seth Lloyd. Capacity of the noisy quantum channel. *Physical Review A*, 55(3):1613–1622, March 1997. arXiv:quant-ph/9604015.

20 Ciara Morgan and Andreas Winter. "Pretty strong" converse for the quantum capacity of degradable channels. *IEEE Transactions on Information Theory*, 60(1):317–333, January 2014. arXiv:1301.4927.

21 Martin Müller-Lennert, Frédéric Dupuis, Oleg Szehr, Serge Fehr, and Marco Tomamichel. On quantum Rényi entropies: a new definition and some properties. June 2013. arXiv:1306.3142.

22 Tomohiro Ogawa and Hiroshi Nagaoka. Strong converse to the quantum channel coding theorem. *IEEE Transactions on Information Theory*, 45:2486–2489, November 1999. arXiv:quant-ph/9808063.

23 Benjamin Schumacher and Michael A. Nielsen. Quantum data processing and error correction. *Physical Review A*, 54(4):2629–2635, October 1996. arXiv:quant-ph/9604022.

24 Benjamin Schumacher and Michael D. Westmoreland. Optimal signal ensembles. *Physical Review A*, 63:022308, January 2001.

25 Naresh Sharma and Naqueeb Ahmad Warsi. On the strong converses for the quantum channel capacity theorems. June 2012. arXiv:1205.1712.

26 Peter W. Shor. Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52(4):R2493–R2496, October 1995.

27 Peter W. Shor. The quantum channel capacity and coherent information. In *Lecture Notes, MSRI Workshop on Quantum Computation*, 2002.

28 Graeme Smith and Jon Yard. Quantum communication with zero-capacity channels. *Science*, 321:1812–1815, September 2008. arXiv:0807.4935.

29 Marco Tomamichel. *A Framework for Non-Asymptotic Quantum Information Theory*. PhD thesis, ETH Zurich, 2012. arXiv:1203.2142.

30 Mark M. Wilde and Andreas Winter. Strong converse for the classical capacity of the pure-loss bosonic channel. *To appear in Problems of Information Transmission*, August 2013. arXiv:1308.6732.

31 Mark M. Wilde, Andreas Winter, and Dong Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels. June 2013. arXiv:1306.1586.

32 Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999.

33 Jacob Wolfowitz. *Coding Theorems of Information Theory*, volume 31. Springer, 1964.

A Strong converse for the classical capacity of the quantum erasure channel

In this appendix, we detail a proof that the strong converse holds for the classical capacity of the quantum erasure channel. To our knowledge, a proof of this statement has not yet appeared in the literature. This result was obtained in collaboration with Naresh Sharma.

Using the generalized divergence framework established in [25] and reviewed in [31] (or even the method of Koenig-Wehner [18]), we obtain the following bound on the success probability when transmitting a classical message through the quantum erasure channel

$$p_{\text{succ}} \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\frac{1}{n}\chi_\alpha\left(\mathcal{N}^{\otimes n}\right)\right)}, \tag{67}$$

where

$$\frac{1}{n}\chi_\alpha\left(\mathcal{N}^{\otimes n}\right) \tag{68}$$

is the regularized Rényi-Holevo information of the erasure channel. So our goal is to prove that this quantity is additive as a function of the quantum erasure channel. First recall that this quantity can be written as an information radius [24, 31]:

$$\chi_\alpha\left(\mathcal{N}^{\otimes n}\right) = \min_{\sigma_{B^n}} \max_{\rho_{A^n}} D_\alpha\left(\mathcal{N}^{\otimes n}(\rho_{A^n})||\sigma_{B^n}\right). \tag{69}$$

With this, we see that we can upper bound this quantity simply by choosing σ_{B^n} to be the output of the erasure channel when the tensor-power maximally mixed state is input:

$$\chi_\alpha\big(\mathcal{N}^{\otimes n}\big) \leq \max_{\rho_{A^n}} D_\alpha\big(\mathcal{N}^{\otimes n}(\rho_{A^n})\|[\mathcal{N}(\pi)]^{\otimes n}\big). \tag{70}$$

As discussed in Section 3.2, the output of the quantum erasure channel is rather special, in the sense that it can be written as a linear combination of 2^n density operators which are supported on orthogonal subspaces. We can index these by a binary string i (where ones in this string represent the systems that get erased and zeros represent systems that do not get erased), and we denote the density operators for $\mathcal{N}^{\otimes n}(\rho_{A^n})$ by $\omega^i_{B^n}$ and those for $[\mathcal{N}(\pi)]^{\otimes n}$ by $\tau^i_{B^n}$. Furthermore, let $\{i\}$ be the set of indices for the systems that get erased, so that we denote the systems that get erased by $A^{\{i\}}$ and those that do not by $A^{\{i\}^c}$. We then find that

$$\begin{aligned}
&\max_{\rho_{A^n}} D_\alpha\big(\mathcal{N}^{\otimes n}(\rho_{A^n})\|[\mathcal{N}(\pi)]^{\otimes n}\big)\\
&= \frac{1}{\alpha-1}\log\max_{\rho_{A^n}} \mathrm{Tr}\big\{\big[\mathcal{N}^{\otimes n}(\rho_{A^n})\big]^\alpha\big([\mathcal{N}(\pi)]^{\otimes n}\big)^{1-\alpha}\big\} &(71)\\
&= \frac{1}{\alpha-1}\log\max_{\rho_{A^n}} \sum_{i\in\{0,1\}^n}(1-p)^{n-|i|}p^{|i|}\mathrm{Tr}\big\{\big[\omega^i_{B^n}\big]^\alpha\big[\tau^i_{B^n}\big]^{1-\alpha}\big\} &(72)\\
&= \frac{1}{\alpha-1}\log\max_{\rho_{A^n}} \sum_{i\in\{0,1\}^n}(1-p)^{n-|i|}p^{|i|}\mathrm{Tr}\big\{[\rho_{A^{\{i\}^c}}]^\alpha[\pi_{A^{\{i\}^c}}]^{1-\alpha}\big\} &(73)
\end{aligned}$$

The above equalities follow simply by substitution and some algebra. Continuing, the last line above is equal to

$$\begin{aligned}
&= \frac{1}{\alpha-1}\log\max_{\rho_{A^n}} \sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|}\mathrm{Tr}\{[\rho_{A^{\{i\}^c}}]^\alpha\} &(74)\\
&\leq \frac{1}{\alpha-1}\log \sum_{i\in\{0,1\}^n}\big[(1-p)d^{\alpha-1}\big]^{n-|i|}p^{|i|} &(75)\\
&= \frac{1}{\alpha-1}\log \sum_{k=0}^{n}\big[(1-p)d^{\alpha-1}\big]^{n-k}p^{k}\binom{n}{k} &(76)\\
&= \frac{1}{\alpha-1}\log\Big((1-p)d^{(\alpha-1)}+p\Big)^n &(77)\\
&= n\left[\frac{1}{\alpha-1}\log\Big((1-p)d^{(\alpha-1)}+p\Big)\right] &(78)
\end{aligned}$$

The inequality follows because $\mathrm{Tr}\{[\rho_{A^{\{i\}^c}}]^\alpha\} \leq 1$ for all $\alpha \geq 1$ (and we are considering $\alpha \in (1,2]$ here). The next few equalities are straightforward. Returning to (67), all of this development implies that we get the following upper bound on success probability

$$p_{\text{succ}} \leq 2^{-n\left(\frac{\alpha-1}{\alpha}\right)\left(R-\left[\frac{1}{\alpha-1}\log\left((1-p)d^{(\alpha-1)}+p\right)\right]\right)} \tag{79}$$

The last line above is a single-letter upper bound. Now, let us set $\alpha = 1+t$, so that the above is

$$\frac{1}{t}\log\big((1-p)d^t+p\big). \tag{80}$$

The limit of this quantity as $t \to 0$ is given by

$$\left.\frac{(1-p)d^t\log d}{(1-p)d^t+p}\right|_{\varepsilon=0} = (1-p)\log d, \tag{81}$$

which is exactly the classical capacity of the quantum erasure channel. Thus, whenever the classical communication rate $R > (1-p)\log d$, we can always find a value of α in a neighborhood of one such that

$$\left(\frac{\alpha-1}{\alpha}\right)\left(R-\left[\frac{1}{\alpha-1}\log\Big((1-p)d^{(\alpha-1)}+p\Big)\right]\right)>0. \tag{82}$$

This concludes the proof.

Interestingly, the proof above demonstrates that tensor-product pure-state codewords are the optimal choice in order to saturate the bound given above. That is, for pure-state codewords, we have the equality $\mathrm{Tr}\{[\rho_{A^{\{i\}^c}}]^\alpha\} = 1$, so that the upper bound is saturated by this choice.

Graph-theoretical Bounds on the Entangled Value of Non-local Games

André Chailloux[1], Laura Mančinska[2], Giannicola Scarpa[3], and Simone Severini[4]

1 **SECRET Project Team, INRIA Paris-Rocquencourt,Paris, France**
`andre.chailloux@inria.fr`

2 **Center for Quantum Technologies, Singapore**
`cqtlm@nus.edu.sg`

3 **Universitat Autònoma de Barcelona, Barcelona, Spain**
`giannicola.scarpa@uab.cat`

4 **University College London, London, United Kingdom**
`simoseve@gmail.com`

Abstract

We introduce a novel technique to give bounds to the entangled value of non-local games. The technique is based on a class of graphs used by Cabello, Severini and Winter in 2010. The upper bound uses the famous Lovász theta number and is efficiently computable; the lower one is based on the quantum independence number, which is a quantity used in the study of entanglement-assisted channel capacities and graph homomorphism games.

1998 ACM Subject Classification G.2.3 Applications

Keywords and phrases Graph theory, non-locality, entangled games

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.67

1 Introduction

In non-local games, two non-communicating players cooperate in order to achieve a task. Each player receives an input and produces an output, and they must satisfy the task's requirements.

In physics, this class of games is also known as "entangled games". They are mostly used to investigate the power of entanglement, by designing intuitive Bell inequalities. One designs a non-local game and proves an upper bound on the winning probability of the classical players (the Bell inequality). Later, one shows that there exists a quantum strategy that by using entanglement can beat that winning probability. Two famous examples of such approach are the CHSH game (based on [3]) and the magic square game (based on [14]).

Non-local games are also important in computer science, where they are usually called "two-prover one-round games". Their intuitive nature has been used in complexity theory to approach the difficult problem of P vs. NP, by defining probabilistically checkable proofs and ultimately leading to the famous unique games conjecture [9, 10].

Estimating or bounding the value of a game given its description is an important task, and much effort has been devoted to the question. For example, the entangled value for the class of XOR games has been shown to be easy to compute with a semidefinite program by Cirel'son [4]. Also the entangled value of unique games turns out to be easy to compute, therefore falsifying the unique games conjecture in the quantum world [11].

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 67–75
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Here, we propose a general approach to bound the value of a non-local game based on graph theory. Given the description of a game, we construct a graph that contains all the information about the game, and we call it the "game graph". The construction is based on the techniques in [7]. Such techniques have also been extended and used in [1].

We first show that the classical value of any game is equal to the independence number of its game graph (renormalized). This reflects the fact that computing exactly the classical value of a game is NP-hard. We then show an efficiently computable upper bound on the quantum value of a game (and therefore on the classical value), given by the celebrated Lovász theta number. We then give lower bound for the games on the uniform distribution given by the quantum independence number, a graph parameter introduced in [5] and futher discussed in [13, 15]. To conclude, we give a class of games for which this upper bound is tight.

We believe this graph-theoretical approach is an important and a fertile field for improvements. We discuss these in the conclusions section.

2 Preliminaries

2.1 Non-local games

We now briefly describe the setting of a non-local game $\mathcal{G}$.

Alice and Bob are separated and forbidden to communicate. They receive inputs x and y from some input sets X and Y, according to some fixed and known probability distribution π, and are required to produce outputs a and b from output sets A and B, respectively. The game rules are encoded in a predicate $\lambda : X \times Y \times A \times B \to \{0,1\}$, which specifies which outputs a, b are correct on inputs x, y. In other words, players win the game on inputs x, y if they output some a, b such that $\lambda(x, y, a, b) = 1$. The goal of the players is to maximize the winning probability.

A *classical strategy* for the game is without loss of generality a pair of functions, $f_A : X \to A$ for Alice and $f_B : Y \to B$ for Bob. (Shared randomness between the two players is easily seen not to be beneficial.) The winning probability of a strategy is calculated as follows:

$$\sum_{x,y} \pi(x,y)\lambda(x,y,f_A(x),f_B(y)).$$

The *classical value* $\omega(\mathcal{G})$ of the game is the maximum winning probability among all classical strategies.

In *entangled strategies* (a.k.a. quantum strategies), players share a fixed (*i.e.*, independent of the inputs) entangled state $|\psi\rangle$. For each input x, Alice has a projective measurement $\{P_a^x\}_{a\in A}$, and for each input y, Bob has a projective measurement $\{Q_b^y\}_{b\in B}$. Upon receiving the inputs, they apply the corresponding measurements to their parts of the entangled state and produce classical outputs a and b, respectively. The winning probability of a strategy is calculated as follows:

$$\sum_{xy} \pi(x,y)\lambda(x,y,a,b)\langle\psi|P_a^x \otimes Q_b^y|\psi\rangle.$$

The *entangled value* $\omega^*(\mathcal{G})$ of the game is the supremum of the winning probability, taken over all entangled strategies.

A Bell inequality for a game is an upper bound on its classical value. We have a Bell inequality violation for a game $\mathcal{G}$ if the entangled value is strictly larger than the classical one. The violation is quantified by the ratio $\omega^*(\mathcal{G})/\omega(\mathcal{G})$.

The CHSH game is one particularly famous example [3]. Here, the inputs $x \in \{0, 1\}$ and $y \in \{0, 1\}$ are uniformly distributed, and Alice and Bob win the game if their respective outputs $a \in \{0, 1\}$ and $b \in \{0, 1\}$ satisfy $a \oplus b = x \wedge y$; in other words, a should equal b unless $x = y = 1$. The classical value of this game is easily seen to be $\omega(\mathcal{G}) = 3/4$, while the entangled value is known to be $\omega^*(\mathcal{G}) = 1/2 + 1/(2\sqrt{2}) \approx 0.85$.

A non-local game is said to be a *pseudo-telepathy* game if the quantum value is 1 while the classical value is strictly less than 1.

2.2 Notions of graph theory

A *simple graph* $G = (V, E)$ consists of a finite vertex set V and its edge set $E \subsetneq V \times V$ (the inclusion here is strict because there are no edges of the form (v, v)). Two vertices $(v, w) \in E$ are "adjacent" or equivalently "form an edge". All graphs considered here are simple graphs. For a graph $G = (V, E)$, we also denote its vertex set with $V(G)$ and its edge set with $E(G)$ whenever confusion has to be avoided.

An *independent set* of a graph is a subset I of $V(G)$ such that no two elements of I are adjacent. The *independence number* of a graph G, denoted by $\alpha(G)$, is the maximum size of an independent set of G.

A d-dimensional orthogonal representation of $G = (V, E)$ is a map $\phi : V \to \mathbb{C}^d$ such that for all $(v, w) \in E$, $\langle \phi(v)|\phi(w)\rangle = 0$. (If all the vectors have unit norm, this is called orthonormal representation.)

We finally introduce an important graph parameter: the *theta number* (a.k.a. Lovász number or theta function). It was originally defined by Lovász [12] to solve a long-standing problem posed by Shannon [16]: computing the Shannon capacity of the five-cycle. There are many equivalent formulations of the theta number (see [8] for a detailed survey). The one that we use in this paper is the following:

$$\vartheta(G) = \max \sum_{v \in V(G)} |\langle \psi|\psi_v\rangle|^2, \tag{1}$$

where the maximum is over unit vectors ψ and orthonormal representations $\{\psi_v\}_{v \in V(G)}$. Lovász [12] proved that $\alpha(G) \leq \vartheta(G)$ holds (this inequality is part of the so-called "sandwich theorem" [8]). The theta number can be approximated to within arbitrary precision in polynomial time, hence it gives a tractable and in many cases useful bound for α.

2.3 Quantum Independence Number

In this section we define the quantum independence number and state some of its properties.

First, let us briefly give some historical background. In [13] the concept of quantum independence number is presented in the context of zero-error information theory. This quantity is usually called in literature "one-shot zero-error entanglement-assisted channel capacity" and denoted as α^*. A new definition of quantum independence number, denoted as α_q, came in [15], in the context of graph homomorphisms. As of today, it is not known if the two quantities are equal for all graphs. In this paper we use the second quantity, but for simplicity we omit the details about homomorphisms and provide a direct definition.

As with the quantum chromatic number (see [6]), the quantum independence number can be defined in terms of a non-local game. Informally, the *independent set game* with parameter t for a graph $G = (V, E)$ is as follows. Two players, Alice and Bob, claim that they know an independent set I of G consisting of t vertices. A referee wants to test this claim with a non-local game. He forbids communication between the players, generates two uniformly random numbers $x, y \in [t]$ and separately asks Alice to provide the x-th vertex of I and Bob to provide the y-th vertex of I. The players are required to output the same vertex if $x = y$, and to output non-adjacent vertices if $x \neq y$. A formal definition follows.

▶ **Definition 1.** The *independent set game with parameter* t on the graph $G = (V, E)$ is a non-local game with input sets $X = Y = [t]$, output sets $A = B = V$. The probability distribution π is the uniform distribution on the input pairs. Alice gets input x and outputs v, Bob gets input y and outputs w. The players *lose the game* in the following two cases:

1. $x = y$ and $v \neq w$
2. $x \neq y$ and $(v, w) \in E$ or $v = w$

A classical strategy consists w.l.o.g. of two deterministic functions $f_A : [t] \to V$ for Alice and $f_B : [t] \to V$ for Bob. Shared randomness, as seen for the coloring game, is not beneficial. A little thought will show that to win with probability 1, we must have $f_A = f_B$ (to avoid the first losing condition) and that $\{f_A(1), \ldots, f_A(t)\}$ must be a valid independent set of the graph of size t (to avoid the second losing condition). It follows that the classical players cannot win the game with probability 1 when $t > \alpha(G)$.

It is proven in [15] that w.l.o.g. quantum strategies for the independent set game consist of projective measurements on a maximally entangled state, that the projective measurements of Alice and Bob are the same and that all the projectors can be real-valued. Therefore we can define a *quantum independent set* of size t as a collection of t projective measurements $\{P_v^x\}_{v \in V}$ for all $x \in [t]$ that have the whole vertex set as outputs, with the following consistency condition:

$$\text{for all } (u, v) \in E \text{ or } u = v \text{ and for all } x \neq x', \quad P_u^x P_v^{x'} = 0. \tag{2}$$

▶ **Definition 2.** For all graphs G, the *quantum independence number* $\alpha_q(G)$ is the maximum number t such that there exists a quantum independent set of G of size t.

3 Game graphs

3.1 Definition and relation to $\omega(\mathcal{G})$

Consider a non-local game $\mathcal{G}$ with input sets X, Y, output sets A, B, predicate $\lambda : X \times Y \times A \times B \to \{0, 1\}$ and uniform distribution on the inputs.

▶ **Definition 3.** A graph $G = (V, E)$ *associated* to the game $\mathcal{G}$ has:

1. $V = \{xyab \mid x \in X, y \in Y, a \in A, b \in B \text{ and } \lambda(x, y, a, b) = 1\}$,
2. $E = \{\{xyab, x'y'a'b'\} \mid (x = x' \wedge a \neq a') \vee (y = y' \wedge b \neq b')\}$.

This definition is inspired by a construction in [7] in the framework of contextuality of physical theories. The authors used something similar to Definition 3 for the special case of the CHSH game. Here we generalize to all games.

For simplicity, we prove the results in this section for the case where the game has the uniform distribution on the inputs and λ is a boolean function. It is straightforward to generalize to games with real-valued predicate and any probability distribution π of the inputs, as follows. Consider the (vertex) weighted graph with all the quadruples $xyab$ in the vertex set, labelled with weight$(xyab) = \lambda(x, y, a, b) \cdot \pi(x, y)$, and the same edge set as before. The classical bound and the Lovász theta bound that we will prove later can be adapted by considering the weighted versions of these parameters. However, we do not know how to generalize our last result because we do not define the quantum independence number for a weighted graph.

Now we prove that that the classical value of a game can be expressed in terms of the independence number of its game graph.

▶ **Theorem 4.** *Let $\mathcal{G}$ be a non-local game with input sets X and Y, uniform input distribution and associated graph G. Then*

$$\omega(\mathcal{G}) = \frac{\alpha(G)}{|X \times Y|}.$$

Proof. Let $k = |X \times Y|$. We begin by proving that $\omega(\mathcal{G}) \geq \alpha(G)/k$. Namely, we show that given a maximal independent set $I \subseteq V$ of size ℓ, we can exhibit a strategy for $\mathcal{G}$ that answers correctly to at least ℓ of the k questions. By the structure of G, the independent set I cannot contain vertices $xyab$ and $xy'a'b'$ such that $a \neq a'$. Similarly, I cannot contain vertices $xyab$ and $x'ya'b'$ such that $b \neq b'$. Hence, we have the following strategy: on input x, Alice outputs the unique a determined by the vertices in the independent set I. Bob behaves similarly. Since V contains only winning quadruples $xyab$, the size ℓ of the independent set means Alice and Bob answer correctly to at least ℓ input pairs. Hence, $\omega(\mathcal{G}) \geq \ell/k$.

Now we show that $\omega(\mathcal{G}) \leq \alpha(G)/k$, *i.e.*, if there exists a strategy that wins on ℓ of the k input pairs, then there exists an independent set with weight ℓ. We have that w.l.o.g. classical strategies consist of a pair of functions. Fix Alice and Bob's functions f_A and f_B that win on ℓ input pairs. Now take the set of quadruples $S = \{(x, y, f_A(x), f_B(y))\}_{x \in X, y \in Y}$. We have that $I = S \cap V$ is a set of ℓ vertices of G. Since f_A and f_B are deterministic, I cannot contain vertices $xyab$ and $xy'a'b'$ such that $a \neq a'$ nor vertices $xyab$ and $x'ya'b'$ such that $b \neq b'$. Therefore, there cannot be an edge between any pair of the elements of I and we have that I is an independent set of G of size ℓ. Hence, $\alpha(G) \geq \ell$. Combining the two directions proves the theorem. ◀

3.2 Bounds on the entangled value of a game

Cabello, Severini and Winter [7] observe that the quantum value of the CHSH game is equal to the theta number of its associated graph divided by the number of questions. We have found by direct calculation that this is not always true for general games, for example in the case of the 2-fold parallel repetition of CHSH. The same conclusion follows from the results of Acín *et al.* in [1]. Here we prove the upper bound directly for our specific constructions.

▶ **Theorem 5.** *Let $\mathcal{G}$ be a non-local game with input sets X and Y, uniform input distribution and associated graph $G = (V, E)$. Then*

$$\omega^*(\mathcal{G}) \leq \frac{\vartheta(G)}{|X \times Y|}.$$

Proof. Let $k = |X \times Y|$. Consider a quantum strategy for $\mathcal{G}$ that achieves the value $\omega^*(\mathcal{G})$. It consists of a shared entangled state $|\psi\rangle$ and a collection of projective measurements $\{P_a^x\}, \{Q_b^y\}$, such that

$$\sum_{xyab} \frac{1}{k} \lambda(x, y, a, b) \langle\psi| P_a^x \otimes Q_b^y |\psi\rangle = \frac{1}{k} \sum_{xyab \in V} \langle\psi| P_a^x \otimes Q_b^y |\psi\rangle = \omega^*(\mathcal{G}).$$

For each quadruple $xyab$ let $|\psi_{xyab}\rangle = P_a^x \otimes Q_b^y |\psi\rangle$. This is an orthogonal representation of G, since for every edge $(xyab, x'y'a'b')$ either $P_a^x P_{a'}^{x'} = 0$ or $Q_b^y Q_{b'}^{y'} = 0$. Now for each $xyab$ consider the normalized vector

$$|\psi'_{xyab}\rangle = \frac{|\psi_{xyab}\rangle}{||\psi_{xyab}||} = \frac{|\psi_{xyab}\rangle}{\sqrt{\langle\psi| P_a^x \otimes Q_b^y |\psi\rangle}}.$$

We have that $\{\psi'_{xyab}\}_{xyab\in V}$ and ψ are a feasible solution for the formulation (1) of $\vartheta(G)$. We conclude

$$\begin{aligned}\vartheta(G) &\geq \sum_{xyab\in V} |\langle\psi|\psi_{xyab}\rangle|^2\\ &= \sum_{xyab\in V} \left|\frac{\langle\psi|\psi_{xyab}\rangle}{||\psi_{xyab}||}\right|^2\\ &= \sum_{xyab\in V} \frac{\langle\psi|P_a^x\otimes Q_b^y|\psi\rangle^2}{\langle\psi|P_a^x\otimes Q_b^y|\psi\rangle}\\ &= \sum_{xyab\in V} \langle\psi|P_a^x\otimes Q_b^y|\psi\rangle\\ &= k\cdot\omega^*(\mathcal{G}).\end{aligned}$$ ◀

We now have the following lower bound in terms of the quantum independence number.

▶ **Theorem 6.** *Let $\mathcal{G}$ be a non-local game with input sets X and Y, uniform input distribution and associated graph $G=(V,E)$. Then*

$$\omega^*(\mathcal{G}) \geq \frac{\alpha_q(G)}{|X\times Y|}$$

To prove the theorem, we will use the following lemma.

▶ **Lemma 7.** *Let M, N be positive semidefinite matrices. Then for any vector $|v\rangle$, we have that*

$$\langle v|\operatorname{supp}(M+N)|v\rangle \geq \langle v|\operatorname{supp}(M)|v\rangle,$$

where supp(M) *denotes the projector onto the support (*i.e., *the column space) of M.*

Proof. If P is a projector onto a subspace Π then $\langle v|P|v\rangle$ is the squared length of the projection of $|v\rangle$ into Π. Hence, to prove the lemma it suffices to show that $\operatorname{supp}(M)\subseteq\operatorname{supp}(M+N)$, where by abusing the notation we use supp to denote the support itself (rather than the projection onto it).

For contradiction, suppose that $\operatorname{supp}(M)\not\subseteq\operatorname{supp}(M+N)$. Then the orthogonal complement of supp(M) (i.e. the nullspace Null(M)) does not contain Null$(M+N)$. Hence we can pick a vector $|w\rangle$ such that $(M+N)|w\rangle=0$ but $M|w\rangle\neq 0$. This further implies that

$$\langle w|N|w\rangle = \langle w|(M+N)|w\rangle - \langle w|M|w\rangle = -\langle w|M|w\rangle < 0,$$

since M is positive semidefinite and $M|w\rangle\neq 0$. This completes the proof as we have reached a contradiction with the initial assumption that N is positive semidefinite. ◀

Proof of Theorem 6. Given a quantum strategy $\{P^i_{xyab}\}$ for the independent set game on G with parameter t, we construct a strategy to win the game $\mathcal{G}$ with probability at least $t/|X\times Y|$, as follows.

Players share a maximally entangled state with local dimension d (which is the dimension of the projectors above). On input x, Alice measures her half of the state using the projective measurement$\{P_a^x\}_{a\in A}\bigcup\{I-\sum_a P_a^x\}$, where the individual elements are defined as follows:

$$P_a^x = \operatorname{supp}\left(\sum_{\substack{yb\\ xayb\in V}}\sum_i P^i_{xayb}\right),$$

where we use $\mathrm{supp}(M)$ to denote the projector onto the image of M. We show that this is a valid projective measurement. For all y, b, y', b' there is an edge $(xyab, xy'a'b') \in E$. Therefore in the strategy for the independent set game we have that for all i, j each projector P^i_{xyab} is orthogonal to $P^j_{xy'a'b'}$. Hence, for all $a \neq a'$ we have $P^x_a \cdot P^x_{a'} = 0$. Bob constructs projectors P^y_b similarly.

Now we lower bound the quantum value of $\mathcal{G}$ as follows:

$$\begin{aligned} |X \times Y| \cdot \omega^*(\mathcal{G}) &\geq \sum_{xyab \in V} \langle\psi| P^x_a \otimes P^y_b |\psi\rangle \\ &= \sum_{xyab \in V} \langle\psi| \,\mathrm{supp}\left(\sum_{i,j} \sum_{\substack{y'b' \\ xay'b' \in V}} \sum_{\substack{x'a' \\ x'a'yb \in V}} P^i_{xay'b'} \otimes P^j_{x'a'yb} \right) |\psi\rangle, \end{aligned}$$

where we have used the fact that $\mathrm{supp}(M \otimes N) = \mathrm{supp}(M) \otimes \mathrm{supp}(N)$ for all matrices M, N to obtain the last equality. Now by applying Lemma 7, we drop all the terms except the ones with $i = j, a = a', b = b', x = x'$ and $y = y'$, and we have that

$$|X \times Y| \cdot \omega^*(\mathcal{G}) \geq \sum_{xyab \in V} \langle\psi| \,\mathrm{supp}\left(\sum_i P^i_{xayb} \otimes P^i_{xayb} \right) |\psi\rangle \tag{3}$$

$$= \sum_{xyab \in V} \langle\psi| \left(\sum_i P^i_{xayb} \otimes P^i_{xayb} \right) |\psi\rangle \tag{4}$$

$$= \sum_{xyab \in V} \sum_i \frac{1}{d} \mathrm{Tr}(P^i_{xayb}) \tag{5}$$

$$= \sum_i \frac{1}{d} \mathrm{Tr}(I_d) \tag{6}$$

$$= \alpha_q(G). \tag{7}$$

In the above we have observed that $\mathrm{supp}(P + Q) = P + Q$ for mutually orthogonal projectors P and Q to get Expression (4). We have used properties of $|\psi\rangle = \frac{1}{\sqrt{d}} \sum_i |i, i\rangle$ to obtain Expression (5). We have used the fact that, for all i, $\{P^i_{xayb} : \lambda(x, a, y, b) = 1\}$ forms a measurement to obtain Expression (6). ◀

3.2.1 Tightness of the lower bound

Here we obtain an equality relation between the value of the game and the quantum independence number of the game graph, for a class of pseudo-telepathy games.

▶ **Theorem 8.** Let $\mathcal{G}$ be a pseudo-telepathy game with a 0-1 valued predicate λ, admitting a quantum strategy consisting of a maximally entangled state $|\psi\rangle$ and pairwise commuting projectors. Let G be the corresponding game graph. Then,

$$\omega^*(\mathcal{G}) = \frac{\alpha_q(G)}{|X \times Y|} = 1.$$

Proof. From Theorem 6 we have $\alpha_q(G) \leq |X \times Y| \cdot \omega^*(\mathcal{G})$. We need to prove the other direction.

Let $\{P^x_a\}, \{Q^y_b\}$ be the strategies that win the game $\mathcal{G}$ on $|\psi\rangle$. We have:

$$\sum_{xy} \pi(x, y) \sum_{ab : \lambda(xyab) = 1} \langle\psi| P^x_a \otimes Q^y_b |\psi\rangle = 1,$$

so for all (x, y) we must have

$$\sum_{ab:\lambda(xyab)=1} \langle\psi|P_a^x \otimes Q_b^y|\psi\rangle = 1$$

and for all quadruples (x, y, a, b) such that $\lambda(xyab) = 0$ we have $P_a^x Q_b^y = 0$.

Let $\Pi_{xyab} = P_a^x Q_b^y$. These are projectors thanks to the commutativity assumption. We observe:

1. For all (x, y) we have

$$\sum_{ab:\lambda(xyab)=1} P_a^x Q_b^y = \sum_{ab} P_a^x Q_b^y = \sum_a P_a^x \sum_b Q_b^y = I,$$

where the second equality follows from $Q_b^y Q_{b'}^y = \delta_{bb'}$.

2. For each edge $(x, y, a, b), (x', y', a', b')$ we have a collection of t real-valued projective measurements $\{P_v^x\}_{v\in V}$ for all $x \in [t]$ that have the whole vertex set as outputs,

$$\Pi_{xyab}\Pi_{x'y'a'b'} = 0,$$

because if $x = x'$ and $a \neq a'$ then $P_a^x P_{a'}^x = 0$, and if $y = y'$ and $b \neq b'$ then $Q_b^y Q_{b'}^y = 0$.

Therefore, we can construct $|X \times Y|$ projective measurements that are a winning strategy for the independent set game with $t = |X \times Y|$ as follows. For each pair (x, y) consider the projective measurement $\{\Pi_{xyab}\}_{a,b:\lambda(xyab)=1}$ (and zero matrices for the other vertices of the graph). The first observation above proves that those are valid projective measurements; the second observation shows that they respect the consistency condition (2). ◀

4 Concluding remarks and open problems

We have formalized and discussed a novel approach for the study of non-local game in a combinatorial fashion. Work in progress on this approach relate to the easy generalization to more than 2 players, and the less-easy computation of graphs for the parallel repetition of games.

Our approach has ample room for improvement. Open questions include:

1. Can we find a tighter lower bound for the entangled value of all games by using some variant of the quantum independence number, such as the one in [2]? Alternatively, can we prove tightness of the current lower bound?
2. Can we find better lower bounds, for example using one of the variants of Lovász theta number?
3. Can we characterize the class of games for which the Lovász bound is tight? We know that the value of CHSH is exactly the theta number of its game graph (see [7]). Is this true for all the XOR games? This would reflect the fact that their value is easy to compute.
4. Are there other graph parameters related to the classical and entangled values of specific classes of games, for example unique games?
5. We have shown that for a class of pseudo-telepathy games that quantum players can win using commutative projective measurements on maximally entangled state, this bound is tight. A similar class of games is shown in [13] to be in one-to-one correspondence with a generalization of Kochen-Specker sets. It is not clear to us if those two results together could be used to prove something stronger. Perhaps the whole class could be interpreted as pseudo-telepathy games based on some graph parameter (maybe the homomorphism games in [15]) and the relationship to the quantum independence number would be a consequence of this.

Acknowledgements. The authors thank the referees of TQC 2014 for useful comments. Laura Mančinska is supported by the MOE Tier 3 Grant "Random numbers from quantum processes" (MOE2012-T3-1-009). G. Scarpa is supported by the European Commission. Part of this work was done while G. Scarpa was a PhD student at CWI, supported by Ronald de Wolf's VIDI grant from NWO. S. Severini is supported by the Royal Society and EPSRC.

References

1 Antonio Acín, Tobias Fritz, Anthony Leverrier and Ana Belén Sainz. A Combinatorial Approach to Nonlocality and Contextuality. December 2012. arXiv:1212.4084.

2 Jop Briët, Harry Buhrman, Monique Laurent, Teresa Piovesan, and Giannicola Scarpa. Zero-error source-channel coding with entanglement. In *The Seventh European Conference on Combinatorics, Graph Theory and Applications*, volume 16 of *CRM Series*, pages 157–162. Scuola Normale Superiore, 2013.

3 John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969.

4 B.S. Cirel'son. Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100, 1980.

5 T. S. Cubitt, D. Leung, W. Matthews, and A. Winter. Improving zero-error classical communication with entanglement. *Phys. Rev. Lett.*, 104:230503–230506, 2010.

6 P. J. Cameron, A. Montanaro, M. W. Newman, S. Severini, and A. Winter. On the quantum chromatic number of a graph. *Electr. J. Comb.*, 14(1), 2007.

7 Adán Cabello, Simone Severini, and Andreas Winter. Graph-theoretic approach to quantum correlations. *Phys. Rev. Lett.*, 112:040401, January 2014. Full version on arXiv:1010.2163.

8 D. E. Knuth and Stanford University. Computer Science Dept. *The sandwich theorem*. Stanford University, Dept. of Computer Science, 1993.

9 Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing*, STOC'02, pages 767–775, New York, NY, USA, 2002. ACM.

10 Subhash Khot. In *Proceedings of IEEE 25th Annual Conference on Computational Complexity*, pages 99–121, June 2010.

11 Julia Kempe, Oded Regev, and Ben Toner. Unique games with entangled provers are easy. *SIAM Journal on Computing*, 39(7):3207–3229, 2010. Preliminary version in FOCS'08. arXiv:0710.0655.

12 L. Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inf. Theory*, 25(1):1–7, 1979.

13 L. Mančinska, G. Scarpa, and S. Severini. New separations in zero-error channel capacity through projective Kochen-Specker sets and quantum coloring. *IEEE Trans. Inf. Theory*, 59(6):4025–4032, 2013.

14 A. Peres. Two simple proofs of the Kochen-Specker theorem. *J. Phys. A*, 24:L175–L178, 1991.

15 D. E. Roberson and L. Mancinska. Graph Homomorphisms for Quantum Players. December 2012. arXiv:1212.1724.

16 Claude E. Shannon. The zero error capacity of a noisy channel. IT-2(3):8–19, September 1956.

Optimal Bounds for Parity-Oblivious Random Access Codes with Applications

André Chailloux[1], Iordanis Kerenidis[2], Srijita Kundu[3], and Jamie Sikora[4]

1 **INRIA, Paris Rocquencourt, SECRET Project Team**
2 **Laboratoire d'Informatique Algorithmique: Fondements et Applications, Université Paris Diderot, Paris, France and Centre for Quantum Technologies, National University of Singapore, Singapore**
3 **Chennai Mathematical Institute, Chennai, India**
4 **Laboratoire d'Informatique Algorithmique: Fondements et Applications, Université Paris Diderot, Paris, France**

Abstract

Random Access Codes is an information task that has been extensively studied and found many applications in quantum information. In this scenario, Alice receives an n-bit string x, and wishes to encode x into a quantum state ρ_x, such that Bob, when receiving the state ρ_x, can choose any bit $i \in [n]$ and recover the input bit x_i with high probability. Here we study a variant called parity-oblivious random acres codes, where we impose the cryptographic property that Bob cannot infer any information about the parity of any subset of bits of the input, apart form the single bits x_i.

We provide the optimal quantum parity-oblivious random access codes and show that they are asymptotically better than the optimal classical ones. For this, we relate such encodings to a non-local game and provide tight bounds for the success probability of the non-local game via semi-definite programming. Our results provide a large non-contextuality inequality violation and resolve the main open question in [22].

1998 ACM Subject Classification E.4 Coding and Information Theory: Data compaction and compression

Keywords and phrases quantum information theory, contextuality, semidefinite programming

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.76

1 Introduction

Quantum Information theory studies how information is encoded in quantum mechanical systems and how it can be transmitted through quantum channels. A main question is whether quantum information is more powerful than classical information. A celebrated result by Holevo [13], shows that quantum information cannot be used to compress classical information. In high level, in order to transmit n uniformly random classical bits, one needs to transmit no less than n quantum bits. This might imply that quantum information is no more powerful than classical information. This however is wrong in many situations. In the model of communication complexity, one can show that transmiting quantum information may result in exponential savings on the communication needed to solve specific problems ([20, 5, 3, 11, 21]).

One specific information task that has been extensively studied in quantum information is the notion of *random access codes* (RACs) [1, 16]. In this scenario, Alice receives an n-bit string x, drawn from the uniform distribution, and wishes to encode x into a quantum state ρ_x, such that Bob, when receiving the state ρ_x, can choose any bit $i \in [n]$ and recover the input bit x_i with high probability by performing some general quantum operation on ρ_x.

RACs have been used in various situations in quantum information and computation, including in communication complexity, non-locality, extractors and divide-independence cryptography. [4, 14, 19, 10, 15]. Even though this task seems easier than transmitting the entire input string x, it is known that both in the classical and the quantum world, the length of the encoding must be at least $\Omega(n)$ and in fact, there is no gain between classical and quantum encodings [16].

On the other hand, a well-known example of the superiority of quantum information is the example of *dense coding*, or equivalently a RAC of length 1 for uniform inputs of length $n = 2$. In this case, the optimal classical encoding can achieve success probability $3/4$, while there exists a quantum encoding that achieves strictly higher success probability, in fact $\cos^2(\pi/8)$ [8, 23]. An advantage can also be proven for the case of encoding three bits into one qubit, but not for $n \geq 4$ [12].

Nevertheless, a question remained of whether there are variants of random access codes, for which we can have an asymptotically significant advantage in the quantum case. We show that this is indeed the case for the so-called *parity-oblivious* RACs. These are the usual RACs with the extra cryptographic property that the receiver cannot infer any information about the parity of any subset of bits of the input, apart from the single bits x_i.

Random acres codes that are parity-oblivious have been considered before. For example, the dense coding examples for encoding two or three classical bits in one qubit have this property. It is not hard to check, that for the 2-to-1 encoding, Bob's reduced density matrix is exactly the same for the cases where the inputs have parity 0 or 1, in other words, Bob has no information about $x_1 \oplus x_2$. Moreover, Spekkens, Buzacott, Keehn, Toner, and Pryde [22] used parity-oblivious RACs to provide non-contextuality inequalities.

1.1 Our results

In this paper, we provide the optimal quantum parity-oblivious RAC and show that it is asymptotically better than the optimal classical one. We say that an encoding with success probability $\frac{1}{2}(1+\alpha)$ has bias α. More precisely, we prove the following theorem.

▶ **Theorem 1.** *For any $n \in \mathbb{N}$, the optimal quantum parity-oblivious random access code for inputs of size n, denoted here as* PO-RAC^n*, has bias $\frac{1}{\sqrt{n}}$.*

The main idea of the proof is that quantum encodings can be studied through their relation to non-local games. Such equivalences between encodings and non-local games were previously noted in [17, 7]. A non-local game is a game between two non-communicating parties, who receive some inputs and must produce outputs that satisfy some known predicate. The best-known example is the CHSH game, where the two parties must output bits a and b, whose parity is equal to the logical and of their inputs x and y. The important quantity of such games is the optimal success probability when the two parties are allowed to share an arbitrary entangled state in the beginning of the protocol. In [7], it was shown that certain variants of the CHSH game are equivalent to some variants of quantum RACs and their respective success probabilities are equal.

In order to show an upper bound on the bias of quantum PO-RACs, we first define a weaker variant where only the parities of even-size subsets are hidden, denoted as EPO-RAC^n. An upper bound on the bias of these codes would imply an upper bound on the bias of the general PO-RACs.

Then, we study a natural non-local game which we call the INDEX game and show that EPO-RAC with *average* bias are equivalent to the INDEX game. In other words, the bias of any INDEX game strategy and the *average decoding bias* of an EPO-RAC are equal. In the INDEX^n game (parameterized by n here), Alice receives an n-bit string x, Bob receives an index t, and Alice and Bob are supposed to output bits a and b such that $a \oplus b = x_t$.

▶ **Theorem 2** (Equivalence). *For any $n \in \mathbb{N}$, there exists a quantum* EPO-RAC^n *with* average decoding bias α *if and only if there exists a quantum* INDEX^n *strategy with bias α.*

Last, noting that the INDEX game is an XOR game, i.e. the winning condition depends on the XOR of Alice and Bob's one-bit answers, we use a tight semidefinite programming characterization due to [9] and provide the exact optimal quantum bias.

▶ **Theorem 3** (Optimal INDEX game biases). *For any $n \in \mathbb{N}$, the optimal quantum bias of an* INDEX^n *strategy is $1/\sqrt{n}$ and the optimal classical bias is $\sqrt{\frac{2}{\pi n}}(1 + O(1/n))$.*

Since the *worst case bias* of a quantum PO-RAC is obviously upper bounded by the optimal *average case bias* of a quantum EPO-RAC, Theorems 2 and 3 show that every PO-RAC^n has bias at most $1/\sqrt{n}$. On the other hand, we give an explicit construction of a PO-RAC^n with bias $1/\sqrt{n}$ that uses $\lfloor n/2 \rfloor$ qubits. First, we provide a parity-oblivious encoding where Alice and Bob share $\lfloor n/2 \rfloor$ EPR pairs and then Alice sends one classical bit of communication.

▶ **Theorem 4** (Optimal PO-RAC^n). *For any integer n, there exists a* PO-RAC^n *with bias $1/\sqrt{n}$ that uses $\lfloor n/2 \rfloor$ qubits and* 1 *classical bit.*

We also remark that even though quantum PO-RAC^n and EPO-RAC^n both share the same optimal bias, the same is not true if we consider *odd-parity-oblivious* encodings where the S-parities are hidden for $|S|$ odd and strictly greater than 1. Consider encoding a six-bit string $(x_1, \ldots, x_6)$ where the first three bits are encoded using the optimal PO-RAC^3, and similarly for the last three bits. It is a straightforward exercise to verify this is odd-parity oblivious with bias $1/\sqrt{3} > 1/\sqrt{6}$.

1.2 Application to non-contextuality

The basic primitives in an operational theory are preparations and measurements. A hidden variable model is *preparation and measurement non-contextual*, if whenever two preparations yield the same statistics for all possible measurements then they have an equivalent representation in the model; and whenever two measurements have the same statistics for all preparations then they have an equivalent representation in the model [22]. Similar to non-locality, a non-contextuality inequality is any inequality on probability distributions that follows from the assumption that there exists a hidden variable model that is preparation or measurement non-contextual.

Spekkens, Buzacott, Keehn, Toner, and Pryde [22] proved the following *non-contextuality inequality* (or NC inequality, for short): In an operational theory that admits a preparation non-contextual hidden variable model, the *average case bias* for any PO-RAC^n is at most $1/n$.

Then, they noted that quantum mechanics violates this non-contextuality (NC) inequality for $n \in \{2, 3\}$, since there exists a quantum parity-oblivious encoding of two and three classical bits into one qubit, with average decoding probability $\frac{1}{2}(1 + \frac{1}{\sqrt{2}})$ and $\frac{1}{2}(1 + \frac{1}{\sqrt{3}})$, respectively [1, 12]. It was left as an open question whether quantum mechanics violates this

NC inequality for larger n. The main difficulty to extend these results for larger input size is that we pose no bound on the dimension of the encoding.

Through our Theorem 1 that provides the optimal bias for PO-RAC^{n}s, we resolve the main open question in [22] and provide a family of non-contextuality inequality violations that grow with the input size n. More precisely, we show an explicit non-contextuality violation of order $\sqrt{n}$.

▶ **Theorem 5.** *For any integer n, there exists an explicit non-contextuality inequality that provides a violation of order $\sqrt{n}$.*

2 Preliminaries

We provide the definitions of the different variants of random access codes that we use and of the non-local game we consider.

2.1 Random Access Codes

▶ **Definition 6** (Random access code)**.** For an integer $n \geq 2$, a quantum *random access code* of n bits, denoted RAC^n, with bias α consists of an encoding map of $x \in \{0,1\}^n$ into quantum states ρ_x together with a sequence of n possible measurements such that the result of the i'th measurement is x_i with probability at least $\frac{1}{2}(1+\alpha)$.

Note that the usual treatment of RACs is to analyze the relationships between n, α, and the encoding dimension (i.e., the dimension of ρ_x). In this paper, we are not concerned with the encoding dimension, but rather the optimal bias when we enforce certain cryptographic properties to RACs. For example, we enforce that Bob remains oblivious of some information about the string x, meaning that he cannot infer any information about it from the encoding. In particular, we consider for each subset S of bits of x the S-parity, which is defined as $\bigoplus_{i \in S} x_i$.

▶ **Definition 7** (Parity-oblivious random access codes)**.** For an integer $n \geq 2$, a quantum *parity-oblivious random access code*, denoted as PO-RAC^n, is a RAC^n with the cryptographic constraint that the receiver is oblivious of every S-parity, for $|S| \geq 2$.

For classical codes, the optimal bias of a PO-RAC^n is known to be $\frac{1}{n}$ (Proposition 1).

In our proofs, we also use a weaker variant of parity-oblivious random access codes, where only the S-parities of even-size remain oblivious.

▶ **Definition 8** (Even-parity-oblivious random access codes)**.** For an integer $n \geq 2$, a quantum *even-parity-oblivious random access code*, denoted as EPO-RAC^n, is a RAC^n with the cryptographic constraint that the receiver is oblivious of every S-parity, for $|S|$ even.

▶ Remark. In the definition of RAC^{n}s, we have that *every bit* is decode with bias α. We have occasion to study EPO-RAC^{n}s with *average case bias* α, that is, the average over all $i \in [n]$ of the decoding probabilities. When we consider average case biases, it is explicitly mentioned, otherwise, worst-case bias is assumed.

2.2 Non-local games

In a non-local game, two non-communicating parties, Alice and Bob, receive some inputs x and y, respectively, and must output a and b, respectively, such that (x, y, a, b) satisfy some specific condition. For example in the CHSH game, the condition is $a \oplus b = x \cdot y$. The goal is to find the optimal quantum (classical) success probability of satisfying the condition when Alice and Bob are allowed to share some initial quantum state (shared randomness).

We define the following non-local game.

▶ **Definition 9** (Index game). The *Index game*, denoted here as INDEX^n, is the following XOR game:

- Alice's input: Alice receives a random s from the set $S := \{0,1\}^n$.
- Bob's input: Bob receives a random index t from the set $T := [n]$.
- Winning condition: They win if Alice's output bit a and Bob's output bit b satisfy $a \oplus b = s_t$.

The choice of initial resource state and local measurement operators (that depend on the respective inputs) comprise a *strategy*. We say that a strategy has *bias* α if it succeeds with probability $\frac{1}{2}(1+\alpha)$.

Note that our game is similar to the retrieval games studied in [17].

3 Equivalence of EPO-RAC^n decoding and INDEX^n strategies

In this section we prove the equivalence in Theorem 2.

▶ **Theorem 2** (Equivalence). *For any $n \in \mathbb{N}$, there exists a quantum* $\mathrm{EPO\text{-}RAC}^n$ *with bias α if and only if there exists a quantum* INDEX^n *strategy with bias α.*

3.1 From EPO-RAC^n to INDEX^n

Let us fix an $\mathrm{EPO\text{-}RAC}^n$ $\{\rho_x\}_{x\in\{0,1\}^n}$ with bias α. Let $\mathcal{B}$ the Hilbert space used for the encoding. Our goal is to construct a strategy for INDEX^n with bias α. For each ρ_x, we fix a purification $|\psi_x\rangle$ of ρ_x in the space $\mathcal{A}\otimes\mathcal{B}$. For $a \in \{0,1\}$, let $\boldsymbol{a}$ be the n-bit string $(a,\ldots,a)$ and $\bar{s}$ is the complement string of s. We define

$$|\Omega_s\rangle = \frac{1}{\sqrt{2}}\sum_{a\in\{0,1\}} |a\rangle_{\mathcal{O}}|\psi_{s\oplus\boldsymbol{a}}\rangle_{\mathcal{AB}} = \frac{1}{\sqrt{2}}\left(|0\rangle|\psi_s\rangle + |1\rangle|\psi_{\bar{s}}\rangle\right) .$$

We would like to show that if Bob has the register $\mathcal{B}$ of the above state, then he has no information about s. Note that his reduced state is the state $\sigma_s = \frac{1}{2}\rho_s + \frac{1}{2}\rho_{\bar{s}}$.

The first step is to see that Bob has no information about any parity (odd or even) of the string s. For the even parities, note that we started with an $\mathrm{EPO\text{-}RAC}^n$ encoding and that the strings s and $\bar{s}$ have the same even parities. Hence, Bob has with half probability the state ρ_s from which he cannot get any information about the even parities of s and with half probability the state $\rho_{\bar{s}}$ from which he cannot get any information about the even parities of $\bar{s}$ and consequently s.

For the odd parities: fix an subset $S \subseteq \{1,\ldots,n\}$ of odd size and let $s_S = \oplus_{i\in S} s_i$. Let $M = \{M_0, M_1\}$ be any two outcome POVM. Let $P_b = \{s \in \{0,1\}^n : s_S = b\}$. Each P_b has size 2^{n-1} and $s \in P_b \Leftrightarrow \bar{s} \in P_{\bar{b}}$ since S is an odd subset. We have

$$\begin{aligned}
\Pr[\text{Bob guesses } s_S \text{ using } M] &= \frac{1}{2^n}\sum_{s\in P_0} tr(M_0\sigma_s) + \sum_{s\in P_1} tr(M_1\sigma_s) \\
&= \frac{1}{2^n}\sum_{s\in P_0} tr(M_0\sigma_s) + \sum_{s\in P_0} tr(M_1\sigma_{\bar{s}}) \\
&= \frac{1}{2^n}\sum_{s\in P_0} tr((M_0+M_1)\sigma_s) \qquad\qquad \text{using } \forall s, \sigma_s = \sigma_{\bar{s}} \\
&= \frac{1}{2^n}\sum_{s\in P_0} tr(I\sigma_s) = \frac{|P_0|}{2^n} = 1/2\,.
\end{aligned}$$

This means that for any measurement M, Bob has probability $1/2$ to guess s_S which means that Bob has no information about this bit.

In the following lemma we prove that if someone has no information about any parity of subsets of bits of a string x, then he has no information about the string x. This is intuitively an obvious statement that we rigorously prove below.

▶ **Lemma 10.** *Let X be the uniform distribution on $x \in \{0,1\}^n$. If Bob has no information about any parity of subsets of bits of x, then he has no information about x.*

Proof. If Bob has some information about x, then the states ρ_x cannot be all the same, which in turn implies that there exists a subset $T \in \{0,1\}^n$ of size 2^{n-1} such that $\rho_T = \frac{1}{2^{n-1}}\sum_{x\in T}\rho_x$ is not equal to $\rho_{\bar{T}} = \frac{1}{2^{n-1}}\sum_{x\in\bar{T}}\rho_x$. This means that there exists a two-outcome measurement that outputs 1 if $x \in T$ and -1 otherwise, with positive bias. We now show for a contradiction that this measurements must also output a parity of some subset with positive bias. Define the function $f : \{0,1\}^n \to \{-1,+1\}$, as the indicator function of T and let b the measurement outcome. Then

$$\mathbb{E}[b \cdot f(x)] > 0\,.$$

By taking the Fourier representation of the function and denoting $x_S = \bigoplus_{i\in S} x_i$ we have

$$\mathbb{E}[b \cdot \sum_S \hat{f}(S)x_S] > 0\,,$$

$$\sum_S \hat{f}(S)\mathbb{E}[b \cdot x_S] > 0\,.$$

Since for the empty set we have $\hat{f}(\emptyset) = \mathbb{E}[f(x)] = 0$, the above implies that there exists a parity S for which $E[b \cdot x_S] > 0$, which is a contradiction. ◀

The above statement means that for each s, we have $Tr_{\mathcal{OA}}|\Omega_s\rangle\langle\Omega_s| = Tr_{\mathcal{OA}}|\Omega_0\rangle\langle\Omega_0|$. In particular, this means that there exist unitaries $\{U_s\}$ acting on $\mathcal{AO}$ such that $(U_s \otimes I)|\Omega_0\rangle = |\Omega_s\rangle$.We use the state $|\psi_0\rangle$ to define the INDEX^n strategy:

- Alice and Bob share the state $|\Omega_0\rangle \in \mathcal{A} \otimes \mathcal{B}$.
- Upon receiving $s \in \{0,1\}^n$, Alice applies U_s on $\mathcal{OA}$ such that Alice and Bob share $|\Omega_s\rangle$. Alice measures register $\mathcal{O}$ in the computational basis and outputs the corresponding a.
- For Alice's input s and output a, Bob has an encoding ρ_x where $x = s \oplus \boldsymbol{a}$. Upon receiving $t \in [n]$, Bob measures $\mathcal{B}$ just as in the EPO-RAC^n to learn x_t. He outputs b equal to his guess.

- Alice and Bob win the game if $b = s_t \oplus a = x_t$ meaning that they win the game if and only if Bob correctly guesses x_t.

Since our encoding has bias α, we see that with this INDEX^n strategy, they succeed with probability

$$\frac{1}{n}\sum_{i=1}^{n} \Pr[\text{Bob outputs } a \oplus s_t] = \frac{1}{n}\sum_{i=1}^{n} \Pr[\text{Bob outputs } x_t] = \frac{1}{2}(1+\alpha),$$

as desired. ◀

3.2 From INDEX^n to EPO-RAC^n

Suppose Alice and Bob have a strategy to win the INDEX^n game with bias α with starting state $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$. On input $s \in \{0,1\}^n$, Alice performs her side of the optimal strategy for INDEX^n and has some output a. We have:

$$\frac{1}{n}\sum_{i=1}^{n} \Pr[\text{Bob outputs } a \oplus s_t] = \frac{1}{2}(1+\alpha)\,.$$

Let $\rho_{s,a}$ the state that Bob has when Alice inputs s and outputs a. Let x such that $\forall i, x_i = s_i \oplus a$. When Alice has inputs satisfying $s \oplus \boldsymbol{a} = x$, Bob has the state $\sigma_x = \frac{1}{2}(\rho_{x,0} + \rho_{\bar{x},1})$. We show that $\{\sigma_x\}_x$ is an EPO-RAC^n with average bias α.

1. It's a EPO-RAC^n: for every even parity S, we have $\bigoplus_{i\in S} x_i = \bigoplus_{i\in S}(s_i \oplus a) = \bigoplus_{i\in S} s_i$. Bob has no information about s from non signalling so Bob has no information about $\bigoplus_{i\in S} s_i$.
2. It has average bias α: Alice and Bob win the INDEX^n game with bias α hence

$$\frac{1}{n}\sum_{i=1}^{n} \Pr[\text{Bob outputs } x_t] = \frac{1}{n}\sum_{i=1}^{n} \Pr[\text{Bob outputs } a \oplus s_t] = \frac{1}{2}(1+\alpha)\,.$$

▶ Remark. Note that the above equivalence also holds in the classical setting.

4 On the structure of optimal Index Game strategies

In this section, we prove Theorem 3, below.

▶ **Theorem 11** (Optimal Index Game biases). *For any $n \in \mathbb{N}$, the optimal quantum bias of an* INDEX^n *strategy is $1/\sqrt{n}$ and the optimal classical bias is $\sqrt{\frac{2}{\pi n}}(1 + O(1/n))$.*

4.1 The quantum value

The quantum bias of any XOR game can be found efficiently by solving a semidefinite program (SDP) [9]. Specifically, the quantum bias of the INDEX^n game can be calculated as the optimal value of either SDP below

Primal problem (P)		Dual problem (D)	
supremum:	$\langle B, X\rangle$	infimum:	$\langle e, y\rangle$
subject to:	$\text{diag}(X) = e,$	subject to:	$\text{Diag}(y) \succeq B,$
	$X \succeq 0,$		

where

- $\text{diag}(X)$ is the vector on the diagonal of the square matrix X,
- e is the vector of all ones,
- $\text{Diag}(y)$ is the diagonal matrix with the vector y on the diagonal,
- $B := \frac{1}{2}\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$, where $A_{s,t} := \frac{(-1)^{s_t}}{n2^n}$.

For (P), consider the positive semidefinite matrix $X := YY^\top$, where

$$Y := \begin{bmatrix} \sqrt{n}\,2^n A \\ I_T \end{bmatrix}.$$

To show X is feasible in (P), one can check that each diagonal entry of X is equal to 1 from the definition of A above. Note that $\langle B, X\rangle := \sqrt{n}\,2^n\langle A, A\rangle = 1/\sqrt{n}$ proving that the quantum bias is at least $1/\sqrt{n}$ (since the quantum bias is the maximum of $\langle B, X\rangle$ over all feasible X).

For (D), let $y := \begin{bmatrix} u\, e_S \\ v\, e_T \end{bmatrix}$ where $u, v > 0$ and e_S and e_T are the vectors of all ones indexed by entries in S and T, respectively. Then

$$\text{Diag}(y) \succeq B \iff \begin{bmatrix} uI_S & -\frac{1}{2}A \\ -\frac{1}{2}A^T & vI_T \end{bmatrix} \succeq 0 \iff uvI_T \succeq \frac{1}{4}A^\top A = \frac{1}{4n^2 2^n} I_T.$$

From above, if we set $v := \frac{1}{2n\sqrt{n}}$ and $u := \frac{1}{2\sqrt{n}2^n}$, then y is feasible in (D). Since $\langle e, y\rangle = 2^n u + nv = \frac{1}{\sqrt{n}}$, we know the quantum bias is at most $1/\sqrt{n}$ (since the quantum bias is equal to the minimum of $\langle e, y\rangle$ over all feasible y).

Therefore, the quantum bias is exactly $1/\sqrt{n}$, as required.

4.2 The classical value

We can assume without loss of generality that Alice and Bob's strategies are deterministic. Define $b \in \{0,1\}^n$ as the string of potential answers Bob gives where b_t is the bit that Bob outputs on input $t \in [n]$. Now let us examine Alice's strategy. For a fixed input s, if she outputs 0, they win the game with probability $\frac{1}{n}|b \oplus s|_H$, where $|x|_H$ denotes the Hamming weight of a string $x \in \{0,1\}^n$. If she outputs 1, they win the game with probability $\frac{1}{n}|b \oplus \overline{s}|_H = n - \frac{1}{n}|b \oplus s|_H$. This means that they win the game with probability at most

$$\begin{aligned}\mathop{\mathbb{E}}_{s\in\{0,1\}^n}\left[\frac{1}{n}\max\{|b\oplus s|_H, n - \frac{1}{n}|b\oplus s|_H\}\right] &= \frac{1}{n}\mathop{\mathbb{E}}_{s}\left[\frac{n}{2} + \left|\frac{n}{2} - |b\oplus s|_H\right|\right] \\ &= \frac{1}{2} + \frac{1}{n}\mathop{\mathbb{E}}_{s}\left[\left|\frac{n}{2} - |b\oplus s|_H\right|\right].\end{aligned}$$

The quantity $\mathbb{E}_s[|n/2 - |b \oplus s|_H|]$ corresponds to the expected deviation that the uniform binomial distribution has from the average. This is a well studied quantity and we know that

$$\mathop{\mathbb{E}}_{s}\left[\left|\frac{n}{2} - |b\oplus s|_H\right|\right] = \frac{2}{\sqrt{\pi}}\frac{\Gamma(n+1/2)}{\Gamma(n)} = \sqrt{\frac{2n}{\pi}}\left(1 + O\left(\frac{1}{n}\right)\right).$$

Therefore, any strategy has success probability bounded above by

$$\frac{1}{2} + \frac{1}{n}\mathop{\mathbb{E}}_{s}\left[\left|\frac{n}{2} - |b\oplus s|_H\right|\right] = \frac{1}{2} + \sqrt{\frac{2}{\pi n}}\left(1 + O\left(\frac{1}{n}\right)\right).$$

Now, consider the following strategy: Alice outputs a which equals the majority of s, and Bob outputs 0. This strategy has success probability precisely

$$\frac{1}{2} + \frac{1}{n} \mathop{\mathbb{E}}_{s} \left[\left| \frac{n}{2} - |b \oplus s|_H \right| \right]$$

which is optimal.

5 A construction of a quantum PO-RAC^n with optimal bias

In this section we give an explicit construction of an quantum PO-RAC^n with optimal bias.

▶ **Theorem 12** (Optimal PO-RAC^n). *For any integer $n \geq 2$, there exists a* PO-RAC^n *with bias $1/\sqrt{n}$ that uses $\lfloor n/2 \rfloor$ qubits and 1 classical bit.*

Our construction builds upon the well-known RACs for sending 2 (resp. 3) bits with bias $1/\sqrt{2}$ (resp. $1/\sqrt{3}$) [24, 2, 12]. These are the vertices from the corners of a square inscribed in an equatorial plane in the Bloch sphere, and the corners of a cube inscribed in the Bloch sphere, respectively. To generalize this idea to an n-cube inscribed in an n-dimensional sphere, we use the intuition of *hyperbits* which is a way to visualize such unit vectors in a quantum mechanical setting. A full discussion of hyperbits and their equivalence to certain quantum protocols is beyond the scope of this paper, but we refer the interested reader to the work of Pawlowski and Winter [18].

5.1 The construction

Our construction is very similar to the proof of Tsirelson's theorem [23]. We start by recursively defining the observables $G_{n,1}, \ldots, G_{n,n}$ which are used to define the actions of Alice and Bob in the PO-RAC^n.

For $n = 2$ and $n = 3$, we define

$$G_{2,1} := X, \quad G_{2,2} := Y \qquad \text{and} \qquad G_{3,1} := X, \quad G_{3,1} := Y, \quad G_{3,3} := Z.$$

We use the $n = 3$ observables as a base case for a recursive formula: for n even, we define

$$G_{n,i} := G_{n-1,i} \otimes X, \quad \text{for} \quad i \in \{1, \ldots, n-1\}, \qquad \text{and} \qquad G_{n,n} = \mathbb{I} \otimes Y$$

and for n odd, we define

$$G_{n,i} := G_{n-2,i} \otimes X, \quad \text{for} \quad i \in \{1, \ldots, n-2\}, \quad G_{n,n-1} = \mathbb{I} \otimes Y, \qquad \text{and} \qquad G_{n,n} = \mathbb{I} \otimes Z.$$

Note that these act on $\lfloor n/2 \rfloor$ qubits, have eigenvalues ± 1, and satisfy the anti-commutation relation

$$\{G_{n,i}, G_{n,j}\} = 2\delta_{i,j}\mathbb{I}.$$

Define the following operators for $x \in \{0,1\}^n$ and $t \in [n]$:

$$A_x := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-1)^{x_i} G_i \qquad \text{and} \qquad B_t := G_t^\top.$$

Note that $A_x^2 = \mathbb{I}$, for all $x\ in \{0,1\}^n$, and $B_t^2 = \mathbb{I}$, for all $t \in [n]$, so each have ±1 eigenvalues.

The PO-RAC^n protocol is defined below.

- Encoding states: Alice chooses a uniformly random $x \in \{0,1\}^n$, creates $\lfloor n/2 \rfloor$ EPR pairs, and measures the first "halves" with the observable A_x to get an outcome $a \in \{-1,+1\}$. She sends the second "halves" and a to Bob. Bob now has a quantum state encoding the string x.
- Decoding procedure: If Bob wishes to learn x_t, he measures his EPR halves with the observable B_t to get an outcome $b \in \{-1,+1\}$. He computes $c = ab$ and outputs 0 if $c = +1$, and 1 otherwise.

In the next two lemmas, we show that the bias of this RAC^n is $\frac{1}{\sqrt{n}}$ and that it is parity-oblivious, thereby proving Theorem 4.

▶ **Lemma 13.** *This* RAC^n *has bias* $1/\sqrt{n}$*.*

Proof. We can assume at the beginning of the protocol, Alice and Bob share the maximally entangled state

$$|\psi\rangle := \frac{1}{\sqrt{2^{\lfloor \frac{n}{2} \rfloor}}} \sum_{j=1}^{2^{\lfloor \frac{n}{2} \rfloor}} |j\rangle_{\mathcal{A}} |j\rangle_{\mathcal{B}}.$$

The expectation value of the observable $C = A_x \otimes B_t$ in this state is given by:

$$\langle C \rangle = \langle \psi | A_x \otimes B_t | \psi \rangle = \frac{1}{\sqrt{n}} \frac{1}{2^{\lfloor \frac{n}{2} \rfloor}} \sum_{i=1}^{n} (-1)^{x_i} \underbrace{\sum_{j,k=1}^{2^{\lfloor \frac{n}{2} \rfloor}} \langle j|_{\mathcal{A}} \langle j|_{\mathcal{B}}\, G_i \otimes G_t^{\top}\, |k\rangle_{\mathcal{A}} |k\rangle_{\mathcal{B}}}_{=2^{\lfloor \frac{n}{2} \rfloor} \delta_{i,t}} = \frac{(-1)^{x_t}}{\sqrt{n}}.$$

where the third equality is derived from the anti-commutation relation.
Now, $\langle C \rangle = \Pr[c = +1] - \Pr[c = -1] = \langle \psi | A_x \otimes B_t | \psi \rangle$, so

$$\Pr[\text{Bob outputs } 0] = \Pr[c = +1] = \frac{1}{2}\left[1 + \frac{(-1)^{x_t}}{\sqrt{n}}\right]$$

$$\Pr[\text{Bob outputs } 1] = \Pr[c = -1] = \frac{1}{2}\left[1 - \frac{(-1)^{x_t}}{\sqrt{n}}\right]$$

implying

$$\Pr[\text{Bob outputs } x_t] = \frac{1}{2}\left(1 + \frac{1}{\sqrt{n}}\right),$$

as desired. ◀

▶ **Lemma 14.** *This* RAC^n *is parity-oblivious.*

Proof. Protocols involving shared entanglement and sending one bit of classical information have limited guessing probabilities for functions such as parity, as shown in [18]. In particular, it can be been shown that the biases of learning $\bigoplus_{i \in S} x_i$, denoted here as α_S, satisfy

$$\sum_{S \subseteq \{0,1\}^n \setminus \text{Empty Set}} \alpha_S^2 \leq 1.$$

For our protocol,

$$\sum_{S:|S|=1} \alpha_S^2 = n \cdot \left(\frac{1}{\sqrt{n}}\right)^2 = 1$$

implying $\alpha_S = 0$ for all S of size 2 or greater, implying it is parity-oblivious. ◀

6 Large non-contextuality inequality violations

Spekkens et al. [22] constructed a family of non-contextuality inequalities from the notion of parity-oblivious random access codes. More precisely, they showed that

▶ Proposition 1 ([22], NC inequality). In any operational theory that admits a preparation non-contextual hidden variable model, the *average case bias* for any PO-RAC^n is at most $1/n$.

In order to quantify the violation of this NC inequality, we consider the ratio of the average case bias of quantum PO-RAC^n and PO-RAC^n of any operational theory that admits a preparation non-contextual hidden variable model.

Note, that if three exists a game for which the winning probability of any classical strategy cannot deviate from $1/2$ by more than δ_1 and, moreover, there is a quantum strategy obtaining winning probability at least $1/2 + \delta_2$, then we can obtain a violation of order δ_2/δ_1 (see [6] for details).

Then, Theorem 5 is a direct consequence of Proposition 1 and our Theorem 1.

▶ **Theorem 15.** *For any $n \in \mathbb{N}$, there exists an explicit non-contextuality inequality that provides a violation of order $\sqrt{n}$.*

References

1 A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and a lower bound for 1-way quantum automata. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing*, pages 376 – 383, 1999.

2 A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani. Dense quantum coding and quantum finite automata. *Journal of the ACM*, 49(4):496–511, 2002.

3 Z. Bar-Yossef, T. S. Jayram, and I. Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *Proceedings of 36th ACM STOC*, pages 128–137, 2004.

4 Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and ldcs. In *FOCS'08*, pages 477–486, 2008.

5 H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87:167902, Sep 2001.

6 Harry Buhrman, Oded Regev, Giannicola Scarpa, and Ronald de Wolf. Near-optimal and explicit bell inequality violations. *Theory of Computing*, 8(27):623–645, 2012.

7 A. Chailloux, I. Kerenidis, and J. Sikora. Strong connections between quantum encodings, non-locality and quantum cryptography. *PRA, to appear.*, 2014.

8 J. Clauser, M. Horne, A. Shimony, and R. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969.

9 R. Cleve, W. Slofstra, F. Unger, and S. Upadhyay. Perfect parallel repetition theroem for quantum XOR proof systems. *Computational Complexity*, 17(2):282–299, 2008.

10 Anindya De and Thomas Vidick. Near-optimal extractors against quantum storage. In *STOC'10*, pages 161–170, 2010.

11 D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2008.

12 M. Hayashi, K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita. (4,1)-quantum random access coding does not exist – one qubit is not enough to recover one of four bits. *New Journal of Physics*, 8(8):129, 2006.

13 A. Holevo. Some estimates of the information transmitted by quantum communication channels. *Problemy Peredachi Informatsii*, 9:3–11, 1973.

14 Kazuo Iwama, Harumichi Nishimura, Rudy Raymond, and Shigeru Yamashita. Unbounded-error one-way classical and quantum communication complexity. In *ICALP'07*, pages 110–121, 2007.

15 Hong-Wei Li, Marcin Pawłowski, Zhen-Qiang Yin, Guang-Can Guo, and Zheng-Fu Han. Semi-device-independent randomness certification using $n \to 1$ quantum random access codes. *Phys. Rev. A*, 85:052308, May 2012.

16 A. Nayak. Optimal lower bounds for quantum automata and random access codes. *Proceedings of 40th IEEE Symposium on Foundations of Computer Science*, 0:369–376, 1999.

17 J. Oppenheim and S. Wehner. The uncertainty principle determines the non-locality of quantum mechanics. *Science*, 330:6007:1072–1074, 2010.

18 M. Pawlowski and A. Winter. From qubits to hyperbits. *Phys. Rev. A*, 85:022331, 2012.

19 Marcin Pawłowski and Marek Żukowski. Entanglement-assisted random access codes. *Phys. Rev. A*, 81:042326, Apr 2010.

20 Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proc. 31st Annual ACM Symposium on Theory of Computing*, pages 358–367, New York, NY, USA, 1999. ACM.

21 O. Regev and B. Klartag. Quantum one-way communication can be exponentially stronger than classical communication. In *STOC'11*, pages 31–40, 2011.

22 R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde. Preparation contextuality powers parity-oblivious multiplexing. *Physical Review Letters*, 102:010401, 2009.

23 B. Tsirelson. Quantum analogues of the bell inequalities: The case of two spatially separated domains. *Journal of Soviet Mathematics*, 36:557–570, 1987.

24 S. Wiesner. Conjugate coding. *SIGACT News*, 15(1):78–88, 1983.

Convexity Properties of the Quantum Rényi Divergences, with Applications to the Quantum Stein's Lemma *

Milán Mosonyi[1,2]

1 Física Teòrica: Informació i Fenomens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain
2 Mathematical Institute, Budapest University of Technology and Economics, Egry József u 1., Budapest, 1111 Hungary

Abstract

We show finite-size bounds on the deviation of the optimal type II error from its asymptotic value in the quantum hypothesis testing problem of Stein's lemma with composite null-hypothesis. The proof is based on some simple properties of a new notion of quantum Rényi divergence, recently introduced in [Müller-Lennert, Dupuis, Szehr, Fehr and Tomamichel, J. Math. Phys. **54**, 122203, (2013)], and [Wilde, Winter, Yang, arXiv:1306.1586].

1998 ACM Subject Classification E.4 Coding and information theory, H.1.1 Information theory

Keywords and phrases Quantum Rényi divergences, Stein's lemma, composite null-hypothesis, second-order asymptotics

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.88

1 Introduction

Rényi defined the α-divergence [36] of two probability distributions p, q on a finite set $\mathcal{X}$ as

$$D_\alpha(p\|q) := \frac{1}{\alpha-1}\log\sum_{x\in\mathcal{X}} p(x)^\alpha q(x)^{1-\alpha},$$

where $\alpha \in (0,+\infty)\setminus\{1\}$. These divergences have various desirable mathematical properties; they are strictly positive, non-increasing under stochastic maps, and jointly convex for $\alpha \in (0,1)$ and jointly quasi-convex for $\alpha > 1$. For fixed p and q, $D_\alpha(p\|q)$ is a monotone increasing function of α, and the limit $\alpha \to 1$ yields the relative entropy (a.k.a. Kullback-Leibler divergence), probably the single most important quantity in information theory. Even more importantly, the Rényi divergences have great operational significance, as quantifiers of the trade-off between the relevant operational quantities in many information theoretic tasks, including hypothesis testing, source compression, and information transmission through noisy channels [12]. A direct operational interpretation of the Rényi divergences as generalized cutoff rates has been shown in [12].

In the view of the above, it is natural to look for an extension of the Rényi divergences for pairs of quantum states. One such extension has been known in quantum information theory for quite some time, defined for states ρ and σ as [34]

$$D_\alpha^{(\mathrm{old})}(\rho\|\sigma) := \frac{1}{\alpha-1}\log\operatorname{Tr}\rho^\alpha\sigma^{1-\alpha}.$$

* This work was partially supported by the European Research Council Advanced Grant "IRQUAT".

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 88–98
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

These divergences also form a monotone increasing family, with the Umegaki relative entropy $D_1(\rho\|\sigma) := \operatorname{Tr}\rho(\log\rho - \log\sigma)$ as their limit at $\alpha \to 1$. They are also strictly positive; however, monotonicity under stochastic (i.e., completely positive and trace-preserving) maps only holds for $\alpha \in [0,2]$. Recently, a new quantum Rényi divergence has been introduced in [28, 41], defined as

$$D_\alpha^{(\text{new})}(\rho\|\sigma) := \frac{1}{\alpha-1}\log\operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha.$$

Again, these new divergences yield the Umegaki relative entropy in the limit $\alpha \to 1$, and monotonicity only holds on a restricted domain, in this case for $\alpha \in [1/2, +\infty)$.

Operational interpretation has been found for both definitions in the setting of binary hypothesis testing for different and matching domains of α. The goal in hypothesis testing is to decide between two candidates, ρ and σ, for the true state of a quantum system, based on a measurement on many identical copies of the system. The quantum Stein's lemma [19, 32] states that it is possible to make the probability of erroneously choosing ρ (type II error) to vanish exponentially fast in the number of copies, with the exponent being the relative entropy $D_1(\rho\|\sigma)$, while the probability of erroneously choosing σ (type I error) goes to zero asymptotically. If the type II error is required to vanish with a suboptimal exponent $r < D_1(\rho\|\sigma)$ (this is called the direct domain) then the type I error can also be made to vanish exponentially fast, with the optimal exponent being the Hoeffding divergence $H_r := \sup_{\alpha\in(0,1)}\frac{\alpha-1}{\alpha}[r - D_\alpha^{(\text{old})}(\rho\|\sigma)]$ [4, 18, 30]. Thus, the $D_\alpha^{(\text{old})}$ with $\alpha \in (0,1)$ quantify the trade-off between the rates of the type I and the type II error probabilities in the direct domain. Based on this trade-off relation, a more direct operational interpretation was obtained in [25] as generalized cutoff rates in the sense of Csiszár [12]. On the other hand, if the type II error is required to vanish with an exponent $r > D_1(\rho\|\sigma)$ (this is called the strong converse domain) then the type I error goes to 1 exponentially fast, with the optimal exponent being the converse Hoeffding divergence $H_r^* := \sup_{\alpha>1}\frac{\alpha-1}{\alpha}[r - D_\alpha^{(\text{new})}(\rho\|\sigma)]$ [26]. Thus, the $D_\alpha^{(\text{new})}$ with $\alpha > 1$ quantify the trade-off between the rates of the type I success probability and the type II error probability in the strong converse region. Based on this, a direct operational interpretation of the $D_\alpha^{(\text{new})}$ as generalized cutoff rates was also given in [26] for $\alpha > 1$.

In the view of the above results, it seems that the old and the new definitions provide the operationally relevant quantum extension of Rényi's divergences in different domains: for $\alpha \in (0,1)$, the operationally relevant definition seems to be the old one, corresponding to the direct domain of hypothesis testing, whereas for $\alpha > 1$, the operationally relevant definition seems to be the new one, corresponding to the strong converse domain of hypothesis testing.

This is the picture at least when one wants to describe the full trade-off curve; most of the time, however, one is interested in one single point of this curve, corresponding to $\alpha = 1$, where the transition from exponentially vanishing error probability to exponentially vanishing success probability happens. It is known that using the "wrong" divergence can be beneficial to obtaining coding theorems at this point. Indeed, the strong converse property for hypothesis testing and classical-quantum channel coding has been proved using $D_\alpha^{(\text{old})}$ for $\alpha > 1$ in [29, 32, 33] ("wrong" divergence with the "right" values of α), while a proof for the direct part of these problems was obtained recently in [8], using $D_2^{(\text{new})}$ ("'wrong" divergence with a "wrong" value of α).

Further examples of coding theorems based on the "wrong" Rényi divergence were given in [27], where it was shown that a certain concavity property of the new Rényi divergences, which the old ones don't have, make them a very convenient tool to prove the direct part of various coding theorems in composite/compound settings. This was demonstrated by giving short and simple proofs for the direct part of Stein's lemma with composite null-hypothesis and for classical-quantum channel coding with compound channels. Although the optimal rates for these problems have already been known [10, 11, 13, 31], the proofs in [27] are different from the previous ones, and offer considerable simplifications. The general approach is the following:

1. We start with a single-shot coding theorem that gives a trade-off relation between the relevant quantities of the problem in terms of Rényi divergences. For Stein's lemma, this is Audenaert's trace inequality [3], while for channel coding we use the Hayashi-Nagaoka random coding theorem from [17].
2. We then use general properties of the Rényi divergences to decouple the upper bounds from multiple to a single null-hypothesis/channel and to derive the asymptotics.

The main advantage of this approach is that the second step only relies on universal properties of the Rényi divergences and is largely independent of the concrete problem at hand. In particular, the coding theorems for the composite/compound settings can be obtained with the same amount of effort as for a simple null-hypothesis/single channel.

In this paper we present a variant for the proof of Stein's lemma with composite null-hypothesis. While in [27] exponential bounds on the error probabilities were given, here we study the asymptotics of the optimal type II error probability for a given threshold ε on the type I error probability. Building on results from [6] and [27], we derive finite-size bounds on the deviation of the optimal type II error from its asymptotic value. Such bounds are of practical importance, since in real-life scenarios one always works with finitely many copies.

The structure of the paper is as follows. Section 2 is a summary of notations. In Section 3 we review some properties of the quantum Rényi divergences, including two inequalities from [27]: Lemma 4, which gives quantitative bounds between the old and the new definitions of the quantum Rényi divergences, and Corollary 6, which shows that the convexity of the new Rényi divergence in its first argument can be complemented in the form of a weak quasi-concavity inquality. For readers' convenience, we include the proof of these inequalities. In Section 4 we prove the above mentioned finite-size version of Stein's lemma.

2 Notations

For a finite-dimensional Hilbert space $\mathcal{H}$, let $\mathcal{B}(\mathcal{H})_+$ denote the set of all non-zero positive semidefinite operators on $\mathcal{H}$, and let $\mathcal{S}(\mathcal{H}) := \{\rho \in \mathcal{B}(\mathcal{H})_+\,;\, \operatorname{Tr}\rho = 1\}$ be the set of all density operators (states) on $\mathcal{H}$.

We define the powers of a positive semidefinite operator A only on its support; that is, if $\lambda_1, \ldots, \lambda_r$ are the strictly positive eigenvalues of A, with corresponding spectral projections $P_1, \ldots, P_r$, then we define $A^\alpha := \sum_{i=1}^r \lambda_i^\alpha P_i$ for all $\alpha \in \mathbb{R}$. In particular, $A^0 = \sum_{i=1}^r P_i$ is the projection onto the support of A, and we use $A^0 \leq B^0$ as a shorthand for $\operatorname{supp} A \subseteq \operatorname{supp} B$.

By a *POVM (positive operator-valued measure)* T on a Hilbert space $\mathcal{H}$ we mean a map $T : \mathcal{Y} \to \mathcal{B}(\mathcal{H})$, where $\mathcal{Y}$ is some finite set, $T(y) \geq 0$ for all y, and $\sum_{y\in\mathcal{Y}} T(y) = I$. In particular, a binary POVM is a POVM with $\mathcal{Y} = \{0, 1\}$.

We denote the natural logarithm by log, and use the convention $\log 0 := -\infty$ and $\log +\infty := +\infty$.

3 Rényi divergences

For non-zero positive semidefinite operators ρ, σ, the *Rényi α-divergence* of ρ w.r.t. σ with parameter $\alpha \in (0, +\infty) \setminus \{1\}$ is traditionally defined as [34]

$$D_\alpha^{(\text{old})}(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log \operatorname{Tr}\rho^\alpha\sigma^{1-\alpha} - \frac{1}{\alpha-1}\log\operatorname{Tr}\rho, & \alpha \in (0,1) \text{ or } \rho^0 \le \sigma^0, \\ +\infty, & \text{otherwise.} \end{cases}$$

For the mathematical properties of $D_\alpha^{(\text{old})}$, see, e.g. [22, 25, 35]. Recently, a new notion of Rényi divergence has been introduced in [28, 41], defined as

$$D_\alpha^{(\text{new})}(\rho\|\sigma) := \begin{cases} \frac{1}{\alpha-1}\log \operatorname{Tr}\left(\sigma^{\frac{1-\alpha}{2\alpha}}\rho\sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha - \frac{1}{\alpha-1}\log\operatorname{Tr}\rho, & \alpha \in (0,1) \text{ or } \rho^0 \le \sigma^0, \\ +\infty, & \text{otherwise.} \end{cases}$$

For the mathematical properties of $D_\alpha^{(\text{new})}$, see, e.g. [7, 15, 26, 28, 41].

An easy calculation shows that for fixed ρ and σ, the function $\alpha \mapsto \log\operatorname{Tr}\rho^\alpha\sigma^{1-\alpha}$ is convex, which in turn yields immediately that $\alpha \mapsto D_\alpha^{(\text{old})}(\rho\|\sigma)$ is monotone increasing. Moreover, the limit at $\alpha = 1$ can be easily calculated as

$$D_1(\rho\|\sigma) := \lim_{\alpha\to 1} D_\alpha^{(\text{old})}(\rho\|\sigma) = \begin{cases} \frac{1}{\operatorname{Tr}\rho}\operatorname{Tr}\rho(\log\rho - \log\sigma), & \rho^0 \le \sigma^0, \\ +\infty, & \text{otherwise,} \end{cases} \tag{1}$$

where the latter expression is *Umegaki's relative entropy* [40]. The same limit relation for $D_\alpha^{(\text{new})}(\rho\|\sigma)$ has been shown in [28, Theorem 5]. The following Lemma, due to [37] and [38], complements the above monotonicity property around $\alpha = 1$, and in the same time gives a quantitative version of (1):

▶ **Lemma 1.** *Let $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$ be such that $\rho^0 \le \sigma^0$, let $\kappa := \log(1 + \operatorname{Tr}\rho^{3/2}\sigma^{-1/2} + \operatorname{Tr}\rho^{1/2}\sigma^{1/2})$, let $c > 0$, and $\delta := \min\left\{\frac{1}{2}, \frac{c}{2\kappa}\right\}$. Then*

$$D_1(\rho\|\sigma) \ge D_\alpha^{(\text{old})}(\rho\|\sigma) \ge D_1(\rho\|\sigma) - 4(1-\alpha)\kappa^2\cosh c, \qquad 1-\delta < \alpha < 1,$$

and the inequalities hold in the converse direction for $1 < \alpha < 1 + \delta$.

▶ **Remark 2.** *Assume that ρ and σ are states. The function $f(\alpha) := \operatorname{Tr}\rho^\alpha\sigma^{1-\alpha}$ is convex in α, and $\rho^0 \le \sigma^0$ implies that $f(1) = 1$. Hence, $\alpha \mapsto (f(\alpha)-1)/(\alpha-1)$ is monotone increasing. Comparing the values at $1/2$ and $3/2$, we see that $\operatorname{Tr}\rho^{3/2}\sigma^{-1/2} + \operatorname{Tr}\rho^{1/2}\sigma^{1/2} \ge 2$, and thus $\kappa > 1$.*

▶ **Remark 3.** *The Rényi entropy of a positive semidefinite operator $\rho \in \mathcal{B}(\mathcal{H})_+$ with parameter $\alpha \in (0, +\infty)$ is defined as*

$$S_\alpha(\rho) := -D_\alpha^{(\text{old})}(\rho\|I) = -D_\alpha^{(\text{new})}(\rho\|I) = \frac{1}{1-\alpha}\log\operatorname{Tr}\rho^\alpha - \frac{1}{1-\alpha}\log\operatorname{Tr}\rho.$$

By the above considerations, $\alpha \mapsto S_\alpha(\rho)$ is monotone decreasing, and comparing its values at α and at 0, we get

$$\operatorname{Tr}\rho^\alpha \le (\operatorname{Tr}\rho^0)^{(1-\alpha)}(\operatorname{Tr}\rho)^\alpha, \qquad \alpha \in (0,1). \tag{2}$$

According to the Araki-Lieb-Thirring inequality [2, 23], for any positive semidefinite operators A, B, $\operatorname{Tr} A^\alpha B^\alpha A^\alpha \leq \operatorname{Tr}(ABA)^\alpha$ for $\alpha \in (0,1)$, and the inequality holds in the converse direction for $\alpha > 1$. A converse to the Araki-Lieb-Thirring inequality was given in [5], where it was shown that $\operatorname{Tr}(ABA)^\alpha \leq \left(\|B\|^\alpha \operatorname{Tr} A^{2\alpha}\right)^{1-\alpha} (\operatorname{Tr} A^\alpha B^\alpha A^\alpha)^\alpha$ for $\alpha \in (0,1)$, and the inequality holds in the converse direction for $\alpha > 1$. Applying these inequalities to $A := \rho^{\frac{1}{2}}$ and $B := \sigma^{\frac{1-\alpha}{\alpha}}$, we get

$$\operatorname{Tr} \rho^\alpha \sigma^{1-\alpha} \leq \operatorname{Tr} \left(\rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \rho^{\frac{1}{2}}\right)^\alpha \leq \|\sigma\|^{(1-\alpha)^2} (\operatorname{Tr} \rho^\alpha)^{1-\alpha} \left(\operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}\right)^\alpha \tag{3}$$

for $\alpha \in (0,1)$, and the inequalities hold in the converse direction for $\alpha > 1$. In terms of the Rényi divergences, the above inequalities yield the ones in the following Lemma, the first of which has already been pointed out in [41] and [14].

► **Lemma 4.** *Let $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ be states. For any $\alpha \in (0, +\infty)$,*

$$D_\alpha^{(\mathrm{old})}(\rho\|\sigma) \geq D_\alpha^{(\mathrm{new})}(\rho\|\sigma) \geq \alpha D_\alpha^{(\mathrm{old})}(\rho\|\sigma) - |\alpha - 1| \log \dim \mathcal{H}. \tag{4}$$

Proof. The first inequality is immediate from the first inequality in (3). Taking into account (2), and that $\|\sigma\| \leq 1$, the second inequality in (3) yields the second inequality in (4) for $\alpha \in (0,1)$. For $\alpha > 1$, we have $\operatorname{Tr}(\rho/\|\rho\|)^\alpha \leq \operatorname{Tr}(\rho/\|\rho\|$, and hence we get $\operatorname{Tr}\left(\rho^{\frac{1}{2}} \sigma^{\frac{1-\alpha}{\alpha}} \rho^{\frac{1}{2}}\right)^\alpha \geq \|\sigma\|^{(1-\alpha)^2} \|\rho\|^{-(\alpha-1)^2} \left(\operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}\right)^\alpha$. Using that $\|\rho\| \leq 1$ and that $\|\sigma\| \geq 1/\dim \mathcal{H}$, we get the second inequality in (4) for $\alpha > 1$. ◄

For $\rho, \sigma \in \mathcal{B}(\mathcal{H})_+$, let

$$Q_\alpha^{(\mathrm{old})}(\rho\|\sigma) := \operatorname{Tr} \rho^\alpha \sigma^{1-\alpha}, \qquad Q_\alpha^{(\mathrm{new})}(\rho\|\sigma) := \operatorname{Tr} \left(\sigma^{\frac{1-\alpha}{2\alpha}} \rho \sigma^{\frac{1-\alpha}{2\alpha}}\right)^\alpha \tag{5}$$

be the core quantities of the Rényi divergences $D_\alpha^{(\mathrm{old})}$ and $D_\alpha^{(\mathrm{new})}$, respectively. $Q_\alpha^{(\mathrm{old})}$ is jointly concave in (ρ, σ) for $\alpha \in [0,1]$ (see [22, 35]) and jointly convex for $\alpha \in [1,2]$ (see [1, 35]). The general concavity result in [20, Theorem 2.1] implies as a special case that $Q_\alpha^{(\mathrm{new})}(\rho\|\sigma)$ is jointly concave in (ρ, σ) for $\alpha \in [1/2, 1)$. (See also [15] for a different proof of this). In [28, 41], joint convexity of $Q_\alpha^{(\mathrm{new})}$ was shown for $\alpha \in [1,2]$, which was later extended in [15], using a different proof method, to all $\alpha > 1$. These results are equivalent to the monotonicity of the Rényi divergences under completely positive trace-preserving maps, for $\alpha \in [0,2]$ in the case of $D_\alpha^{(\mathrm{old})}$, and for $\alpha \geq 1/2$ in the case of $D_\alpha^{(\mathrm{new})}$.

The next lemma shows that the concavity of $Q_\alpha^{(\mathrm{new})}$ in its first argument can be complemented by a subadditivity inequality for $\alpha \in (0,1)$:

► **Lemma 5.** *Let $\rho_1, \ldots, \rho_r \in \mathcal{S}(\mathcal{H})$ be states and $\sigma \in \mathcal{B}(\mathcal{H})_+$, and let $\gamma_1, \ldots, \gamma_r$ be a probability distribution. For every $\alpha \in (0,1)$,*

$$\sum_i \gamma_i Q_\alpha^{(\mathrm{new})}(\rho_i\|\sigma) \leq Q_\alpha^{(\mathrm{new})}\left(\sum_i \gamma_i \rho_i \middle\| \sigma\right) \leq \sum_i \gamma_i^\alpha Q_\alpha^{(\mathrm{new})}(\rho_i\|\sigma). \tag{6}$$

Proof. The function $x \mapsto x^\alpha$ is operator concave on $[0, +\infty)$ for $\alpha \in (0,1)$ (see Theorems V.1.9 and V.2.5 in [9]), from which the first inequality in (6) follows immediately. To prove the second inequality, we use a special case of the Rotfel'd inequality, for which we provide a proof below. First let $A, B \in \mathcal{B}(\mathcal{H})_+$ be invertible. Then

$$\begin{aligned}\operatorname{Tr}(A+B)^\alpha - \operatorname{Tr} A^\alpha &= \int_0^1 \frac{d}{dt} \operatorname{Tr}(A+tB)^\alpha \, dt = \int_0^1 \alpha \operatorname{Tr} B(A+tB)^{\alpha-1} \, dt \\ &\leq \int_0^1 \alpha \operatorname{Tr} B(tB)^{\alpha-1} \, dt = \operatorname{Tr} B^\alpha \int_0^1 \alpha t^{\alpha-1} \, dt = \operatorname{Tr} B^\alpha,\end{aligned} \tag{7}$$

where in the first line we used the identity $(d/dt)\operatorname{Tr} f(A+tB) = \operatorname{Tr} Bf'(A+tB)$, and the inequality follows from the fact that $x \mapsto x^{\alpha-1}$ is operator monotone decreasing on $(0,+\infty)$ for $\alpha \in (0,1)$. By continuity, we can drop the invertibility assumption, and (7) holds for any $A,B \in \mathcal{B}(\mathcal{H})_+$. Obviously, (7) extends to more than two operators, i.e., $\operatorname{Tr}(A_1+\ldots+A_r)^\alpha \le \operatorname{Tr} A_1^\alpha + \ldots + \operatorname{Tr} A_r^\alpha$ for any $A_, \ldots, A_r \in \mathcal{B}(\mathcal{H})_+$ and $\alpha \in (0,1)$. Choosing now $A_i := \sigma^{\frac{1-\alpha}{2\alpha}}\gamma_i\rho_i\sigma^{\frac{1-\alpha}{2\alpha}}$ yields the second inequality in (6). ◀

▶ **Corollary 6.** *Let $\rho_1,\ldots,\rho_r \in \mathcal{S}(\mathcal{H})$ be states and $\sigma \in \mathcal{B}(\mathcal{H})_+$, and let $\gamma_1,\ldots,\gamma_r$ be a probability distribution. For every $\alpha \in (0,1)$,*

$$\min_i D_\alpha^{(\mathrm{new})}(\rho_i\|\sigma) + \log\min_i \gamma_i \le D_\alpha^{(\mathrm{new})}\left(\sum_i \gamma_i\rho_i \middle\| \sigma\right) \le \sum_i \gamma_i D_\alpha^{(\mathrm{new})}(\rho_i\|\sigma).$$

Proof. Immediate from Lemma 5. ◀

4 Stein's lemma with composite null-hypothesis

In the general formulation of binary quantum hypothesis testing, we assume that for every $n \in \mathbb{N}$, a quantum system with Hilbert space $\mathcal{H}_n$ is given, together with two subsets $H_{0,n}$ and $H_{1,n}$ of the state space of $\mathcal{H}_n$, corresponding to the *null-hypothesis* and the *alternative hypothesis*, respectively. Our aim is to guess, based on a binary POVM, which set the true state of the system falls into. Here we consider the i.i.d. case with composite null-hypothesis and simple alternative hypothesis. That is, for every $n \in \mathbb{N}$, $\mathcal{H}_n = \mathcal{H}^{\otimes n}$ for some finite-dimensional Hilbert space $\mathcal{H}$; the null-hypothesis is represented by a set of states $\mathcal{N} \subseteq \mathcal{S}(\mathcal{H})$, and the alternative hypothesis is represented by a single state $\sigma \in \mathcal{S}(\mathcal{H})$. For every $n \in \mathbb{N}$, we have $H_{0,n} = \mathcal{N}^{(\otimes n)} := \{\rho^{\otimes n} : \rho \in \mathcal{N}\}$, and $H_{1,n} = \{\sigma^{\otimes n}\}$.

Given a binary POVM $T_n = (T_n(0), T_n(1))$, with $T_n(0)$ corresponding to accepting the null-hpothesis and $T_n(1)$ to accepting the alternative hypothesis, there are two possible ways of making an erroneous decision: accepting the alternative hypothesis when the null-hypothesis is true, called the type I error, or the other way around, called the type II error. The probabilities of these two errors are given by

$$\alpha_n(T_n) := \sup_{\rho\in\mathcal{N}} \operatorname{Tr}\rho^{\otimes n}T_n(1), \quad \text{(type I)} \qquad \text{and} \qquad \beta_n(T_n) := \operatorname{Tr}\sigma^{\otimes n}T_n(0), \quad \text{(type II)}.$$

Note that in the definition of α_n, we used a worst-case error probability.

In the setting of Stein's lemma, one's aim is to keep the type I error below a threshold ε, and to optimize the type II error under this condition. For any set $\mathcal{M} \subseteq \mathcal{S}(\mathcal{H}^{\otimes n})$ and any $\varepsilon \in (0,1)$, let

$$\beta_\varepsilon(\mathcal{M}\|\sigma^{\otimes n}) := \inf\left\{\operatorname{Tr}\sigma^{\otimes n}T_n(0) : \sup_{\omega\in\mathcal{M}} \operatorname{Tr}\omega T_n(1) \le \varepsilon\right\},$$

where the infimum is taken over all binary POVM T_n on $\mathcal{H}^{\otimes n}$. When $\mathcal{M}$ consists of one single element ω, we simply write $\beta_\varepsilon(\omega\|\sigma^{\otimes n})$. The quantum Stein's lemma states that

$$\lim_{n\to+\infty} \frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right) = -D_1(\mathcal{N}\|\sigma) := -\inf_{\rho\in\mathcal{N}} D_1(\rho\|\sigma). \tag{8}$$

This has been shown first in [19, 33] for the case where $\mathcal{N}$ consists of one single element ρ. Theorem 2 in [16] uses group representation techniques to give an approximation of the relative entropy in terms of post-measurement relative entropies, which, when combined with

Stein's lemma for probability distributions, yields (8) for finite $\mathcal{N}$. A direct proof for the case of infinite $\mathcal{N}$, also based on group representation theory, has recently been given in [31]. A version of Stein's lemma with infinite $\mathcal{N}$ has been previously proved in [10], however, with a weaker error criterion.

Here we give a different proof of the quantum Stein's lemma with possibly infinite composite null-hypothesis. Our proof is based on the results of [6], where bounds on β_ε were obtained in terms of Rényi divergences, and general properties of the Rényi divergences from Section 3. Moreover, we give a refined version of (8) in Theorem 9 by providing finite-size corrections to the deviation of $\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)$ from its asymptotic value $-D_1(\mathcal{N}\|\sigma)$ for every $n\in\mathbb{N}$.

We will need the following results from [6]:

▶ **Lemma 7.** *Let $\rho,\sigma\in\mathcal{S}(\mathcal{H})$. For every $\varepsilon\in(0,1)$ and every $\alpha\in(0,1)$,*

$$\log\beta_\varepsilon(\rho\|\sigma)\le -D_\alpha^{(\mathrm{old})}(\rho\|\sigma)+\frac{\alpha}{1-\alpha}\log\varepsilon^{-1}-\frac{h_2(\alpha)}{1-\alpha}, \tag{9}$$

where $h_2(\alpha):=-\alpha\log\alpha-(1-\alpha)\log(1-\alpha)$ is the binary entropy function. Moreover, for every $n\in\mathbb{N}$,

$$\frac{1}{n}\log\beta_\varepsilon\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right)\ge -D_1(\rho\|\sigma)-\frac{1}{\sqrt{n}}4\sqrt{2}\kappa\log(1-\varepsilon)^{-1}, \tag{10}$$

where κ is given in Lemma 1.

Proof. The upper bound (9) is due to [6, Proposition 3.2], while the lower bound in (10) is formula (19) in [6, Theorem 3.3]. ◀

When $\mathcal{N}$ is infinite, we will need the following approximation lemma, which is a special case of [24, Lemma 2.6]:

▶ **Lemma 8.** *For every $\delta>0$, let $\mathcal{N}_\delta\subset\mathcal{N}$ be a set of minimal cardinality such that $\sup_{\rho\in\mathcal{N}}\inf_{\rho'\in\mathcal{N}_\delta}\|\rho-\rho'\|_1\le\delta$. Then $|\mathcal{N}_\delta|\le\min\{|\mathcal{N}|,(1+2\delta^{-1})^D\}$, where $D=(\dim\mathcal{H}+1)(\dim\mathcal{H})/2$, and*

$$\sup_{\rho\in\mathcal{N}}\inf_{\rho'\in\mathcal{N}_\delta}\left\|\rho^{\otimes n}-(\rho')^{\otimes n}\right\|_1\le n\sup_{\rho\in\mathcal{N}}\inf_{\rho'\in\mathcal{N}_\delta}\|\rho-\rho'\|_1\le n\delta,\qquad n\in\mathbb{N}. \tag{11}$$

Now we are ready to prove our main result:

▶ **Theorem 9.** *Let $\varepsilon\in(0,1)$, and for every $n\in\mathbb{N}$, let $0\le\delta_n\le\varepsilon/(2n)$. Then*

$$\begin{aligned}\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)\le&-D_1(\mathcal{N}\|\sigma)\\&+\sqrt{\frac{\log\left(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1}\right)}{n}}\cdot 2\left[8\kappa_{\max}^2+\log\dim\mathcal{H}+D_1(\mathcal{N}\|\sigma)\right]^{\frac{1}{2}}\\&+\frac{\log\left(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1}\right)}{n}\cdot 4\kappa_{\max},\end{aligned} \tag{12}$$

$$\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)\ge -D_1(\mathcal{N}\|\sigma)-\frac{1}{\sqrt{n}}4\sqrt{2}\log(1-\varepsilon)^{-1}\kappa_{\max}, \tag{13}$$

where $\kappa_{\max}:=\sup_{\rho\in\mathcal{N}}\{\log(1+\operatorname{Tr}\rho^{3/2}\sigma^{-1/2}+\operatorname{Tr}\rho^{1/2}\sigma^{1/2})\}\le\log(2+\operatorname{Tr}\sigma^{-1/2})<+\infty$.

In (12), the slowest decaying term after $-D_1(\mathcal{N}\|\sigma)$ is of the order $1/\sqrt{n}$ when $\mathcal{N}$ is finite, and when $\mathcal{N}$ is infinite, it can be chosen to be of the order $\sqrt{\frac{\log n}{n}}$.

Proof. The lower bound in (13) is immediate from (10), and hence we only have to prove (12). We have

$$
\begin{aligned}
\log \beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right) &\le \log \beta_{\varepsilon-n\delta_n}\left(\mathcal{N}_{\delta_n}^{(\otimes n)}\|\sigma^{\otimes n}\right) \le \log \beta_{\frac{\varepsilon-n\delta_n}{|\mathcal{N}_{\delta_n}|}}\left(\sum_{\rho\in\mathcal{N}_{\delta_n}} \frac{1}{|\mathcal{N}_{\delta_n}|}\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right)\\
&\le -D_\alpha^{(\mathrm{old})}\left(\sum_{\rho\in\mathcal{N}_{\delta_n}} \frac{1}{|\mathcal{N}_{\delta_n}|}\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) + \frac{\alpha}{1-\alpha}\log\frac{|\mathcal{N}_{\delta_n}|}{\varepsilon-n\delta_n}\\
&\le -D_\alpha^{(\mathrm{new})}\left(\sum_{\rho\in\mathcal{N}_{\delta_n}} \frac{1}{|\mathcal{N}_{\delta_n}|}\rho^{\otimes n}\middle\|\sigma^{\otimes n}\right) + \frac{\alpha}{1-\alpha}\log\frac{|\mathcal{N}_{\delta_n}|}{\varepsilon-n\delta_n},
\end{aligned}
$$

where the first inequality is due to (11), the second inequality is obvious, the third one follows from (9), and the last one is due to Lemma 4. Note that $\varepsilon-n\delta_n\ge\varepsilon/2$ by assumption. Using Corollary 6, we can continue the above upper bound as

$$
\begin{aligned}
&\log \beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)\\
&\quad\le -\min_{\rho\in\mathcal{N}_{\delta_n}} D_\alpha^{(\mathrm{new})}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) + \log|\mathcal{N}_{\delta_n}| + +\frac{\alpha}{1-\alpha}\log|\mathcal{N}_{\delta_n}| + \frac{\alpha}{1-\alpha}\log\frac{2}{\varepsilon}\\
&\quad\le -n\inf_{\rho\in\mathcal{N}} D_\alpha^{(\mathrm{new})}\left(\rho\|\sigma\right) + \frac{1}{1-\alpha}\log|\mathcal{N}_{\delta_n}| + \frac{1}{1-\alpha}\log\frac{2}{\varepsilon},
\end{aligned}
$$

where in the last line we used the additivity property $D_\alpha^{(\mathrm{new})}\left(\rho^{\otimes n}\|\sigma^{\otimes n}\right) = nD_\alpha^{(\mathrm{new})}\left(\rho\|\sigma\right)$.

By Lemmas 4 and 1, for every $\alpha\in(1/2,1)$ such that $\alpha>1-\frac{c}{2\kappa_{\max}}$,

$$
\begin{aligned}
\inf_{\rho\in\mathcal{N}} D_\alpha^{(\mathrm{new})}\left(\rho\|\sigma\right) &\ge \alpha\inf_{\rho\in\mathcal{N}} D_\alpha^{(\mathrm{old})}\left(\rho\|\sigma\right) - (1-\alpha)\log\dim\mathcal{H}\\
&\ge \alpha\inf_{\rho\in\mathcal{N}} D_1\left(\rho\|\sigma\right) - 4\alpha(1-\alpha)\kappa_{\max}^2\cosh c - (1-\alpha)\log\dim\mathcal{H},
\end{aligned}
$$

where c is an arbitrary positive constant. Now choose $\alpha := 1-a/\sqrt{n}$. Then

$$
\begin{aligned}
\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right) \le -\left(1-\frac{a}{\sqrt{n}}\right)D_1(\mathcal{N}\|\sigma) + \frac{a}{\sqrt{n}}\left(4\kappa_{\max}^2\cosh c+\log\dim\mathcal{H}\right)\\
+\frac{1}{a\sqrt{n}}\left(\log|\mathcal{N}_{\delta_n}|+\log\frac{2}{\varepsilon}\right).
\end{aligned}
$$

Optimizing over a yields

$$
\begin{aligned}
&\frac{1}{n}\log\beta_\varepsilon\left(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}\right)\\
&\quad\le -D_1(\mathcal{N}\|\sigma) + \frac{2}{\sqrt{n}}\left[4\kappa_{\max}^2\cosh c+\log\dim\mathcal{H}+D_1(\mathcal{N}\|\sigma)\right]^{\frac{1}{2}}\cdot\left[\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1})\right]^{\frac{1}{2}}.
\end{aligned}
\tag{14}
$$

The optimum is reached at

$$
a^* = \left[\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1})\right]^{\frac{1}{2}}\cdot\left[4\kappa_{\max}^2\cosh c+\log\dim\mathcal{H}+D_1(\mathcal{N}\|\sigma)\right]^{-\frac{1}{2}},
$$

and we need $a^*/\sqrt{n}\le 1/2$ and $a^*/\sqrt{n}\le c/(2\kappa_{\max})$, which is satisfied if

$$
\kappa_{\max}^2\cosh c \ge \frac{1}{n}\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1}) \qquad\text{and}\qquad c^2\cosh c\ge\frac{1}{n}\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1}).
$$

Let us choose $c > 0$ such that $\cosh c = 2 + \frac{1}{n}\log(2|\mathcal{N}_{\delta_n}|\varepsilon^{-1})$. By Remark 2, $\kappa_{\max} > 1$, and hence the first inequality is satisfied. Moreover, with this choice $c > 1$, and thus the second inequality is satisfied as well.

Substituting this choice of c into (14), and using the subadditivity of the square root, we get (12).

When $\mathcal{N}$ is finite, we can choose $\delta_n = 0$, and hence $\mathcal{N}_{\delta_n} = \mathcal{N}$, for every n. This shows that the second term in (12) is of the order $1/\sqrt{n}$, while the third term is of the order $1/n$. When $\mathcal{N}$ is infinite, we can choose $\delta_n = \varepsilon/(2n^2)$, whence the order of the second term in (12) is $\sqrt{\frac{\log n}{n}}$, and the order of the third term is $\frac{\log n}{n}$. ◀

▶ **Remark 10.** *In the case of a simple null-hypothesis $\mathcal{N} = \{\rho\}$, the limit*

$$\lim_{n\to+\infty} \sqrt{n}\left(\frac{1}{n}\log\beta_\varepsilon(\mathcal{N}^{(\otimes n)}\|\sigma^{\otimes n}) + D_1(\mathcal{N}\|\sigma)\right), \tag{15}$$

called the second-order asymptotics, has been determined in [21, 39]. Their results show that the finite-size bounds of [6] are not asymptotically optimal, and hence the same holds for the bounds in Theorem 9. The merit of these latter results, on the other hand, is that the correction terms are easily computable, and the bounds are valid for any finite n. To the best of our knowledge, the value of the limit (15) *has not yet been determined when* $|\mathcal{N}| > 1$, *and our bounds in Theorem 9 give bounds on the second-order asymptotics in this case.*

Acknowledgements. The author is grateful to Professor Fumio Hiai and Nilanjana Datta for discussions.

References

1 T. Ando. *Concavity of certain maps and positive definite matrices and applications to Hadamard products*. Linear Algebra Appl. **26**, 203–241 1979

2 H. Araki. *On an inequality of Lieb and Thirring*. Letters in Mathematical Physics, Volume 19, Issue 2, pp. 167–170, 1990

3 K.M.R. Audenaert, J. Calsamiglia, Ll. Masanes, R. Munoz-Tapia, A. Acin, E. Bagan, F. Verstraete. *Discriminating states: the quantum Chernoff bound*. Phys. Rev. Lett. **98** 160501, 2007

4 K.M.R. Audenaert, M. Nussbaum, A. Szkoła, F. Verstraete. *Asymptotic error rates in quantum hypothesis testing*. Commun. Math. Phys. **279**, 251–283, 2008

5 K.M.R. Audenaert. *On the Araki-Lieb-Thirring inequality*. Int. J. of Information and Systems Sciences **4**, pp. 78–83, 2008)

6 Koenraad M.R. Audenaert, Milan Mosonyi, Frank Verstraete. *Quantum state discrimination bounds for finite sample size*. J. Math. Phys. **53**, 122205, 2012

7 Salman Beigi. *Quantum Rényi divergence satisfies data processing inequality*. J. Math. Phys., **54**, 122202, 2013

8 Salman Beigi, Amin Gohari. *Quantum Achievability Proof via Collision Relative Entropy*. arXiv:1312.3822, 2013

9 R. Bhatia. *Matrix Analysis*. Graduate Texts in Mathematics **169**, Springer, 1997

10 I. Bjelakovic, J.-D. Deuschel, T. Krüger, R. Seiler, R. Siegmund-Schultze, A. Szkoła. *A quantum version of Sanov's theorem*. Commun. Math. Phys. **260**, pp. 659–671, 2005

11 I. Bjelakovic, H. Boche. *Classical capacities of compound and averaged quantum channels*. IEEE Trans. Inform. Theory **55**, 3360–3374, 2009

12 I. Csiszár. *Generalized cutoff rates and Rényi's information measures*. IEEE Trans. Inf. Theory **41**, 26–34, 1995

13 N. Datta, T.C. Dorlas. *The Coding Theorem for a Class of Quantum Channels with Long-Term Memory.* Journal of Physics A: Mathematical and Theoretical, vol. 40, 8147, 2007

14 Nilanjana Datta and Felix Leditzky. *A limit of the quantum Rényi divergence.* J. Phys. A: Math. Theor. **47** 045304, 2014

15 Rupert L. Frank and Elliott H. Lieb. *Monotonicity of a relative Rényi entropy.* J. Math. Phys. 54 , 122201, 2013

16 Masahito Hayashi. *Asymptotics of quantum relative entropy from a representation theoretical viewpoint.* J. Phys. A: Math. Gen. **34** 3413, (2001)

17 M. Hayashi, H. Nagaoka. *General Formulas for Capacity of Classical-Quantum Channels.* IEEE Trans. Inf. Theory **49**, 2003

18 M. Hayashi. *Error exponent in asymmetric quantum hypothesis testing and its application to classical-quantum channel coding.* Phys. Rev. A **76**, 062301, 2007

19 F. Hiai, D. Petz. *The proper formula for relative entropy and its asymptotics in quantum probability.* Comm. Math. Phys. **143**, 99–114, 1991

20 F. Hiai. *Concavity of certain matrix trace and norm functions.* Linear Algebra and Appl. **439**, 1568–1589, 2013

21 Ke Li. *Second-order asymptotics for quantum hypothesis testing.* Annals of Statistics, Vol. 42, No. 1, pp. 171–189, 2014

22 E.H. Lieb. *Convex trace functions and the Wigner-Yanase-Dyson conjecture.* Adv. Math. **11**, 267–288, 1973

23 E.H. Lieb, W. Thirring. *Studies in mathematical physics.* pp. 269–297. Princeton University Press, Princeton, 1976

24 Vitali D. Milman, Gideon Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces.* Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg, 1986

25 M. Mosonyi, F. Hiai. *On the quantum Rényi relative entropies and related capacity formulas.* IEEE Trans. Inf. Theory, **57**, 2474–2487, 2011

26 M. Mosonyi, T. Ogawa. *Quantum hypothesis testing and the operational interpretation of the quantum Rényi relative entropies.* arXiv:1308.3228, 2013

27 M. Mosonyi. *Inequalities for the quantum Rényi divergences with applications to compound coding problems.* arXiv:1310.7525; submitted to IEEE Transactions on Information Theory

28 M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr, M. Tomamichel. *On quantum Renyi entropies: a new definition and some properties.* J. Math. Phys. **54**, 122203, 2013

29 H. Nagaoka. *Strong converse theorems in quantum information theory.* in the book "Asymptotic Theory of Quantum Statistical Inference" edited by M. Hayashi, World Scientific, 2005

30 H. Nagaoka. *The converse part of the theorem for quantum Hoeffding bound.* quant-ph/0611289, 2006

31 J. Nötzel. *Hypothesis testing on invariant subspaces of the symmetric group, part I - quantum Sanov's theorem and arbitrarily varying sources.* arXiv:1310.5553, 2013

32 T. Ogawa, H. Nagaoka. *Strong converse to the quantum channel coding theorem.* IEEE Transactions on Information Theory, vol. 45, no. 7, pp. 2486-2489, 1999

33 T. Ogawa, H. Nagaoka. *Strong converse and Stein's lemma in quantum hypothesis testing.* IEEE Trans. Inform. Theory **47**, 2428–2433, 2000

34 M. Ohya, D. Petz. *Quantum Entropy and its Use.* Springer, 1993

35 D. Petz. *Quasi-entropies for finite quantum systems.* Rep. Math. Phys. **23**, 57–65, 1986

36 A. Rényi. *On measures of entropy and information.* Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, pp. 547–561, Univ. California Press, Berkeley, California, 1961

37 M. Tomamichel, R. Colbeck, R. Renner. *A fully quantum asymptotic equipartition property.* IEEE Trans. Inform. Theory **55**, 5840–5847, 2009

38 M. Tomamichel. *A framework for non-asymptotic quantum information theory*. PhD thesis, ETH Zürich, 2012

39 M. Tomamichel, M. Hayashi. *A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks*. IEEE Transactions on Information Theory **59**, pp. 7693–7710, 2013

40 H. Umegaki. *Conditional expectation in an operator algebra*. Kodai Math. Sem. Rep. **14**, 59–85, 1962

41 Mark M. Wilde, Andreas Winter, Dong Yang. *Strong converse for the classical capacity of entanglement-breaking and Hadamard channels*. arXiv:1306.1586, 2013

Quantum Learning of Classical Stochastic Processes: The Completely-Positive Realization Problem

Alex Monras*[1] and Andreas Winter†[1,2]

1 Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain

2 ICREA – Institució Catalana de Recerca i Estudis Avançats, Pg. Lluis Companys 23, ES-08010 Barcelona, Spain

Abstract

Among several tasks in Machine Learning, is the problem of inferring the latent variables of a system and their causal relations with the observed behavior. A paradigmatic instance of such problem is the task of inferring the Hidden Markov Model underlying a given stochastic process. This is known as the positive realization problem (PRP) [3] and constitutes a central problem in machine learning. The PRP and its solutions have far-reaching consequences in many areas of systems and control theory, and is nowadays an important piece in the broad field of positive systems theory [21].

We consider the scenario where the latent variables are quantum (e. g., quantum states of a finite-dimensional system), and the system dynamics is constrained only by physical transformations on the quantum system. The observable dynamics is then described by a quantum instrument, and the task is to determine which quantum instrument – if any – yields the process at hand by iterative application.

We take as a starting point the theory of quasi-realizations, whence a description of the dynamics of the process is given in terms of linear maps on state vectors and probabilities are given by linear functionals on the state vectors. This description, despite its remarkable resemblance with the Hidden Markov Model, or the iterated quantum instrument, is however devoid from any stochastic or quantum mechanical interpretation, as said maps fail to satisfy any positivity conditions. The Completely-Positive realization problem then consists in determining whether an equivalent quantum mechanical description of the same process exists.

We generalize some key results of stochastic realization theory, and show that the problem has deep connections with operator systems theory, giving possible insight to the lifting problem in quotient operator systems. Our results have potential applications in quantum machine learning, device-independent characterization and reverse-engineering of stochastic processes and quantum processors, and more generally, of dynamical processes with quantum memory [16, 17].

1998 ACM Subject Classification G.3 Probability and statistics

Keywords and phrases quantum instrument, hidden Markov model, machine learning, quantum measurement

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.99

* This research was partly undertaken while A. M. was at the Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117542, Singapore.

† This research was partly undertaken while A. W. was at the Department of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom.

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 99–109
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Let $\mathcal{M}$ be an alphabet with $|\mathcal{M}| = m$ symbols and let $\mathcal{M}^\ell$ be the set of words of length ℓ. Let $\mathcal{M}^*$ be the free monoid generated by $\mathcal{M}$

$$\mathcal{M}^* = \bigcup_{\ell \geq 0} \mathcal{M}^\ell. \tag{1}$$

We will be concerned with stochastic processes defined on sequences of random variables over $\mathcal{M}$, i. e., stationary probability measures over $\mathcal{M}^*$. We assume throughout that p is a stationary stochastic process on $\mathcal{M}^\infty$, namely,

$$p(\mathbf{u}) \equiv p(\mathcal{Y}_t = u_1, \mathcal{Y}_{t+1} = u_2, \ldots, \mathcal{Y}_{t+\ell-1} = u_\ell), \quad \mathbf{u} = (u_1, \ldots, u_\ell) \in \mathcal{M}^\ell \tag{2}$$

is independent of t. We will use ℓ to denote a generic length of a word $\mathbf{u}$, so that $\mathbf{u}$ can be written as $\mathbf{u} = (u_1, \ldots, u_\ell)$ instead of the more cumbersome $\mathbf{u} = (u_1, \ldots, u_{|\mathbf{u}|})$. Let p be a stationary stochastic process defined on the alphabet $\mathcal{M}$.

▶ **Definition 1.** A quasi-realization of a stochastic process is a quadruple $(\mathcal{V}, \pi, D, \tau)$ where $\mathcal{V}$ is a vector space, $\tau \in \mathcal{V}$, $\pi \in \mathcal{V}^*$ and $D : \mathcal{M}^* \to \mathcal{L}(\mathcal{V})$ is a representation of $\mathcal{M}^*$ over $\mathcal{V}$,

$$D^{(\mathbf{u})} D^{(\mathbf{v})} = D^{(\mathbf{uv})}, \quad \forall \mathbf{u}, \mathbf{v} \in \mathcal{M}^*. \tag{3}$$

In addition, the following relations hold,

$$\pi^\top \left[\sum_{u \in \mathcal{M}} D^{(u)} \right] = \pi^\top, \qquad \left[\sum_{u \in \mathcal{M}} D^{(u)} \right] \tau = \tau \tag{4}$$

and

$$p(\mathbf{u}) = \pi^\top D^{(\mathbf{u})} \tau \quad \forall \mathbf{u} \in \mathcal{M}^*. \tag{5}$$

The smallest dimensional quasi-realization admitted by p is called *regular* realization, and its dimension is the *order* of p. The regular realization is efficiently computable given the probabilities of words of length $2r - 1$, where r is the order of p [10, 24].

2 The classical learning problem

A central task in machine learning is to obtain the latent variables that account for the apparent complexity of a given process p. These variables, although not directly accessible to the observable dynamics summarize past behavior while still providing complete information about future probabilities of events. To accomplish this, one aims to find a random variable X such that the future is independent of the past, given X,

$$p(\mathbf{v}|\mathbf{u}) = \sum_X P(\mathbf{v}|X) P(X|\mathbf{u}). \tag{6}$$

However, such a decomposition to exist at any given time we require that state transition probabilities are only dependent on the generated output, $P(X_t, u_t|X_{t-1})$ in a time-invariant manner. This implies that X is markovian, and we say that p is a Hidden Markov Process. In such case, $\{X_t\}$ represents the latent variables of p, and an important problem in machine learning consists in recovering the probabilities $P(X_t, u_t|X_{t-1})$.

A process' quasi-realization constitutes an abstract model of the behavior of p. However this does not suffice to identify its latent variables, as the vector $\pi D^{(\mathbf{u})}$ does not necessarily satisfy any positivity criterion, and the maps need not be related to any stochastic transition probabilities. Moreover, the vectors $\pi D^{(\mathbf{u})}$ will potentially acquire an unbounded number of distinct values over $\mathcal{V}$, giving little insight on the essential mechanisms driving p.

A *positive* realization of p is a quasi-realization $(\mathcal{V}, \pi, D, \tau)$, such that $D^{(\mathbf{u})}$ are substochastic matrices (nonnegative matrices such that $\sum_{u\in\mathcal{M}} D^{(u)}$ is stochastic), π is the stationary distribution, and $\pi = (1, 1, \ldots, 1)$. **The Positive Realization Problem** (PRP) is the problem of finding a positive realization of a process p, given its regular realization [24].

3 The quantum learning problem

We address the natural quantum generalization of this problem, namely, when the relevant information about the past can be synthesized by a quantum state, rather than a classical random variable. This requirement, less impositive than the classical one [22], has been considered from the perspective of ϵ-machines [15], where it was shown that the statistical complexity of the system could be reduced by a quantum model. Instead, our approach focuses on the dimension of the quantum system, which can be drastically reduced once one allows for quantum states. A highly relevant example in a not too distant scenario can be found in [25].

In the quantum mechanical setting, the factorization condition Eq. (6) is replaced by

$$p(\mathbf{v}|\mathbf{u}) = \rho_{\mathbf{u}}[M^{(\mathbf{v})}], \tag{7}$$

where $\rho_{\mathbf{u}}$ represents a quantum state, and $M^{(\mathbf{v})}$ the POVM element associated with outcome $\mathbf{v}$. Future probabilities are obtained by the Born rule applied to state $\rho_{\mathbf{u}}$. The minimum dimension by which this description can be achieved is given by the positive semidefinite rank [13]. However, in addition, in order to have a physically meaningful description of the mechanisms at work, one expects that the state transition probabilities are given by physical transformations,

$$\rho_{\mathbf{u}v} = \rho_{\mathbf{u}} \circ \mathcal{E}^{(v)}, \tag{8}$$

where $\mathcal{E}^{(v)}$ are completely-positive maps, and $\sum_{v\in\mathcal{M}} \mathcal{E}^{(v)}$ is unital. The set $\{\mathcal{E}^{(v)}\}$ is called a *quantum instrument*. This problem has received little attention in the literature. It arises naturally – albeit in slight disguise – in [5], and more generally in systems identification [6, 17, 2, 7].

The completely positive realization problem (CPRP): *Given a quasi-realization of process $p(\boldsymbol{u})$, determine whether there exist a quantum instrument $\{\mathcal{E}^{(u)}\}$, and positive semidefinite ρ such that*

$$p(\boldsymbol{u}) = \rho[\mathcal{E}^{(u_1)} \circ \cdots \circ \mathcal{E}^{(u_\ell)}(\mathcal{I})], \tag{9}$$

such that $\mathcal{E} = \sum_u \mathcal{E}^{(u)}$ is completely positive and unital, and $\rho \circ \mathcal{E} = \rho$. Stochastic processes admitting a completely-positive realization are called finitely correlated or algebraic [1, 11].

In order to obtain necessary and sufficient conditions for p to be a finitely correlated process, we first generalize a classical result by Ito, Amari and Kobayashi [18]. The latter is the stochastic equivalent to a classic result on linear systems theory [19], *i. e.*, minimal realizations are always related by similarity transformations, and are quotients of higher-dimensional ones.

Define the $\mathcal{W} = \operatorname{span}\{\mathcal{E}^{(\mathbf{u})}(\mathcal{I})\}_{\mathbf{u}\in\mathcal{M}^*}$ as the *accessible* subspace. It is trivially stable under the action of $\mathcal{E}^{(\mathbf{u})}$, $\forall \mathbf{u} \in \mathcal{M}^*$,

$$\mathcal{E}^{(\mathbf{u})}(\mathcal{W}) \subseteq \mathcal{W}. \tag{10}$$

Analogously, we consider the span of states $\widetilde{\mathcal{W}} = \operatorname{span}\{\rho \circ \mathcal{E}^{(\mathbf{u})}\}_{\mathbf{u}\in\mathcal{M}^*}$. Its annihilator, $\widetilde{\mathcal{W}}^{\perp} = \bigcap_{\sigma\in\widetilde{\mathcal{W}}} \ker\sigma$, is the *null* space, *i. e.*, the subspace which has no effect whatsoever for computing word probabilities. Also $\widetilde{\mathcal{W}}^{\perp}$ is stable under $\mathcal{E}^{(\mathbf{u})}$, $\forall \mathbf{u} \in \mathcal{M}^*$,

$$\mathcal{E}^{(\mathbf{u})}(\widetilde{\mathcal{W}}^{\perp}) \subseteq \widetilde{\mathcal{W}}^{\perp}. \tag{11}$$

Define the *quotient* space $\mathcal{V}$ as the accessible space modulo its null component $K = \mathcal{W} \cap \widetilde{\mathcal{W}}^{\perp}$,

$$\mathcal{V} \equiv \frac{\mathcal{W}}{K}. \tag{12}$$

The elements of $\mathcal{V}$ are of the form $a + K$, $a \in \mathcal{W}$. Let $L : \mathcal{W} \to \mathcal{V}$ be the canonical projection onto $\mathcal{V}$,

$$\begin{array}{rccc} L: & \mathcal{W} & \to & \mathcal{V} \\ & v & \mapsto & v + K. \end{array} \tag{13}$$

Since $\mathcal{E}^{(\mathbf{u})}(K) \subseteq (K)$ let $\mathcal{D}$ be the induced quotient map $\mathcal{D}^{(\mathbf{u})} : \mathcal{V} \to \mathcal{V}$, as defined by $\mathcal{D}^{(\mathbf{u})} \circ L = L \circ \mathcal{E}^{(\mathbf{u})}$. Also, define $\tau = L(\mathcal{I})$ and π as the induced quotient functional $\pi \circ L = \rho$. Using the fact that $\rho[\ker L] = 0$ we factor through the entire set of maps $\mathcal{E}^{(u)}$,

$$\begin{aligned} p(\mathbf{u}) &= \rho \circ \mathcal{E}^{(\mathbf{u})}(\mathcal{I}) && \text{(14a)} \\ &= \pi \circ L \circ \mathcal{E}^{(\mathbf{u})}(\mathcal{I}) && \text{(14b)} \\ &= \pi \circ \mathcal{D}^{(\mathbf{u})}(\tau). && \text{(14c)} \end{aligned}$$

This, together with easily shown eigenvector relations (4) illustrate that $(\mathcal{V}, \pi, \mathcal{D}^{(u)}, \tau)$ constitute a perfectly valid quasi-realization. We call such quasi-realization the *quotient* realization. An important step is to realize that the quotient spaces of equivalent quasi-realizations are minimal and hence isomorphic.

▶ **Theorem 2.** *[18] Two quasi-realizations $\mathcal{R}_1 = (\mathcal{V}_1, \pi_1, D_1^{(u)}, \tau_1)$ and $\mathcal{R}_2 = (\mathcal{V}_2, \pi_2, D_2^{(u)}, \tau_2)$ of the same stochastic process p, not necessarily of the same dimension, have isomorphic quotient realizations $\overline{\mathcal{R}}_i = (\overline{\mathcal{V}}_i, \overline{\pi}_i, \overline{D}_i^{(u)}, \overline{\tau}_i)_{i=1,2}$, $\bar{\mathcal{V}}_1 \overset{T}{\cong} \bar{\mathcal{V}}_2$,*

$$\overline{\pi}_1^{\top} = \overline{\pi}_2^{\top} T, \tag{15}$$

$$\overline{\mathcal{D}}_1^{(u)} = T^{-1} \overline{\mathcal{D}}_2^{(u)} T, \tag{16}$$

$$\overline{\tau}_1 = T^{-1} \overline{\tau}_2. \tag{17}$$

This result follows from [18], which proves it only for the Hidden Markov Model case. The proof, however, only relies on the nonnegativity of the process' probabilities, and applies to any pair of equivalent and well-defined (in the sense that they yield the same nonnegative measure on $\mathcal{M}^*$) quasi-realizations.

This result is important in that it establishes the uniqueness of the quotient space $\mathcal{V}$, up to basis changes. Let d be the dimension of $\mathcal{W}$. As can be seen from the definition $d = \dim \mathcal{V} \leq n$, where n is the original realization's dimension. By considering the quotient of a regular realization of dimension r we get $d \leq r$. On the other hand r is a lower bound to the dimension of any quasi-realization. Thus we conclude that $d = r$, hence quotient realizations are indeed regular, and all regular realizations can be regarded as quotient realizations.

4 Semidefinite representable cones and quotient operator systems

The CPRP aims at providing a completely-positive lifting of a regular realization $\mathcal{R} = (\pi, D^{(u)}, \tau)$. As it will be shown, a necessary and sufficient condition is the existence of certain stable cones of a particular kind, containing the vector τ, and whose dual contains π. We focus on finite-dimensional liftings from an r-dimensional regular realization $\mathcal{R}$ acting on $\mathcal{V} \cong \mathbb{R}^r$ to a completely positive realization acting on $\mathcal{B}(\mathcal{H})$ where $\mathcal{H}$ is a finite-dimensional Hilbert space, $\mathcal{H} = \mathbb{C}^n$. We use S^+ to denote the positive semidefinite cone in $\mathcal{B}(\mathcal{H})$. All cones we deal with are convex. A cone $\mathcal{C}$ is pointed iff $x \in \mathcal{C}$ and $-x \in \mathcal{C}$ implies $x = 0$ and $\mathcal{C}$ is generating if $\operatorname{span} \mathcal{C} = \mathcal{V}$. We will use calligraphic letters for subspaces of $\mathcal{B}(\mathcal{H})$, and for any given subspace $\mathcal{W}$, $\mathcal{W}^+$ will denote its intersection with S^+, $\mathcal{W}^+ = \mathcal{W} \cap \mathrm{S}^+$.

▶ **Definition 3.** Let $\mathcal{V}$ be a finite dimensional real vector space. A *semidefinite representable cone* (SDR) is a set $\mathcal{C} \in \mathcal{V}$ such that

$$\mathcal{C} = L(\mathcal{W}^+) \tag{18}$$

where $\mathcal{W} \subseteq \mathcal{B}(\mathcal{H})$ is a subspace and $L : \mathcal{W} \to \mathcal{V}$ is a linear map.

It is easy to see that pointed and generating SDR cones can always be described by subspaces $\mathcal{W}$ such that $\mathcal{W} = \operatorname{span}(\mathcal{W}^+)$ and L is a quotient map from $\mathcal{W}$ to $\mathcal{W}/K \cong \mathcal{V}$, with $K \cap \mathrm{S}^+ = \{0\}$. SDR cones are homogeneous versions of semidefinite representable sets, the feasibility regions of semidefinite programs [4].

▶ **Lemma 4.** *Let $\mathcal{I} \in \mathcal{W} \subseteq \mathcal{B}(\mathcal{H})$ and $\widetilde{\mathcal{W}} \subseteq \mathcal{B}(\mathcal{H})^*$, such that $\mathcal{W} = \operatorname{span}(\mathcal{W}^+)$ and $K = \mathcal{W} \cap \widetilde{\mathcal{W}}^\perp$ satisfies $K \cap \mathrm{S}^+ = \{0\}$. Let L be the canonical projection $L : \mathcal{W} \to \mathcal{W}/K$. Then $\mathcal{C} = L(\mathcal{W}^+)$ is a pointed, generating SDR cone, and its dual is given by*

$$\mathcal{C}^* = \widetilde{L}\big((\widetilde{\mathcal{W}} + \mathcal{W}^\perp)^+\big) \tag{19}$$

where $\widetilde{L}$ is the canonical projection $\widetilde{L} : \widetilde{\mathcal{W}} + \mathcal{W}^\perp \to (\widetilde{\mathcal{W}} + \mathcal{W}^\perp)/\mathcal{W}^\perp \cong \mathcal{V}^*$.

Since $\mathcal{I} \in \mathcal{W} \subseteq \mathcal{B}(\mathcal{H})$, $\mathcal{W}$ can be regarded as an operator system [23]. Let $\mathcal{W}_n = \mathcal{W} \otimes \mathcal{B}(\mathbb{C}^n)$ and $\mathcal{W}_n^+$ its positive cone. Likewise, given a linear map $L : \mathcal{W} \to \mathcal{V}$, let $L_n \equiv L \otimes \mathcal{I}_n : \mathcal{W}_n \to \mathcal{V}_n$. We define cones $\mathcal{C}_n$ as

$$\mathcal{C}_n = L_n(\mathcal{W}_n^+) \subset \mathcal{V}_n. \tag{20}$$

Since $K \cap \mathrm{S}^+ = \{0\}$ then $(\mathcal{V}, \mathcal{C}_n, L(\mathcal{I}))$ define a quotient operator system [20].

5 Regular quasi-realizations as quotient realizations

From Theorem 2 it follows that given a regular quasi-realization $\mathcal{R} = (\mathcal{V}, \pi, \mathcal{D}^{(u)}, \tau)$, for an equivalent completely-positive realization $\mathcal{Q} = (\mathcal{B}(\mathcal{H}), \rho, \mathcal{E}^{(u)}, \mathcal{I})$ to exist, the former must be a quotient realization of the latter. This implies several constraints on the structure of the stable subspaces of $\mathcal{Q}$, and provide necessary conditions for the feasibility of the CPRP.

For a hypothetical completely-positive realization for p, the accessible subspace $\mathcal{W} = \operatorname{span}\{\mathcal{E}^{(u)}(\mathcal{I})\}$ is an operator system in $\mathcal{B}(\mathcal{H})$, and complete positivity of $\mathcal{E}$ in $\mathcal{W}$ suffices, by virtue of Arveson's theorem, to ensure complete positivity in $\mathcal{B}(\mathcal{H})$,

$$\mathcal{E}_n(\mathcal{W}_n^+) \subseteq \mathcal{W}_n^+. \tag{21}$$

The null space $\widetilde{\mathcal{W}}^{\perp}$ of $\mathcal{Q}$, and more precisely its restriction to $\mathcal{W}$, $K = \mathcal{W} \cap \widetilde{\mathcal{W}}^{\perp}$ must also be stable under the action of $\mathcal{E}^{(\mathbf{u})}$. The quotient space is $\mathcal{V} = \mathcal{W}/K$ and the canonical projection $L : \mathcal{W} \to \mathcal{V}$ brings $\mathcal{Q}$ to $\mathcal{R}$. In particular, we have the following relations

$$\tau = L(\mathcal{I}), \tag{22a}$$

$$\pi \circ L = \rho \tag{22b}$$

which relate $\mathcal{R}$ to $\mathcal{Q}$. Under the quotient construction, the induced maps satisfy the relation

$$\mathcal{D} \circ L = L \circ \mathcal{E}. \tag{23}$$

Using the definitions of the previous section, we have

$$\mathcal{D}_n(\mathcal{C}_n) \subseteq \mathcal{C}_n, \quad \forall n \geq 1. \tag{24}$$

This is precisely the condition of complete positivity in the quotient operator system $(\mathcal{V}, \mathcal{C}_n, L(\mathcal{I}))$. Hence a necessary condition for the existence of a CP realization is that the regular realization is completely-positive with respect to a quotient operator system, together with relations

$$\tau \in \mathcal{C} \tag{25}$$

$$\pi \in \mathcal{C}^*, \tag{26}$$

which follow from (22). However, as it turns out, this condition does not suffice to guarantee existence of a completely positive lift in $\mathcal{W}$. In fact, there exist completely-positive maps in $\mathcal{V}$ which are not induced quotients of completely-positive maps in $\mathcal{W}$. To overcome this difficulty, we will not impose complete positivity in the standard operator systems sense, but instead impose a stronger condition that guarantees complete positivity in the quotient operator system $\mathcal{V}$ as well as in $\mathcal{W}$.

Let us denote $\mathcal{E}$ for an arbitrary element $\mathcal{E}^{(u)}$ and regard it as an element in $\mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H})^*$. Maps satisfying $\mathcal{E}(\mathcal{W}) \subseteq \mathcal{W}$ and $\mathcal{E}(K) \subseteq K$ are in the subspace $\mathcal{S} \subseteq \mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H})^*$,

$$\mathcal{S} = \mathcal{W} \otimes \widetilde{\mathcal{W}} + K \otimes \mathcal{B}(\mathcal{H})^* + \mathcal{B}(\mathcal{H}) \otimes \mathcal{W}^{\perp}. \tag{27}$$

Let $\varphi : K^{\perp} \to (\mathcal{W}/K)^* = \mathcal{V}^*$ be the natural isomorphism between these two spaces, and let $\phi : \mathcal{B}(\mathcal{H})^* \to \mathcal{B}(\mathcal{H})^*/\mathcal{W}^{\perp}$ be the canonical quotient map modulo $\mathcal{W}^{\perp}$. Then define

$$\widetilde{L} : \mathcal{W}^{\perp} + \widetilde{\mathcal{W}} \xrightarrow{\phi} K^{\perp} \xrightarrow{\varphi} \mathcal{V}^*. \tag{28}$$

Now, consider the map $L \otimes \widetilde{L}$. In principle, the range of this map is not well defined in the entire $\mathcal{B}(\mathcal{H}) \otimes \mathcal{B}(\mathcal{H})^*$, and arbitrary extensions would be required. However, for each of these spaces is well-defined,

$$L \otimes \widetilde{L} : \quad K \otimes \mathcal{B}(\mathcal{H})^* \longrightarrow 0 \quad (K = \ker L) \tag{29}$$

$$\mathcal{B}(\mathcal{H}) \otimes \mathcal{W}^{\perp} \longrightarrow 0 \quad (\mathcal{W}^{\perp} = \ker \widetilde{L}) \tag{30}$$

$$\mathcal{W} \otimes \widetilde{\mathcal{W}} \longrightarrow \mathcal{V} \otimes \mathcal{V}^*. \tag{31}$$

We thus have that $\mathcal{D} = L \otimes \widetilde{L}(\mathcal{E}) \in \mathcal{V} \otimes \mathcal{V}^*$ is the induced quotient map. Also, completely-positive maps with these stable subspaces form a cone $\mathcal{S}^{\mathrm{CP}}$, where CP denotes intersection with the completely positive cone. Finally, we conclude that

$$\mathcal{D} \in \mathcal{P} = L \otimes \widetilde{L}\left(\mathcal{S}^{\mathrm{CP}}\right). \tag{32}$$

By the Choi-Jamiolkowski isomorphism, $\mathcal{S}^{\mathrm{CP}}$ is isomorphic the positive semidefinite subcone of some subspace of $\mathcal{B}(\mathcal{H}\otimes\mathcal{H})$, hence, $\mathcal{P}$ is semidefinite representable. One can check that $\mathcal{D}\in\mathcal{P}$ implies complete positivity in the operator system $(\mathcal{V},\mathcal{C}_n,L(\mathcal{I}))$.

Notice that the identity map is in $\mathcal{P}$ since it just corresponds to the induced map of the identity in $\mathcal{S}^{\mathrm{CP}}$, which is completely positive and satisfies all the stability conditions. In addition, other useful properties hold for $\mathcal{P}$. In particular,

- $\mathcal{P}$ is pointed.
- $\mathcal{P}$ is closed under composition, *i. e.* it is a semigroup.
- $\mathcal{C}\otimes_{\max}\mathcal{C}^*\subseteq\mathcal{P}$, where $\otimes_{\max}$ denotes the maximal tensor product, *i. e.* the convex hull of pairs of elements $\rho\otimes\sigma$, $\rho\in\mathcal{C},\sigma\in\mathcal{C}^*$.

Notice also, that given $\mathcal{P}$ and π,τ, one can obtain $\mathcal{C}$ from $\mathcal{P}$, $\mathcal{C}=\mathcal{P}\tau$.

In conclusion, the necessary conditions for the CPRP can be stated as

- $\tau\in\mathcal{C}$,
- $\pi\in\mathcal{C}^*$,
- $\mathcal{D}\in\mathcal{P}$.

with $\mathcal{P}$ of the type (32). The next section shows that these conditions are also sufficient.

6 Sufficiency of the conditions

So far we have derived a set of necessary conditions which follow from the hypothesis that an underlying completely-positive realization exists. In this section we show that these are also sufficient.

▶ **Theorem 5** (Removing spurious eigenvectors). *Let $\{\mathcal{E}^{(u)}\}$ be a set of completely positive maps on $\mathcal{B}(\mathcal{H})$ with $\mathcal{E}=\sum_u\mathcal{E}^{(u)}$, and let ρ, $\mathcal{I}$ be positive semidefinite operators in $\mathcal{B}(\mathcal{H})$ such that $\mathrm{tr}[\rho\mathcal{I}]=1$. If ω is a positive semidefinite eigenvector of $\mathcal{E}$ such that $\mathrm{tr}[\rho\omega]=0$, then there is always another set of CP maps $\{\hat{\mathcal{E}}^{(u)}\}$ on $\mathcal{B}(\ker(\omega))$ and positive semidefinite operators $\hat\rho$, $\hat{\mathcal{I}}\in\mathcal{B}(\ker(\omega))$ such that*

$$\mathrm{tr}[\rho\,\mathcal{E}^{(\boldsymbol{u})}(\mathcal{I})]=\mathrm{tr}[\hat\rho\,\hat{\mathcal{E}}^{(\boldsymbol{u})}(\hat{\mathcal{I}})]\qquad\forall\boldsymbol{u}\in\mathcal{M}^*.\tag{33}$$

Proof. Let $\mathcal{P}=\ker(\omega)$ and $\mathcal{Q}=\mathrm{range}(\omega)=\mathcal{P}^\perp$ its orthogonal complement. Let P (resp. Q) be the corresponding orthogonal projection in $\mathcal{H}$, and $\Pi_{\mathcal{P}}=P\cdot P$, (resp. $\Pi_{\mathcal{Q}}$) the hereditary projection on $\mathcal{B}(\mathcal{H})$. Since ω is a positive semidefinite eigenvector, we have that $\mathcal{E}\circ\Pi_{\mathcal{Q}}=\Pi_{\mathcal{Q}}\circ\mathcal{E}\circ\Pi_{\mathcal{Q}}$. From positivity, this extends to all $\mathcal{E}^{(\mathbf{u})}$ and thus

$$\Pi_{\mathcal{P}}\circ\mathcal{E}^{(\mathbf{u})}=\Pi_{\mathcal{P}}\circ\mathcal{E}^{(\mathbf{u})}\circ\Pi_{\mathcal{P}},\quad\forall\mathbf{u}\in\mathcal{M}^*.\tag{34}$$

From orthogonality of $\rho\geq0$ and $\omega\geq0$ it follows that $\rho=\Pi_{\mathcal{P}}(\rho)$ and we can write

$$\begin{aligned}p(\mathbf{u})&=\mathrm{tr}[\rho\,\Pi_{\mathcal{P}}\mathcal{E}^{(\mathbf{u})}(\mathcal{I})]\\&=\mathrm{tr}[\rho\,\Pi_{\mathcal{P}}\mathcal{E}^{(u_1)}\Pi_{\mathcal{P}}\circ\Pi_{\mathcal{P}}\mathcal{E}^{(u_2)}\Pi_{\mathcal{P}}\cdots\Pi_{\mathcal{P}}\mathcal{E}^{(u_\ell)}\Pi_{\mathcal{P}}(\mathcal{I})].\end{aligned}\tag{35}$$

Replace $\mathcal{H}\leftarrow\mathcal{P}$, $\mathcal{B}(\mathcal{H})\leftarrow\mathcal{B}(\mathcal{P})$ and

$$\mathcal{E}^{(u)}\leftarrow\Pi_{\mathcal{P}}\mathcal{E}^{(u)}\Pi_{\mathcal{P}}\tag{36a}$$

$$\mathcal{I}\leftarrow\Pi_{\mathcal{P}}(\mathcal{I})\tag{36b}$$

$$\rho\leftarrow\Pi_{\mathcal{P}}(\rho).\tag{36c}$$

The resulting maps are still completely positive and ρ, $\mathcal{I}$ are positive semidefinite with support in $\mathcal{B}(\mathcal{P})$, thus the new $\mathcal{I}$ has $\mathrm{tr}[\mathcal{I}\omega]=0$. In addition, from Eq. (35), they generate the same process. ◀

▶ **Theorem 6.** *Given a pseudo-realization $\mathcal{R} = (\mathcal{V}, \pi, \mathcal{D}, \tau)$, an equivalent, finite-dimensional, unital, completely-positive realization exists $(\mathcal{B}(\mathcal{H}), \rho, \mathcal{E}, \mathcal{I})$ if and only if there is an SDR cone $\mathcal{P} \subset \mathcal{V} \otimes \mathcal{V}^*$ such that*

1. $\mathcal{D}^{(u)} \in \mathcal{P},\ \forall u \in \mathcal{M}$,
2. $\tau \in \mathcal{C}$,
3. $\pi \in \mathcal{C}^*$.

where $\mathcal{C}$, $\mathcal{C}^*$ and $\mathcal{P}$ are of type (18)*,* (19) *and* (32)*, respectively.*

Proof. That the conditions are necessary was proven in the previous section. It follows from condition 1 that CP maps $\mathcal{E}^{(u)} : \mathcal{B}(\mathcal{H}) \to \mathcal{B}(\mathcal{H})$ can be defined such that $\mathcal{E}^{(u)}(K) \subseteq K$ and $\mathcal{E}^{(u)}(\mathcal{W}) \subseteq \mathcal{W}$, and that

$$L \circ \mathcal{E}^{(u)} = \mathcal{D}^{(u)} \circ L, \quad \forall u \in \mathcal{M}. \tag{37}$$

To lift the vectors τ and π, notice that since $\tau \in \mathcal{C}$ and $\pi \in \mathcal{C}^*$, there is $\mathcal{I} \in \mathcal{W}^+$ and $\rho \in (\mathcal{W}^\perp + \widetilde{\mathcal{W}})^+$ such that

$$\tau = L(\mathcal{I}) \tag{38}$$

$$\rho = \pi \circ L. \tag{39}$$

At this point it is easy to check that $\mathcal{D}^{(\mathbf{u})}(\tau) = \mathcal{D}^{(\mathbf{u})} L(\mathcal{I}) = L\mathcal{E}^{(u)}(\mathcal{I})$, so that

$$\pi \cdot \mathcal{D}^{(\mathbf{u})}(\tau) = \rho \circ \mathcal{E}^{(\mathbf{u})}(\mathcal{I}), \quad \forall \mathbf{u} \in \mathcal{M}^*, \tag{40}$$

However, the operators ρ and $\mathcal{I}$ are not left- and right-eigenvectors of $\mathcal{E} = \sum_{u \in \mathcal{M}} \mathcal{E}^{(u)}$, so they $(\mathcal{B}(\mathcal{H}), \mathcal{I}, \mathcal{E}, \rho)$ does not form a realization. In order to find a proper completely-positive realization, we will iteratively replace them by suitable projections by making use of Theorem 5, until the desired properties are obtained. In the process, we remove all spurious contributions to ρ and $\mathcal{I}$ until only relevant contributions to Eq. (40) remain.

STEP 1: Consider the Cesàro mean $\omega_n = \frac{1}{n}\sum_{k=1}^n \mathcal{E}^k(\mathcal{I})$. Clearly, $\omega_n \geq 0\ \forall n$. Define the ratio $\lambda = \lim_{n\to\infty} \frac{\|\omega_{n+1}\|}{\|\omega_n\|}$ so that the limit is well-defined,

$$\omega = \lim_{n\to\infty} \frac{\omega_n}{\lambda^n}. \tag{41}$$

Clearly, $\omega \geq 0$, and

$$\mathcal{E}(\omega) = \lim_{n\to\infty} \frac{1}{n\lambda^n} \sum_{k=1}^n \mathcal{E}^{k+1}(\mathcal{I}) = \lambda\omega. \tag{42}$$

At this point, two different scenarios may occur. Either $\lambda = 1$ or $\lambda > 1$. Consider first the case when $\lambda > 1$. This means that there is a contribution to $\mathcal{I}$ which grows under the action of $\mathcal{E}$, and ω captures its asymptotic behavior. One can see that

$$\begin{aligned} \mathrm{tr}[\rho\omega] &= \lim_{n\to\infty} \frac{1}{n\lambda^n} \sum_{k=1}^n \mathrm{tr}[\rho\mathcal{E}^k(\mathcal{I})] \\ &= \lim_{n\to\infty} \frac{1}{\lambda^n} \\ &= 0. \end{aligned} \tag{43}$$

Hence, by making use of Theorem 5, we can obtain a new set of CP maps $\{\mathcal{E}^{(u)}\}$, ρ and $\mathcal{I}$ such that $\mathrm{tr}[\mathcal{I}\omega] = 0$. However, ρ and $\mathcal{I}$ are still not eigenvectors. Repeat STEP 1 until $\lambda = 1$.

If $\lambda = 1$ then $\omega = \lim_{n\to\infty} \omega_n$ is well defined. Replace $\mathcal{I} \leftarrow \omega$ and proceed to STEP 2.

At each iteration of STEP 1 a new ω is obtained, orthogonal to all previous ones, and the associated eigenvalue can only be equal or decrease. The aim of this iteration is to capture the eigenspace of $\mathcal{E}$ with the largest eigenvalue and remove it without altering the resulting stochastic process $p(\mathbf{u})$.

Because $\mathcal{E}$ has only finitely many eigenvalues, eventually λ will equal 1. In that case, the resulting ω is strictly positive. Proceed to PART 2.

STEP 2: At this point $\mathcal{I}$ is an eigenvector but ρ is not. Rerurn STEP 1 with the dual realization, *i. e.*, with $(\mathcal{B}(\mathcal{H})^*, \mathcal{I}, \mathcal{E}^*, \rho)$, interchanging the roles of ρ and $\mathcal{I}$.

After STEP 2, ρ is an eigenvelue of $\mathcal{E}$ but $\mathcal{I}$ may not be. A further iteration of steps 1 and 2 will lead to further dimension reductions. Since the dimension is finite, eventually no further truncations will be necessary and both $\mathcal{I}$ and ρ will be proper left- and right-eigenvalues of $\mathcal{E}$.

Once one has iterated through STEPS 1 and 2, one has a completely-positive realization $(\rho, \mathcal{E}^{(u)}, \mathcal{I})$ with the required stability properties for ρ and $\mathcal{I}$. It just remains to ensure that $\mathcal{I} > 0$. The procedure is very similar to the one just exposed.

STEP 3: Let $\mathcal{Q} = \ker(\mathcal{I})$ and $\mathcal{P} = \mathcal{Q}^\perp = \mathrm{range}(\mathcal{I})$ its orthogonal complement. Since $\mathcal{I} \geq 0$ is an eigenvector of $\mathcal{E}$, we have that $\mathcal{E}^{(\mathbf{u})}(\mathcal{I}) \in \mathcal{B}(\mathcal{P})$, $\forall \mathbf{u} \in \mathcal{M}^*$. Hence we can make the substitutions $\mathcal{H} \leftarrow \mathcal{P}$, $\mathcal{B}(\mathcal{H}) \leftarrow \mathcal{B}(\mathcal{P})$ and

$$\mathcal{E}^{(u)} \leftarrow \Pi_{\mathcal{P}} \mathcal{E}^{(u)} \Pi_{\mathcal{P}} \tag{44a}$$

$$\mathcal{I} \leftarrow \Pi_{\mathcal{P}}(\mathcal{I}) \tag{44b}$$

$$\rho \leftarrow \Pi_{\mathcal{P}}(\rho). \tag{44c}$$

With this, now $\mathcal{I} > 0$. One can define the completely positive map $\mathcal{N}(x) = \mathcal{I}^{-1/2} x \mathcal{I}^{-1/2}$. Finally, replace

$$\mathcal{E}^{(u)} \leftarrow \mathcal{N} \mathcal{E}^{(u)} \mathcal{N}^{-1} \tag{45a}$$

$$\mathcal{I} \leftarrow \mathcal{N}(\mathcal{I}) = \mathbb{1} \tag{45b}$$

$$\rho \leftarrow \mathcal{N}^{-1}(\rho). \tag{45c}$$

This substitution makes $\sum_{u \in \mathcal{M}} \mathcal{E}^{(u)}(\mathbb{1}) = \mathbb{1}$, while preserving complete positivity and the resulting ρ is the stationary state of the system. This concludes the proof. ◀

Note that several steps in the reduction algorithm could be avoided by imposing further conditions on the properties of the subspaces defining $\mathcal{P}$, but to explore these relations is beyond the scope of this work.

This constructive algorithm shows that not only appropriate completely positive maps can be obtained from the condition $\mathcal{D} \in \mathcal{P}$, but also that their structure can be cast into the form of a quantum instrument, where ρ is a fixed point of $\sum_{u \in \mathcal{M}} \mathcal{E}^{(u)}$. The fact that a dimension smaller than that of $\mathcal{B}(\mathcal{H})$ is capable of reproducing the model described by $(\mathcal{B}(\mathcal{H}), \rho, \mathcal{E}, \mathcal{I})$ is ultimately to the non-primitivity of $\mathcal{E}^*$ and the lack of information completeness of the POVM elements $M^{(\mathbf{u})} = \mathcal{E}^{(\mathbf{u})}(\mathcal{I})$. This theorem establishes under which that this explanation is the only possible one, revealing the essential traits that a quasi-realization should exhibit in order to be equivalent to a higher-dimensional quantum model.

7 Discussion

This result represents a generalization of Dharmadhiraki's polyhedral cone condition [9] and establishes the type of positivity that needs to be respected at the level of the regular realization for there to exist a certain lifting in $\mathcal{B}(\mathcal{H})$. The result, highlights a central issue that goes unnoticed in the commutative case. Unlike in the formulation of Dharmadhiraki's cone condition, the truly fundamental object is the set of cones $\mathcal{P}$, from which the cones $\mathcal{C}$ and $\mathcal{C}^*$ can be derived. This shifts the focus from the geometry of the cone of states, and sets it on the nature of the semigroup of transformations corresponding to a given process p.

Of course this is far from a full solution to the problem. Although condition (32) can be verified by a semidefinite program, finding the suitable cone $\mathcal{P}$ for a given process is still a formidable challenge. Our result highlights significant departures from the PRP, so that novel approaches may be possible. In particular, the CPRP turns out to be deeply related to lifting properties for quotient operator systems. Aspects of this theory are deeply connected with several open questions in operator theory [12], such as Connes Embedding Problem and Kirchberg's conjecture. In addition, classical algorithms for learning Hidden Markov Models using matrix factorizations [8] may be extended to semidefinite factorizations [13, 14] thus establishing links between the computational complexity of the CPRP and that of other relevant problems in Quantum Information science. An interesting question, from the operator systems theory point of view, is to identify the abstract operator system in $\mathcal{V}$ for which $\mathcal{P}$ is *the cone* of completely positive maps, and to determine its nuclearity properties.

Just as the positive realization problem, the completely-positive realization problem is highly relevant in systems identification and quantum control. It addresses the problem of finding compact models for systems with quantum memory and a classical readout interface. In particular, modeling stochastic processes which are generated by quantum devices will be the primary application of our results. The positive description of a process not only provides insight into the physical mechanisms underlying a process, but allows to identify *latent* variables, e. g., variables that are not directly observed but allow to draw order and simplicity in otherwise apparently chaotic and highly unpredictable behavior. In this sense, accounting for hidden quantum mechanical mechanisms, and more importantly, quantum memory to an information source, is potentially the difference between obtaining a simple description of a process or a highly complex one.

Acknowledgements. The authors acknowledge financial support by the Ministry of Education, Singapore, and thank the hospitality of the Centre for Quantum Technologies where this work originated. We also acknowledge financial support from ERDF: European Regional Development Fund, the Spanish MICINN, through contract FIS2008-01236 and the ERC (Advanced Grant "IRQUAT"). AW also acknowledges financial support by the EC STREP "RAQUEL" and the Philip Leverhulme Trust.

References

1 Luigi Accardi. Noncommutative markov chains associated to a pressigned evolution: An application to the quantum theory of measurement. *Advances in Mathematics*, 29(2):226–243, February 1978.

2 Christian Arenz, Giulia Gualdi, and Daniel Burgarth. Control of open quantum systems: case study of the central spin model. *New Journal of Physics*, 16(6):065023, June 2014.

3 Luca Benvenuti and Lorenzo Farina. A tutorial on the positive realization problem. *IEEE Transactions on Automatic Control*, 49(5):651 – 664, May 2004.

4 Grigoriy Blekherman, Pablo A. Parrilo, and Rekha R. Thomas. *Semidefinite Optimization and Convex Algebraic Geometry.* SIAM, 2012.

5 Robin Blume-Kohout, John King Gamble, Erik Nielsen, Jonathan Mizrahi, Jonathan D. Sterk, and Peter Maunz. Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit. *arXiv:1310.4492 [quant-ph]*, October 2013.

6 Daniel Burgarth and Kazuya Yuasa. Quantum system identification. *Physical Review Letters*, 108(8):080502, February 2012.

7 Daniel Burgarth and Kazuya Yuasa. Identifiability of open quantum systems. *arXiv:1401.5240 [quant-ph]*, January 2014.

8 G. Cybenko and V. Crespi. Learning hidden markov models using nonnegative matrix factorization. *IEEE Transactions on Information Theory*, 57(6):3963–3970, June 2011.

9 Sudhakar W. Dharmadhikari. Sufficient conditions for a stationary process to be a function of a finite markov chain. *The Annals of Mathematical Statistics*, 34(3):1033–1041, 1963.

10 Roy V. Erickson. Functions of markov chains. *The Annals of Mathematical Statistics*, 41(3):843–850, June 1970.

11 Mark Fannes, Bruno Nachtergaele, and Reinhard F. Werner. Finitely correlated states on quantum spin chains. *Communications in Mathematical Physics (1965-1997)*, 144(3):443–490, 1992.

12 Douglas Farenick and Vern I. Paulsen. Operator system quotients of matrix algebras and their tensor products. *Mathematica Scandinavica*, 111(2):210–243, December 2012.

13 Samuel Fiorini, Serge Massar, Sebastian Pokutta, Hans Raj Tiwary, and Ronald de Wolf. Linear vs. semidefinite extended formulations: Exponential separation and strong lower bounds. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 95–106, New York, NY, USA, 2012. ACM.

14 João Gouveia, Pablo A. Parrilo, and Rekha R. Thomas. Lifts of convex sets and cone factorizations. *Mathematics of Operations Research*, 38(2):248–264, May 2013.

15 Mile Gu, Karoline Wiesner, Elisabeth Rieper, and Vlatko Vedral. Quantum mechanics can reduce the complexity of classical models. *Nature Communications*, 3:762, March 2012.

16 Madalin Guta. Fisher information and asymptotic normality in system identification for quantum markov chains. *Physical Review A*, 83(6):062324, June 2011.

17 Madalin Guta and Naoki Yamamoto. Systems identification for passive linear quantum systems: the transfer function approach. *arXiv:1303.3771*, March 2013.

18 Hisashi Ito, Shun-Ichi Amari, and Kingo Kobayashi. Identifiability of hidden markov processes and their minimum degrees of freedom. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 74(7):77–84, January 1991.

19 Rudolf E. Kalman. Irreducible realizations and the degree of a rational matrix. *Journal of the Society for Industrial and Applied Mathematics*, 13(2):520–544, June 1965.

20 Ali S. Kavruk, Vern I. Paulsen, Ivan G. Todorov, and Mark Tomforde. Quotients, exactness, and nuclearity in the operator system category. *Advances in Mathematics*, 235:321–360, March 2013.

21 David G. Luenberger. Positive linear systems. In *Introduction to Dynamic Systems: Theory, Models, and Applications*. Wiley, 1 edition, May 1979.

22 Alex Monras, Almut Beige, and Karoline Wiesner. Hidden quantum markov models and non-adaptive read-out of many-body states. *Applied Mathematical and Computational Sciences*, 3:93, 2011.

23 Vern Paulsen. *Completely Bounded Maps and Operator Algebras*. Cambridge University Press, February 2003.

24 Mathukumalli Vidyasagar. The complete realization problem for hidden markov models: a survey and some new results. *Mathematics of Control, Signals, and Systems*, 23(1-3):1–65, October 2011.

25 Michael M. Wolf and David Perez-Garcia. Assessing quantum dimensionality from observable dynamics. *Physical Review Letters*, 102(19):190504, May 2009.

Hidden Subgroup Quantum Algorithms for a Class of Semi-Direct Product Groups

Wim van Dam[1] and Siladitya Dey[2]

1 Department of Computer Science, Department of Physics, University of California, Santa Barbara, California, 93106, United States of America
vandam@cs.ucsb.edu

2 Department of Computer Science, University of California, Santa Barbara California, 93106, United States of America
siladitya_dey@cs.ucsb.edu

Abstract

A quantum algorithm for the Hidden Subgroup Problem over the group $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$ is presented. This algorithm, which for certain parameters of the group qualifies as 'efficient', generalizes prior work on related semi-direct product groups.

1998 ACM Subject Classification F.1.2 Modes of Computation, F.2.2 Nonnumerical Algorithms and Problems

Keywords and phrases quantum algorithms, quantum complexity theory, computational group theory

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.110

1 Introduction and Related Work

1.1 Introduction

The quantum algorithm to factorize integers as given by Shor [8] in 1994 is exponentially faster than any known classical algorithm. The success of Shor's algorithm resulted in a great deal of interest in quantum computing, subsequently resulting in the design of several more quantum algorithms that are exponentially faster than their classical counterparts. Several of these algorithms solve the problem of finding subgroup generators of a group using evaluations of a function that "hides" the subgroup [2]. This generalized framework is captured by the Hidden Subgroup Problem (referred henceforth as HSP) and has been successful in admitting quantum algorithms that are exponentially faster than their classical counterparts. It is known that there exists an efficient solution to the HSP for finite Abelian groups, but this is not known to hold for non-Abelian groups. The motivation for research in this area stems from knowledge that an efficient solution to the HSP over the symmetric group (dihedral group) will result in an efficient quantum algorithm for graph isomorphism (shortest vector in a lattice). In this article, we present an algorithm to solve the hidden subgroup in the specific class of non-Abelian groups, i.e. the semi-direct product groups of the form $G := \mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$, where p, q are prime with $p \neq q$ and $r, s \in \mathbb{Z}^+$ with the relative sizes of the subgroups bounded by $p^r/q^{t-j} \in O(\text{poly}(\log p^r))$ where $j \in \{0, \ldots, t-1\}$ is a parameter specific to the group. For certain parameters of G and its subgroup, this algorithm has running time $O(\text{poly}(\log |G|))$, and hence qualifies as 'efficient'.

In Section 2 we clarify the structure of the group $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$ and its subgroups. The quantum algorithm that will help solve for the hidden subgroup, H within this specific class of non-Abelian groups is presented in Section 3.

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 110–117
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1.2 Related Work

There has been considerable work in trying to solve the HSP in semi-direct product groups. In this article we discuss the case $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$. It was shown in [1] that the HSP in $\mathbb{Z}/N\mathbb{Z} \rtimes \mathbb{Z}/q\mathbb{Z}$, for positive integers N, q such that $N/q \in O(\mathrm{poly}(\log N))$, reduces to finding cyclic subgroups of order q and can be efficiently solved. This work was extended in [6], which developed an efficient HSP algorithm in $(\mathbb{Z}/p^r\mathbb{Z})^m \rtimes \mathbb{Z}/p\mathbb{Z}$, with p prime and integers r, m. Following this, in 2009 an efficient quantum algorithm to solve the HSP in $\mathbb{Z}/p\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$ for distinct odd primes and $s > 0$ such that $p/q \in O(\mathrm{poly}(\log p))$ was shown [4]. More recently in [5], the HSP problem was considered in $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$ where p, q are distinct primes such that $p^r/q \in O(\mathrm{poly}(\log p^r))$. The current article extends this previous result [5]. Specifically, the group $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$ has a parameter t (as explained in the next section) that characterizes the group. In [5] an algorithm was presented for the $t = 1$ case; here we deal with all possible values $t \in \{0, \dots, s\}$. Whether or not our algorithm qualifies as efficient depends on the specific parameters of G and its subgroup, which will be explained in Section 3.

2 The Group $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$ and Its Subgroups

2.1 Some Properties of the Group $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$

In this section we discuss and prove various properties of the semi-direct product group $G := \mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$, with p, q prime and $r, s \in \mathbb{Z}^+$. We know that $\mathbb{Z}/p^r\mathbb{Z}$ and $\mathbb{Z}/q^s\mathbb{Z}$ are finite, cyclic, Abelian groups. Let $\phi\colon \mathbb{Z}/q^s\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/p^r\mathbb{Z})$ be the group homomorphism that defines G, for all $a, c \in \mathbb{Z}/p^r\mathbb{Z}$ and all $b, d \in \mathbb{Z}/q^s\mathbb{Z}$:

$$(a, b)(c, d) = (a + \phi(b)(c), b + d). \tag{1}$$

As $\mathbb{Z}/q^s\mathbb{Z}$ is cyclic, we have for all b that $\phi(b) = \phi(1 + \cdots + 1) = \phi(1)^b$. In a similar vein, since $\mathbb{Z}/p^r\mathbb{Z}$ is also cyclic, we have $\phi(1)(c) = \phi(1)(1 + \cdots + 1) = \phi(1)(1) + \cdots + \phi(1)(1) = c\phi(1)(1)$. We thus see that ϕ is completely determined by the single value $\phi(1)(1)$, which from now on will be denoted by $\alpha := \phi(1)(1) \in (\mathbb{Z}/p^r\mathbb{Z})^*$. The group operation in $G = \mathbb{Z}/p^r\mathbb{Z} \rtimes_\alpha \mathbb{Z}/q^s\mathbb{Z}$ thus simplifies to

$$(a, b)(c, d) = (a + \alpha^b c, b + d). \tag{2}$$

The identity in G is $(0, 0)$ and the inverse is expressed by $(a, b)^{-1} = (-\alpha^{-b}a, -b)$.

Because it must hold that $1 = \alpha^0 = \alpha^{(q^s)}$ we have that there exists a smallest integer $t \in \{0, \dots, s\}$ such that $\alpha^{(q^t)} = 1$. As explained in [5], if the groups $G = \mathbb{Z}/p^r \rtimes_\alpha \mathbb{Z}/q^s\mathbb{Z}$ and $G' = \mathbb{Z}/p^r \rtimes_{\alpha'} \mathbb{Z}/q^s\mathbb{Z}$ have the same t-parameter ($t = t'$), then these groups are isomorphic. Additionally, if $t = 0$ we have that $\alpha = 1$, making G Abelian. From now on we will thus assume that $t \in \{1, \dots, s\}$. It can be shown that $q^t \mid (p - 1)$. We also note that it can be shown that G is supersolvable.

2.2 Subgroups of $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$

Following [5, Theorem 2], the subgroups of the group $G = \mathbb{Z}/p^r\mathbb{Z} \rtimes_\alpha \mathbb{Z}/q^s\mathbb{Z}$ are from either one of three types. With $t \in \{1, \dots, s\}$ the parameter of G as explained in the previous section, these types are as follows.

Type I: $H^{\mathrm{I}}_{i,j} = \langle (p^i, q^j) \rangle$, for each $i \in \{0, \dots, r\}$ and $j \in \{t, \dots, s\}$.

Type II: $H^{\mathrm{II}}_{j,\eta} = \langle (\eta, q^j) \rangle$, for each $j \in \{0, \dots, t-1\}$ and $\eta \in \mathbb{Z}/p^r\mathbb{Z}$.

Type III: $H^{\mathrm{III}}_{i,j,\eta} = \langle (p^i, 0), (\eta, q^j) \rangle$, for each $i \in \{0, \dots, r-1\}$, $j \in \{0, \dots, t-1\}$, and $\eta \in \{0, \dots, p^i - 1\}$.

We point out that [5, Theorem 2] allows the η parameter for Type III subgroups to be from the whole set $\{0,\dots,p^r-1\}$ but that this creates ambiguity as, for example, $H^{\mathrm{III}}_{0,0,0} = \langle(1,0),(0,1)\rangle = G$ and $H^{\mathrm{III}}_{0,0,1} = \langle(1,0),(1,1)\rangle = G$ as well. By limiting η to the set $\{0,\dots,p^i-1\}$ each triplet of parameters (i,j,η) defines a unique Type III subgroup.

Next, we will describe the parameterization of the elements of these three types of subgroups. The elements of the Type I subgroup are of the form $(p^i,q^j)^z = (zp^i,zq^j)$ where $z\in\mathbb{Z}$ and where we used the fact that $\alpha^{(q^j)}=1$ as $j\ge t$. Because $\gcd(p^r,q^s)=1$ we can further simplify this description to

$$H^{\mathrm{I}}_{i,j} = \{(xp^i,yq^j) : x\in\{0,\dots,p^{r-i}-1\}, y\in\{0,\dots,q^{s-j}-1\}\}, \tag{3}$$

showing that $H^{\mathrm{I}}_{i,j}$ has $p^{r-i}q^{s-j}$ elements.

The subgroups of Type II and III are less trivial to describe. To better understand the elements of the subgroup $\langle(\eta,q^j)\rangle$, consider first some small powers of the generating element (η,q^j):

$$\begin{cases} (\eta,q^j)^{-1} & = (-\eta\alpha^{(-q^j)}, -q^j) \\ (\eta,q^j)^0 & = (0,0) \\ (\eta,q^j)^1 & = (\eta,q^j) \\ (\eta,q^j)^2 & = (\eta+\eta\alpha^{(q^j)}, 2q^j) \\ (\eta,q^j)^3 & = (\eta+\eta\alpha^{(q^j)}+\eta\alpha^{(2q^j)}, 3q^j) \end{cases} \tag{4}$$

and so on. In general we have the following characterization.

▶ **Lemma 1.** *Let $G = \mathbb{Z}/p^r\mathbb{Z} \rtimes_\alpha \mathbb{Z}/q^s\mathbb{Z}$ and let $t\in\{1,\dots,s\}$ be the smallest positive integer such that $\alpha^{(q^t)}=1$. For any $\eta\in\mathbb{Z}/p^r\mathbb{Z}$ and $j\in\{0,\dots,t-1\}$ consider the cyclic subgroup $H=\langle(\eta,q^j)\rangle$. For each exponent $y\in\mathbb{Z}$, the elements of H can be described by $(\eta,q^j)^y = (\eta S(y), yq^j)$ where $S\colon\mathbb{Z}\to\mathbb{Z}/p^r\mathbb{Z}$ is defined by*

$$S(y) := \frac{\alpha^{(yq^j)}-1}{\alpha^{(q^j)}-1}. \tag{5}$$

As a result, the subgroup has q^{s-j} elements such that

$$H^{\mathrm{II}}_{j,\eta} := \langle(\eta,q^j)\rangle = \{(\eta S(y), yq^j) : y\in\{0,\dots,q^{s-j}-1\}\}. \tag{6}$$

Proof. In [5, Lemma A2] it is shown that $\alpha^{(q^j)}-1$ is invertible in $\mathbb{Z}/p^r\mathbb{Z}$ hence the definition of S in Equation 5 does indeed make sense. Assuming for a given y that $(\eta,q^j)^y = (\eta S(y), yq^j)$ we get $(\eta,q^j)^{y+1} = (\eta,q^j)(\eta S(y),yq^j) = (\eta(1+\alpha^{(q^j)}S(y)),(y+1)q^j)$. With this relation $S(y+1) = 1+\alpha^{(q^j)}S(y)$ and $S(0)=0$ the Equality 5 can be proven by induction on y.

From the $\mathbb{Z}/q^s\mathbb{Z}$ part of G it is obvious that the values y such that $(\eta S(y),yq^j)=(0,0)$ must obey that y is a multiple of q^{s-j}. Conversely, if $y=\lambda q^{s-j}$, then $S(y) = (\alpha^{(\lambda q^s)} - 1)/(\alpha^{(q^j)}-1) = 0$. Hence $(\eta S(y),yq^j)=(0,0)$ if and only if $y = 0 \bmod q^{s-j}$. ◀

Upon further inspection it is clear that S has period q^{t-j}, which will be helpful in the reduction of the complexity of finding the hidden subgroup H in G.

The Type III subgroups are obviously extensions of the previous type. As we have $(p^i,0)(\eta S(y),yq^j) = (p^i+\eta S(y),yq^j)$ and $(\eta S(y),yq^j)(p^i,0) = (\eta S(y)+\alpha^{(yq^j)}p^i,yq^j)$ it is clear that the elements of H^{III} can be described by

$$H^{\mathrm{III}}_{i,j,\eta} = \{(xp^i+\eta S(y),yq^j) : x\in\{0,\dots,p^{r-i}-1\}, y\in\{0,\dots,q^{s-j}-1\}\}, \tag{7}$$

which also shows that it has $p^{r-i}q^{s-j}$ elements, and hence that $H^{\mathrm{III}}_{0,0,\eta}=G$, regardless of η. More generally, it is only the value $\eta \bmod p^i$ that matters in the definition of this subgroup.

3 Quantum Algorithm for HSP in $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$

3.1 Overview of Algorithm

In this section we will present a quantum algorithm that solves the hidden subgroup problem in $G = \mathbb{Z}/p^r\mathbb{Z} \rtimes_\alpha \mathbb{Z}/q^s\mathbb{Z}$, but before doing so we will reduce the problem significantly. As in the previous section, the group operation is defined by $(a, b)(c, d) = (a + \alpha^b c, b + d)$ where $\alpha \in (\mathbb{Z}/p^r\mathbb{Z})^*$ and for which there exists a smallest integer $t \in \{1, \ldots, s\}$ such that $\alpha^{(q^t)} = 1$. Let f be the subgroup hiding function on G, which obeys

$$f((a, b)) = f((a', b')) \text{ if and only if } (a, b)^{-1}(a', b') \in H. \tag{8}$$

In other words, f is constant on the left cosets of H and f is different between different cosets of H.

Recall from Section 2.2 that the subgroups of G are one of three types with potentially unknown parameters i, j, η. In [5, Section 3] it was claimed that it was sufficient to solve the HSP for Type II subgroups but the current authors were unable to reproduce this result. Instead we will present an alternative way of finding the hidden subgroup.

We assume that all the parameters (p, r, q, s, α, and t) of the group G are known. For our purposes, an algorithm is considered efficient if its running time is bounded by $O(\text{poly}(\log(|G|)) = O(\text{poly}(r \log p + s \log q))$. Note that when an algorithm suggests that a group H' is the hidden subgroup, then that suggestion can be checked by querying f on $(0, 0)$ and on the generators of H'. If H' passes this check we can conclude that $H' \leq H$; otherwise a mistake was made and the algorithm should be executed again to find another suggestion for H. Repeating the above procedure will give a 'largest' subgroup that with high probability will equal the true hidden subgroup.

Because of the just described approach to solve the HSP, it is sufficient to use an algorithm that finds the hidden subgroup with a success rate that is significant enough. For the current case of possible subgroups of G it is therefore sufficient to simply guess the parameters $i \in \{0, \ldots, r\}$ and $j \in \{0, \ldots, s\}$ as the probability of doing so correctly equals $1/rs \in \Omega(1/\text{poly}(\log |G|))$. If the subgroup is of Type I this will have answered the HSP completely. In the case of Type II or Type III subgroups the following quantum algorithms will have to be employed to find the unknown parameter $\eta \in \mathbb{Z}/p^r\mathbb{Z}$ (Type II) or $\eta \in \{0, \ldots, p^i - 1\}$ (Type III).

3.2 Quantum Algorithm for Finding HSP in $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$

▶ **Theorem 2.** *Let p and q be distinct primes and let r and s be positive integers. Define the semi-direct product group $G := \mathbb{Z}/p^r\mathbb{Z} \rtimes_\alpha \mathbb{Z}/q^s\mathbb{Z}$ by the non-commuting group operation that, for all $a, c \in \mathbb{Z}/p^r\mathbb{Z}$ and all $b, d \in \mathbb{Z}/q^s\mathbb{Z}$, has $(a, b)(c, d) = (a + \alpha^b c, b + d)$ for an $\alpha \in (\mathbb{Z}/p^r\mathbb{Z})^*$. Let $t \in \{1, \ldots, s\}$ be the smallest positive integer such that $\alpha^{(q^t)} = 1$.*

Let the function f on G hide a Type II subgroup $H = \langle(\eta, q^j)\rangle$ and assume that the parameter $j \in \{0, \ldots, t - 1\}$ is known. There exists a probabilistic quantum algorithm that determines the unknown parameter $\eta \in \mathbb{Z}/p^r\mathbb{Z}$ with success probability $(1 - 1/p)(q^{t-j}/p^r)$ using only one query to f.

Proof. This proof is inspired by the PGM algorithm described in [1], but it uses several additional ingredients specific to the properties of this group G and its subgroups (for which see Section 2).

1. Initialize the register in the state,

$$|\psi_1\rangle = \frac{1}{\sqrt{p^r q^{t-j}}} \sum_{x\in\mathbb{Z}/p^r\mathbb{Z}} \sum_{y=0}^{q^{t-j}-1} |x, yq^j, f((x, yq^j))\rangle.$$

   Note how the second register contains only multiples of q^j and how the range of yq^j goes only to q^t and not q^s.

2. The p^r different left cosets of H that are relevant for this algorithm are described by $(\ell, 0)H = \{(\ell + \eta S(y), yq^j) : y \in \{0, \dots, q^{t-j} - 1\}\}$, for each $\ell \in \mathbb{Z}/p^r\mathbb{Z}$. After measuring (and ignoring) the third register of $|\psi_1\rangle$ in the computational basis we thus get the state

$$|\psi_2\rangle = \frac{1}{\sqrt{q^{t-j}}} \sum_{y=0}^{q^{t-j}-1} |\ell + \eta S(y), yq^j\rangle,$$

   for an unknown and irrelevant $\ell \in \mathbb{Z}/p^r\mathbb{Z}$.

3. Applying the Fourier Transform over $\mathbb{Z}/p^r\mathbb{Z}$ to the first register of $|\psi_2\rangle$ we get, with $\omega := \exp(\mathrm{i}2\pi/p^r)$,

$$\begin{aligned}|\psi_3\rangle &= \frac{1}{\sqrt{p^r q^{t-j}}} \sum_{k\in\mathbb{Z}/p^r\mathbb{Z}} \sum_{y=0}^{q^{t-j}-1} \omega^{k(\ell+\eta S(y))}|k, yq^j\rangle \\ &= \frac{1}{\sqrt{p^r q^{t-j}}} \sum_{k\in\mathbb{Z}/p^r\mathbb{Z}} \omega^{k\ell}|k\rangle \sum_{y=0}^{q^{t-j}-1} \omega^{k\eta S(y)}|yq^j\rangle.\end{aligned}$$

4. Measure the first register in the computational basis and assume the result is some invertible $k \in (\mathbb{Z}/p^r\mathbb{Z})^*$ (which occurs with probability $(1 - 1/p)$). Tracing out this k register gives us the remaining superposition

$$|\psi_4\rangle = \frac{1}{\sqrt{q^{t-j}}} \sum_{y=0}^{q^{t-j}-1} \omega^{k\eta S(y)}|yq^j\rangle.$$

5. We now take the yq^j register in $|\psi_4\rangle$ and use it to append a second register with the value $kS(y) = k(\alpha^{(yq^j)} - 1)/(\alpha^{(q^j)} - 1) \bmod p^r$. As α, q, j, p^r, k are known and $(\alpha^{(q^j)} - 1)$ is invertible, this transformation can be done efficiently, yielding

$$|\psi_5\rangle = \frac{1}{\sqrt{q^{t-j}}} \sum_{y=0}^{q^{t-j}-1} \omega^{k\eta S(y)}|yq^j, kS(y)\rangle.$$

6. Because k is invertible and the function S is injective on $\{0, \dots, q^{t-j}-1\}$, we can determine a unique solution y from the value $kS(y)$. Using Shor's discrete logarithm algorithm we can hence efficiently implement the unitary mapping $|yq^j, kS(y)\rangle \mapsto |0, kS(y)\rangle$, giving

$$|\psi_6\rangle = \frac{1}{\sqrt{q^{t-j}}} \sum_{y=0}^{q^{t-j}-1} \omega^{k\eta S(y)}|kS(y)\rangle.$$

7. Finally, we perform an inverse Fourier transform over $\mathbb{Z}/p^r\mathbb{Z}$ in the hope of observing the unknown η. To calculate the probability of this occurring, consider the ideal state $|\hat{\eta}\rangle := \sum_{z\in\mathbb{Z}/p^r\mathbb{Z}} \omega^{z\eta}|z\rangle/\sqrt{p^r}$, which is guaranteed to give η. The fidelity squared between this perfect state and our actual state is $|\langle\psi_6|\hat{\eta}\rangle|^2 = (q^{t-j})/p^r$, which is thus the probability of observing η at the end of this step.

The above algorithm requires one f-query and poly$(\log|G|)$ time and space. Its overall success probability equals $(1-1/p)(q^{t-j}/p^r)$. ◀

The algorithm for finding Type III subgroups is an adaptation of the just described algorithm. Crucially, the unknown parameter η is an element of the set $\{0,\ldots,p^i-1\}$, which influences the first register of the algorithm.

▶ **Theorem 3.** *Let p and q be distinct primes and let r and s be positive integers. Define the semi-direct product group $G := \mathbb{Z}/p^r\mathbb{Z} \rtimes_\alpha \mathbb{Z}/q^s\mathbb{Z}$ by the non-commuting group operation that, for all $a,c \in \mathbb{Z}/p^r\mathbb{Z}$ and all $b,d \in \mathbb{Z}/q^s\mathbb{Z}$ has $(a,b)(c,d) = (a+\alpha^b c, b+d)$ for an $\alpha \in (\mathbb{Z}/p^r\mathbb{Z})^*$. Let $t \in \{1,\ldots,s\}$ be the smallest positive integer such that $\alpha^{(q^t)} = 1$.*

Let the function f hide a Type III subgroup $H = \langle (p^i,0),(\eta,q^j)\rangle$ in G and assume that the parameters $i \in \{0,\ldots,r-1\}$ and $j \in \{0,\ldots,t-1\}$ are known. There exists a probabilistic quantum algorithm that can determine the unknown parameter $\eta \in \{0,\ldots,p^i-1\}$ with success probability $(1-1/p)(q^{t-j}/p^i)$ using only one query to f.

Proof. This algorithm is quite similar to the one of the previous theorem, except for the fact that the first register will be restricted to elements of $\mathbb{Z}/p^i\mathbb{Z}$.

1. Initialize the register in the state
$$|\psi_1\rangle = \frac{1}{\sqrt{p^i q^{t-j}}}\sum_{x=0}^{p^i-1}\sum_{y=0}^{q^{t-j}-1} |x, yq^j, f((x,yq^j))\rangle.$$
Note how the second register contains only multiples of q^j, how the range of yq^j goes only to q^t and not q^s, and how the first register contains only p^i elements.
2. The p^i different left cosets of H that are relevant for this algorithm are described by $(\ell,0)H = \{(\ell + xp^i + \eta S(y), yq^j) : x \in \{0,\ldots,p^{r-i}-1\}, y \in \{0,\ldots,q^{t-j}-1\}\}$, for each $\ell \in \{0,\ldots,p^i-1\}$. As the first register contains values from the set $\{0,\ldots,p^i-1\}$ this description further reduces to $\{(\ell+\eta S(y) \bmod p^i, yq^j) : y \in \{0,\ldots,q^{t-j}-1\}\}$. Measuring the third register of $|\psi_1\rangle$ in the computational basis we get the state
$$|\psi_2\rangle = \frac{1}{\sqrt{q^{t-j}}}\sum_{y=0}^{q^{t-j}-1} |\ell + \eta S(y) \bmod p^i, yq^j\rangle,$$
for an unknown and irrelevant $\ell \in \{0,\ldots,p^i-1\}$.
3. From now on we interpret the first register of $|\psi_2\rangle$ as containing values from $\mathbb{Z}/p^i\mathbb{Z}$ and we apply the Fourier Transform over $\mathbb{Z}/p^i\mathbb{Z}$ to it. With $\omega := \exp(\mathrm{i}2\pi/p^i)$, we get
$$\begin{aligned}|\psi_3\rangle &= \frac{1}{\sqrt{p^i q^{t-j}}}\sum_{k\in\mathbb{Z}/p^i\mathbb{Z}}\sum_{y=0}^{q^{t-j}-1} \omega^{k(\ell+\eta S(y))}|k, yq^j\rangle \\ &= \frac{1}{\sqrt{p^i q^{t-j}}}\sum_{k\in\mathbb{Z}/p^i\mathbb{Z}} \omega^{k\ell}|k\rangle \sum_{y=0}^{q^{t-j}-1} \omega^{k\eta S(y)}|yq^j\rangle.\end{aligned}$$
4. Measure the first register in the computational basis and assume the result is some invertible $k \in (\mathbb{Z}/p^i\mathbb{Z})^*$, which occurs with probability $(1-1/p)$. Tracing out this k register gives us the remaining superposition
$$|\psi_4\rangle = \frac{1}{\sqrt{q^{t-j}}}\sum_{y=0}^{q^{t-j}-1} \omega^{k\eta S(y)}|yq^j\rangle.$$

5. We now take the yq^j register in $|\psi_4\rangle$ and use it to append a second register with the value $kS(y) = k(\alpha^{(yq^j)} - 1)/(\alpha^{(q^j)} - 1) \bmod p^i$. As α, q, j, p^i, k are known and $(\alpha^{(q^j)} - 1)$ is invertible, this transformation can be done efficiently, yielding

$$|\psi_5\rangle = \frac{1}{\sqrt{q^{t-j}}} \sum_{y=0}^{q^{t-j}-1} \omega^{k\eta S(y)} |yq^j, kS(y)\rangle.$$

6. Because k is invertible and the function S is injective on $\{0, \ldots, q^{t-j}-1\}$, we can determine a unique solution y from the value $kS(y)$. Using Shor's discrete logarithm algorithm we can hence efficiently implement the unitary mapping $|yq^j, kS(y)\rangle \mapsto |0, kS(y)\rangle$, giving

$$|\psi_6\rangle = \frac{1}{\sqrt{q^{t-j}}} \sum_{y=0}^{q^{t-j}-1} \omega^{k\eta S(y)} |kS(y)\rangle.$$

7. Finally we perform an inverse Fourier transform over $\mathbb{Z}/p^i\mathbb{Z}$ in the hope of observing the unknown η. To calculate the probability of this occurring consider the ideal state $|\hat{\eta}\rangle := \sum_{z\in\mathbb{Z}/p^i\mathbb{Z}} \omega^{z\eta}|z\rangle/\sqrt{p^i}$, which is guaranteed to give η. The fidelity squared between this perfect state and our actual state is $|\langle\psi_6|\hat{\eta}\rangle|^2 = (q^{t-j})/p^i$, which is thus the probability of observing η at the end of this step.

The above algorithm requires one f-query and poly$(\log|G|)$ time and space. Its overall success probability equals $(1 - 1/p)(q^{t-j}/p^i)$. ◀

Summarizing the above theorems, we have the following result.

▶ **Corollary 4.** *Let p and q be distinct primes and let r and s be positive integers. Define the semi-direct product group $G := \mathbb{Z}/p^r\mathbb{Z} \rtimes_\alpha \mathbb{Z}/q^s\mathbb{Z}$ by the non-commuting group operation that, for all $a, c \in \mathbb{Z}/p^r\mathbb{Z}$ and all $b, d \in \mathbb{Z}/q^s\mathbb{Z}$, has $(a, b)(c, d) = (a + \alpha^b c, b + d)$ for an $\alpha \in (\mathbb{Z}/p^r\mathbb{Z})^*$. Let $t \in \{1, \ldots, s\}$ be the smallest positive integer such that $\alpha^{(q^t)} = 1$. Let the function f on G hide a subgroup H. There exists a quantum algorithm that determines H with a time complexity depending on the type of H in the following manner.*

Type I: *If $H^{\mathrm{I}}_{i,j} = \langle(p^i, q^j)\rangle$ for some unknown $i \in \{0, \ldots, r\}$ and $j \in \{t, \ldots, s\}$, then H will be found efficiently in time $O(\mathrm{poly}(\log|G|))$.*

Type II: *If $H^{\mathrm{II}}_{j,\eta} = \langle(\eta, q^j)\rangle$ for some unknown $j \in \{0, \ldots, t-1\}$ and $\eta \in \mathbb{Z}/p^r\mathbb{Z}$, then H will be found in time $O(\mathrm{poly}(\log|G|, p^r/q^{t-j}))$.*

Type III: *If $H^{\mathrm{III}}_{i,j,\eta} = \langle(\eta, q^j)\rangle$ for some unknown $i \in \{0, \ldots, r-1\}$, $j \in \{0, \ldots, t-1\}$ and $\eta \in \{0, \ldots, p^i - 1\}$, then H will be found in time $O(\mathrm{poly}(\log|G|, p^i/q^{t-j}))$.*

The quantum algorithm can be considered efficient, i.e. has running time $O(\mathrm{poly}(\log|G|))$, if the subgroup is of Type I, or if the Type II subgroup $H^{II}_{j,\eta}$ has $p^r/q^{t-j} \in \mathrm{poly}(\log|G|)$, or if the Type III subgroup $H^{III}_{i,j,\eta}$ has $p^i/q^{t-j} \in \mathrm{poly}(\log|G|)$.

These running times should be compared to the classical algorithm of repeatedly simply guessing the parameters i, j, η of the hidden subgroup. For Type I, II, and III subgroups this approach gives a running time of $O(\mathrm{poly}(\log|G|))$, $O(\mathrm{poly}(\log|G|, p^r))$, and $O(\mathrm{poly}(\log|G|, p^i))$ respectively. Hence we see that the presented quantum algorithm provides a speed-up of order $\Omega(q^{t-j})$ for subgroups of Type II and III.

4 Conclusion

In this paper, we consider the Hidden Subgroup Problem in the semi-direct product group $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$ with p, q distinct primes. Our result generalizes the work in [5], which imposed a restriction on the kind of homomorphism that the semi-direct product uses. The result here holds for all possible $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/q^s\mathbb{Z}$. While our algorithm is efficient for certain cases of the parameters of G and H, it is not so in other cases. This partial result is not unexpected as the design of an efficient algorithm for the HSP for the dihedral group $\mathbb{Z}/p^r\mathbb{Z} \rtimes \mathbb{Z}/2\mathbb{Z}$ remains a major open problem in the theory of quantum algorithms.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. 0747526.

References

1 Dave Bacon, Andrew M. Childs, Wim van Dam, *From optimal measurement to efficient quantum algorithms for the hidden subgroup problem over semidirect product groups*, FOCS'05: Proceedings of the 46th IEEE Symposium on Foundations of Computer Science, pp. 469–478, 2005.

2 Andrew M. Childs, Wim van Dam, *Quantum Algorithms for Algebraic Problems*, Reviews of Modern Physics, **82**(1):1–52, 2010.

3 Mark Ettinger, Peter Høyer, *On quantum algorithms for noncommutative hidden subgroups*, Advances in Applied Mathematics, 25(3):239–251, 2000.

4 Demerson N. Gonçalves, Renato Portugal, Carlos M. M. Cosme, *Solutions to the hidden subgroup problem on some metacyclic groups*, 4th Workshop on Theory of Quantum Computation, Communication, and Cryptography, Lecture Notes in Computer Science, Vol. 5906, pp. 1–9, Springer, 2009.

5 Demerson N. Gonçalves, Renato Portugal, *Solution to the Hidden Subgroup Problem for a Class of Noncommutative Groups*, arXiv:1104.1361, 2011.

6 Yoshifumi Inui, François Le Gall, *An efficient quantum algorithm for the hidden subgroup problem over a class of semi-direct product groups*, Quantum Information and Computation, **7**(5&6):559–570, 2007.

7 A. Yu. Kitaev, *Quantum measurements and the Abelian Stabilizer Problem*, arXiv:quant-ph/9511026, 1995.

8 Peter W. Shor, *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*, SIAM Journal on Computing, **26**(5):1484–1509, 1997.

9 Daniel R. Simon, *On the power of quantum computation*, SIAM Journal on Computing, **26**:116–123, 1994.

Difficult Instances of the Counting Problem for 2-quantum-SAT are Very Atypical *

Niel de Beaudrap

**Centrum Wiskunde & Informatica (CWI), Science Park 123, Amsterdam,
The Netherlands**
beaudrap@cwi.org

Abstract

The problem *2-quantum-satisfiability* (2-QSAT) is the generalisation of the 2-CNF-SAT problem to quantum bits, and is equivalent to determining whether or not a spin-1/2 Hamiltonian with two-body terms is frustration-free. Similarly to the classical problem #2-SAT, the counting problem #2-QSAT of determining the size (*i.e.* the dimension) of the set of satisfying states is #P-complete. However, if we consider random instances of 2-QSAT in which constraints are sampled from the Haar measure, intractible instances have measure zero. An apparent reason for this is that almost all two-qubit constraints are entangled, which more readily give rise to long-range constraints.

We investigate under which conditions product constraints also give rise to efficiently solvable families of #2-QSAT instances. We consider #2-QSAT involving only discrete distributions over tensor product operators, which interpolates between classical #2-SAT and #2-QSAT involving arbitrary product constraints. We find that such instances of #2-QSAT, defined on Erdős–Rényi graphs or bond-percolated lattices, are asymptotically almost surely efficiently solvable except to the extent that they are biased to resemble monotone instances of #2-SAT.

1998 ACM Subject Classification F.2 Analysis Of Algorithms And Problem Complexity, G.2.1 Combinatorics, J.2 Physical Sciences And Engineering

Keywords and phrases Frustration-free, Hamiltonian, quantum, counting, satisfiability

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.118

1 Introduction

Local spin Hamiltonians are simplified models for physical systems, in which the system is approximated by finite-range interactions between particle sites in a fixed network. We consider problems which involve the minimum eigenvalue of two-body Hamiltonians, $H = \sum_{\langle u,v \rangle} h_{u,v}$, for projectors $h_{u,v}$ acting on pairs of qubits (*i.e.* spin-1/2 particles) u and v drawn from some set V. When each $h_{u,v}$ is a projector onto standard basis states, finding the minimum energy of H is in effect MAX-2-SAT, or the problem of finding an assignment to boolean variables which satisfies as many constraints as possible, from a given list of constraints on pairs of bits. Minimum eigenspace problems are therefore at least NP-hard in general, and are even NP-hard to approximate to within a small percentage error [15]. Even if the minimum energy is known, determining the degeneracy (the dimension of the lowest-energy eigenspace) is #P-hard in general, or as difficult as determining the number of satisfying solutions to an instance of 3-SAT [17]. Thus, such problems should be considered to be intractable, barring major and unexpected advances in technique.

* Supported by a Vidi grant from the Netherlands Organisation for Scientific Research (NWO) and the European Commission project QALGO.

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 118–140
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

This article concerns the conditions under which computing the degeneracy of local Hamiltonians on spin-1/2 particles is possible in polynomial time, as opposed to its worst-case complexity of being #P-hard. We make this question more precise below.

1.1 Counting problems for frustration-free spin-1/2 Hamiltonians

A special case of interest are *frustration-free* Hamiltonians, for which there are states $|\psi\rangle$ which minimize all of the terms $\langle\psi|\, h_{u,v} \,|\psi\rangle$ simultaneously. Finding ground states of such systems may still be hard, but one may at least certify succinct descriptions of ground states, *e.g.* by direct calculation of energy contributions from each term $h_{u,v}$. These models are therefore a potentially useful proving ground for analytical techniques in many-body theory. Indeed, there is a wide class of such Hamiltonians on qubits, for which one may efficiently characterise the ground-state manifold [5].

Bravyi [2] defines the *quantum satisfiability problem*, or k-QSAT (for any fixed $k \geqslant 1$), to be essentially the problem of determining whether a Hamiltonian consisting of a sum of projectors, each acting non-trivially on at most k spin-1/2 particles, is frustration-free. Bravyi shows that 2-QSAT is efficiently solvable; by contrast, 3-QSAT may not have any efficient solutions, even if it were somehow shown that $\mathsf{P} = \mathsf{NP}$ [9].

A natural problem for frustration-free systems is to determine the "degeneracy" of their ground-state energy levels. Given a two-body spin-1/2 Hamiltonian H as input, let #2-QSAT denote the problem of computing the dimension of the subspace of states which minimizes the energy contributions of each interaction term of H independently. We refer to this dimension as the *value* of the instance of #2-QSAT. This value is positive if and only if H is frustration-free, and greater than one if H is also degenerate. The name #2-QSAT is chosen (see also Ref. [13]) in analogy to the problem #2-SAT of counting the satisfying assignments to an instance of 2-SAT . The dimension of the ground-state manifold of a frustration-free spin-1/2 Hamiltonian is simply the size of a basis for the solution space: if the projectors $h_{u,v}$ are all diagonal operators, this problem is #2-SAT. Thus #2-QSAT may be construed as a counting problem in the traditional sense.

While 2-SAT is efficiently solvable, the counting problem #2-SAT is #P-complete [17], *i.e.* polynomial-time equivalent to counting satisfying assignments for instances of 3-CNF-SAT. As #2-QSAT generalizes #2-SAT, the former problem is at least as hard in the worst case. (Ji, Wei, and Zeng [13] show that in fact #2-QSAT $\in$ #P.) One may ask if there are broad subfamilies of #2-QSAT which are considerably easier than #P to compute, and if so, whether such conditions can themselves be easily decided.

1.2 Entanglement and worst case vs. typical counting complexity

Though #2-QSAT is #P-complete, there is a sense in which "generic" instances of #2-QSAT are easily solved. Fix any graph G on n vertices. If we assign a qubit to each vertex, and a term $h_{u,v} = |\eta_{u,v}\rangle\langle\eta_{u,v}|$ for each edge $uv \in G$, where $|\eta_{u,v}\rangle$ is distributed according to the Haar measure, the resulting #2-QSAT instance can be easily solved (except with probability 0) from the structure of G [14, 3]. Specifically, if the graph is a tree, the #2-QSAT instance has value $n + 1$; if the graph has a single cycle, it has value 2; and if it has two or more cycles, it has value zero (*i.e.* it is unsatisfiable, or frustrated as a Hamiltonian).

The apparent reason for this is because a Haar-random state $|\eta_{u,v}\rangle$ is almost certainly entangled. Following Refs. [2, 14, 5], if $h_{u,v}$ and $h_{v,w}$ project onto entangled states $|\eta_{u,v}\rangle$ and $|\eta_{v,w}\rangle$, a single-spin state on u determines the feasible single-spin states at both v and w similarly to an instance of classical 2-XOR-SAT, in which the states of each interacting pair of bits strongly restrict each other. Typical instances of 2-QSAT thus have effective long-range constraints between qubits within any connected component. As a result, any graph which is dense enough to contain multiple cycles almost certainly gives rise to an overconstrained instance of 2-QSAT, corresponding to a frustrated Hamiltonian. This is in contrast to 2-CNF-SAT formulae, which as instances of 2-QSAT have constraints given by standard-basis vectors $|\eta_{u,v}\rangle = |e_u\rangle \otimes |e_v\rangle$ for $e_u, e_v \in \{0,1\}$. Such constraints on qubit-pairs $\{u,v\}$ and $\{v,w\}$ may fail to impose any constraints between the next-nearest neighbour qubits u and w. This is particularly important in the monotone special case of #2-SAT, which corresponds to #2-QSAT instances in which $|\eta_{u,v}\rangle = |00\rangle_{u,v}$ for all edges uv (corresponding to the constraint $x_u \vee x_v$ on boolean strings $x \in \{0,1\}^n$), which is itself #P-complete [17].

1.3 The typical difficulty of #2-QSAT with product constraints

To obtain instances of #2-QSAT which resist solution by polynomial-time algorithms, there must be a substantial chance of obtaining tensor product constraints on each edge. That this does not happen for Haar random constraints (a natural analogue to uniformly random constraints on pairs of bits) is a feature of quantum information theory, but does not shed much light on the range of difficulty of #2-QSAT. We ask: which random graph families, and which distributions of constraints, yield difficult instances of #2-QSAT? Specifically, if only product constraints are involved, when is #2-QSAT likely to be polynomial-time solvable?

We show, both for Erdős–Rényi graphs and for bond-percolated rectangular lattices in two and three dimensions, that difficult instances of #2-QSAT are rare if we select *i.i.d* product constraints from a distribution which differs substantially from monotone constraints. In particular, on bond-percolated lattices, we expect the value of any #2-QSAT instance to be efficiently solvable almost surely; and for Erdős–Rényi graphs, the difficult-to-compute regime vanishes as the "monotonicity" of the constraint distribution decreases.

We state our results more precisely, as follows. A property which holds *asymptotically almost certainly (or surely)* is one which holds with probability $1 - O(1/\mathrm{poly}(n))$. Following the usual terminology associated with the study of random graphs, we often omit the word "asymptotically" in connection with properties which hold almost surely/certainly: statements about discrete distributions which are "almost" certain or sure, are intended to be interpreted in the limit $n \to \infty$. Considering (families of) Hamiltonians on n qubits, we say that a system is *highly disconnected* if its connected subsystems almost surely all have size $O(\log n)$; similarly, if it can almost surely be decomposed into subsystems of size $O(\log n)$ which are independent of one another (despite chains of intermediate interactions), we say that the system is *highly decoupled*. The following Lemma follows easily from the definitions of these terms: we discuss this in Section 2.4.

▶ **Lemma 1.** *Instances of* #2-QSAT *which are highly disconnected, frustrated, or highly decoupled are* easy *(solvable in time $O(\mathrm{poly}\, n)$ on* e.g. *a deterministic Turing machine).*

We consider constraint models interpolating between monotone #2-SAT on one hand, and continuous probability density functions of product constraints on the other. For $f \geqslant 1$, let $\mathbf{q} = (q_1, q_2, \ldots, q_f)$ be a distribution on f distinct single-qubit states $|\alpha_1\rangle, |\alpha_2\rangle, \ldots, |\alpha_f\rangle$, used to generate constraints $|\eta_{u,v}\rangle = |\alpha_u\rangle \otimes |\alpha_v\rangle$, where the factors are independently sampled from $\mathbf{q}$. For example, $\mathbf{q} = (1,0,0,\ldots)$ for monotone 2-SAT, and $\mathbf{q} = (\frac{1}{2}, \frac{1}{2}, 0, \ldots)$ for uniformly-random 2-SAT. If $\mathbf{q} = (1/f, 1/f, \ldots, 1/f, 0, \ldots)$, we have $\|\mathbf{q}\|_2 = \|\mathbf{q}\|_\infty = 1/f$, which approaches 0 as $f \to \infty$; this limiting distribution is precisely that of single-qubit constraints chosen from the Haar measure.[1] Vector norms of $\mathbf{q}$ thus measure how monotone the random constraints typically are. Let $Q_2 = 1 - \|\mathbf{q}\|_2^2$, and let $Q_\infty = 1 - \|\mathbf{q}\|_\infty$.

▶ **Theorem 2** (Erdős–Rényi models). *For an Erdős–Rényi graph on n vertices with $m = \gamma n$ edges, instances of #2-QSAT with $\gamma < \frac{1}{2}$ are almost certainly highly disconnected, and instances with $\gamma > \frac{1}{2Q_2}$ are almost certainly frustrated; while if $2\gamma Q_\infty - \ln(2\gamma) > 1$, frustration-free instances are almost certainly highly decoupled.*

– thus, in the $\mathbf{q} \to \mathbf{0}$ limit, a phase of typically "difficult" problems exists only for $m/n \sim \frac{1}{2}$.

▶ **Theorem 3** (Bond-percolated lattice models). *Let $d \in \{2, 3\}$, and consider a d-dimensional square or cubic lattice on n vertices: a segment of the rectangular grid $\mathbb{Z} \times \mathbb{Z}$ of dimensions $O(\sqrt{n}) \times O(\sqrt{n})$, or of the cubic grid $\mathbb{Z} \times \mathbb{Z} \times \mathbb{Z}$ with dimensions $O(\sqrt[3]{n}) \times O(\sqrt[3]{n}) \times O(\sqrt[3]{n})$, in which edges are present between nearest neighbours independently with some probability p. Let p_c denote the critical percolation probability, at which there asymptotically almost surely exists a component of size $\Omega(n)$. For bond-percolated vertices with m edges, if Q_∞ is bounded away from 0, there is a transition at $\frac{m}{dn} \in \Theta(n^{-1/7})$ from being almost certainly highly disconnected and frustration-free to being almost certainly frustrated. If we condition on frustration-free instances, we find instead that instances for which the percolation probability is subcritical (that is when $\frac{m}{dn} \leqslant p_c$) are almost certainly highly disconnected, while instances for which Q_∞ is greater than some constant $p_{fin} < 1$ (which depends on d) are almost certainly highly decoupled.*

– thus, a typical instance is almost surely solvable in polynomial-time even for $\mathbf{q}$ which deviates from monotonicity by only a finite amount.

The above results suggest that the only difficult instances of #2-QSAT must be specially constructed to resemble monotone instances of #2-SAT. Specifically: **(a)** hard instances of #2-QSAT are atypical, and **(b)** the reason for this does not have to do with entangled constraints, but rather that an instance of #2-QSAT is only likely to be difficult if its constraints are not very diverse and it is relatively sparsely constrained.

Structure of this article

Section 2 contains preliminary definitions and discussion, including types of easily solved instances of #2-QSAT, and techniques to infer long-range constraints and to count solutions to instances of #2-QSAT. Section 3 presents the conditions under which #2-QSAT is easily solvable for instances whose interaction graphs are generated according to either the Erdős–Rényi distribution or percolated rectangular/cubic lattice models. In Section 4 we suggest some ways in which this work might be extended.

2 Preliminaries

We consider *simple graphs*, containing no parallel edges or single-vertex loops. We denote the state-space of a generic qubit by $\mathcal{H}_2 \cong \mathbb{C}^2$, and space of a particular qubit u by $\mathcal{H}_u$. For the sake of brevity we occasionally neglect error terms which are decreasing in n: for instance, we write $f(n) \sim g(n)$ when $f(n) = g(n)\big[1 \pm o(1)\big]$ (which is an equivalence relation) and $f(n) \gtrsim g(n)$ when $f(n) > g(n)\big[1 \pm o(1)\big]$ (which is a quasi-order).

[1] As $\mathbf{q}$ contains no information about the states $|\alpha_j\rangle$, we are glossing over how well-defined the limit $\mathbf{q} \to \mathbf{0}$ is. We do not consider this here, but propose that $|\langle \alpha_j | \alpha_k \rangle| \leqslant 1 - \Omega(1/f)$ for all $j \neq k$ should be sufficient to maintain a promise gap between the ground-state energy level and excited energy levels.

While 2-QSAT allows for a broader range of constraints, in this article we consider only Hamiltonians $H = \sum h_{u,v}$, where $h_{u,v}$ is a rank-1 projector on $\mathbb{C}^2 \otimes \mathbb{C}^2$ and the sum ranges over pairs of vertices $\{u, v\}$ which are adjacent in some graph (usually a typical graph from a given probability distribution on graphs). It should be easy to see by extending the results below that instances of 2-QSAT whose constraints correspond to projectors of rank 2 or more will only increase the probability that the instance is efficiently solvable, by reason of the emergence of long-range constraints on the marginals of satisfying states.

For each rank-1 projector $h_{u,v}$, we consider the state $|\eta_{u,v}\rangle \in \mathcal{H}_u \otimes \mathcal{H}_v$ such that

$$h_{u,v} = |\eta_{u,v}\rangle \langle\eta_{u,v}| \otimes \mathbb{1}_{V \smallsetminus \{u,v\}} \,. \tag{1}$$

For H frustration-free, the operator $\langle\eta_{u,v}|$ is a constraint on any ground-state $|\psi\rangle$ of H: for $\rho_{u,v}$ the density operator of $|\psi\rangle$ on $\{u, v\}$, we have $\langle\eta_{u,v}|\, \rho_{u,v} = 0$ by hypothesis. Thus, as with the classical decision problem 2-SAT, we describe instances of 2-QSAT by a list of local "forbidden" configurations $\langle\eta_{u,v}| : \mathbb{C}^2 \to \mathbb{C}$ on pairs of qubits $u, v \in V$ (implicitly taking the tensor product with the identity on all other qubits) for a global state to avoid.

2.1 Constraint induction

Let $|\Psi^-\rangle \propto |01\rangle - |10\rangle$ be the singlet state. Following Ref. [2], given constraints $\langle\eta_{u,v}| , \langle\eta_{v,w}|$ for $u \neq w$ which both act on a qubit $v \in V$, we may infer a further implicit constraint $\langle\tilde{\eta}_{u,w}|$, such that $\langle\tilde{\eta}_{u,w}|\, \rho_{u,w} = 0$ whenever both $\langle\eta_{u,v}|\, \rho_{u,v} = 0$ and $\langle\eta_{v,w}|\, \rho_{v,w} = 0$ hold:

$$\langle\tilde{\eta}_{u,w}| \;\propto\; \Big[\langle\eta_{u,v}| \otimes \langle\eta_{v,w}|\Big]\Big[\mathbb{1}_u \otimes \big|\Psi^-\big\rangle \otimes \mathbb{1}_w\Big]. \tag{2}$$

We may renormalise $\langle\tilde{\eta}_{u,w}|$ so that $\langle\tilde{\eta}_{u,w}|\tilde{\eta}_{u,w}\rangle = 1$, provided that the operator is non-zero. We may induce further implicit constraints recursively. For two operators $\langle\eta_{u,v}|$ and $\langle\eta_{v,w}|$, we may write the operator obtained via Eqn. (2) by $\langle\eta_{u,v}| * \langle\eta_{v,w}|$. It is easy to show that the binary operator "$*$" is associative, so that

$$\langle\eta_{u,v}| * \langle\eta_{v,w}| * \langle\eta_{w,x}| \;\propto\; \Big[\langle\eta_{u,v}| \otimes \langle\eta_{v,w}| \otimes \langle\eta_{w,x}|\Big]\Big[\mathbb{1}_u \otimes \big|\Psi^-\big\rangle \otimes \big|\Psi^-\big\rangle \otimes \mathbb{1}_x\Big], \tag{3}$$

and so forth for longer chains, so that we may write $\langle\tilde{\eta}_{u,z}| = \langle\eta_{u,v}| * \langle\eta_{v,w}| * \cdots * \langle\eta_{y,z}|$ for an operator acting on $\{u, z\}$ induced by a chain of constraints from the input instance of 2-QSAT. This is similar, in the classical setting, to computing the transitive closure of the implication graph defined by Aspvall, Plass, and Tarjan [1], in which case we may find multiple constraints between a pair of variables which tightly constrain their values. Similarly, in the more general quantum setting, we may obtain multiple constraints $\langle\eta^{(1)}_{u,v}| , \langle\eta^{(2)}_{u,v}| , \ldots$ which may allow us to represent their joint state-space as a two-dimensional subspace $S \leqslant \mathcal{H}_u \otimes \mathcal{H}_v$, allowing us to reduce the number of qubits involved in the problem by a renormalisation step [5] without affecting the dimension of the space of satisfying states $|\psi\rangle$.

With respect to the operation "$*$" of induction of constraints, there are two significantly different constraint types: product constraints $\langle\eta_{u,v}| = \langle\alpha_u| \otimes \langle\beta_v|$, and entangled constraints which do not factor in this manner. It is immediate that for $\langle\tilde{\eta}_{u,w}| = \langle\eta_{u,v}| * \langle\eta_{v,w}|$, the constraint $\langle\tilde{\eta}_{u,w}|$ is a product constraint if either $\langle\eta_{u,v}|$ or $\langle\eta_{v,w}|$ is; and that $\langle\tilde{\eta}_{u,w}| = 0$ only if both $\langle\eta_{u,v}| = \langle\alpha_i| \otimes \langle\alpha_j|$ and $\langle\eta_{v,w}| = \langle\alpha_k| \otimes \langle\alpha_\ell|$ satisfy $|\alpha_j\rangle \propto |\alpha_k\rangle$. When this occurs, the marginal state of u cannot indirectly constrain the marginal on w, or vice-versa, through the interaction with v: by setting v to the state $|\bar{\alpha}_j\rangle$ in the kernel of $\langle\alpha_j|$, we extend any marginal on $\{u, w\}$ to one on $\{u, v, w\}$ which satisfies the constraints $\langle\eta_{u,v}|$ and $\langle\eta_{v,w}|$.

2.2 Randomly generated instances of #2-QSAT

A "random instance" of #2-QSAT is a sample from a probability distribution over instances of #2-QSAT, generally with a fixed number n of qubits and m of constraints. We consider a generation process in which one first generates a random graph, either by selecting a fixed number m of edges from the set of all possible pairs of edges (the *Erdős–Rényi graph model*), or by considering a subgraph of some lattice in which each lattice-edge is included with a probability p such that the expected number of edges is m, associating a qubit to each vertex of the graph. At each edge uv in the random graph, we assign an operator $\langle \eta_{u,v} | : \mathbb{C}^4 \to \mathbb{C}$ according to some probability distribution, representing two-body constraints on the qubits.

We would like to also consider instances of #2-QSAT which are guaranteed to have a non-zero value, corresponding to a distribution on two-body frustration-free Hamiltonians. This requires a subtler random generation procedure. For a model of random graphs (*e.g.* either an Erdős–Rényi model or a percolated lattice model), we select a random order for the edge-set of the graph. Adding these edges sequentially to graph, we assign a constraint to each, restricting the choice of constraint so that the resulting instance of 2-QSAT is satisfiable. In any continuous distribution (such as the Haar measure), any non-trivial restriction of the constraint model typically will be to a set of measure zero; the notion of restriction we intend is limit as $\varepsilon \to 0$, of the Haar measure conditioned on being within an ε-neighbourhood (in the Euclidean norm on $\mathbb{C}^4$) of the valid choices of constraint. (For instance, if only a finite set of constraints avoid making the instance unsatisfiable, such a restriction yields the uniform distribution over those constraints.) For the Haar measure, as well as for the product-constraint model of our article, there is always a choice of constraint for which the instance is satisfiable at each step: this is easy to show in the Haar random case by a minor extension of Ref. [14], and can be established for the constraint model of this article without difficulty (see *e.g.* the beginning of Section 3).

2.3 Remarks on the counting complexity of instances of #2-QSAT

Given a randomly generated instance of #2-QSAT, we ask: with what probability is it a "difficult" instance? Our notion of "difficulty" is defined relative to some fixed algorithm A: a family of instances for which A can successfully compute the answer in polynomial time are "easy", and families for which A has no such upper bound are "difficult". Such statements depend on the state of the art in combinatorics: an improved analysis of random graphs may show that some family of formerly "difficult" instances happen to be solvable by A in polynomial time. If one accepts standard complexity-theoretic assumptions such as $\mathsf{P} \neq \mathsf{NP}$, there are families of instances of 2-QSAT which are inherently "easy" or "difficult" for any algorithm implemented *e.g.* on Turing machines. The aim of this article is to establish bounds on the extent of any such "difficult" regime for certain distributions on #2-QSAT.

An instance of 2-QSAT is *monotone* if there is a state $|\alpha_0\rangle \in \mathbb{C}^2$ such that $\langle \eta_{u,v} | = \langle \alpha_0 | \otimes \langle \alpha_0 |$ for each $uv \in E(G)$. This is equivalent to there being a local unitary operator U such that $\langle \eta_{u,v} | (U \otimes U) = \langle 00 |$ for all $uv \in E(G)$: the classical monotone instances of #2-SAT are a special case in which we may take U to be the identity. As monotone #2-SAT is #P-complete [17], it follows that #2-QSAT is at least #P-hard. Ji, Wei, and Zeng [13] show that #2-QSAT is also contained in #P, by a simple transformation of instances of #2-QSAT which preserves the solution space and puts the interaction graph into a standard form.

Even monotone instances of #2-QSAT may have structural properties which may render it "easy". For instance, instances whose interaction graphs G have bounded *tree-width* [16] (see Ref. [6] for an introductory reference) may be solved in poly(n) time,[2] albeit with a constant factor which grows exponentially with the tree-width [8]. This algorithm is useful in particular for tree graphs or connected graphs which have a single cycle, which respectively have tree-width 1 and 2. Conversely, instances of #2-SAT which are not monotone may still be "difficult": for a fixed graph G, if we assign a uniformly random clause to each $uv \in E(G)$, represented in the format of constraint operators for an instance of #2-QSAT as one of the operators $\langle \eta_{u,v} | \in \bigl\{ \langle 00| , \langle 01| , \langle 10| , \langle 11| \bigr\}$ then the non-trivial constraints arising between pairs of bits by the induction procedure of Eqn. (2) only extend over paths of expected length $O(1)$ in G. Then only for sets of nodes where the constraints are relatively dense can there be a chance of giving rise to long-range constraints of order the size of a given connected component: this is necessary to impose enough structure to obtain an instance of #2-SAT substantially different in complexity from a monotone instance on $n^{O(1)}$ variables.

2.4 Three types of easily solved cases of #2-QSAT

We now remark on the simple observations presented in Lemma 1: this will allow us to reduce the task of proving that instances of #2-QSAT are easy, to showing that they fall into one of three structural classes of Hamiltonian – *frustrated*, *highly disconnected*, or *highly decoupled*, in the senses described preceding Lemma 1.

Following Chvatal and Reed [4] concerning phase transitions in the satisfiability of random instances of 2-CNF-SAT, one may obtain results concerning random classical #2-SAT on Erdős–Rényi graphs with n vertices and m clauses. Specifically, an instance of 2-SAT with density $\frac{m}{n} > 1$ is almost certainly unsatisfiable, and so by definition has value zero as an instance of #2-SAT; and this can be determined in polynomial time by detecting certain unsatisfiable substructures. Similar remarks apply for *frustrated* instances of #2-QSAT: if one can efficiently determine that it is frustrated, this suffices to show that it has value zero.

As for easily solvable instances of #2-SAT with positive values, if $\frac{m}{n} < \frac{1}{2}$, the underlying graph is almost certainly composed of components of size $O(\log n)$ having at most one cycle. One can solve each such component in polynomial time using brute-force techniques (testing all possible assignments for each component); using dynamic programming and taking advantage of the existence of a tree decomposition for the component, one can even solve them in time linear in the component size (up to a logarithmic factor due to handling vertex labels for a graph of size n). These represent a *disconnected* regime in random #2-SAT; and again, similar techniques apply for #2-QSAT if we can establish that the components scale as $O(\log n)$, and/or have treewidth bounded by a constant as we have described above. It then suffices to multiply the #2-QSAT values for each component together: for random graph models (such as the ones we consider) where small components dominate, this may be done efficiently, *e.g.* using an algorithm which we describe in Appendix A.

Finally, we may consider *highly decoupled* instances, in which a subsystem which is contiguous nevertheless decomposes into independent subsystems of size $O(\log n)$. These may arise in instances which have been constructed to be frustration-free, due to the proliferation of qubits whose states are "fixed" by their constraints. When a qubit x can only occupy a unique state in a satisfying state, we refer to this as the *fixed state* of the qubit x (which we

[2] The approach here, for instances having tree-width at most $w > 0$, is essentially to use dynamic programming to count the partially-satisfying solutions for each of 2^w possible assignments (in some local basis) for each qubit indexed by a vertex in a tree-decomposition. A more complete description can be found in [8].

denote $|\bar{\psi}_x\rangle$). As we add constraints to a satisfiable instance of 2-QSAT, there are at least two ways in which an added constraint can increase the number of qubits with fixed states:

- either by adding a constraint $\langle\eta_{x,y}|$ between some qubit x, and a qubit y which already has a fixed state such that $\langle\eta_{x,y}|\bigl(\mathbb{1}_x \otimes |\bar{\psi}_y\rangle\bigr) \neq \mathbf{0}^\dagger$,
- or by adding a constraint which closes a chain of constraints starting and ending at x, which is only satisfiable by a single state $|\bar{\psi}_x\rangle$.

Any constraint $\langle\eta_{x,y}|$ acting on a qubit x with a fixed state will either be satisfied by $|\bar{\psi}_x\rangle$ regardless of the state of y, or will serve to fix the state of y. Thus, interactions between qubits with fixed states with *non*-fixed qubits will, by construction, fail to give rise to any long-range constraints between qubits without fixed states. If there are enough qubits with fixed states, these may then effectively partition the set of *non*-fixed qubits into independent subsystems; if these subsystems are of size $O(\log n)$, the system is then *highly decoupled*. Thus, to solve an instance of #2-QSAT, it also suffices to identify enough fixed qubits to partition the remainder into systems whose degeneracy may be efficiently computed.

Our result is to show how in two different random graph models, for random instances of 2-QSAT with enough diversity in the constraints to differ substantially from monotone instances, there is (at most) a narrow range in which the density of constraints may give rise to instances which are neither highly disconnected, nor frustrated, nor highly decoupled almost surely.

3 Discrete probabilistic models

We consider a constraint model of *independent factor distributions*, in which constraints are product operators $\langle\alpha|\otimes\langle\beta|$ for some *i.i.d.* single-qubit operators $\langle\alpha|, \langle\beta| : \mathbb{C}^2 \to \mathbb{C}$ distributed over some set of operators $\{\langle\alpha_1|, \langle\alpha_2|, \ldots, \langle\alpha_f|\}$ for some $f \geqslant 1$, where $\langle\alpha_j| \not\propto \langle\alpha_k|$ for $j \neq k$. Given an edge which represents a product constraint, the probability of obtaining $\langle\eta_{u,v}| = \langle\alpha_h|_u \otimes \langle\alpha_j|_v$ is given by $q_h q_j$, where $\mathbf{q} = (q_1, q_2, \ldots, q_f)$ is a fixed probability distribution. Throughout the following, we suppose that $1 > q_1 \geqslant q_2 \geqslant \cdots \geqslant q_f > 0$, so that there is some probability of obtaining non-monotone instances of 2-QSAT.

Independent factor distributions have convenient features for analysis. Following Ref. [13], the ground-state manifold for an instance of 2-QSAT having only product constraints has a basis consisting of product states. Furthermore, non-zero induced constraints $\langle\eta_{u,v}| * \langle\eta_{v,w}|$ range over the same two-qubit operators as the individual edge-constraints themselves (albeit with a different probability distribution than $\mathbf{q}\otimes\mathbf{q}$). As with Haar-random models, when we wish to consider only random *frustration-free* Hamiltonians, we must specially select the constraints to meet that restriction. We construct the random graph in the same manner as described in Section 2.2, this time restricting the choice of constraints according to the condition of not giving rise to a frustrated (*i.e.* an unsatisfiable) instance of 2-QSAT. Frustration can only arise if both qubits on which the constraint are each restricted to some "fixed" state to satisfy the earlier constraints placed on it: a "non-frustrating" choice of constraint can then be made simply by having it be satisfied by one of the two fixed states.

We may consider how likely long-range constraints (as described in Section 2.1) are for such a constraint model. Let $x_0, x_\ell \in V(G)$ be two vertices connected by a path $P = x_0 x_1 \cdots x_\ell$ in the interaction graph of a random instance of #2-QSAT. We may consider what constraints may exist on the joint state of x_0 and x_ℓ by virtue of the inducted constraint $\mathcal{C}_P = \langle\eta_{x_0,x_1}|*\langle\eta_{x_1,x_2}|*\cdots*\langle\eta_{x_{\ell-1},x_\ell}|$. One may show by induction that $\mathcal{C}_P$ is non-zero if and only if $\langle\eta_{x_{h-1},x_h}| * \langle\eta_{x_h,x_{h+1}}| \neq 0$ for each index $0 < h < \ell$ of internal vertices of the path. For each such h, we have $\langle\eta_{x_{h-1},x_h}| * \langle\eta_{x_h,x_{h+1}}| = 0$ if and only if $\langle\eta_{x_{h-1},x_h}| = \langle\alpha_i| \otimes \langle\alpha_j|$

and $\langle \eta_{x_{h-1},x_h} | = \langle \alpha_k | \otimes \langle \alpha_\ell |$ for some $j \neq k$. Because the right-factor of $\langle \eta_{x_{h-1},x_h} |$ and the left-factor of $\bigl\langle \eta_{x_h,x_{h+1}} \bigr|$ are independently distributed, this occurs with probability

$$Q_2 \;:=\; 1 - \|\mathbf{q}\|_2^2 \;=\; \sum_{j=1}^{f} q_j (1 - q_j) \;\leqslant\; 1 - \frac{1}{f}\,, \tag{4}$$

with equality if and only if $\mathbf{q}$ is uniform. Note that $Q_2 > 0$, where the lower bound is the infimum as $\mathbf{q} \to (1, 0, 0, \ldots)$. As the probabilities of having identical factors at each vertex are independent, we then have

$$\Pr\Bigl[\mathcal{C}_P \neq 0\Bigr] \;=\; \prod_{h=1}^{\ell-1} \bigl(1 - \|\mathbf{q}\|_2^2\bigr) \;=\; Q_2^{\ell-1}\,. \tag{5}$$

Thus, $\mathcal{C}_P$ is non-zero and proportional to $\langle \alpha_h | \otimes \langle \alpha_j |$ with probability $q_h q_j Q_2^{\ell-1}$ for each $1 \leqslant h, j \leqslant f$, and equal to zero with probability $1 - Q_2^{\ell-1}$. Because the long-range constraints which involve a particular vertex as a *mid-point* are not independent of one another, it may be useful in some cases to bound this probability from below by $Q_\infty^{\ell-1}$, where $Q_\infty = 1 - \|\mathbf{q}\|_\infty$, where $\|\mathbf{q}\|_\infty = q_1$ is an upper bound on the probability that the single-qubit operators $\langle \alpha_j | , \langle \alpha_k |$ with which two different constraints act on x are the same.

3.1 Erdős–Rényi interaction graphs

The attenuation of the probability of long-range constraints described in Eqn. (5) is similar to what occurs in uniformly random 2-SAT. For Erdős–Rényi interaction graphs on n vertices and m edges – a distribution on labelled graphs which may be sampled by listing each of the $\binom{n}{2}$ potential edges in a random order, and selecting the first m edges for inclusion – this motivates an analysis which follows closely to that of Chvatal and Reed [4], adapting it for counting problems and to involve more general constraint distributions. We show that, except for a "difficult phase" in the regime $\frac{1}{2} \leqslant \frac{m}{n} \leqslant \frac{1}{2Q_2}$, a random instance of #2-QSAT is almost certainly either highly disconnected or frustrated, according to whether $\frac{m}{n}$ is below or above the boundaries of the difficult phase. In particular, the difficult phase shrinks to a band of zero width at $\frac{m}{n} \sim \frac{1}{2}$ as $Q_2 \to 1$. In the special case of frustration-free instances, this band expands to $\frac{1}{2} \leqslant \frac{m}{n} \leqslant \frac{1}{2Q_\infty}(1+\delta)$ for some small δ which vanishes as $Q_\infty \to 1$; this band also converges to $\frac{m}{n} \sim \frac{1}{2}$ as $Q_\infty \to 1$. Thus in the "completely non-monotonic" limit $\mathbf{q} \to \mathbf{0}$, #2-QSAT is always easy; and there is a substantial band of instances which may be difficult to solve only if the constraint distribution shows a corresponding bias towards a small, finite number of constraints.

3.1.1 The highly disconnected phase in Erdős–Rényi models

Whether or not we restrict to frustration-free instances of 2-QSAT, the existence of a highly disconnected regime in instances of 2-QSAT on Erdős–Rényi graphs G follows directly from the random graph model itself. For $\frac{m}{n} < \frac{1}{2}$, almost certainly G contains only components of size $O(\log n)$, and almost certainly contains no components having more than one cycle [7]. Any instance of 2-QSAT on such a graph will thus be highly disconnected, regardless of the constraint distribution. For our results on Erdős–Rényi models, it thus suffices to establish upper bounds on the extent of any difficult phase.

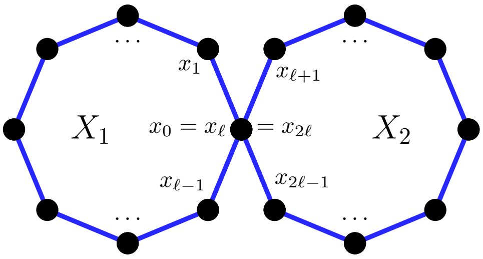


Figure 1 Example of a "figure eight" graph on $2\ell - 1$ vertices, for $\ell = 8$. By Eqn. (7), the probability of such a graph describing a frustrated figure-eight subsystem scales as $O\bigl(Q_2^{2\ell}\bigr)$.

3.1.2 The frustrated phase in unconditional Erdős–Rényi models

For a random graph with $m \in \Omega(n)$ edges, we adapt the analysis of Chvatal and Reed [4, Theorem 4] to consider the probability that the giant component Γ contains a "frustrated figure eight" (corresponding to a "snake" in Ref. [4]): a subsystem X such that

1. Its interaction graph contains a *figure eight graph*, which we define as a pair of cycles $X_1 = x_0 x_1 \cdots x_{\ell-1} x_\ell$ and $X_2 = x_\ell x_{\ell+1} \cdots x_{2\ell-1} x_{2\ell}$ of the same length, where $x_0 = x_\ell = x_{2\ell}$, and where X_1 and X_2 intersect only at the vertex $x_0 = x_\ell$. (See Fig. 1 for an example.) There may be additional edges connecting vertex-pairs $x_j x_k$ (though these will typically be unlikely), and $X = X_1 \cup X_2$ may be connected to other vertices.
2. For each $0 \leqslant j < 2\ell$, the constraints $\bigl\langle \eta_{x_j, x_{j+1}} \bigr| = \langle \beta_j | \otimes \langle \gamma_j |$ satisfy $\langle \gamma_j | \neq \langle \beta_{j+1} |$.
3. We have $\bigl\{ \langle \beta_0 | , \langle \gamma_{\ell-1} | \bigr\} \cap \bigl\{ \langle \beta_\ell | , \langle \gamma_{2\ell-1} | \bigr\} = \varnothing$, so that the constraints imposed by X_1 and X_2 on their common spin x_0 are not simultaneously satisfiable.

The cycles X_1 and X_2 are either "alternating loops" or "quasi-alternating loops" in the terminology of Ref. [13], and impose constraints on x_0 is which cannot be simultaneously satisfied. Thus a frustrated figure eight is unsatisfiable by construction. We consider the probability of a *large* frustrated figure-eight arising in a random instance of 2-QSAT with constraints given by an independent factor distribution, which in particular implies that it is part of the largest contiguous subsystem of the Hamiltonian.

In a system with a figure-eight subgraph, the probability of $\langle \gamma_{j-1} | \neq \langle \beta_j |$ is simply Q_2 for each of the $2\ell - 2$ sites x_j of the two cycles, excluding the shared vertex $x_0 = x_\ell = x_{2\ell}$. The conditions at the node x_ℓ, where we require $\langle \beta_0 | = \langle \gamma_{\ell-1} | \neq \langle \beta_\ell | = \langle \gamma_{2\ell-1} |$, occur with a probability Q_{crux} which also depends only on $\mathbf{q}$. (By a routine calculation, one may show that

$$\begin{aligned} Q_{\text{crux}} &= \sum_h q_h \Biggl(\Biggl[q_h \sum_{j,k \neq h} q_j q_k \Biggr] + \Biggl[\sum_{i \neq h} q_i \sum_{j,k \notin \{h,i\}} q_j q_k \Biggr] \Biggr) \\ &= 1 - 4\|\mathbf{q}\|_2^2 + 2\|\mathbf{q}\|_2^4 + 4\|\mathbf{q}\|_3^3 - 3\|\mathbf{q}\|_4^4 \,. \end{aligned} \tag{6}$$

Then $Q_{\text{crux}} \to 1$ as $\mathbf{q} \to \mathbf{0}$, and $Q_{\text{crux}} \in \Theta(1)$ for $\|\mathbf{q}\|_\infty$ bounded away from 1.) Given a fixed figure-eight graph X on $2\ell - 1$ vertices, the probability that it gives rise to a *frustrated* figure-eight system is then

$$\Pr\Bigl[X \text{ a frustrated subsystem} \Bigr] = Q_2^{2\ell-2} Q_{\text{crux}}. \tag{7}$$

Let $m = \gamma n$ for some constant $\gamma > 0$. Using a second moment probabilistic argument, adapting the proof of Ref. [4, Theorem 4], we show that the largest contiguous subsystem almost certainly contains a frustrated figure eight so long as $\gamma > \frac{1}{2Q_2}$.

Let φ_ℓ denote the number of frustrated figure eight subsystems in G on $2\ell - 1$ vertices. The mean $\mathbb{E}(\varphi_\ell)$ over all random graphs on n vertices and m edges can be evaluated by considering all sets S of $2\ell - 1$ vertices, and summing the probability of S being such figure eight subsystem for all such subsets. We will make use of the equality

$$\frac{n!}{(n-t)!} \;\sim\; n^t \exp\bigl(-\alpha(n,t)\bigr), \qquad \text{where } \; \alpha(n,t) := t + (n - t + \tfrac{1}{2}) \ln\!\left(1 - \frac{t}{n}\right) \tag{8}$$

which holds for $t \in o(n)$,[3] ignoring a relative error term of $O(\frac{1}{n})$ using the notation defined at the beginning of Section 2. By considering (i) the number of ways that we may choose the common vertex, (ii) the number of distinguishable ways that we may construct two cycles on ℓ vertices (built in either order) which incorporate the common node, (iii) the number of ways of allocating the remaining edges after having built X, and (iv) the probability that X is a frustrated figure eight given that it is present in the random graph, we may obtain

$$\begin{aligned}\mathbb{E}(\varphi_\ell) \;&=\; Q_2^{2\ell-2} Q_{\mathrm{crux}} \cdot \frac{n}{2} \left[\tfrac{1}{2}\binom{n-1}{\ell-1}(\ell-1)!\right] \left[\tfrac{1}{2}\binom{n-\ell}{\ell-1}(\ell-1)!\right] \left[\frac{\dbinom{\binom{n}{2} - 2\ell + 1}{m - 2\ell + 1}}{\dbinom{\binom{n}{2}}{m}}\right] \\ &\sim\; \frac{Q_{\mathrm{crux}}}{8Q_2}\left(\frac{2Q_2 m}{n}\right)^{2\ell-1} \frac{\exp\Bigl(-\alpha\bigl(n, 2\ell-1\bigr) + \alpha\bigl(\binom{n}{2}, 2\ell-1\bigr)\Bigr)}{\exp\Bigl(\alpha\bigl(m, 2\ell-1\bigr)\Bigr)}\,.\end{aligned} \tag{9}$$

For $\ell \in o(n^{1/2})$, we have $\alpha(n, 2\ell - 1) \in o(1)$; then we can easily show that $\varphi_\ell > 0$ with non-zero probability, provided that $m = \frac{1}{2Q_2}\bigl(1 + \Omega(\frac{1}{\ell})\bigr)n$.

Next, we show that φ_ℓ almost surely doesn't differ substantially from its mean. Define a random variable $\varphi_X \in \{0, 1\}$ such that $\varphi_X = 1$ for instances of 2-QSAT whose constraint subgraph contains a frustrated figure-eight on a given subgraph X. We compare $\mathbb{E}(\varphi_X)^2$ against $\mathbb{E}(\varphi_X\varphi_Y)$, where $X = x_0x_1\cdots x_{2\ell-1}x_0$ and $Y = y_0y_1\cdots y_{2\ell-1}y_0$ are both figure-eight graphs on $2\ell - 1$ vertices, but which may have vertices and edges in common. By definition, we have $\mathrm{Var}(\varphi_\ell) = \mathbb{E}(\varphi_\ell^2) - \mathbb{E}(\varphi_\ell)^2$. We have

$$\mathbb{E}(\varphi_\ell) \;=\; \sum_X \Pr\bigl[\varphi_X = 1\bigr], \qquad\qquad \mathbb{E}(\varphi_\ell^2) \;=\; \sum_{X,Y} \Pr\bigl[\varphi_X\varphi_Y = 1\bigr], \tag{10}$$

where we sum over all possible figure-eight subgraphs X, Y on $2\ell - 1$ vertices selected from n vertices. We show that $\mathbb{E}(\varphi_\ell^2) \sim \mathbb{E}(\varphi_\ell)^2$, which implies that $\mathrm{Var}(\varphi_\ell) \in o\bigl(\mathbb{E}(\varphi_\ell)^2\bigr)$.

Consider the probability that a given subgraph g on t edges occurs as a subgraph of G. Accounting for how we can distribute t edges among the first m elements of a random sequence of edges, we have

$$f(t) \;:=\; \Pr\bigl[g \subseteq G\bigr] \;=\; \binom{m}{t} t! \left[\frac{\bigl(\binom{n}{2} - t\bigr)!}{\binom{n}{2}!}\right] \;\sim\; \left(\frac{2\gamma}{n}\right)^t \exp\Bigl(\alpha\bigl(\tbinom{n}{2}, t\bigr) - \alpha(\gamma n, t)\Bigr)\,. \tag{11}$$

We suppose that $\ell \in o(n^{1/2})$, so that $f(2\ell + \delta(\ell)) \;\sim\; (2\gamma/n)^{2\ell+\delta(\ell)}$ for $\delta(\ell) \in \pm O(\ell)$, again using $\mathrm{e}^{\alpha(N,t)} \sim 1$ for $t \in o(N^{1/2})$. For figure-eight subgraphs X, Y on $2\ell - 1$ vertices

[3] This may be easily recovered using Stirling's approximation.

each, write $\Phi(X,Y) := \Pr\bigl[\varphi_X\varphi_Y = 1 \,\big|\, X \cup Y \subseteq G\bigr]$ for the probability of the frustration conditions on $X \cup Y$. Then if $\bigl|E(X) \cap E(Y)\bigr| = i$,

$$\Pr\Bigl[\varphi_X\varphi_Y = 1\Bigr] \;=\; \Pr\Bigl[X \cup Y \subseteq G\Bigr]\Phi(X,Y) \;=\; f(4\ell - i)\Phi(X,Y). \tag{12}$$

For X fixed, define $\Phi_i(X)$ to be the sum of $\Pr\bigl[\varphi_X\varphi_Y = 1\bigr]$ over all figure-eight subgraphs Y of the same size, for which $|E(X) \cap E(Y)| = i$ as above (*i.e.* the probability of obtaining two frustrated figure-eight subsystems which intersect in this way, one of which is X). The probability of having any pair of isomorphic frustrated figure eight subgraphs, of which one is X, is then given by $\Phi(X) := \sum_i \Phi_i(X)$.

We may show that for a fixed X, the contribution of $\Phi_0(X)$ is the only significant contribution to $\Phi(X)$. Note that if none of the edges of X and Y overlap, the frustration conditions for X and for Y are completely independent, even if X and Y share vertices: that is, $\Phi(X,Y) = \bigl[Q_2^{2\ell-2}Q_{\mathrm{crux}}\bigr]^2$ in this case. We can then upper bound $\Phi_0(X)$ roughly by removing the restriction on Y that $X \wedge Y$ have no edges. Let $F_{2\ell-1}$ denote the number of possible frustrated figure eight graphs on $2\ell - 1$ vertices selected from n vertices: then

$$\Phi_0(X) \;<\; \sum_Y f(4\ell)\Phi(X,Y) \;\sim\; F_{2\ell-1}\,\bigl(2\gamma/n\bigr)^{4\ell}\,Q_2^{4\ell-4}Q_{\mathrm{crux}}^2. \tag{13a}$$

For all other $0 < i \leqslant 2\ell$, we consider the number $N(i,j)$ of figure-eight subgraphs Y on $2\ell - 1$ vertices, for which $X \wedge Y$ has i edges and j vertices, and consider an upper bound $\Phi(i,j)$ for the frustration probabilities $\Phi(X,Y)$ for all such subgraphs Y. Then we have

$$\Phi_i(X) \;\leqslant\; \sum_j N(i,j)f(4\ell - i)\Phi(i,j) \tag{13b}$$

for $i > 0$. We bound the parameters $\Phi(i,j)$ and $N(i,j)$ by considering bounds on the frustration conditions holding at each site in $X \cup Y$, and by considering how the number of components in $X \wedge Y$ affects both $N(i,j)$ and the probability of all the local frustration conditions holding.

Local frustration conditions

If X and Y intersect at all, the probabilities of the frustration conditions holding for any shared vertex only differs from what it would be independently for X and for Y if they also share edges. For instance, if $x_j = y_k$ for $j,k \notin \{0,\ell,2\ell\}$, and $e_{x,j}, e_{x,j+1} \notin E(Y)$, then the frustration conditions for X and for Y at x_j are independent of one another and obtain with probability Q_2^2, as if x_j and y_k were actually distinct vertices. Similarly, if $x_j = y_k$ for $j,k \notin \{0,\ell,2\ell\}$, and $e_{x,j}, e_{x,j+1} \in E(Y)$, then the frustration conditions are identical and they obtain with probability Q_2. The most interesting cases are for the "crux" vertices x_ℓ and y_ℓ, and for the "junction" vertices of degree 3 in $X \cup Y$ arising from $x_j = y_k$ for $j,k \notin \{0,\ell,2\ell\}$.

- Vertices in $X \cup Y$ of degree 3 correspond to vertices $x_j = y_k$ for $j,k \notin \{0,\ell,2\ell\}$, where one of the edges $e_{x,j}$ or $e_{x,j+1}$ is equal to one of the edges $e_{y,k}$ or $e_{y,k+1}$. To satisfy the frustration conditions, the common edge of X and Y which is adjacent to x_j must act on x_j differently from the remaining two edges, but the other two edges may act on x_j in either distinct or identical ways to each other. Routine calculation shows that the probability of this occurring is $Q_{\mathrm{junct}} := \|\mathbf{q}\|_2^2 - \|\mathbf{q}\|_3^3$.
- The probability that the frustration conditions for X holds at x_ℓ, when $x_\ell = y_k$ for some $0 < k < 2\ell$, may be somewhat complicated if some of the edges of Y incident to y_k

overlap some of the edges $\{e_{x,1}, e_{x,\ell}, e_{x,\ell+1}, e_{x,2\ell}\}$ incident to x_ℓ. Similar remarks apply to the other crux vertex y_ℓ. As there are at most two crux vertices in $X \cup Y$, we may ultimately subsume the probability that these conditions hold at x_ℓ or at y_ℓ as a constant factor, and simply bound the probability from above by 1.

Vertex types and simultaneous frustration

The probability of X and Y both being frustrated depends on the number of junction vertices, crux vertices, and other vertices in $X \cup Y$, which are closely related to the number of components. Extending the observation made with respect to the probability of frustration conditions holding at the crux vertices, we adopt an approach of avoiding case analysis, by sweeping various scalar factors under the rug when they depend only on a constant number of vertices. To do so, we define a scalar factor c (which we do not explicitly calculate) to bound from above any contributions by constant factors in the various cases.

In most cases, the components of $X \wedge Y$ (if it is non-empty) will consist of paths, and possibly one non-path tree component in the case that $x_\ell = y_\ell$ (with at least three of the edges of X and Y overlapping at that vertex). In rare cases, $X \wedge Y$ may have a component which contains an entire cycle, or indeed two cycles if $X = Y$. In the typical case where $X \wedge Y$ is cycle-free, the number of components will be the difference $j - i$; Otherwise, $X \wedge Y$ has one or two cycles, so that it has $j - i + 1$ or $j - i + 2$ components. In any case, the number of components is $j - i + O(1)$. We may then make the following remarks concerning vertices of different types:

- As we note above, $X \cup Y$ has at most $O(1)$ distinct crux vertices, for which frustration conditions occur with constant probability regardless of the number of edges of X and Y which overlap at those vertices.
- The number of junction vertices is minimized when each component of $X \wedge Y$ is a path segment, with each component having two junction vertices at its endpoints; the largest number of junction vertices a component may have is four, in the case that the the two crux vertices coincide so that one component of $X \wedge Y$ has four leaf nodes. (Three junction nodes are possible as well if the two crux nodes coincide, but where only three of the edges of X and of Y coincide.) Thus the number of junction vertices is $2(j-i)+O(1)$ in all cases.
- The frustration conditions elsewhere are governed by edge-pairs meeting at some vertex, where either both edges are common to X and Y or both belong to one figure-eight graph X and Y (the same one), but not to both. Considering the edges x_0x_1, x_1x_2, *etc.* in sequence and pairing each with the one that follows it, we may count these edge-pairs by considering those edges x_jx_{j+1} for which x_{j+1} is not a junction or crux vertex. The number of edges in X which meet at non-junction, non-crux vertices is $2\ell - 2(j-i) - O(1)$, and similarly for Y; and the number of such edges in $X \wedge Y$ is $i - 2(j-i) - O(1)$, yielding a total of $4\ell - 2(j-i) - i \pm O(1)$.

Thus for $0 < i < 2\ell$ we have

$$\Phi(X,Y) \leqslant cQ_2^{4\ell-i}\left(\frac{Q_\text{junct}}{Q_2}\right)^{2(j-i)} \tag{14a}$$

for some constant c depending only on the probability distribution $\mathbf{q}$ of constraint probabilities. For $i = 2\ell$, we have $X = Y$ and $j = 2\ell - 1$: then following Eqn. (7) we may explicitly evaluate

$$\Phi(X,X) = \Pr\big[\varphi_X = 1 \,\big|\, X \subseteq G\big] = Q_\text{crux}Q_2^{2\ell-2}. \tag{14b}$$

Ways to overlap at i edges

Following the analysis of Ref. [4], we may bound $N(i,j)$ by considering upper bounds on (*i*) the number of ways a fixed shape for the graph $X \wedge Y$ could be mapped injectively into X and into Y, (*ii*) the number of ways that the components of $X \wedge Y$ could be arranged into the vertex-order of Y, and (*iii*) the number vertices which may belong to $Y \smallsetminus (X \wedge Y)$. The number of subgraphs Y such that $X \wedge Y$ has i edges and j vertices can then be bounded by

$$\begin{aligned} N_\ell(i,j) &< 4\binom{2\ell+2}{2j-2i+2}^2 \ell\,(j-i)!\,2^{j-i}\,n^{2\ell-j-1} \\ &\leqslant 4\ell(2\ell+2)^{4(j-i)+4}\,2^{j-i}\,n^{2\ell-j-1} \end{aligned} \tag{15a}$$

in the case $0 < i < \ell$, and

$$\begin{aligned} N_\ell(i,j) &< 4\binom{2\ell+2}{2j-2i+2}^2 \ell\,(j-i)!\,2^{j-i+1}\,n^{2\ell-j-1} \\ &\leqslant 8\ell(2\ell+2)^{4(j-i)+4}\,2^{j-i}\,n^{2\ell-j-1} \end{aligned} \tag{15b}$$

for $0 < i < 2\ell$ more generally. If for the sake of brevity we define $\Lambda = 2(2\ell+2)^4/n$, we then have

$$N_\ell(i,j) \leqslant \begin{cases} 2\ell\Lambda^{j-i+1}n^{2\ell-i} & \text{if } 0 < i < \ell, \\ 4\ell\Lambda^{j-i+1}n^{2\ell-i} & \text{if } \ell \leqslant i < 2\ell. \end{cases} \tag{15c}$$

Again, we have $X = Y$ if $i = 2\ell$, so that $N_\ell(2\ell, j) = 1$.

Suppose that $\ell \in o(n^{1/4})$, so that $\Lambda \in o(1)$. We may then use the above remarks to bound $\Phi_i(X)$ for $i > 0$. For $0 < i < \ell$, the graph $X \wedge Y$ has no cycles, so that $i+1 \leqslant j \leqslant 2\ell - 1$; we may then bound

$$\begin{aligned} \Phi_i(X) &\leqslant \sum_{j=i+1}^{2\ell-1} N(i,j)\Phi(i,j)f(4\ell-i) \\ &< \sum_{j=i+1}^{2\ell-1} \Bigl[2\ell\Lambda^{j-i+1}n^{2\ell-i}\Bigr] \left[cQ_2^{4\ell-i}\left(\frac{Q_\text{junct}}{Q_2}\right)^{2(j-i)}\right] (2\gamma/n)^{4\ell-i} \\ &= 2c\ell\Lambda n^{2\ell-i}Q_2^{4\ell-i}\,(2\gamma/n)^{4\ell-i} \sum_{j=i+1}^{2\ell-1}\left(\frac{\Lambda Q_\text{junct}^2}{Q_2^2}\right)^{j-i} \\ &< 2c\ell\Lambda n^{-2\ell}\,(2\gamma Q_2)^{4\ell-i}\left(\frac{\Lambda Q_\text{junct}^2}{Q_2^2}\right)\left(\frac{1}{1-\Lambda Q_\text{junct}^2 Q_2^{-2}}\right) \\ &\sim \left(\frac{2cQ_\text{junct}^2}{Q_2^2}\right)\ell\Lambda^2 n^{-2\ell}\,(2\gamma Q_2)^{4\ell-i}\,. \end{aligned} \tag{16a}$$

For $\ell < i < 2\ell$, we may only bound $i \leqslant j \leqslant 2\ell - 1$, and for $i = 2\ell$ we have $j = 2\ell - 1 = i - 1$; we may then obtain similar bounds

$$\Phi_i(X) \lesssim 4c\ell\Lambda n^{-2\ell}\,(2\gamma Q_2)^{4\ell-i} \qquad \text{for } \ell \leqslant i < 2\ell, \tag{16b}$$

$$\Phi_{2\ell}(X) \sim Q_\text{crux}Q_2^{-2}n^{-2\ell}\,(2\gamma Q_2)^{2\ell} \qquad \text{for } i = 2\ell. \tag{16c}$$

Expanding the formulas for $\Phi_i(X)$ for $i > 0$ and eliding the constant factors, we may obtain

$$\Phi(X) = \Phi_0(X) + \ell n^{-2\ell}(2\gamma Q_2)^{4\ell} O\Biggl(\Lambda^2 \sum_{i=1}^{\ell-1}(2\gamma Q_2)^{-i} + \Lambda \sum_{i=\ell}^{2\ell-1}(2\gamma Q_2)^{-i} + \ell^{-1}(2\gamma Q_2)^{-2\ell}\Biggr). \tag{17}$$

For $\ell \in \omega(1)$, the asymptotic expression of the previous equation is bounded from above by $O(\Lambda^2)$, provided that $\mathrm{poly}(\ell)(2\gamma Q_2)^{-\Theta(\ell)} \subseteq o(1)$. For the latter to hold, it suffices that $2\gamma Q_2 - 1 \in \omega\bigl(\ell^{-1}\log(\ell)\bigr)$. We then obtain the upper bound

$$\Phi(X) \;=\; \Phi_0(X) \;+\; O\Bigl(\ell \Lambda^2 n^{-2\ell}(2\gamma Q_2)^{4\ell}\Bigr). \tag{18}$$

We may show that $\Phi(X) = \Phi_0(X)\bigl[1 + o(1)\bigr]$: using Eqn. (8), we may estimate

$$\begin{aligned} F_{2\ell-1} \;&=\; n \cdot \frac{1}{2}\left[\tfrac{1}{2}\binom{n-1}{\ell-1}(\ell-1)!\right]\left[\tfrac{1}{2}\binom{n-\ell}{\ell-1}(\ell-1)!\right] \\ &=\; \frac{n!}{8(n-2\ell+1)!} \;\sim\; \tfrac{1}{8}n^{2\ell-1}, \end{aligned} \tag{19}$$

so that we have

$$\Phi_0(X) \;\lesssim\; F_{2\ell-1}\,\bigl(2\gamma/n\bigr)^{4\ell}\,Q_2^{4\ell-4}Q_{\mathrm{crux}}^2 \;=\; \left(\frac{Q_{\mathrm{crux}}^2}{8Q_2^4}\right) n^{-2\ell-1}\,(2\gamma Q_2)^{4\ell}, \tag{20}$$

whereas by $\ell \in o(n^{1/9})$ and $\Lambda \in \Theta(\ell^4/n) \subseteq o(n^{-5/9})$ we have

$$O\Bigl(\ell\Lambda^2 n^{-2\ell}(2\gamma Q_2)^{4\ell}\Bigr) \;\subseteq\; o\Bigl(n^{-2\ell-1}(2\gamma Q_2)^{4\ell}\Bigr). \tag{21}$$

We then have $\Phi(X) \sim \Phi_0(X)$ as promised. Thus we have $\mathbb{E}(\varphi_\ell^2) \sim \mathbb{E}(\varphi_\ell)^2$, so that $\mathrm{Var}(\varphi_\ell) \in o(\mathbb{E}(\varphi_\ell)^2)$. By Chebyshev's inequality, the probability that φ_ℓ^2 varies from its mean by $\omega(\mathrm{Var}(\varphi_\ell))$ is zero; then in particular φ_ℓ is almost surely greater than 1 provided that $\mathbb{E}(\varphi_\ell) > 1$.

Frustrated subsystems may be efficiently detected when they are present, as follows. For each vertex $x \in V(G)$, constraint-pair $(\langle\alpha_h|\,, \langle\alpha_j|)$, and $\ell > 1$, we may enumerate the number of alternating paths (in the terminology of Ref. [13]) of length ℓ which begin an end at x whose first constraint is of the form $\langle\alpha_h| \otimes \langle\gamma|$ and whose final constraint is of the form $\langle\beta| \otimes \langle\alpha_j|$. We may do so by traversing all alternating paths starting at x by a breadth-first search, and noting at each step whether in one step we may reach a visited vertex which could be used to close an alternating path back to x. Any one such path represents an alternating or quasi-alternating loop at x. If for any $\ell > 1$ there are two such loops with inconsistent constraints, then the constraints at x are unsatisfiable. Exploring all of the alternating paths from x for any one constraint pair $(\langle\alpha_h|\,, \langle\alpha_j|)$ can be done in time $O(m)$; doing so for all constraint-pairs and all $x \in V(G)$ can be done in time $O(nmf^2)$. The frustrated pair of constraints may not represent a frustrated figure eight (*e.g.* if the alternating paths starting and ending at x are of different lengths), but nevertheless serve to certify that the instance of #2-QSAT is frustrated, and are present for all frustrated instances.

Thus for $m \geqslant \frac{1+\varepsilon}{2Q_2}n$ for positive $\varepsilon \in \omega(n^{-1/9}\log(n))$, an instance of 2-QSAT constructed on G selected according to the Erdős–Rényi distribution will be frustrated almost surely, due to the presence of multiple frustrated figure-eight subsystems of size $O(\mathrm{poly}(n))$. Furthermore, one may determine that such frustrations exist in polynomial time, when they are present.

3.1.3 The highly decoupled phase in frustration-free Erdős–Rényi models

In constructing frustration-free instances of 2-QSAT from a discrete distribution, we may suppose that constraints are repeatedly sampled for each new constraint until we obtain one which does not render the instance unsatisfiable. Any constraint which on the first "try" would have resulted in a frustrated instance, we call a *would-be* frustration. We may then consider the structures in the Hamiltonian which *would have* arisen, had we taken the constraint which was first selected for any interaction, and thus speak counterfactually of such features as "would-be" frustrated figure-eight subsystems.

In frustrated figure-eight subsystems $X = X_1 \cup X_2$, the common qubit x_ℓ has conflicting constraints imposed on it by the two cycles X_1 and X_2. If we condition on frustration-free instances, this becomes a *would-be* frustrated figure-eight. As X is being constructed, one of the cycles (without loss of generality, X_1) must be completed before the other: this is either a loop or quasi-alternating loop at x (in the terminology of Ref. [13]). A quasi-alternating loop at x fixes the state of x, which by construction do not by themselves satisfy the constraints imposed on x by X_2. Similar remarks apply when X_1 is an alternating loop, which allows two possible single-qubit states for x which on their own satisfy the constraints imposed by X_1. In the case that X_1 is an alternating loop, x may be in one of two states $|\psi_x^0\rangle$ or $|\psi_x^1\rangle$ in a product with the rest of X_1, in which case all of the other spins of X_1 are in a product state $|\Phi^0\rangle$ or $|\Phi^1\rangle$ (respectively) determined by that state, or it may be entangled with the rest of the loop in some superposition $u_0|\psi_x^0\rangle|\Phi^0\rangle + u_1|\psi_x^1\rangle|\Phi^1\rangle$. In either case, the marginal of any satisfying state on x is a mixture of $|\psi_x^0\rangle$ or $|\psi_x^1\rangle$, neither of which on their own satisfy the constraints imposed by X_2 on x. Then in any case, upon the completion of the cycle X_1, the states of all qubits in X_2 which are accessible from x at that time are uniquely fixed. Each subsequent edge of X_2 which connects more qubits to x_ℓ also fixes the state of those qubits. This means in particular that every one of the ℓ qubits $v \in V(X_2)$ have fixed states $\big|\bar{\psi}_v\big\rangle$. We call such a subsystem of fixed qubits a *frozen* subsystem. Thus, a would-be frustrated figure-eight on $2\ell - 1$ qubits contains an (actually) frozen cycle of ℓ qubits.

The analysis of the preceding section concerning frustrated figure-eight subsystems $X = X_1 \cup X_2$ can be used to demonstrate the the existence of a "frozen core", or a subgraph of the giant component which itself contains $\Omega(n)$ vertices. The growth of this frozen core will gradually start to obstruct long-range constraints within the giant component, until eventually it renders the #2-QSAT problem highly decoupled.

To describe the growth of large frozen subsystems in frustration-free Erdős–Rényi models, we consider a random graph model for qubits with fixed states. Define a directed graph F defined by the 2-QSAT instance consisting of frozen subsystems, including only vertices representing qubits with fixed states, and with arcs $x \to y$ for qubits connected by constraints $\langle \eta_{x,y}|$ such that $\langle \eta_{x,y}| \big(|\bar{\psi}_x\rangle \otimes \mathbb{1}\big) \ne \mathbf{0}^\dagger$. We call this digraph the *frozen subgraph* of G.

We may establish lower bounds on the growth of F in terms of an Erdős–Rényi graph U, where edges of G belong to $E(U)$ independently with some probability $\tilde{Q} \leqslant Q$, and where all edges of U are covered by arcs of F. We consider $Q_\infty = 1 - \|\mathbf{q}\|_\infty$, and let $p_\infty = mQ_\infty/\binom{n}{2}$. We then let U be an Erdős–Rényi graph having $m_\infty \approx \binom{n}{2}p_\infty$ edges: we treat this as a subgraph of the Erdős–Rényi interaction graph G,[4] including each edge

[4] We may simulate randomly sampling over graphs with m edges, by considering graphs in which edges are present *i.i.d* with probability $p = m/\binom{n}{2}$ – the $\sqrt{n}$ variance in the number of edges is smaller than the scales at which phase transitions such as the emergence of the giant component occur.

of G with probability Q_∞. Consider a random colouring $c : V \to \{1, 2, \ldots, f\}$, in which $\Pr[c(x) = j] = q_j$. For a given qubit x which has a fixed state $|\bar{\alpha}_{c(x)}\rangle$, and a newly added edge $xy \in E(G)$, the probability that $x \to y$ is an arc of the frozen subgraph F is $1 - q_{c(x)} \geqslant Q_\infty$. From an initial set S of fixed qubits, we then simulate the construction of F as follows:

1. For each newly included vertex $x \in V(F)$ or $x \in S$, assign its colour $c(x)$;
2. For each neighbour y of x in G: If $xy \in U$, include $x \to y$ in F; otherwise include $x \to y$ in F with probability $(q_1 - q_{c(x)})/q_1$; otherwise exclude it.
3. Repeat the above until all $x \in S$ have been traversed, and no new vertices have been included in F.

This construction reproduces the probability distribution of arcs in F, with the random colouring of the vertex $c(y)$ taking the place of the action of constraints $\langle \eta_{x,y}| = \langle \beta|_x \otimes \langle \alpha_{c(y)}|$ which fixes the state of the qubit y.

From the above, we may show that the largest (weakly connected) component of F grows at least as quickly as that of the Erdős–Rényi graph U having $m_\infty \sim mQ_\infty$ edges. In particular, if $\frac{m}{n} > \gamma_\infty$ for $\gamma_\infty := \frac{1}{2Q_\infty}$, then U has a giant connected component $\Gamma^{(U)}$; if any vertices of $\Gamma^{(U)}$ are in F, then the entire component $\gamma^{(U)}$ is a subgraph of F. As we have noted, there are frozen cycles (arising from would-be frustrated figure eights) of size $\ell \in \mathrm{poly}(n)$ for $(1+\varepsilon)/2Q_2 \leqslant \frac{m}{n} \leqslant \gamma_\infty$: and almost surely a constant fraction of these vertices are subsumed into $\Gamma^{(U)}$, which has size $O(n)$. Then for $\frac{m}{n} > \gamma_\infty$, the giant component of U is almost surely contained in some weakly-connected component of F. Thus F almost surely contains a frozen core $\Gamma^{(F)}$ for $\gamma > \gamma_\infty$, which is at least as large as $\Gamma^{(U)}$.

Because the qubits in the frozen core cannot mediate non-trivial long-range constraints between non-fixed qubits, and do not contribute to the value of the #2-QSAT instance, they in effect play no role in the solution and may be removed. Let $\gamma = \frac{m}{n}$. By Ref. [7, Theorem 9b], the subgraph $\Gamma^{(U)}$ contains $(1 - \frac{1}{2\gamma Q_\infty}\xi(\gamma Q_\infty))n + o(n)$ vertices, where

$$\xi(\rho) = \sum_{k \geqslant 1} \frac{k^{k-1}}{k!}(2\rho \mathrm{e}^{-2\rho})^k \tag{22}$$

and where $\frac{1}{2\rho}\xi(\rho)$ expresses (almost surely and up to $o(1)$ error) the fraction of vertices which are contained in tree components in an Erdős–Rényi graph with ρn edges. Following Ref. [7, Theorem 4b], the function $\xi : [0, \infty) \to [0, 1]$ has the property that $\xi(\rho)\mathrm{e}^{-\xi(\rho)} = 2\rho\mathrm{e}^{-2\rho}$. We may show that for any super-critical edge-density $\rho > \frac{1}{2}$, there is a sub-critical edge-density $\tilde{\rho} := \frac{1}{2}\xi(\rho) < \frac{1}{2}$ such that the distribution of the sizes of tree-components for the edge-densities ρ and $\tilde{\rho}$ are the same up to a normalization factor.[5] Thus deleting the giant component from the Erdős–Rényi graph with density ρ gives rise to a graph indistinguishable from an Erdős–Rényi graph with density $\tilde{\rho}$, albeit on $\frac{1}{2\rho}\xi(\rho)n$ vertices. More generally, deleting the subgraph $\Gamma^{(U)}$ from the graph G yields a graph indistinguishable from an Erdős–Rényi graph on $\frac{1}{2\gamma Q_\infty}\xi(\gamma Q_\infty)n$ vertices, with edge-density given by

$$\tilde{\gamma} := \tfrac{1}{2}\xi(\gamma Q_\infty) + \gamma(1 - Q_\infty)\left[\frac{\xi(\gamma Q_\infty)}{2\gamma Q_\infty}\right]^2 = \tfrac{1}{2}\xi(\gamma Q_\infty) + \tfrac{1 - Q_\infty}{4\gamma Q_\infty^2}\xi(\gamma Q_\infty)^2, \tag{23}$$

[5] Consider a randomly selected tree component T, and let $\tau_\rho(t) = \frac{1}{2\rho t!}t^{t-2}(2\rho\mathrm{e}^{-2\rho})^t$. The probability $P_\rho(t)$ that T has size t, when selecting tree-components from the Erdős–Rényi graph with ρn edges, is then $P_\rho(t) \sim \tau_\rho(t) \Big/ \sum_k \tau_\rho(k)$ by Ref. [7, Eqn. 2.22]. From $\tilde{\rho} := \frac{1}{2}\xi(\rho)$ and Ref. [7, Eqn. 4.4] we may immediately see that $P_\rho(t) = P_{\tilde{\rho}}(t)$ for all t. As all but an insignificant number of vertices are contained in either the giant component or in trees, the two distributions on graphs are indistinguishable.

where the first term accounts for the density of $U \smallsetminus \Gamma^{(U)}$, and the second term accounts for the contribution of edges $e \in E(G) \smallsetminus E(U)$ which are also not incident to $\Gamma^{(U)}$.

As the frozen core $\Gamma^{(F)} \supseteq \Gamma^{(U)}$ grows, the subgraph of G that remains after removing $\Gamma^{(F)}$ becomes more sparse, and eventually becomes highly disconnected. That is to say, the instance with the frozen subsystems included is highly decoupled. Note that $\xi(\rho) = 2\rho$ for $\rho \in [0, \frac{1}{2}]$, achieving a maximum of 1 and then decreasing for $\rho \geqslant \frac{1}{2}$. It follows that $\tilde{\gamma} = \gamma$ for $\gamma Q_\infty \leqslant \frac{1}{2}$, achieving a maximum of $1/2Q_\infty$ and then subsequently bounded by

$$\begin{aligned}\tilde{\gamma} \;\leqslant\; \left[\tfrac{1}{2} + \tfrac{1-Q_\infty}{2Q_\infty}\right]\xi(\gamma Q_\infty) \;&\leqslant\; \tfrac{1}{2Q_\infty}\xi(\gamma Q_\infty) \\ &\leqslant\; \gamma \mathrm{e}^{\xi(\gamma Q_\infty)}\mathrm{e}^{-2\gamma Q_\infty} \;\leqslant\; \gamma \mathrm{e}^{1-2\gamma Q_\infty}. \end{aligned} \tag{24}$$

If $2\gamma Q_\infty - \ln(2\gamma) > 1$, we then have $\tilde{\gamma} < \frac{1}{2}$. In this case $G \smallsetminus \Gamma^{(U)}$ becomes subcritical and thus highly disconnected; the same is then true of $G \smallsetminus \Gamma^{(F)}$.

Thus for γ sufficiently large, frustration-free instances of #2-QSAT almost surely contain a frozen core pervasive enough to cause the problem to be highly decoupled. It is easy to show that such a frozen core can be easily detected, as well, using the same techniques as described in the preceding section for frustrated figure-eights. We may detect the existence of alternating and quasi-alternating loops at each vertex x in the graph, and then consider the constraints on x and its neighbours to discover an initial set of frozen spins. Following this, using a single breadth-first traversal, we may discover the entire frozen subgraph and its largest component in particular. Discovering the frozen core is therefore possible in polynomial time using standard techniques.

3.2 Bond-percolated lattice graphs

The analysis for random 2-QSAT is much simpler for bond-percolated square or cubic lattices. In this graph model, we take vertices labelled either $(a, b) \in \{0, 1, \ldots, L-1\}^2$ or $(a, b, c) \in \{0, 1, \ldots, L-1\}^3$, and connect each pair of vertices which differ by 1 in a single co-ordinate, independently with some probability p. We let d denote the dimension of the lattice, let $n = L^d$ be the number of vertices and $m \sim dpn$ be the expected number of edges.

The analysis of phase transitions in the difficulty of #2-QSAT for independent factor constraints is simpler for percolated lattices than for Erdős–Rényi graphs, as cycles arise in the percolated lattice much more easily and as the degree of each vertex is necessarily bounded. Furthermore, we only expect the largest components to grow with n if p is greater than a "percolation threshold" p_c [10],[6] in which case the largest component is unique and scales as $O(n)$. For #2-QSAT with independent factor constraints, this allows one to show:

- #2-QSAT is almost certainly efficiently solvable for any value of p, as there are overlapping phases of frustrated and highly disconnected instances, occurring respectively for $p \in \omega(n^{-1/7})$ and $p \leqslant p_c \in O(1)$;
- For frustration-free instances of #2-QSAT, provided that $Q_\infty := 1 - \|\mathbf{q}\|_\infty > p_c$, there is a transition directly from highly disconnected instances for $p < p_c$ to highly decoupled instances for $p \geqslant p_c$, due to the emergence of frozen subgraph whose components decouple the system into small non-interacting components (in a way which is similar to, but more straightforward than, the analogous phenomenon in models on Erdős–Rényi graphs.)

[6] For d_2, we have $p_c = \frac{1}{2}$; for $d = 3$, we have $p_c \approx 0.24881$; *c.f.* Ref [10]. N.B. For $d = 3$ it is not yet known whether there exists an infinite component when $p = p_c$; this is known not to occur for $d = 2$ or $d \geqslant 19$, and the same is conjectured for $d = 3$ [10, Section 9.4].

In this Section we outline these results in enough detail to indicate how the results may be shown more completely. Furthermore, results which are similar in quality could also be shown for any lattice model, depending in practise only on the size of the smallest cycles and the percolation threshold p_c of the lattice.

3.2.1 Critical thresholds for unconditional percolated lattice models

If each edge in a d-dimensional rectangular lattice (for $d \in \{2, 3\}$) is present independently with probability $p \in o(1)$, then the first components with cycles to emerge as p increases are the ones with the fewest edges. That is, if the probability of there being a component in G which is isomorphic to a graph g is $\Omega(1)$, then G will contain infinitely many isomorphic copies of any component g' for which $|E(g')| < |E(g)|$. The first components with cycles to emerge are therefore individual square facets of the lattice, which are almost surely absent for $p \in o(n^{-1/4})$, and present in infinite abundance for $p \in \omega(n^{-1/4})$.

The smallest subgraph of a rectangular lattice which contains two cycles is a *domino graph*, as pictured in Fig. 2, which has seven edges. These are therefore almost certainly absent for $p \in o(n^{-1/7})$, and almost certainly abundantly present for $p \in \omega(n^{-1/7})$. It is not difficult to show that each of these has a constant probability of being a *frustrated domino*: a system similar to a frustrated-figure eight in which the constraints give rise to unsatisfiable restrictions on the state of the two central qubits. Consider the three independent paths between the central vertices of a domino subgraph (also depicted in Fig. 2). Given that each edge represents a non-zero constraint (which happens with constant probability), the two outer paths in the domino each give rise to a non-zero path constraint with probability $Q_2^2 = (1 - \|\mathbf{q}\|_2^2)^2$. With some probability, the three path constraints will act on each of their endpoints in a different way from the others. This remains true even for classical instances of #2-SAT, if the constraint-operators are chosen from a probability distribution over a distribution on $\{\langle 00| , \langle 01| , \langle 10| , \langle 11|\}$ in which each element occurs with probability $\Omega(1)$, each such domino is unsatisfiable with constant probability, in which case the entire instance of #2-SAT which contains it has value zero. (This would occur, for instance, for an independent factor distribution $\mathbf{q} = (q_1, q_2)$ in which $\langle \alpha_1| = \langle 0|$ and $\langle \alpha_2| = \langle 1|$, where q_1 and q_2 are both bounded away from zero.) Thus, there is a phase transition at $p \in \Theta(n^{-1/7})$ from almost certain satisfiability to almost certain unsatisfiability, due to the probable emergence of frustrated dominoes, of which there are almost surely infinitely many once $p \in \omega(n^{-1/7})$.

The components in a bond-percolated lattice for $p \in O(n^{-1/7})$ almost certainly have size $O(1)$: specifically, they will almost surely have seven vertices or fewer. Thus the complexity of computing #2-QSAT is almost surely governed by that of multiplying $O(n)$ "small" integers. A simple algorithm to do so is described in Appendix A. Thus, #2-QSAT is almost surely easy for p increasing up to, and even through, the phase transition at $p \in \Theta(n^{-1/7})$; afterwards, of course, the value is almost surely zero. Difficult instances of #2-QSAT on percolated lattices are thus either ones which are asymptotically monotone – that is, for which Q_2 decreases with n – or ones which almost surely never occur. Similar phenomena will occur for any lattice model, with a phase transition at $p \in \Theta(n^{-1/\beta})$, where β is the number of edges in the smallest subgraph having more than one cycle.

3.2.2 Critical thresholds for frustration-free percolated lattice models

To obtain interesting instances of #2-SAT or #2-QSAT on a percolated rectangular lattice, we must condition on models which are frustration-free. However, for p less than the percolation probability p_c, almost surely the resulting graph G contains only components of size $o(f(n))$

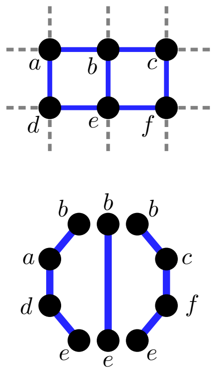


Figure 2 (*Top:*) An isolated "domino" subgraph of a square lattice. Dashed lines indicate missing edges incident to the subgraph. A domino subgraph in a 3D lattice may also occur with the two cycles meeting at a right angle. (*Bottom:*) Illustration of the three independent paths between the central qubits of a domino subgraph. If the constraints acting on b do so with different tensor factors $\langle\alpha|, \langle\alpha'|, \langle\alpha''| : \mathbb{C}^2 \to \mathbb{C}$ and similarly for the constraints $\langle\beta|, \langle\beta'|, \langle\beta''| : \mathbb{C}^2 \to \mathbb{C}$ acting on e, and the path-constraints are all non-zero, then these form an infeasible system of constraints on the states of b and e. Similar remarks apply for any pair of qubits connected by more than two independent paths.

for any $f \in \omega(1)$.[7] This implies that for $p < p_c$, it again suffices to compute the values of #2-QSAT for each component individually,[8] so that #2-QSAT is almost surely efficiently solvable so long as $p \leqslant p_c$. It thus suffices for us to consider the regime $p > p_c$.

We may proceed similarly to the analysis of the giant component in frustration-free Erdős–Rényi models in Section 3.1.3. Would-be frustrated subsystems – such as frustrated figure-eights on seven vertices (consisting of two square cells intersecting at one qubit) or would-be frustrated dominoes – will arise in abundance for $p \in \Theta(1)$. Each one gives rise to several qubits with fixed states, which contribute to the presence of a non-empty frozen subgraph F. If there is a giant component $\Gamma^{(G)}$, then there are almost certainly would-be frustrated subsystems inside it: we ask to what extent these give rise to frozen subsystems which decouple $\Gamma^{(G)}$.

As with the Erdős–Rényi case, we may let $Q_\infty = 1 - \|\mathbf{q}\|_\infty$ be a lower bound on the probability that any two constraints coinciding at a qubit give rise to a non-zero constraint on a path of length two, such that we may treat this as as independent events even for various pairs of constraints meeting acting on the same qubit. For instance, the probability that any domino subgraph is a would-be-frustrated domino is at least Q_∞^7. For any qubit $x \in V(F)$, the probability that some neighbour y in G is also subsumed into $V(F)$ is also at least Q_∞. We may then consider a percolated lattice model U in which edges are present with probability Q_∞, and any such component which contains a frozen seed gives rise to a component in the frozen subgraph F.

When does the frozen core $\Gamma^{(F)}$ decouple an instance of #2-QSAT? That is: when does $G \smallsetminus V(\Gamma^{(F)})$ decompose as a collection of small components? This relates to the problem, when U has a giant component $\Gamma^{(U)}$, of whether the complement of $\Gamma^{(U)}$ in the complete

[7] For $d = 2$ (for which $p_c = \frac{1}{2}$) or $d = 3$ (for which $p_c \approx 0.24881$), the distribution of component sizes decreases geometrically for $p < p_c$ [10, Section 6.3].

[8] As the components all have essentially constant size, this may be done for each component in $O(\log n)$ time, dominated merely by the time required to process the labels of vertices.

(square or cubic) lattice has any infinite components (in the limit $n \to \infty$). For both $d \in \{2,3\}$, there exists a threshold $p_{\text{fin}} < 1$ [11] such that the complement of $\Gamma^{(U)}$ in the lattice decomposes into components of finite size when $Q_\infty > p_{\text{fin}}$.[9] Consider the case $Q_\infty > p_c$:

- If $p = 1$ (that is, G is simply the entire $O(n)$-vertex square or cubic lattice segment), then by construction $G \smallsetminus U$ is a collection of small components. As $\Gamma^{(U)}$ is almost surely subsumed by a frozen core $\Gamma^{(F)}$ of qubits with fixed states, which do not contribute to the value of the #2-QSAT instance. As the complete lattice with $\Gamma^{(F)}$ removed consists of components of finite size, the resulting instance of #2-QSAT is highly decoupled.
- If $p < 1$, then we may model the resulting 2-QSAT instance on the percolated lattice by reducing from the previous case (in which the instance is highly decoupled), and removing each constraint in the complete lattice with probability $1-p$: doing so does not make the instance any less decoupled.

Thus, for $Q_\infty > p_{\text{fin}}$ (which occurs for $\|\mathbf{q}\|_\infty$ below some constant), there is a phase transition for random frustration-free instances of #2-QSAT from highly disconnected instances to highly decoupled instances. This means that for $d = 2$, difficult instances of #2-QSAT are only likely if the constraint model is "at least as monotone" as some distribution of classical #2-SAT constraints; for $d = 3$, a bias towards monotonicity which would be substantial even for #2-SAT is necessary to obtain difficult instances.[10]

As a final remark, note that even in the case that $Q_\infty \leqslant p_{\text{fin}}$, there is a chance that frozen subsystems will decouple the largest component $\Gamma^{(G)}$ into small subsystems. Any domino-shaped subsystem of $\Gamma^{(G)}$ has a finite probability of containing a frozen cycle, which can be treated in the giant component as nodes which are removed from $\Gamma^{(G)}$ with some finite probability $1 - P_{\text{site}} > 0$. Using results on mixed site- and bond-percolation [12], if $P_{\text{site}}\, p < p_c$, the giant component $\Gamma^{(G)}$ still decouples into small subsystems whose degeneracy may be efficiently computed. We do not present any quantitative results for $Q_\infty \leqslant p_c$, but mention this to indicate that it likely that #2-QSAT may remain easy even for some values $Q_\infty < p_c$, for reasons similar to what we have shown for $Q_\infty > p_c$.

4 Open questions

The results of this article may allow for some improvements, which would further bound any "difficult" regime in random distributions of #2-QSAT on random graphs.

- For frustration-free instances, $Q_\infty = \min_j (1 - q_j)$ is used as a percolation probability on an existing random graph, to obtain lower bounds on the transition to a highly decoupled phase; whereas $Q_2 = \mathbb{E}_j\big[1 - q_j\big]$ is used for potentially frustrated models (where we take $\Pr[j] = q_j$). Can we replace bounds involving Q_∞ with tighter bounds involving Q_2?
- If we remove the condition of frustration-freeness from #2-QSAT altogether, we are left with the problem of computing the degeneracy of the ground-state manifold of a potentially frustrated Hamiltonian. Physical intuition suggests that this is typically "1", but as with #2-QSAT, the classical problem of determining how many boolean strings satisfy a maximum number of constraints is a hard problem in general. Under what

[9] A simple duality argument shows that $p_{\text{fin}} = p_c = \frac{1}{2}$ for $d = 2$ [11]. For $d = 3$, only know the more general result $p_c \leqslant p_{\text{fin}} < 1$ is currently known. While no numerical results are known about p_{fin} for $d = 3$, the growth of infinite clusters in each planar cross-section of the cubic lattice suggests that p_{fin} is closer to $1 - p_c$ than to 1.

[10] This implies, for instance, that uniformly random #2-SAT on bond-percolated cubic lattices is almost surely efficiently solvable whether or not we condition on satisfiability.

conditions is it provably easy to compute the ground-state degeneracy of random local Hamiltonians?

Acknowledgements. This work was partly performed at the University of Cambridge, with support from the EC project QCS. I would like to thank Ronald de Wolf, as well as an anonymous referee, for helpful comments on the preliminary drafts.

References

1 B. Aspvall, M. F. Plass, and R. E. Tarjan. A linear-time algorithm for testing the truth of certain quantified boolean formulas. *Information Processing Letters*, 8(3):121–123, 1979.

2 S. Bravyi. Efficient algorithm for a quantum analogue of 2-SAT. arXiv:quant-ph/0602108, 2006.

3 S. Bravyi, C. Moore, and A. Russell. Bounds on the Quantum Satisfiability threshold. In Andrew Chi-Chih Yao, editor, *ICS*, pages 482–489. Tsinghua University Press, 2010. arXiv:0907.1297.

4 V. Chvatal and B. Reed. Mick gets some (the odds are on his side). In *Proc. 33rd Annual FOCS*, pages 620–627, 1992.

5 N. de Beaudrap, T. J. Osborne, and J. Eisert. Ground states of unfrustrated spin hamiltonians satisfy an area law. *New J. Phys.*, 12:095007, 2010. arXiv:1009.3051.

6 R. Diestel. *Graph Theory, 4th Edition*, volume 173 of *Graduate texts in mathematics*. Springer, 2012.

7 P. Erdős and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungary. Acad. Sci.*, 5:17–61, 1960.

8 E. Fischer, J. A. Makowsky, and E. V. Ravve. Counting truth assignments of formulas of bounded tree-width or clique-width. *Discrete Applied Mathematics*, 156:511–529, 2008.

9 D. Gosset and D. Nagaj. Quantum 3-SAT is QMA1-complete. arXiv:1302.0290, 2013.

10 G. Grimmett. *Percolation*. Springer, Berlin, 2nd edition, 1999.

11 G. R. Grimmett, A. E. Holroyd, and G. Kozma. Percolation of finite clusters and infinite surfaces. *Mathematical Proceedings of the Cambridge Philosophical Society*, 156:263–279, 2014. arXiv:1303.1657.

12 J. M. Hammersley. A generalization of McDiarmid's theorem for mixed Bernoulli percolation. *Math. Proc. Camb. Phil. Soc.*, 1980.

13 Z. Ji, Z. Wei, and B. Zeng. Complete characterization of the ground space structure of two-body frustration-free hamiltonians for qubits. *Phys. Rev. A*, 84:042338, 2011.

14 C. R. Laumann, R. Moessner, A. Scardicchio, and S. L. Sondhi. Phase transitions and random quantum satisfiability. *Quant. Inf. and Comp.*, 10:1–15, 2010. arXiv:0903.1904.

15 C. Papadimitriou and M. Yannakakis. Optimization, approximation, and complexity classes. *J. Comp. Sys. Sci.*, 43:425–440, 1991.

16 N. Robertson and P. D. Seymour. Graph minors. III. Planar tree-width. *Journal of Combinatorial Theory, Series B*, 36(1):49–64, 1984.

17 L. Valiant. The complexity of enumeration and reliability problems. *SIAM J. Computing*, 8:410–421, 1979.

A An effective technique for multiplying together long lists of mostly small numbers

The value of an instance of #2-QSAT is at most 2^n. We may decompose the value of an instance of #2-QSAT as a product of the values of each connected component. In the easily solved instances which arise either when the interaction graph is highly disconnected, or when a large frozen subsystem decouples the Hamiltonian into small independent subsystems, the value of #2-QSAT for these instances is $O(\log n)$. One might then show that simply multiplying together these values can be performed in polynomial time, by accounting for the increase in size of the integers involved in the multiplication as more and more factors are included in the product. Rather than analyse the growth of the product in an iterative multiplication algorithm, we will show a different algorithm, by which the complexity of evaluating this product is asymptotically no greater than multiplying two n-digit numbers.

By sorting the non-giant components of G in order of size (we assume only non-giant components henceforth), we may construct a binary tree such that

- The leaves represent sets, each of which contains an individual component and having a stored #2-QSAT value of one more than the component size;
- Each node which is not a leaf represents the union of the sets of components represented by its child nodes, and stores the product of the #2-QSAT values of its children;
- The #2-QSAT values of the children of any node are either similar in size (*e.g.* differing by a factor of at most 3), or the degeneracy of one of them is constant (*e.g.* at most 3).

We start by pairing the largest component with the second largest component; in the case that the second-largest component is less than half the size of the largest, we first pair it together with a small component (*e.g.* isolated vertices), and pair the largest component with the parent to these two nodes. We continue similarly for the next two largest components, using the smallest components to compensate for differences in the size of the degeneracies of subtrees. (Because there are $O(n)$ components in the Erdős–Rényi graph for any number of edges m, the components of constant size must dominate, and the smallest ones will occur most frequently as a result of the reduced probability of being merged with other components. For bond-percolated lattices, the distribution of component sizes is monotone decreasing for any bond-percolation probability p, so again small components dominate.) The degeneracy of the root node of the tree then is the degeneracy of the Hamiltonian.

The number of bits required to represent the degeneracy at each level in the tree either remains about constant, or decreases by a factor of 2, with each level down from the parent node. Due to the domination by components of constant size, there will be $\Theta(n)$ leaves on either side of the tree, so that it will have depth $O(\log n)$; most subtrees will be balanced. Thus there will be approximately $O(\log n)$ rounds of (in principle parallelisable) multiplications, where the t^{th} round from the final one is between numbers of size $n/2^t$, and each round involves about 2^t multiplications in total. For any given multiplication algorithm running in some time $O(n^d)$ (*e.g.* where $d = 2$ for the usual straightforward algorithm taught in schools), we can recursively evaluate the value of the entire #2-QSAT instance, corresponding to the root node of the tree, in time

$$\sum_{t=1}^{O(\log n)} 2^t \left(\frac{n}{2^t}\right)^d = \sum_{t=1}^{O(\log n)} 2^{t(1-d)} n^d = \left[\frac{2^{(1-d)} - 2^{O((1-d)\log n)}}{1 - 2^{(1-d)}}\right] n^d \in O(n^d). \tag{25}$$

Circuit Obfuscation Using Braids

Gorjan Alagic[1], Stacey Jeffery[2], and Stephen Jordan[3]

1 Institute for Quantum Information and Matter
California Institute of Technology
Pasadena, CA, USA
galagic@gmail.com

2 Institute for Quantum Computing
University of Waterloo
Waterloo, ON, Canada
smjeffery@gmail.com

3 National Institute of Standards and Technology
Gaithersburg, MD, USA
stephen.jordan@nist.gov

Abstract

An obfuscator is an algorithm that translates circuits into functionally-equivalent similarly-sized circuits that are hard to understand. Efficient obfuscators would have many applications in cryptography. Until recently, theoretical progress has mainly been limited to no-go results. Recent works have proposed the first efficient obfuscation algorithms for classical logic circuits, based on a notion of indistinguishability against polynomial-time adversaries. In this work, we propose a new notion of obfuscation, which we call partial-indistinguishability. This notion is based on computationally universal groups with efficiently computable normal forms, and appears to be incomparable with existing definitions. We describe universal gate sets for both classical and quantum computation, in which our definition of obfuscation can be met by polynomial-time algorithms. We also discuss some potential applications to testing quantum computers. We stress that the cryptographic security of these obfuscators, especially when composed with translation from other gate sets, remains an open question.

1998 ACM Subject Classification D.4.6 Security and Protection

Keywords and phrases obfuscation, cryptography, universality, quantum

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.141

1 Introduction

1.1 Past work on circuit obfuscation

Informally, an obfuscator is an algorithm that accepts a circuit as input, and outputs a hard-to-understand but functionally equivalent circuit. In this subsection, we briefly outline the state of current research in classical circuit obfuscation. To our knowledge, quantum circuit obfuscation has not been considered in any prior published work.

Methods used for obfuscating logic circuits in practice have so far been essentially ad hoc [11, 41]. Until recently, theoretical progress has primarily been in the form of no-go theorems for various strong notions of obfuscation [7, 21]. The ability to efficiently obfuscate certain circuits would have important applications in cryptography. For instance, sufficiently strong obfuscation of circuits of the form "encrypt with a hard-wired private key" could turn a private-key encryption scheme into a public-key encryption scheme. As this example illustrates, one undesirable outcome is when the input circuit can be recovered completely

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 141–160
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

from the obfuscated circuit. In this case, we say that the obfuscator *completely failed* on that circuit [7]. Unfortunately, every obfuscator will completely fail on some circuits (e.g., learnable circuits.) On the other hand, there are trivial obfuscators which will erase at least some information from some circuits, e.g., by removing all instances of $X^{-1}X$ for some invertible gate X.

In order to give a useful formal definition of obfuscation, one must decide on a reasonable definition of "hard-to-understand." The most stringent definition in the literature demands *black-box obfuscation*, i.e., that the output circuit is computationally no more useful than a black box that computes the same function. Barak et al. [8] gave an explicit family of circuits that are not learnable and yet cannot be black-box obfuscated. They also showed that there exist (non-learnable) private-key encryption schemes that cannot be turned into a public-key cryptosystem by obfuscation. Their results do not preclude the possibility of black-box obfuscation for specific families of circuits, or of applying obfuscation to produce public-key systems from private ones in a non-generic fashion. It is an open problem whether quantum circuits can be black-box obfuscated.

A weaker but still quite natural notion is called *best-possible obfuscation*; in this case, we ask that the obfuscated circuit reveals no more information than any other circuit that computes the same function. Goldwasser and Rothblum [21] showed that for efficient obfuscators, best-possible obfuscation is equivalent to *indistinguishability obfuscation*, which is defined as follows. For any circuit C, let $|C|$ be the number of elementary gates, and let f_C be the Boolean function that C computes.

▶ **Definition 1.** A probabilistic algorithm $\mathcal{O}$ is an *indistinguishability obfuscator* for the collection $\mathcal{C}$ of circuits if the following three conditions hold:

1. (functional equivalence) for every $C \in \mathcal{C}$, $f_{\mathcal{O}(C)} = f_C$;
2. (polynomial slowdown) there is a polynomial p such that $|\mathcal{O}(C)| \leq p(|C|)$ for every $C \in \mathcal{C}$;
3. (indistinguishability obfuscation) For any $C_1, C_2 \in \mathcal{C}$ such that $f_{C_1} = f_{C_2}$ and $|C_1| = |C_2|$, the two distributions $\mathcal{O}(C_1)$ and $\mathcal{O}(C_2)$ are indistinguishable.

In the third part of the above definition, one must choose a notion of indistinguishability for probability distributions. Goldwasser and Rothblum [21] consider three such notions: perfect (exact equality), statistical (total variation distance bounded by a constant), and computational (no probabilistic polynomial-time Turing Machine can distinguish samples with better than negligible probability). They show that the existence of an efficient statistical indistinguishability obfuscator would result in a collapse of the polynomial hierarchy to the second level. This result also applies if the condition $|C_1| = |C_2|$ in property (3) of Definition 1 is relaxed to $|C_1| = k|C_2|$ for any fixed constant k [21].

A recent breakthrough has shown that computational indistinguishability may be achievable in polynomial time. Combining a new obfuscation scheme for NC1 circuits with fully homomorphic encryption, Sahai et al. gave an efficient obfuscator which achieves the computational indistinguishability condition under plausible hardness conjectures [19]. Subsequent work outlined a number of cryptographic applications of computational indistinguishability [38].

1.2 Outline of present work

1.2.1 New notion of obfuscation

An exact deterministic indistinguishability obfuscator would yield a solution to the circuit equivalence problem. For general Boolean circuits, this problem is co-NP hard. Therefore, exact deterministic indistinguishability obfuscation of general Boolean circuits cannot be achieved in polynomial time under the assumption $\mathrm{P} \neq \mathrm{NP}$. We propose an alternative route to weakening the exactness condition, by pursuing a notion of "partial-indistinguishability". In partial-indistinguishability obfuscation, we relax condition (3) so that it need only hold for C_1 and C_2 that are related by some fixed, finite set of relations on the underlying gate set.[1]

▶ **Definition 2.** Let G be a set of gates and Γ a set of relations satisfied by the elements of G. An algorithm $\mathcal{O}$ is a (G, Γ)-*indistinguishability obfuscator* for the collection $\mathcal{C}$ of circuits over G if the following three conditions hold:

1. (functionality) for every $C \in \mathcal{C}$, $f_C = f_{\mathcal{O}(C)}$;
2. (polynomial slowdown) there is a polynomial p such that $|\mathcal{O}(C)| \leq p(|C|)$ for every $C \in \mathcal{C}$;
3. ((G, Γ)-indistinguishability) for any $C_1, C_2 \in \mathcal{C}$ that differ by some sequence of applications of the relations in Γ, $\mathcal{O}(C_1) = \mathcal{O}(C_2)$.

The power of the obfuscation is now determined by the power of the relations Γ. If Γ is a complete set of relations, generating all circuit equivalences over G, then a (G, Γ)-indistinguishability obfuscator is a perfect indistinguishability obfuscator according to Definition 1. (Complete sets of relations for $\{\text{Toffoli}\}$ and $\{\text{AND}, \text{OR}, \text{NOT}\}$ are given in [27, 26].) If Γ is the empty set then even the identity map fits the definition, and no obfuscation is taking place. With different sets of relations, one can interpolate between these extremes. The intermediate obfuscators form a partially ordered set, where a (G, Γ')-indistinguishability obfuscator is strictly stronger than a (G, Γ)-indistinguishability obfuscator if Γ' is a strict superset of Γ. We remark that partial-indistinguishability is no stronger than perfect indistinguishability, and appears to be incomparable with statistical and computational indistinguishability. This is part of our motivation in considering this new definition.

In the context of quantum computation, we make only a few minor changes to Definitions 1 and 2. First, the obfuscators will still be classical algorithms. On the other hand, the gates will be unitary and the circuits to be obfuscated will be unitary quantum circuits. Finally, the notion of functional equivalence now simply means that the operator-norm distance between the unitary implemented by C and the unitary implemented by $\mathcal{O}(C)$ is bounded by a small constant $\epsilon > 0$.

1.2.2 Group normal forms

A finitely generated group can be specified by a presentation. This is a list of generators $\sigma_1, \ldots, \sigma_n$ and a list of relations obeyed by these generators. (A relation is simply an identity such as $\sigma_1 \sigma_3 = \sigma_3 \sigma_1$.) All group elements are obtained as products of the generators and their inverses. However, by applying the relations, we can get multiple words in the generators and their inverses that encode the same group element. A normal form specifies, for each group element, a unique decomposition as a product of generators and their inverses. For certain groups, including the braid groups, polynomial time algorithms are known which, given a product of generators and their inverses, can reduce it to a normal form. The word problem is, given two words in the alphabet $\{\sigma_1, \ldots, \sigma_n, \sigma_1^{-1}, \ldots, \sigma_n\}$, to decide whether they specify the same group element. If a normal form can be computed, then this solves the word problem: just reduce both words to normal form and check whether the results are identical. However, an efficient solution for the word problem does not in general imply an efficiently computable normal form.

[1] Our construction for satisfying this definition uses *reversible* gates. The definition of functional equivalence becomes more technical in that context, as discussed in Section 3.1.

1.2.3 Efficient constructions from group representations

In this paper, we propose a general method of designing partial-indistinguishability obfuscators based on groups with efficiently computable normal forms. If a set of gates G obeys the relations Γ of the generators of a group with an efficiently computable normal form, then the reduction to normal form is an efficient (G, Γ)-indistinguishablity obfuscator. The gates may obey additional relations beyond Γ, which is why the obfuscator does not solve the circuit-equivalence problem, which is believed to be intractable for both classical and quantum circuits.

To demonstrate this method, we discuss an implementation using the braid groups B_n, for both classical reversible circuits and unitary quantum circuits. The number of strands n in the braid group depends linearly on the number of dits or qudits on which the circuit acts. In Section 3, we describe a computationally universal reversible classical gate obeying the braid group relations, which was constructed in [34, 37, 31] from the quantum double of A_5. In Section 4.1, we describe a computationally universal quantum gate obeying the braid group relations, which was constructed in [18] from the Fibonacci anyons. Our obfuscation scheme is similar in spirit to previously-proposed obfuscation schemes based on applying local circuit identities [41], but the uniqueness of normal forms adds a qualitatively new feature. One consequence of this feature is that we can satisfy Definition 2 and guarantee the partial-indistinguishability property against computationally unbounded adversaries. The running time of the obfuscator is the same as the running time of the the normal form algorithms, which take time $O(l^2 m \log m)$ for m-strand braids of length l [14].

We remark that these gate sets that obey the braid group relations are not artificial constructions; in fact, they are the most natural choice in many contexts, some of which we list here. In the quantum case, these gates are native to certain proposed physical implementations of quantum computers [31], where the topological braiding property provides inherent fault-tolerance. The problem of approximating the Jones Polynomial invariant of links is complete for polynomial-time quantum computation [2]; an analogous fact is true for a restricted case of quantum computations motivated by NMR implementations [40]. Both of these facts are naturally expressed in the gate set constructed from the Fibonacci representation. In the classical case, the gate set derived from quantum doubles of finite groups was recently used to show BPP-completeness for approximation of certain link invariants [32].

We remark that another potential group family for constructing partial-indistinguishability obfuscators are the mapping class groups $\mathrm{MCG}(\Sigma_g)$ of unpunctured surfaces of genus g. These groups also have quantumly universal representations [5] and an efficiently solvable word problem [23]. It is not known if there are also classically universal permutation representations, or if there are efficiently computable normal forms.

1.2.4 Other gate sets

In some applications the native gate set will be different than the ones used in our construction. It is natural to ask if our obfuscators can be used in these settings as well. By universality (quantum or classical), one has an efficient algorithm B which translates circuits from the native gate set to the braiding gate set, as well as an efficient algorithm C for translation in the opposite direction. We also let N denote the partial-indistinguishability obfuscator. One might then attempt to obfuscate by applying the following:

▶ Algorithm 1.

1. *input: a circuit C on n (qu)dits*
2. *output: The circuit* $\mathsf{C}(\mathsf{N}(\mathsf{B}(C)))$

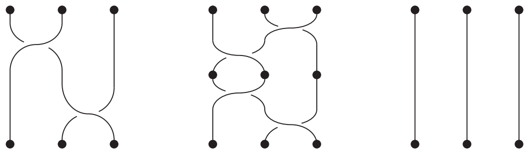


Figure 1 The generator σ_i represents the (clockwise) crossing of strands i and $i+1$ connecting a bottom row of "pegs" to a top row. Multiplication of group elements corresponds to composition of braids. As an example, we show the 3-strand braid $\sigma_1^{-1}\sigma_2$ (left), and the same braid composed with its inverse $\sigma_2^{-1}\sigma_1$ (middle), which is equivalent to the identity element of B_3 (right).

We stress that, unlike the map N, the composed map $\mathsf{N} \circ \mathsf{B}$ does not necessarily satisfy Definition 2. As we discuss in Section 5.1, careless choice of the map B can partially or completely break the security of the obfuscator. Finding translation algorithms securely composable with partial-indistinguishability obfuscators is an area of current investigation.

2 Relevant Properties of the Braid Group

The braid group B_n is the infinite discrete group with generators $\sigma_1, \ldots, \sigma_{n-1}$ and relations

$$\begin{aligned} \sigma_i \sigma_j &= \sigma_j \sigma_i && \forall\ |i-j| \geq 2 \\ \sigma_i \sigma_{i+1} \sigma_i &= \sigma_{i+1} \sigma_i \sigma_{i+1} && \forall\ i. \end{aligned} \tag{1}$$

The group B_n is thus the set of all words in the alphabet $\{\sigma_1, \ldots, \sigma_{n-1}, \sigma_1^{-1}, \ldots, \sigma_{n-1}^{-1}\}$, up to equivalence determined by the above relations. In 1925 Artin proved that the abstract group defined above precisely captures the topological equivalence of braided strings [6], as illustrated in Fig. 1. A charming exposition of this subject can be found in [30].

In the word problem on B_n, we are given words w and z, and our goal is to determine if they are equal as elements of B_n. One solution is to put both w and z into a *normal form*, and then check if they are equal as words. For our purposes, it is enough to describe the normal form and specify the complexity of the algorithm for computing it. The details of the algorithm, along with a thorough and accessible presentation of the relevant facts about braids, can be found in [14].

We first observe that the word problem is easily shown to be decidable if we restrict our attention to an important subset of B_n. Note that the presentation (1) can also be viewed as a presentation of a monoid, which we denote by B_n^+. The elements of B_n^+ are called *positive braids*, and are words in the generators σ_i only (no inverses), up to equivalence determined by the relations in (1). Since all the relations of B_n preserve word length, and there are only finitely many words of any given length, we can decide the word problem (albeit very inefficiently) simply by trying all possible combinations of the relations.

Building upon this, one can give an (inefficient) algorithm for the word problem on B_n itself [22]. First, given two elements a, b of B_n^+, we write $a \preccurlyeq b$ if there exists $z \in B_n^+$ such that $b = az$; in this case we say that a is a *left divisor* of b. Similarly, we write $a \succcurlyeq b$ if there exists $y \in B_n^+$ such that $b = ya$; in this case we say that a is a *right divisor*[2] of b. The center of B_n is the cyclic group generated by Δ_n^2, where

$$\Delta_n := \Delta_{n-1}\sigma_{n-1}\sigma_{n-2}\cdots\sigma_1 \in B_n^+$$

[2] The terminology is not accidental; it turns out that we can also define l.c.m.s and g.c.d.s in B_n^+, and that B_n is the group of fractions of B_n^+. These facts are some of the achievements of Garside theory [20].

(see p.30 of [22] for a simple proof). Geometrically, Δ_n implements a twist by π in the z-plane as the strands move from $z = 0$ to $z = 1$. One can show that $\sigma_i \preccurlyeq \Delta_n$ for all i, i.e. there exists $x_i \in B_n^+$ such that $\sigma_i^{-1} = x_i \Delta_n^{-1}$. Given a word w in the σ_i and their inverses, we first replace the leftmost instance of an inverse generator (say it is σ_i^{-1}) with $x_i \Delta_n^{-1}$. We then insert $\Delta_n^{-1}\Delta_n$ in front of x_i, and observe that conjugating a positive braid x by Δ_n results in another positive braid (specifically, the rotation of x by π in the z-plane). In this way, we can push Δ_n^{-1} all the way to the left. We repeat this process for each inverse generator appearing in the word, resulting in a word of the form $\Delta_n^p b$ where $p \in \mathbb{Z}$ and $b \in B_n^+$. Since we can solve the word problem in B_n^+, we can factor out the maximal power of Δ_n appearing as a left divisor of b. We thus have that, as elements of the braid group, $w = \Delta_n^{p'} b'$ with Δ_n not a left divisor of b' and p' unique. This solves the word problem in B_n.

We can make the above algorithm efficient by finding an efficiently computable normal form for a positive braid word b that does not have Δ_n as a left divisor. Recall that the symmetric group S_n has a remarkably similar presentation to B_n. Indeed, starting with (1), letting $\sigma_i = (i\ i+1)$ and adding the relations $\sigma_i^2 = 1$ for all i results in the standard presentation of S_n. In other words, there is a surjective homomorphism $\pi : B_n \to S_n$ with $\sigma_i \mapsto (i\, i+1)$. In terms of the geometric interpretation, a braid is mapped to the permutation on $[n]$ defined by the connections between the top and bottom "pegs," as in Figure 1. For each $\sigma \in S_n$, there is a unique preimage of σ that can be drawn so that any given pair of strands cross only in the positive direction, and at most once. We call such braids *simple braids*, and they form a subset of B_n^+ of size $n!$.

▶ **Definition 3** (p. 4 of [14]).

1. A sequence of simple braids $(s_1, \ldots, s_p)$ is said to be *normal* if, for each j, every σ_i that is a left divisor of s_{j+1} is a right divisor of s_j.
2. A sequence of permutations $(f_1, \ldots, f_p)$ is said to be *normal* if, for each j, $f_{j+1}^{-1}(i) > f_{j+1}^{-1}(i+1)$ implies $f_j(i) > f_j(i+1)$.

A sequence of simple braids $(s_1, \ldots, s_p)$ is normal if and only if the sequence of permutations $(\pi(s_1), \ldots, \pi(s_p))$ is normal. Given a permutation $f \in S_n$, let $\hat{f}$ denote the simple braid of B_n satisfying $\pi(\hat{f}) = f$.

▶ **Theorem 4** (p. 4 of [14] and Ch. 9 of [15]).

1. *Every braid z in B_n admits a unique decomposition of the form $\Delta_n^m s_1 \ldots s_p$ with $m \in \mathbb{Z}$ and $(s_1, \ldots, s_p)$ a normal sequence of simple braids satisfying $s_1 \neq \Delta_n$ and $s_p \neq 1$.*
2. *Every braid z in B_n admits a unique decomposition of the form $\Delta_n^m \hat{f}_1 \ldots \hat{f}_p$ with $m \in \mathbb{Z}$ and $(f_1, \ldots, f_p)$ a normal sequence of permutations satisfying $f_1 \neq \pi(\Delta_n)$ and $f_p \neq 1$.*

The most efficient algorithms for computing the normal form of a word w in the generators of B_n have complexity $O(|w|^2 n \log n)$ [14].

3 Obfuscation of Classical Reversible Circuits

3.1 Reversible Circuits

In the next section, we will describe a gate R which is universal for classical computation and satisfies Definition 2 when Γ is the set of relations of the braid group. Because group elements are invertible, R must be a reversible gate, that is, it must bijectively map its possible inputs to its possible outputs. We will thus work in the setting of *reversible classical circuits.* These circuits are composed entirely of reversible gates. For more background on reversible computation see [9, 17, 36].

Because reversible circuits cannot erase any information, they operate using ancillary dits ("ancillas") to store unerasable data left over from intermediate steps in the computation. A reversible circuit evaluating a function $f : \{0, \dots, d-1\}^n \to \{0, \dots, d-1\}^m$ thus operates on $r \geq \max(n, m)$ dits, where $r - n$ of the input dits are work dits to be initialized to some fixed value independent of the problem instance, and $r - m$ of the output dits contain unerasable leftover data, to be ignored. Efficient procedures are known for compiling arbitrary logic circuits into reversible form, e.g., by using the Toffoli (or CCNOT) gate [9, 17].

In adapting Definitions 1 and 2 to reversible circuits, one is faced with two natural choices for the notion of functional equivalence. One may either demand that the original and obfuscated circuits implement the same function $f : \{0,1\}^n \to \{0,1\}^m$, ignoring the ancilla dits (*weak equivalence*), or demand that they implement the same transformation on the entire set of r dits, including the ancillas (*strong equivalence*). Our constructions will satisfy the latter. Strong equivalence implies weak equivalence, so our construction proves that both possible definitions of partial-indistinguishability are polynomial-time achievable when Γ is the set of relations of the braid group. We remark that, as with ordinary irreversible circuits, determining if two arbitrary reversible circuits are equivalent (weakly or strongly) is coNP-complete [29].

3.2 Classical computation with braids

We now briefly describe a classical reversible gate R which satisfies the braid relations. The complete details of the construction and the proof of universality of R are given in Appendix A. Taken together with Theorem 4, this yields an obfuscator satisfying Definition 2.

Let G be a finite group and set $d = |G|$. Consider the reversible gate R that acts on pairs of dits encoding group elements by

$$R(a, b) = (b, b^{-1}ab). \tag{2}$$

Let R_i denote R acting on the i and $(i+1)^{\text{th}}$ wires of a circuit. By direct calculation, one can check that the set $\{R_1, \dots, R_{n-1}\}$ satisfies the braid relations, that is,

$$\begin{array}{rcll} R_iR_j & = & R_jR_i & \forall\ |i-j| \geq 2 \\ R_iR_{i+1}R_i & = & R_{i+1}R_iR_{i+1} & \forall\ i. \end{array} \tag{3}$$

In 1997, Kitaev discovered that the gate set $\{R, R^{-1}\}$ is universal for classical reversible computation when G is the symmetric group S_5 [31]. Ogburn and Preskill subsequently showed that the alternating group A_5, which is half as large as S_5, is already sufficient [37]. The universality construction for A_5 was subsequently presented in greater detail and generalized to all non-solvable groups by Mochon [34]. To make our presentation more accessible and self-contained, we give in Appendix A an explicit description of Mochon's universality construction in the the case $G = A_5$. The construction proves computational universality by showing how to efficiently compile Toffoli circuits into R-circuits.

Given any R-circuit, we can apply the algorithm of Theorem 4 by interpreting each R_i as σ_i and each R_i^{-1} as σ_i^{-1}. This leads to partial-indistinguishability obfuscation of R-circuits. A discussion of whether this can also yield meaningful obfuscation for classical circuits constructed from other gate sets is given in Section 5.

4 Quantum Circuits

4.1 Quantum computation with braids

In Section 3.2 and Appendix A, we discuss classical universality of circuits encoded as braids. It turns out that an analogous theory can be developed for quantum circuits, and is well-understood. The family of so-called Fibonacci representations of the braid groups have dense image in the unitary group, and there are efficient classical algorithms for translating any quantum circuit into a braid (and vice-versa) in a way that preserves unitary functionality [18]. A brief synopsis of these facts is given below. We remark that there are in fact many unitary representations of the braid groups that satisfy these properties, and which are physically motivated by the so-called fractional quantum Hall effect. In this setting, the image of these representations consists of unitary operators which describe the braiding of excitations in a 2-dimensional medium [31].

Approachable descriptions of the Fibonacci representation are given in [40, 42]. In [40], what we call the "Fibonacci representation" here, is called the "$\star\star$" irreducible subrepresentation. This is a family of representations $\rho_{\mathrm{Fib}}^{(n)} : B_n \to U(F_{n-4})$, where F_k is the k-th Fibonacci number. For our application, the essential properties of the Fibonacci representation are *locality* and *local density*. These two properties mean that, under a certain qubit encoding, braid generators correspond to local unitaries, and local unitaries correspond to short braid words. Standard arguments from quantum computation tell us that we can achieve the latter to precision ϵ with $O(\log^{2.71}(1/\epsilon))$ braid generators by means of the Solovay-Kitaev algorithm [13].

A natural basis for the space of $\rho_{\mathrm{Fib}}^{(n)}$ can be identified with strings of length n from the alphabet $\{\star, p\}$, which begin with $\star$, end with p, and do not contain "$\star\star$" as a substring[3]. Following [2][4], for n a multiple of four, we identify a particular subspace V_n of $\rho_{\mathrm{Fib}}^{(n)}$ by discarding some basis elements, as follows. Partition a string s into substrings of length four. If each of these substrings is equal to either $\star p \star p$ (this will encode a 0) or $\star ppp$ (this will encode a 1), then the basis element corresponding to s is in V_n; otherwise, it is not. Note that $\dim V_n = 2^{n/4}$. The following theorem follows from [2, 13].

▶ **Theorem 5.** *There is a classical algorithm which, given an $(n/4)$-qubit quantum circuit C and $\epsilon > 0$, outputs a braid $b \in B_n$ of length $O(|C| \log^{2.71}(1/\epsilon))$ satisfying*

$$\left\| C - \rho_{\mathrm{Fib}}^{(n)}(b)\Big|_{V_n} \right\| \leq \epsilon \; ;$$

this algorithm has complexity $O(|b|)$.

For the opposite direction, we can identify a subspace $W_n \subset (\mathbb{C}_2)^{\otimes n}$ by discarding all bitstrings except those that start with 0, end with 1 and do not have "00" as a substring. Then $\dim W_n = \dim \rho_{\mathrm{Fib}}^{(n)}$ and we have the following.

▶ **Theorem 6.** *There is a classical algorithm which, given $b \in B_n$ and $\epsilon > 0$, outputs a quantum circuit C on n qubits of length $O(|b| \log^{2.71}(1/\epsilon))$ such that*

$$\left\| C|_{W_n} - \rho_{\mathrm{Fib}}^{(n)}(b) \right\| \leq \epsilon \; ;$$

this algorithm has complexity $O(|C|)$.

[3] In [40] the $\star\star$ subrepresentation of B_n acts on strings of length $n+1$ that begin and end with $\star$. One can leave the initial and/or final $\star$ implicit as these are left unchanged by all braiding operations. We omit the final $\star$ leaving us strings of length n that begin with $\star$ and end with p.

[4] Reference [2] describes the basis vectors in terms of "paths". The correspondence between the path notation and the $p\star$ notation is given in appendix C of [40].

The two algorithms in the above theorems are described explicitly in [2].

4.2 Obfuscating quantum computations

While the state of knowledge about classical obfuscation is limited, essentially nothing is known about the quantum case. Here we discuss how to use the facts from the previous section to construct a partial-indistinguishability obfuscator for quantum circuits.

In light of Theorem 5, $\{\rho_{\mathrm{Fib}}(\sigma_1), \ldots, \rho_{\mathrm{Fib}}(\sigma_{n-1})\}$ may be regarded as a universal set of elementary quantum gates. By the homomorphism property of ρ_{Fib}, this set satisfies the braid relations. These gates differ from conventional quantum gates in that they do not possess locality defined in terms of a strict tensor product structure. Nevertheless, as shown above, the power of unitary circuits composed from these gates is equivalent to standard quantum computation. By interpreting each $\rho_{\mathrm{Fib}}(\sigma_j)$ as a braid-group generator σ_j, we can apply the algorithm from Theorem 4 directly to circuits from this gate set, resulting in a partial-indistinguishability obfuscator satisfying Definition 2.

With the algorithms from the previous section in hand, we could also attempt to apply the obfuscation algorithm, Algorithm 1, directly to quantum circuits. For an input circuit C on n qubits, the running times of both of this algorithm is $O(|C|^2 n \cdot \mathrm{polylog}(n, 1/\epsilon))$. The length of the output cannot be longer than the running time. We are not aware of a better upper bound for the length of the output. The security of this algorithms is questionable, and some attacks and possible countermeasures are discussed in Section 5.

Note that reduction of arbitrary quantum circuits to a normal form using a *complete* set of gate relations should not be possible in polynomial time; this would yield a polynomial-time algorithm for deciding whether a quantum circuit is equivalent to the identity, which is a coQMA-complete problem [28].

4.3 Testing claimed quantum computers with a quantum obfuscator

It is natural to consider quantum analogues of the applications of obfuscation from classical computer science. We now consider a potential application of quantum circuit obfuscation that does not fit this mold: testing claimed quantum computers. A similar proposal using a restricted class of quantum circuits has been previously made in [39].

Suppose Bob claims to have access to a universal quantum computer with some fixed finite number of qubits. Alice has access to a classical computer only, as well as a classical communication channel with Bob. Can Alice determine if Bob is telling the truth? Barring tremendous advances in complexity theory, a provably correct test is unlikely;[5] can we still design a test in which we have a high degree of confidence? Given the extensive work on classical algorithms for factoring, a reasonable idea is to simply ask Bob to factor a sufficiently large RSA number. However, Shor's algorithm only begins to outperform the best classical algorithms when thousands of logical qubits can be employed. A much smaller universal quantum computer (e.g., a few dozen qubits) is likely to be a far simpler engineering challenge and could still be quite useful, e.g., for simulating certain quantum systems. A test that works in this case would thus be very valuable. We now outline a new proposal for such a test.

[5] Notice that even a proof that BQP $\neq$ BPP would be insufficient; one would have to find specific problems and instance sizes where some quantum strategy provably beats every classical one. We are thus left with a situation analogous to the practical security guarantees of modern cryptographic systems, which tell us how many bit operations it would take to crack a given instance using the fastest known algorithms.

Simply put, we propose asking questions that are classically easy to answer, but posing them in an obfuscated manner. In this test, Alice would repeatedly generate quantum circuits and ask Bob to run them. At least some of the circuits would in fact be quantumly-obfuscated classical reversible circuits, allowing Alice to easily check the answers. Previous work has yielded tests of quantum computers in the case that the verifier can perform some limited quantum operations [10, 3].

We have considerable freedom when designing an obfuscation-based test of quantum computers. How to choose these parameters in a way that makes the test difficult to fool with a classical computer is an open question. For purposes of illustration, we give one example. Let $\mathcal{O}$ be the obfuscation algorithm for quantum circuits described above.

▶ Algorithm 2.

1. *Select a random bitstring s of length k.*
2. *Let C be the $(k+1)$-bit circuit that, on all-zero input, initializes wires 2 through $k+1$ to s and then computes the parity of s into the first wire.*
3. *Compute $\mathcal{O}(C)$, and let n be the number of qubits needed to run $\mathcal{O}(C)$.*
4. *Ask Bob to run D on the all zeros string and return the first bit of output.*

Clearly, k must be chosen so that n is smaller than the number of logical qubits Bob claims to control. To fool Alice, a purely classical Bob must determine the parity of s. The dictionary attack (*i.e.* Bob repeatedly guesses at k, obfuscates the corresponding circuit, and compares the result to the circuit given by Alice) is of no use provided k is reasonably large, e.g., 80 bits, which can be encoded using a braid of 115 strands using the Zeckendorf encoding described in [40].

We now show that there can be no efficient general-purpose algorithm for breaking our test by detecting whether a given quantum circuit is in fact (almost) classical, and if so, simulating it.

▶ **Definition 7.** Let c be a bit string specifying a quantum circuit via a standard universal set Q of quantum gates, and let U_c be the corresponding unitary operator. Fix some constants $r, d, a \in \mathbb{N}$, and fix a set R of reversible gates. The problem $\mathrm{CLASS}(r, d, a, Q, R)$ is to find a reversible circuit of at most $r|c|^d$ gates from R such that the corresponding permutation matrix P satisfies $\|U_c - P\| \leq 2^{-a|c|}$.

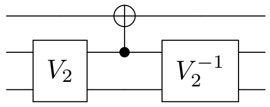
Note that $\mathrm{CLASS}(r, d, a, Q, R)$ is not a decision problem. Thus, to formulate the question of whether this problem can be efficiently solved, we must ask not whether $\mathrm{CLASS}(r, d, a, Q, R)$ is contained in P but whether it is contained in FP. We now provide some formal evidence that this is not the case. Note that the following theorems continue to hold if we change the classicality condition in Definition 7 to $\|U_c - P\| \leq |c|^{-a}$.

▶ **Theorem 8.** *For any fixed $r, d, a \in \mathbb{N}$, any universal reversible gate set R, and any universal quantum gate set Q, if $\mathrm{CLASS}(r, d, a, Q, R) \in \mathrm{FP}$ then $\mathrm{QCMA} \subseteq \mathrm{P}^{\mathrm{NP}}$.*

Note that, $\mathrm{QCMA} \subseteq \mathrm{P}^{\mathrm{NP}}$ would be very surprising because, among other things, it would imply $\mathrm{BQP} \subseteq \mathrm{PH}$, and there is evidence that this is false [1, 16].

Proof. The standard QCMA-complete language $\mathcal{L}$ is as follows. Let $\mathcal{C}$ be the set of all quantum circuits (expressed as a concatenation of bitstrings that index elements of the gate set Q). $\mathcal{C}$ decomposes as the disjoint union of $\mathcal{L}$ and $\bar{\mathcal{L}}$ where $\mathcal{L}$ consists of the quantum circuits that accept at least one classical (*i.e.* computational basis state) input, and $\bar{\mathcal{L}}$ consists of the circuits that reject all inputs. Given a quantum circuit $V_1 \in \mathcal{C}$, (the "verifier") we can amplify it using standard techniques [33, 35] to accept YES instances with probability at least $1 - O(2^{-n})$ and accept NO instances with probability at most $O(2^{-n})$. Let V_2 be such an amplified verifier. Further, let

$V_3 \quad =$



where the second-to-top qubit is the acceptance qubit of V_2. If $V_i \in \bar{\mathcal{L}}$ then $\|V_3 - \mathbb{1}\| = O(2^{-n})$. By assumption, there exists a polynomial time classical algorithm for solving CLASS(r, d, a, Q, R). When presented with V_3, this algorithm will produce a polynomial-size reversible circuit V_4 strongly equivalent to the identity. By querying an oracle for the problem of strong equivalence of reversible circuits, one can decide whether V_4 is equivalent to the circuit of no gates, and hence to the identity operation. If $V_1 \in \bar{\mathcal{L}}$, this oracle will accept. If $V_1 \in \mathcal{L}$ then the algorithm for problem 1 will answer NO or produce a circuit that this oracle rejects. As shown in [29], the problem of deciding strong equivalence of reversible circuits is contained in coNP. Thus, we can decide QCMA in $\mathrm{P}^{\mathrm{coNP}}$, which is equal to the more familiar complexity class P^{NP}. ◀

5 Some Attacks

5.1 Compiler attacks

The security or insecurity of braid-based partial-indistinguishability obfuscation remains an area of current investigation. From a purely information-theoretic point of view, the power of this obfuscation comes from the many-to-one nature of the map N that takes arbitrary braid words to their normal form. If the initial braid words are highly structured because they are obtained by compilation from a different gate set, then this can undermine or destroy the many-to-one feature of N.

In Section 3.2, we describe a reversible gate R on pairs of 60-state dits, corresponding to elements of A_5, that obeys the relations of the braid group and can perform universal classical computation. The gate itself and the proof that it is universal come from the quantum computation literature [31, 37, 34]. Appendix A recounts the universality proof of [34], which can be viewed as a compiler B_R that maps circuits constructed from the well-known universal reversible Toffoli gate into circuits constructed from the R gate. As a cautionary example, we now show that naively obfuscating Toffoli circuits using the composed map $\mathsf{N} \circ \mathsf{B}_R$ is completely insecure.

The construction in Appendix A gives a general mapping from a Toffoli gate to a corresponding braid. We will refer to braids obtained in this way as *Toffoli braids.* Recall that the normal form of a braid in B_n has the form $\Delta_n^m s_1 \dots s_p$ for a normal sequence of simple braids $(s_1, \dots, s_p)$. A Toffoli braid obtained from a Toffoli with controls c_1 and c_2 and target t has normal form

$$\Delta_n^0 s_1(c_1, c_2, t) s_2 s_3 s_4 s_5 s_6 s_7 s_8 s_9(c_1, c_2, t) s_{10} s_{11} s_{12} s_{13}(c_1, c_2, t) s_{14}(t). \tag{4}$$

The factors $s_2, \dots, s_8, s_{10}, s_{11}$ and s_{12} only depend on n, and not on the wires c_1, c_2 or t. Note that this is a positive braid — consisting only of $\sigma_1, \dots, \sigma_{n-1}$ and none of their inverses. Any product of such braids will thus also be a positive braid, so attempting to obfuscate a circuit in Toffoli gates using this construction will yield only positive braids.

Because Toffoli is a 3-bit gate, there are only $\binom{n}{3}$ ways to apply a Toffoli to n bits. Thus, one may, in polynomial time, test each of these $\binom{n}{3}$ possibilities as a guess for the last gate of the obfuscated circuit. One performs the test by compiling the guessed Toffoli gate into a braid, appending the inverse of this braid to the normal form braid produced as the output the obfuscator, and then reducing the resulting braid to normal form. If the guess is correct, then the resulting braid is still a braid corresponding to a circuit — the original obfuscated circuit with its last Toffoli gate removed — and thus this will result in a positive braid. If the guess is incorrect, then appending the inverse of a positive braid, which consists entirely of $\sigma_1^{-1}, \ldots, \sigma_{n-1}^{-1}$, might result in a braid that is no longer positive — that is, has a negative power of Δ_n, and this seems to be the case with *any* wrong guess, based on some limited tests.

Furthermore, the presence of a negative power of Δ_n is efficiently recognizable, so it is immediately clear whether or not the guess was correct.

This attack is related to so-called length-based attacks. These have been introduced in the cryptanalysis of braid based key-exchanged protocols [25]. In the present context, the natural length-based attack is to guess the final gate, append the inverse of the corresponding braid to the normal-form braid produced by the obfuscator, and the reduce the product braid to normal form. If the result is a shorter word in the braid-group generators than the original normal form, then this can be taken as heuristic evidence that the guess was correct. Intuitively, one expects that the longer the braid words are that implement individual gates from the original gate set, then the better such attacks should work.

One can easily propose modifications to the naive obfuscator $\mathsf{N} \circ \mathsf{B}_R$ that thwart guessing-based attacks such as the two attacks described above. In particular, one finds that the gate R described in Appendix A has order 60. Hence, one can start with the positive Toffoli braid in equation (4) and then each generator σ_i can independently, with probability $\frac{1}{2}$, be replaced with σ_i^{-59}, without altering the functionality of the circuit. The number of generators in a Toffoli braid depends on n, and which wires the Toffoli acts on, but there are always at least 124. Thus, each gate will be compiled into one of 2^{124} braid-words uniformly at random. Thus, guessing-based attacks on the composition of this compiler with N may become impractical. Whether such a scheme is vulnerable to other attacks remains an open question for future research.

5.2 Dictionary attacks

The partial-indistinguishability obfuscator described in the preceding sections deterministically maps input circuits to obfuscated circuits. This creates a potential weakness in the obfuscation. Suppose Alice wishes to run a computation C on Bob's server but does not wish Bob to know what computation she is running. Thus, she sends the obfuscated circuit $\mathcal{O}(C)$ to Bob, who executes it, and returns the result. To improve security, Alice may instead use a circuit C' in which her desired input is hard-coded, and which applies a one-time pad at the end of the computation. If the obfuscation is secure, then Bob is unlikely to learn anything about C, the input, or the output. However, if Bob knows that the circuits Alice is likely to want to execute are drawn from some small set S, then Bob can simply compute $\{\mathcal{O}(s) | s \in S\}$ and identify Alice's computation by finding it in this list. Such attacks are sometimes called "dictionary" attacks after the practice of recovering passwords by feeding all words from a dictionary into the hash function and comparing against the hashed password.

Dictionary attacks may or may not be a serious threat to our obfuscation scheme, depending on the the size of the set of likely circuits to be obfuscated. In cryptographic applications where dictionary attacks are a concern, the standard way to protect against them is to append random bits prior to encryption. (In the context of hashing passwords, this practice is called "salting".) Such a strategy can be applied to our obfuscator, but some care must be taken in doing so. The most obvious strategy is to append a random circuit on the output ancillas prior to obfuscation. However, attackers can defeat this countermeasure by using the polynomial-time algorithms for computing left-greatest-common-divisors in the braid group [15]. However, prior to obfuscation, one may introduce extra dits, and apply random circuits before, after, and simultaneously with the computation, in a way so as not to disrupt it. The problem of optimizing the details of this procedure so as to maximize security and efficiency is left to future work.

6 Future Work

6.1 Classical and quantum universality

It is of interest to consider other computationally universal representations of the braid group, which might provide more efficient translations from circuits to braids. One avenue for obtaining such representations is by finding other solutions to the Yang-Baxter equation, besides the operator R from Appendix A. Our investigations so far prove that no permutation matrix solution of dimension up to 16×16 is a universal gate and suggest that no permutation matrix solution of dimension 25×25 is a universal gate. In the quantum case, it has been shown that no 4×4 unitary solution is universal [4].

More generally, one may look for other finitely-generated groups with computationally universal representations and efficiently computable normal forms. One potential candidate family are the mapping class groups $\mathrm{MCG}(\Sigma_g)$ of unpunctured surfaces of genus g. These groups also have quantumly universal representations [5] and an efficiently solvable word problem [23]. It is not known if there are also classically universal permutation representations, or if there are efficiently computable normal forms.

6.2 Expanding the set of indistinguishability relations

By [29], achieving efficient indistinguishability obfuscation for the complete set of relations of a universal gate set is unlikely. However, it is possible that partial-indistinguishability obfuscation on R gates could be achieved with a larger set of relations than the braid relations. For example, the universal reversible gate described in Appendix A has order 60. If we add the relations $\sigma_i^{60} = \mathbb{1}$ for $i = 1, 2, \ldots, n-1$ to B_n, we obtain a "truncated" (but still infinite for large n [12]) factor of the braid group. If a normal form can still be computed in polynomial time for this group then one could construct an efficient obfuscator using the relations of this truncated group, which would be strictly stronger than our braid group obfuscator. This approach also provides motivation for finding a complete set of relations for the gate R.

Acknowledgements. We thank Anne Broadbent, Rainer Steinwandt, Scott Aaronson, Bill Fefferman, Leonard Schulman, Robert König, and Yi-Kai Liu for helpful discussions. We also thank Mariano Suárez-Alvarez and Gjergji Zaimi for leading us to reference [12] via `math.stackexchange` and `mathoverflow`. Portions of this paper are a contribution of NIST, an agency of the US government, and are not subject to US copyright.

References

1 Scott Aaronson. BQP and the polynomial hierarchy. In *STOC'10: Proceedings of the 42nd ACM Symposium on Theory of Computing*, pages 141–150, 2010. arXiv:0910.4698.

2 Dorit Aharonov and Itai Arad. The BQP-hardness of approximating the Jones polynomial. *New Journal of Physics*, 13(3):035019, 2011.

3 Dorit Aharonov, Michael Ben-Or, and Elad Eban. Interactive proofs for quantum computation. In *Proceedings of Innovations in Computer Science (ICS 2010)*, pages 453–469, 2010. arXiv:0810.5375.

4 G. Alagic, S. Jordan, and A. Bapat. Classical simulation of Yang-Baxter gates. To appear in: Proceedings of TQC2014.

5 Gorjan Alagic, Stephen P. Jordan, Robert Koenig, and Ben W. Reichardt. Approximating Turaev-Viro 3-manifold invariants is universal for quantum computation. *Physical Review A*, 82:040302(R), 2010. arXiv:1003.0923.

6 Emil Artin. Theorie der Zöpfe. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 4:42–72, 1925.

7 B. Barak. Can we obfuscate programs? http://www.cs.princeton.edu/~boaz/Papers/obf_informal.html.

8 Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. In *Advances in Cryptology – CRYPTO 2001*, number 2139 in Lecture Notes in Computer Science, pages 1–18. Springer-Verlag, 2001.

9 C. H. Bennett. Logical reversibility of computation. *IBM Journal of Research and Development*, 17(6):525–532, 1973.

10 Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. Universal blind quantum computation. In *Proceedings of the 50th Annual IEEE Symposium on Fountations of Computer Science (FOCS 2008)*, pages 517–526, 2009. arXiv:0807.4154.

11 Christian S. Collberg and Clark Thomborson. Watermarking, tamper-proofing, and obfuscation – tools for software protection. *IEEE Transactions on Software Engineering*, 28(8):735–746, 2002.

12 H. S. M. Coxeter. Factor groups of the braid group. In *Proceedings of the 4th Canadian Mathematical Congress*, pages 95–122, 1959. See http://mathoverflow.net/questions/48849/.

13 Christopher M. Dawson and Michael A. Nielsen. The Solovay-Kitaev algorithm. *Quantum Information and Computation*, 6(1):81–95, 2006. arXiv:quant-ph/0505030.

14 Patrick Dehornoy. Efficient solutions to the braid isotopy problem. *Discrete Applied Mathematics*, 156:3094–3112, 2008. arxiv:math/0703666.

15 D. Epstein, J. Cannon, D. Holt, S. Levy, M. Paterson, and W. Thurston. *Word processing in groups*. Jones and Bartlett Publ., 1992.

16 Bill Fefferman and Chris Umans. Pseudorandom generators and the BQP vs. PH problem, 2010. arXiv:1007.0305.

17 E. Fredkin and T. Toffoli. Conservative logic. *International Journal of Theoretical Physics*, 21(3/4):219–253, 1982.

18 Michael H. Freedman, Michael Larsen, and Zhenghan Wang. A modular functor which is universal for quantum computation. *Communications in Mathematical Physics*, 227:605–622, 2002. arXiv:quant-ph/0001108.

19 Sanjam Garg, Craig Gentry, Shai Halevi, Mariana Raykova, Amit Sahai, and Brent Waters. Candidate indistinguishability obfuscation and functional encryption for all circuits. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 40–49, 2013.

20 F. A. Garside. The braid group and other groups. *Quart. J. Math. Oxford Ser.*, 2, 20:235–254, 1969.

21 Shafi Goldwasser and Guy N. Rothblum. On best-possible obfuscation. In *Theory of Cryptography – TCC 2007*, pages 194–213. Springer, 2007.

22 Juan González-Meneses. Basic results on braid groups, 2010. `arxiv:1010.0321 [math]`.

23 Hessam Hamidi-Tehrani. On complexity of the word problem in braid groups and mapping class groups. *Topology and its Applications*, 105:237–259, 2000.

24 Jarmo Hietarinta. All solutions to the constant quantum Yang-Baxter equation in two dimensions. *Physics Letters A*, 165:245–251, 1992.

25 D. Hofheinz and R. Steinwandt. A practical attack on some braid group based cryptographic primitives. In *Public Key Cryptography*, pages 187–198, 2003.

26 Edward V. Huntington. Sets of independent postulates for the algebra of logic. *Transactions of the American Mathematical Society*, 4:288–309, 1904.

27 Kazuo Iwama, Yahiko Kambayashi, and Shigeru Yamashita. Transformation rules for designing CNOT-based quantum circuits. In *DAC'02: Proceedings of the 39th Annual Design Automation Conference*, pages 419–424, 2002.

28 Dominik Janzing, Pawel Wocjan, and Thomas Beth. "Identity Check" is QMA-complete, 2003. `arXiv:quant-ph/0305050`.

29 Stephen Jordan. Strong equivalence of reversible circuits is coNP-complete. *Quantum Information and Computation*, 14(15/16):1303–1308, 2014. arXiv:1307.0836.

30 Louis H. Kauffman. *Knots and Physics*. Wold Scientific, 1991.

31 A. Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303:2–30, 2003. `arXiv:quant-ph/9707021`.

32 Hari Krovi and Alexander Russell. Quantum fourier transforms and the complexity of link invariants for quantum doubles of finite groups, 2012. `arXiv:quant-ph/1210.1550 [quant-ph]`.

33 Chris Marriott and John Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005. arXiv:cs/0506068.

34 Carlos Mochon. Anyons from nonsolvable finite groups are sufficient for universal quantum computation. *Physical Review A*, 67(2):022315, 2003. `arXiv:quant-ph/0206128`.

35 Daniel Nagaj, Pawel Wocjan, and Yong Zhang. Fast amplification of QMA. *Quantum Information and Computation*, 9(11/12):1053–1068, 2009. arXiv:0904.1549.

36 Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2000.

37 R. Walter Ogburn and John Preskill. Topological quantum computation. In *Quantum Computing and Quantum Communications*, volume 1509 of *Lecture Notes in Computer Science*, pages 341–356. Springer, 1999. First NASA International Conference QCQC'98.

38 Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: Deniable encryption, and more. *IACR Cryptology ePrint Archive*, 2013:454, 2013.

39 Dan Shepherd and Michael J. Bremner. Temporally unstructured quantum computation. *Proceedings of the Royal Society A*, 465:1413–1439, 2009. arXiv:0809.0847.

40 Peter W. Shor and Stephen P. Jordan. Estimating Jones polynomials is complete for one clean qubit. *Quantum Information and Computation*, 8(8/9):681–714, 2008. `arXiv:0707.2831`.

41 Eric D. Simonaire. Sub-circuit selection and replacement algorithms modeled as term rewriting systems. Master's thesis, Air Force Institute of Technology, 2008.

42 Simon Trebst, Matthias Troyer, Zhenghan Wang, and Andreas W. W. Ludwig. A short introduction to Fibonacci anyon models. *Progress in Theoretical Physics Supplement*, 176:384–407, 2008. `arXiv:0902.3275`.

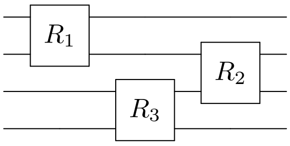


Figure 2 An example of a reversible circuit constructed from a single gate R. As a product of matrices, we write this $R_2 R_3 R_1$, in keeping with the convention [36] that circuit diagrams are to be read left-to-right, whereas the matrix product acts right-to-left. Note that in subsequent circuit diagrams we drop the subscripts from the R gates as these can be read off from the "wires" the gates act on.

A Classical Computation with Braids

In this section, we present a reversible gate R on pairs of 60-state dits that can perform universal computation and obeys the relations of the braid group. The universality construction for this gate comes from the quantum computation literature [31, 37, 34], but we present it here in purely classical language to make it accessible to a broader audience.

Suppose we arrange n dits on a line, and allow R to act only on neighboring dits. Further, we do not allow R to be applied "upside-down". Then, there are $n-1$ choices for how to apply R. We label these $R_1, R_2, \ldots, R_{n-1}$, as illustrated in Figure 2. Each of $R_1, \ldots, R_{n-1}$ corresponds to a $d^n \times d^n$ permutation matrix. Specifically, R_j is obtained by taking the tensor product of R with identity matrices according to $R_j = \mathbb{1}_{d \times d}^{\otimes (j-1)} \otimes R \otimes \mathbb{1}_{d \times d}^{\otimes (n-j-1)}$.

$R_1, \ldots, R_{n-1}$ generate a subgroup of S_{d^n}. Among others, these generators obey the relations

$$R_i R_j = R_j R_i \quad \forall |i-j| \geq 2. \tag{5}$$

If R satisfies

$$R_1 R_2 R_1 = R_2 R_1 R_2 \tag{6}$$

then

$$R_i R_{i+1} R_i = R_{i+1} R_i R_{i+1} \quad \forall i \tag{7}$$

and in this case the gates $R_1, \ldots, R_{n-1}$ satisfy all the relations of the braid group B_n. In other words, the map defined by $\sigma_i \mapsto R_i$ and $\sigma_i^{-1} \mapsto R_i^{-1}$ is a homomorphism from B_n to S_{d^n}, *i.e.* a representation of the braid group. Note that this representation is never faithful as B_n is infinite.

The condition 6 is known as the Yang-Baxter equation[6]. Finding all the matrices satisfying the Yang-Baxter equation at a given dimension has only been achieved at $d = 2$ [24]. However, certain systematic constructions coming from mathematical physics can produce infinite families of solutions. In particular, let G be any finite group, and let R be the permutation on the set $G \times G$ defined by

$$R(a, b) = (b, b^{-1} a b). \tag{8}$$

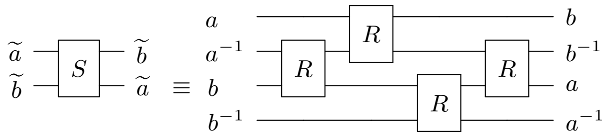
[6] Actually, two slightly different equations go by the name Yang-Baxter in the literature. Careful sources distinguish these as the algebraic Yang-Baxter equation and the braided Yang-Baxter relation (which is sometimes called the quantum Yang-Baxter equation). Equation 6 is the latter. Furthermore, some sources treat a more complicated version of the Yang-Baxter equation in which R depends on a continuous parameter. In such works equation 6 is often referred to as the constant Yang-Baxter equation.

By direct calculation one sees that any such an R satisfies the Yang-Baxter equation. (In physics language, R comes from the braiding statistics of the magnetic fluxes in the quantum double of G.)

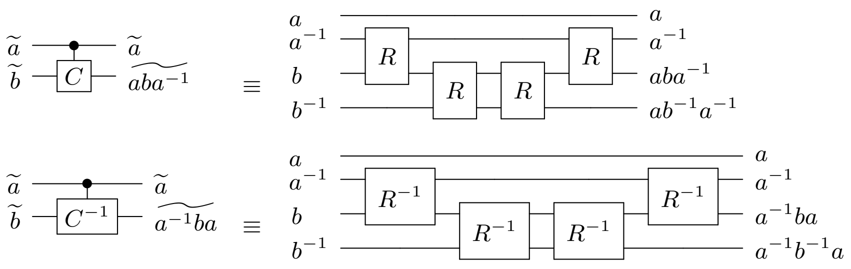
In 1997, Kitaev discovered that choosing G to be the symmetric group S_5 yields an R gate sufficient to perform universal reversible computation [31]. Ogburn and Preskill subsequently showed that the alternating group A_5, which is half as large as S_5, is already sufficient. The universality construction for A_5 was subsequently presented in greater detail and generalized to all non-solvable groups by Mochon [34]. In the remainder of this section we give a self-contained exposition of the universality construction from [34], shorn of physics language.

To obtain a representation of the braid group, we must strictly enforce the requirement that application of R to neighboring dits on a line is the only allowed operation. In particular, we are not given as elementary operations the ability to apply R upside-down, or to non-neighboring dits, or to move dits around. Thus, to prove computational universality, it is helpful to first construct a SWAP gate from R gates, which exchanges neighboring dits. As is well-known, the $n-1$ swaps of nearest neighbors on a line generate the full group S_n of permutations, and thus a SWAP gate enables application of R to any pair of dits.

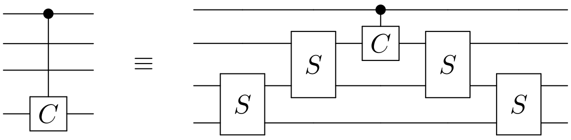
For R gates of the form (2), two pairs of inverse group elements in the order a, a^{-1}, b, b^{-1} can be swapped by applying the product $R_2 R_3 R_1 R_2$. Thus, in the construction of [37, 34], elements of A_5 are always paired with their inverses. This can be regarded as a form of encoding; $|A_5| = 60$, so each 60-state dit is encoded by a corresponding pair of elements of A_5. We introduce the notation $\widetilde{g} \equiv (g, g^{-1})$ for this encoding, and similarly, abbreviate the encoded swap operation as follows.

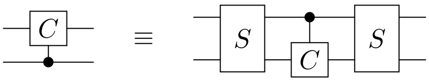


Similarly, the sequence $R_2 R_3 R_3 R_2$ performs the transformation $(\widetilde{a}, \widetilde{b}) \mapsto (\widetilde{a}, \widetilde{aba^{-1}})$ on a pair of encoded dits. We abbreviate this in circuit diagrams as follows.



This notation can easily be extended to provide a shorthand for the sequence of gates needed to implement a C gate between non-neighboring pairs of bits, as illustrated by the following examples.





Next, consider the following product of elements of A_5 (which should be read right-to-left).

$$f(g_1, g_2) = (521)g_1(14352)g_2(124)g_1^{-1}(15342)g_2^{-1}(521) \tag{9}$$

One sees that

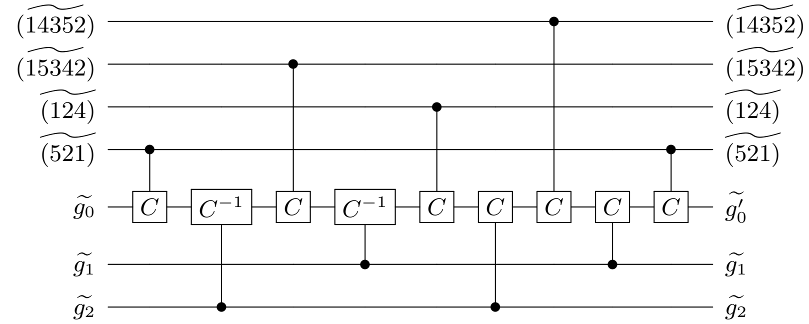
$$\begin{aligned} f((345),(345)) &= \mathbb{1} \\ f((345),(435)) &= \mathbb{1} \\ f((435),(345)) &= \mathbb{1} \\ f((435),(435)) &= (12)(34) \end{aligned}$$

where $\mathbb{1}$ denotes the identity permutation. Furthermore, conjugating (345) by (12)(34) yields (435), and conversely, conjugating (435) by (12)(34) yields (345). Thus, we may think of (345) as an encoded zero and (435) as an encoded one, and we see that

$$f(g_1, g_2)g_0 f(g_1, g_2)^{-1} \tag{10}$$

toggles g_0 between one and zero if g_1 and g_2 are both encoded ones and leaves g_0 unchanged otherwise. Such a doubly-controlled toggling operation is known as a Toffoli gate, which is well-known to be a computationally universal reversible gate [17].

As a circuit diagram, this construction can be expressed as follows.



Here, if g_0, g_1, g_2 encode bits b_0, b_1, b_2 then g_0' encodes $b_0 \oplus b_1 \wedge b_2$. The four ancillary dits $\widetilde{(14352)}$, $\widetilde{(15342)}$, $\widetilde{(124)}$, and $\widetilde{(521)}$, are used to "catalytically" facilitate the construction of a Toffoli gate, and thus computations built from arbitrarily many Toffoli gates can be performed with only one copy of these four dits.

Unpacking the various shorthand notations, one sees that the above circuit represents the following braid of 132 crossings on 14 strands, which encodes a Toffoli gate with the first wire as target, and the second and third wires as controls.

$$\begin{array}{rcllll}
T & = & \sigma_8\sigma_9\sigma_9\sigma_8 & \sigma_{10}\sigma_{11}\sigma_9\sigma_{10} & \sigma_{10}\sigma_{11}\sigma_{11}\sigma_{10} & \sigma_{10}\sigma_{11}\sigma_9\sigma_{10} \\
 & & \sigma_2\sigma_3\sigma_1\sigma_2 & \sigma_4\sigma_5\sigma_3\sigma_4 & \sigma_6\sigma_7\sigma_5\sigma_6 & \sigma_8\sigma_9\sigma_9\sigma_8 \\
 & & \sigma_6\sigma_7\sigma_5\sigma_6 & \sigma_4\sigma_5\sigma_3\sigma_4 & \sigma_2\sigma_3\sigma_1\sigma_2 & \sigma_{12}\sigma_{13}\sigma_{11}\sigma_{12} \\
 & & \sigma_{10}\sigma_{11}\sigma_9\sigma_{10} & \sigma_{10}\sigma_{11}\sigma_{11}\sigma_{10} & \sigma_{10}\sigma_{11}\sigma_9\sigma_{10} & \sigma_{12}\sigma_{13}\sigma_{11}\sigma_{12} \\
 & & \sigma_6\sigma_7\sigma_5\sigma_6 & \sigma_8\sigma_9\sigma_9\sigma_8 & \sigma_6\sigma_7\sigma_5\sigma_6 & \sigma_{10}\sigma_{11}\sigma_9\sigma_{10} \\
 & & \sigma_{10}^{-1}\sigma_{11}^{-1}\sigma_{11}^{-1}\sigma_{10}^{-1} & \sigma_{10}\sigma_{11}\sigma_9\sigma_{10} & \sigma_4\sigma_5\sigma_3\sigma_4 & \sigma_6\sigma_7\sigma_5\sigma_6 \\
 & & \sigma_8\sigma_9\sigma_9\sigma_8 & \sigma_6\sigma_7\sigma_5\sigma_6 & \sigma_4\sigma_5\sigma_3\sigma_4 & \sigma_{12}\sigma_{13}\sigma_{11}\sigma_{12} \\
 & & \sigma_{10}\sigma_{11}\sigma_9\sigma_{10} & \sigma_{10}^{-1}\sigma_{11}^{-1}\sigma_{11}^{-1}\sigma_{10}^{-1} & \sigma_{10}\sigma_{11}\sigma_9\sigma_{10} & \sigma_{12}\sigma_{13}\sigma_{11}\sigma_{12} \\
 & & \sigma_8\sigma_9\sigma_9\sigma_8 & & &
\end{array} \tag{11}$$

Note that we take the convention that this should be read backwards compared to the way one reads English text. This is in keeping with the conventional notation for the composition of functions and our right-to-left multiplication of R matrices. We have used whitespace to divide crossings into groups of four as these correspond to elementary S and R gates.

Given this construction of the Toffoli gate by braid crossings, it is a simple matter to "compile" any given logic circuit into a corresponding braid. A_5 has 60 elements. Thus, encoding a single bit into a a pair of A_5 elements appears somewhat wasteful. It is natural to try to find Yang-Baxter solutions acting on d-state dits for smaller d that achieve universal classical computation. In appendix B, we improve upon the A_5-based construction to show that $d = 44$ suffices. We have also used exhaustive computer search to find all permutation solutions satisfying the Yang-Baxter equation up to $d = 5$ (i.e. up to 25×25 permutation matrices). Our examination of these solutions suggests that none are computationally universal. Where between 5 and 44 lies the minimal d remains an interesting open question.

B Optimizing Classical Braid Gates

In appendix A we have recounted the construction of [34], which shows that the reversible gate R, which acts on pairs of 60-state dits and satisfies the Yang-Baxter equation, can perform universal classical computation. In this section, based on a suggestion of Robert König, we show that R can be modified to obtain a gate acting on pairs of 44-state dits that satisfies the Yang-Baxter equation and can perform universal classical computation. Our computational evidence suggests that no reversible gate on d-state dits satisfying the Yang-Baxter equation can perform universal computation for $d \leq 5$. Where between 5 and 44 the minimal d lies for which computationally universal reversible Yang-Baxter gates acting on d-state qudits exist remains an open question.

The universality construction of [34], recounted in appendix A, starts with all dits initialized to states from the following set.

$$\begin{array}{rcl}
S & = & \{g, g^{-1} | g \in S_0\} \\
S_0 & = & \{(14352), (15342), (124), (521), (345), (435)\}
\end{array}$$

Here we show that the orbit of S under the action of the gate R is not all of A_5, rather the orbit has only 44 elements. Thus the restriction of the matrix R onto this 44-dimensional subspace is a permutation-matrix that satisfies the Yang-Baxter equation and is capable of universal classical computation.

Recalling (2), one sees that the orbit O_R of S under R is

$$O_R = \{b^{-1}ab | a \in S, b \in \langle S \rangle\} \tag{12}$$

where $\langle S\rangle$ is the subgroup of A_5 generated by S. A simple computer algebra calculation shows that $\langle S\rangle = A_5$, thus O_R consists of exactly those elements of A_5 conjugate to S.

It is well known that the conjugacy classes of A_5 are as follows.

1) the identity	(1 element)
2) 3-cycles	(20 elements)
3) conjugates of (12)(34)	(15 elements)
4) conjugates of (12345)	(12 elements)
5) conjugates of (21345)	(12 elements)

One sees that O_R contains 2), and does not contain 1) or 3). The only remaining question is whether O_R contains both 4) and 5) or just one of them. A simple computer algebra calculation shows that (14352) and (15342) are non-conjugate elements of A_5. Hence O_R must contain both 4) and 5). Therefore, $|O_R| = 44$.

Classical Simulation of Yang-Baxter Gates

Gorjan Alagic[1], Aniruddha Bapat[1], and Stephen Jordan[2]

1 Institute for Quantum Information and Matter
California Institute of Technology
Pasadena, CA

2 National Institute of Standards and Technology
Gaithersburg, MD

Abstract

A unitary operator that satisfies the constant Yang-Baxter equation immediately yields a unitary representation of the braid group B_n for every $n \geq 2$. If we view such an operator as a quantum-computational gate, then topological braiding corresponds to a quantum circuit. A basic question is when such a representation affords universal quantum computation. In this work, we show how to classically simulate these circuits when the gate in question belongs to certain families of solutions to the Yang-Baxter equation. These include all of the qubit (i.e., $d = 2$) solutions, and some simple families that include solutions for arbitrary $d \geq 2$. Our main tool is a probabilistic classical algorithm for efficient simulation of a more general class of quantum circuits. This algorithm may be of use outside the present setting.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases Quantum, Yang-Baxter, Braid, Anyon

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.161

1 Introduction

The Yang-Baxter equation, named after C. N. Yang and R. J. Baxter, appears in a number of areas of mathematics and physics. Yang encountered the equation while working on two-dimensional quantum field theory, while Baxter applied it to exactly solvable models in statistical mechanics [2]. An accessible review of some of the many applications of the Yang-Baxter equation can be found in [21]. In this work, we will consider what is typically called the constant quantum Yang-Baxter equation, and is defined as follows. Let V be a finite-dimensional complex Hilbert space and R a linear operator on $V \otimes V$. Then R satisfies the quantum Yang-Baxter equation (YBE) if

$$(R \otimes I)(I \otimes R)(R \otimes I) = (I \otimes R)(R \otimes I)(I \otimes R)\,,$$

where I denotes the identity operator on V. In this case, we say that R is a Yang-Baxter operator. The YBE bears a close resemblance to the relation

$$\sigma_i \sigma_{i+1} \sigma_i = \sigma_{i+1} \sigma_i \sigma_{i+1}$$

of the braid group B_n. Indeed, a Yang-Baxter operator naturally gives the space $V^{\otimes n}$ the structure of a representation $\rho_{(R,n)}$ of B_n. Turaev showed that if R also satisfies the so-called Markov property, then it corresponds to an invariant of links [24]. The invariant is given by the (appropriately scaled) trace of $\rho_{(R,n)}$, evaluated at any braid whose trace closure is equal to the link. More generally, one can derive a link invariant from the trace of any representation of B_n which satisfies the Markov property. This is the case for the famous

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 161–175
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Jones representation and the corresponding Jones Polynomial invariant [15]. Freedman, Kitaev, Larsen and Wang [8, 9, 10] showed that the Jones representation has significant meaning in quantum computation. Informally speaking, the Jones representation provides a functionality-preserving "dictionary" between quantum circuits and braids. One consequence of these results is that additively approximating the Jones Polynomial is a universal problem for quantum computation. It also appears that this dictionary could correspond to a physically plausible implementation of quantum computers by means of exotic particles called non-abelian anyons [22]. One downside of the Jones representation in this context is that topological locality of braiding does not translate naturally into tensor-product locality of the corresponding quantum circuit. In particular, it is not the case that braiding two adjacent strands correponds to applying a Yang-Baxter operator on the space of two adjacent qubits. One might hope that the Jones representation could be made to look this way, e.g., by changing bases or manipulating the multiplicities of its irreducible summands. However, Rowell and Wang recently showed that this is impossible unless the Jones representation in question[1] is in fact *not* quantum-universal (see Corollary 4.2 in [23].)

Alternatively, one may ask if there exist other representations of the braid groups with the desired local structure and which exhibit computational universality. This amounts to finding unitary solutions to the YBE and determining if they are universal gates. In this work, we investigate low-dimensional solutions with this motivation in mind. All of the qubit (i.e., $d = \dim V = 2$) solutions to the YBE were found by Hietarinta [12]; the unitary ones among those were identified by Dye [5]. It was previously known that, when their eigenvalues are roots of unity, these solutions yield braid group representations with finite image [7, 6]. We show how to classically approximate the matrix entries of any quantum circuit constructed from a particular kind of two-qudit gate. Most of the qubit solutions to the YBE, as well as some solution families of arbitrary dimension, are special cases of this gate. For the remaining qubit solutions, we give a different result: how to classically simulate a quantum computation that begins in any product state, and ends with a measurement of an observable on logarithmically many qubits. This is typically considered sufficient to rule out quantum universality. However, some caution is called for: there are gate sets which are known to be classically simulable in this sense but become hard to simulate when one is allowed to measure all the output qubits in the computational basis [17, 3].

We remark that, as pointed out by Lomonaco and Kauffman [18], some qubit solutions to the YBE are entangling gates, and any entangling gate together with arbitrary single-qubit gates is universal [4]. However, in that case we are no longer computing with representations of the braid group. Indeed, a primary motivation for the topological approach to quantum computation is to rely on the topological stability of braiding for fault-tolerance. Applying single-qubit gates fault-tolerantly as part of this approach would require additional ideas. For this reason, we restrict ourselves to just one gate, which acts on two qubits and is a solution to the YBE. Some classes of entangling gates that have previously been shown to be classically simulatable are given in [16, 11].

2 Preliminaries

2.1 Gates, circuits, and universality

We briefly review basic notions about quantum gates, circuits, and computational universality. For more details, we refer the reader to the text of Nielsen and Chuang [20]. Given an integer

[1] Recall that, just like the Jones polynomial, the Jones representation has a parameter (in addition to n) which is typically a root of unity. Quantum universality holds for most but not all values of this parameter.

$d \geq 1$, let $[d] = \{0, 1, \ldots, d-1\}$. Let $V = \mathbb{C}[d]$ be a d-dimensional complex Hilbert space with distinguished orthonormal basis $\{|i\rangle : i \in [d]\}$. We refer to copies of $[d]$ as dits and copies of V as qudits. For any k and any $x \in [d]^k$, set $|x\rangle = |x_1\rangle \otimes |x_2\rangle \otimes \cdots \otimes |x_k\rangle$. The space $V^{\otimes k}$ has a preferred basis $\{|x\rangle : x \in [d]^k\}$, which we will call the computational basis. A unitary operator on $V^{\otimes k}$ is called a k-qudit gate.

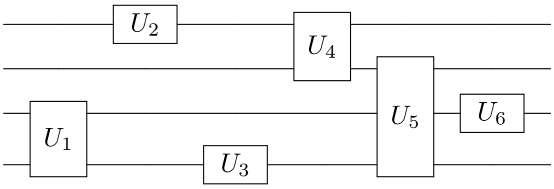
Let $\mathcal{R}$ be a set of gates which act on k or fewer qudits. Fix $n > 0$ and, for each l-qudit gate $R \in \mathcal{R}$, define $R_j \in \mathrm{U}(V^{\otimes n})$ to be the operator that applies R to qudits $j, \ldots, j+l$ and the identity operator I to the rest. Define $\mathcal{R}^{(n)}$ to be the set of all R_j, for every $R \in \mathcal{R}$ and every valid index j. An n-qudit quantum circuit over the gate set $\mathcal{R}$ (or $\mathcal{R}$-circuit for short) is a finite sequence

$$C = (U_1, U_2, \ldots, U_m)$$

where for each i, $U_i \in \mathcal{R}^{(n)}$ or $U_i^{-1} \in \mathcal{R}^{(n)}$. We will sometimes denote the number of gates in the circuit C by $|C| = m$. The circuit defines an operator

$$C = U_m \cdot U_{m-1} \cdots U_1 \in U(V^{\otimes n}).$$

Note that we have overloaded notation so that C refers to both the sequence of gates and the operator implemented by their composition. Pictorially, an $\mathcal{R}$-circuit is represented by a diagram like the following, where each wire corresponds to one qudit.



For pictorial convenience, the gates shown in the figure only act on nearest neighbors. While the nearest-neighbor condition is needed for certain other types of circuits to be classically simulatable (e.g. matchgates [16]), our results do not require it. We adopt here the common convention that circuits are applied from left to right (unfortunately, the opposite of the case for operators.) Of general interest are gate sets which allow for universal quantum computation.

▶ **Definition 1.** A gate set $\mathcal{R}$ is **universal** if there exists $N > 0$ such that N-qudit $\mathcal{R}$-circuits form a dense subset of $U(V^{\otimes N})$.

The Solovay-Kitaev theorem [20] tells us that, for universal $\mathcal{R}$, any unitary operator in $U(V^{\otimes N})$ can be approximated to precision ϵ with an N-qudit $\mathcal{R}$-circuit of length polylog$(1/\epsilon)$. Standard arguments also show that density can be extended from N to any $n \geq N$.

Quantum-computational power can also be defined in terms of complexity classes. The class that is typically associated with efficient quantum computation is called BQP, which stands for bounded-error quantum polynomial time. A drawback of BQP is the lack of known complete problems, i.e., problems which are both in BQP and at least as hard (under classical polynomial-time reduction) as any other problem in BQP. The classical analogue BPP (bounded-error probabilistic polynomial time) suffers from the same drawback. For this reason, we will work with promise versions of these two classes, i.e., PromiseBQP and PromiseBPP. We will not need the formal definitions of these classes (see, e.g., [14]). For us it will suffice to refer to the following.

▶ **Definition 2.** Given a set $\mathcal{R}$ of quantum gates, the problem $\mathcal{I}(\mathcal{R})$ is defined as follows. Given an n-qudit $\mathcal{R}$-circuit C and ditstrings x and y, as well as a promise that either $\langle x|C|y\rangle > 2/3$ or $\langle x|C|y\rangle < 1/3$, decide which is the case.

We may define PromiseBQP as the class of problems which reduce to $\mathcal{I}(\mathcal{R})$ for some universal set of quantum gates $\mathcal{R}$. Interestingly, there are gate sets $\mathcal{R}$ which are not universal in the density sense but for which $\mathcal{I}(\mathcal{R})$ is nonetheless PromiseBQP-hard; an example is $\mathcal{R} = \{\text{Hadamard}, \text{Toffoli}\}$. This gate set is dense over the special orthogonal group, but since the matrix entries are all real, it cannot be dense over the unitary group.

Later on, we will show that when $\mathcal{R}$ consists of a single gate which belongs to certain solution families of the Yang-Baxter equation, then $\mathcal{I}(\mathcal{R}) \in$ PromiseBPP. This means that $\mathcal{R}$ is not quantum universal under either of the above definitions, unless the widely believed conjecture that quantum computation is more powerful than classical computation is false.

2.2 Pauli group and Clifford group

Recall that the single-qubit Pauli operators are defined by

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Each Pauli operator is self-adjoint and unitary. In the n-qubit case, we set

$$X_j = I^{\otimes j-1} \otimes X \otimes I^{\otimes n-j}$$

and likewise for Y_j and Z_j. We define the n-qubit Pauli group $\mathcal{P}_n$ to be the group generated by $\{X_j, Y_j, Z_j : j = 1, \ldots, n\}$. An important property for us is that $\mathcal{P}_n$ spans the space of n-qubit Hermitian operators.

The Clifford group on n qubits is defined to be the normalizer of the Pauli group inside the unitary group, i.e.,

$$\mathcal{C}_n = \{U \in U(2^n) : UPU^\dagger \in \mathcal{P}_n \text{ for all } P \in \mathcal{P}_n\}.$$

By direct computation, it's easy to check that the following gates are elements of $\mathcal{C}_n$ for any $n \geq 2$:

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad P = \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix}, \qquad CNOT = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

It is a theorem (see [11]) that the above gates, when applied to arbitrary qubits or pairs of qubits, actually generate $\mathcal{C}_n$. We will thus call any circuit made up of these gates a Clifford circuit. Since $\mathcal{P}_n \subset \mathcal{C}_n$, we can also add the Pauli operators to this gate set for free. We remark that the conjugation action of a Clifford circuit on an element of $\mathcal{P}_n$ is easy to compute in a direct, gate-by-gate fashion. For details, see [11].

Due to the frequent appearance of $\mathcal{C}_n$ in various areas of quantum information, the computational power of Clifford circuits is well-studied. While $\mathcal{C}_n$ is finite and not universal, adding any gate outside $\mathcal{C}_n$ results in a universal set [19]. A thorough analysis of the computational power of Clifford circuits under various models is performed in [17].

2.3 Yang-Baxter operators and representations of the braid group

Let $V = \mathbb{C}[d]$ and $R \in \mathrm{U}(V \otimes V)$. Then R satisfies the quantum Yang-Baxter equation (YBE) if

$$(R \otimes I)(I \otimes R)(R \otimes I) = (I \otimes R)(R \otimes I)(I \otimes R), \tag{1}$$

where I denotes the identity operator on V. In this case, we say that R is a Yang-Baxter operator. Let $T : |a \otimes b\rangle \mapsto |b \otimes a\rangle$ denote the swap operator on $V \otimes V$. By comparing circuit diagrams, it's not hard to see that R is a solution to (1) if and only if $S = RT$ is a solution to

$$S_{12}S_{13}S_{23} = S_{23}S_{13}S_{12}\,, \tag{2}$$

where

$$S_{12} = S \otimes I\,, \qquad S_{13} = (I \otimes T)(S \otimes I)(I \otimes T)\,, \qquad S_{23} = I \otimes S\,.$$

Equation (2) is sometimes called the algebraic Yang-Baxter equation.

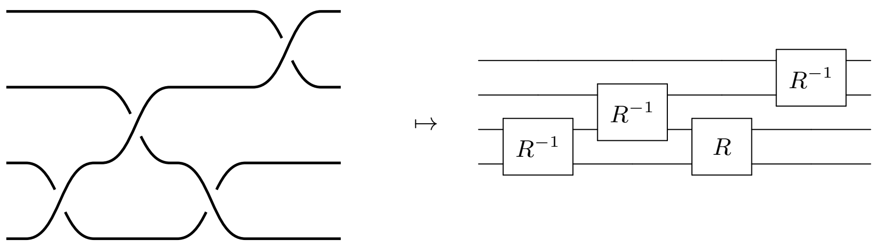
Recall that the braid group B_n is a finitely generated group with generators $\sigma_1, \sigma_2, \cdots, \sigma_{n-1}$ and relations

$$\begin{array}{rcll} \sigma_i\sigma_j & = & \sigma_j\sigma_i & \forall\ |i-j| \geq 2 \\ \sigma_i\sigma_{i+1}\sigma_i & = & \sigma_{i+1}\sigma_i\sigma_{i+1} & \forall\ i. \end{array}$$

In 1925 Artin proved that the abstract group defined above precisely captures the topological equivalence of braided strings [1]. Pictorially, braids are represented with a diagram; an example diagram for $\sigma_3^{-1}\sigma_2^{-1}\sigma_3\sigma_1^{-1}$ is shown below. We read such diagrams left-to-right, keeping the same convention as with circuits. The second generating relation of B_n is known as the Yang-Baxter relation. A solution $R \in \mathrm{U}(V \otimes V)$ of the Yang-Baxter equation yields a unitary representation $\rho_{(R,n)}$ of B_n on the space $V^{\otimes n}$ for every n. It is defined by

$$\rho_{(R,n)}(\sigma_i) = I^{\otimes(i-1)} \otimes R \otimes I^{\otimes n-i-1}\,.$$

The images of braids under $\rho_{(R,n)}$ are precisely the R-circuits on n qudits, where $d = \dim V$. For example, the braid $\sigma_3^{-1}\sigma_2^{-1}\sigma_3\sigma_1^{-1}$ and the corresponding R-circuit are shown below.



Under a plausible physical interpretation, a computation is performed by braiding particle-like excitations whose exchange statistics are described by R. If R is a universal gate, this model would result in universal topological computation. Such a model could provide a basis for a quantum computer architecture with inherent fault-tolerance [22].

3 Classical simulation of certain quantum circuits

In this section, we prove a general result about simulating certain quantum circuits with a classical probabilistic algorithm. We begin with two straightforward lemmas about classical sampling. (See A for proofs).

▶ **Lemma 3.** *Let $\{P_j\}_{j=1}^n$ be probability distributions on $[d]$ and let $P = \Pi_j P_j$ be the corresponding product distribution over $[d]^n$. Suppose that we can calculate $P_j(k)$ for every j and every k in total time* $\mathrm{poly(n,\ d)}$. *Then there's a classical probabilistic algorithm that runs in time* $\mathrm{poly(n,\ d)}$ *and samples from $[d]^n$ according to a probability distribution D such that $|P - D| \leq 1/2^{\mathrm{poly}(n)}$.*

We will also require the following Chernoff-Hoeffding bound for complex-valued random variables.

▶ **Lemma 4.** *Let $X_1, X_2, \dots, X_n$ be independent complex-valued random variables with $\mathbb{E}[X_j] = \mu$ and $|X_j| \leq b$ for all j. Let $S = \sum_j X_j/n$. Then*

$$\Pr\left[|S - \mu| \geq \epsilon\right] \leq 4\exp\left(-n\epsilon^2/8b^2\right) .$$

Let $\mathcal{S}_d$ denote the symmetric group, i.e., the group of permutations of d letters. We denote the action of $\pi \in \mathcal{S}_d$ on an integer $1 \leq j \leq d$ by πj.

▶ **Definition 5.** Let Q be an invertible $d \times d$ matrix over $\mathbb{C}$, and G a subgroup of $\mathcal{S}_d$. Define matrices A, B by setting $A_{ij} = |Q_{ij}|$ and $B_{ij} = |(Q^{-1})_{ij}|$. We say that Q satisfies property (G) if for every $\pi \in G$ and every k, l, we have $\sum_j A_{k,\pi j} B_{jl} \leq 1$.

If Q is unitary, then by Cauchy-Schwarz and the orthonormality of the rows of Q,

$$\sum_j A_{k,\pi j} B_{jl} \leq \left(\sum_j |A_{k,\pi j}|^2 \sum_i |B_{il}|^2 \right)^{1/2} = 1.$$

It follows that unitary matrices satisfy property $(\mathcal{S}_d)$.

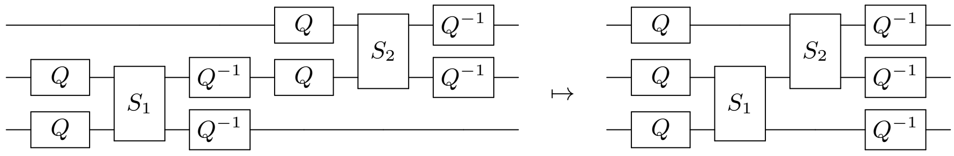
We are now ready to present the main classical simulation algorithm. When we refer to the matrix entries of operators in $\mathrm{GL}(\mathbb{C}[d]) \cong \mathrm{GL}_d(\mathbb{C})$, it will always be in the computational basis. We say that such an operator is computable if its entries can be computed exactly by a classical algorithm in $\mathrm{poly}(d)$ time. Recall that $T : a \otimes b \mapsto b \otimes a$ is the swap operator, and that for a subset S of a group G, $\langle S \rangle$ denotes the subgroup of G generated by S.

▶ **Theorem 6.** *Let $\mathcal{R} = \{R_1, R_2, \dots, R_k\}$ be a set of unitary 2-qudit gates, each one a composition*

$$R_i = (Q \otimes Q) D_i P_i (C_i \otimes C_i)(Q \otimes Q)^{-1} \tag{3}$$

of computable, invertible operators. Suppose that for each i, D_i is a diagonal unitary, C_i is a $d \times d$ permutation matrix, and $P_i = I$ or $P_i = T$. Finally, let Q satisfy property (G) where $G = \langle \{C_i\}_{i=1}^k \rangle \leq \mathcal{S}_d$. Then there exists a classical probabilistic algorithm which, given an n-qudit $\mathcal{R}$-circuit U and strings $x, z \in [d]^n$ and $\epsilon > 0$, outputs a number r in time $\mathrm{poly}(n, |U|, 1/\epsilon)$ *such that $|r - \langle x|U|z\rangle| < \epsilon$ except with probability exponentially small in n and $1/\epsilon$.*

Proof. Set $S_i = D_i P_i (C_i \otimes C_i)$. If we expand each R_i-gate to turn U into a circuit made from S_i-gates and Q-gates, then all of the Q-gates except the initial and final ones are cancelled, as in the example below. We are thus left with a circuit of the form $Q^{\otimes n} V (Q^{-1})^{\otimes n}$ where V is an $\{S_i\}$-circuit. We remark that, in this expanded form, the entire circuit is not necessarily a proper quantum circuit, since Q might not be unitary. The circuit V, on the other hand, is quantum since all of its gates are unitary.



(4)

Before we proceed, note that a non-nearest-neighbor gate can be written as a nearest-neighbor gate conjugated with a swap gate. We depict our gates as acting on nearest neighbors for convenience only, but this condition is not needed for the result to hold. The action of an S_i-gate on the j-th and $(j+1)$-st qudits of a computational basis state is simple to compute. The values of the two qudits are both in $[d]$ initially, and remain in $[d]$ after the action of C_i. Second, these new values are either swapped or left unchanged by P_i. Third, the D_i-gate adds an overall phase factor to the state. By composing these easily-computable actions, the action of V on a computational basis state can be computed in time polynomial in n, d, and $|V|$. Up to phases, this action consists of permuting the n qudits by some $\pi \in S_n$, and applying some bijection $f_j : [d] \to [d]$ to the initial value of the $\pi(j)$-th qudit. Each f_j is a composition of C_i-gates, in the order specified by V. Explicitly, for a basis state $|y\rangle = |y_1 y_2 \dots y_n\rangle$, we write

$$V|y\rangle = e^{i\phi(y)} |f_1 y_{\pi 1} \otimes f_2 y_{\pi 2} \otimes \cdots \otimes f_n y_{\pi n}\rangle \,,$$

where $\phi(y)$ is the overall phase resulting from the D_i-gates. For simplicity of notation, we denoted the image of k under the permutation π as πk, and wrote $f_j y_{\pi j}$ in place of $f_j(y_{\pi j})$.

Next we consider the matrix element

$$\begin{aligned}\langle x|U|z\rangle = \langle x|(Q)^{\otimes n} V (Q^{-1})^{\otimes n}|z\rangle &= \sum_{y\in[d]^n} \langle x|(Q)^{\otimes n} V|y\rangle\langle y|(Q^{-1})^{\otimes n}|z\rangle \\ &= \sum_{y\in[d]^n} e^{i\phi(y)} \prod_{j=1}^{n} \langle x_j|Q|f_j y_{\pi j}\rangle\langle y_j|Q^{-1}|z_j\rangle \,.\end{aligned}$$

We expand the matrix elements of Q and Q^{-1} in terms of magnitudes and phases:

$$\begin{aligned}\langle r|Q|s\rangle &= A(r,s)e^{i\alpha(r,s)} \\ \langle r|Q^{-1}|s\rangle &= B(r,s)e^{i\beta(r,s)}\end{aligned}$$

where A, B, α, β are real-valued and $r, s \in [d]$. Then

$$\begin{aligned}\langle x|(Q)^{\otimes n} V (Q^{-1})^{\otimes n}|z\rangle &= \sum_{y\in[d]^n} e^{i\theta(y)} \prod_{j=1}^{n} A(x_j, f_j y_{\pi j}) B(y_j, z_j) \\ &= \sum_{y\in[d]^n} e^{i\theta(y)} \prod_{j=1}^{n} A(x_j, f_{\sigma j} y_j) B(y_j, z_j) \,,\end{aligned}$$

where $\sigma = \pi^{-1}$ and

$$\theta(y) = \phi(y) + \sum_{j=1}^{n} \big(\alpha(x_j, f_{\sigma j} y_j) + \beta(y_j, z_j)\big) \,.$$

Now we introduce the following normalization factor:

$$\rho = \sum_{y\in[d]^n} \prod_{j=1}^{n} A(x_j, f_{\sigma j} y_j) B(y_j, z_j) = \prod_{j=1}^{n} \sum_{k\in[d]} A(x_j, f_{\sigma j} k) B(k, z_j) \,.$$

This allows us to define a natural probability distribution over $[d]^n$ by

$$P(y) = \frac{1}{\rho} \prod_{j=1}^{n} A(x_j, f_{\sigma j} y_j) B(y_j, z_j) \,,$$

which factorizes as $P(y) = \prod_{j=1}^n P_j(y_j)$, where

$$P_j(l) = \frac{A(x_j, f_{\sigma j} l)B(l, z_j)}{\sum_{k\in[d]} A(x_j, f_{\sigma j} k)B(k, z_j)} .$$

Note that ρ and all of the $P_j(l)$ can be computed in time linear in n and d. By Lemma 3, we can efficiently sample from $[d]^n$ according to P, with error exponentially small in n.

In order to estimate $\langle x|U|z\rangle$, sample repeatedly from this distribution, obtaining outcomes $\xi(j) \in [d]^n$ for $j \in \{1, 2, \ldots\}$ and output the average of the random variables $X_j := \rho \exp(i\theta(\xi(j)))$. Observe that, for each j,

$$\mathbb{E}[X_j] = \sum_{z\in[d]^n} \rho e^{i\theta(z)} P(z) = \langle x|U|z\rangle .$$

To control the absolute value, recall that $f_{\sigma j}$ is a composition of the permutation matrices C_i, and is thus an element of $\langle\{C_i\}_{i=1}^k\rangle \le \mathcal{S}_d$. Since Q satisfies property $(\langle\{C_i\}_{i=1}^k\rangle)$, we have

$$|X_j|^2 = |\rho|^2 = \prod_{j=1}^n \Big| \sum_{k\in[d]} A(x_j, f_{\sigma j} k)B(k, z_j)\Big|^2 \le \prod_{j=1}^n 1^2 \le 1.$$

by Cauchy-Schwarz, for each j. Now set $S(r) = \sum_{j=1}^r X_j/r$. By Lemma 4, for $r \ge 8n/\epsilon^3$ we have

$$\Pr\left[|S(r) - \langle x|U|z\rangle| \ge \epsilon\right] \le 4\exp(-r\epsilon^2/8) \le 4\exp(-n/\epsilon) .$$

◀

An immediate corollary is that, for $\mathcal{R}$ as in the theorem, $\mathcal{I}(\mathcal{R})$ is in PromiseBPP. We will also need the following simple result about simulating circuits constructed from conjugated Clifford gates.

▶ **Theorem 7.** *Let $S \in \mathcal{C}_2$, and $R = (Q \otimes Q)S(Q \otimes Q)^\dagger$ where Q is a single-qubit gate. Let U be a $\{R\}$-circuit on n qubits, M a Hermitian operator on $O(\log(n))$ qubits, and $|\psi\rangle, |\phi\rangle$ arbitrary n-qubit product states. Then $\langle\psi|U^\dagger(M \otimes I)U|\phi\rangle$ can be computed exactly in $O(poly(n))$ classical time.*

Proof. We first apply the procedure from (4) as before, and write

$$U = Q^{\otimes n} V (Q^\dagger)^{\otimes n}$$

where V is described by a circuit consisting only of S gates. The unitary operator implemented by V is an element of $\mathcal{C}_n$. Now let M be a Hermitian operator on $m = c\log(n)$ qubits, and suppose for simplicity that it acts only on the first m qubits. Let I denote the identity operator on the $(m + 1)$st through nth qubits. We write

$$\begin{aligned}\langle\psi|U^\dagger(M \otimes I)U|\phi\rangle &= \langle\psi|Q^{\otimes n}V^\dagger Q^{\dagger\otimes n}(M \otimes I)Q^{\otimes n} V Q^{\dagger\otimes n}|\phi\rangle \\ &= \langle\psi|Q^{\otimes n}V^\dagger(M' \otimes I)V Q^{\dagger\otimes n}|\phi\rangle ,\end{aligned}$$

where $M' = Q^{\otimes m} M Q^{\dagger\otimes m}$.

As discussed earlier, a basis for the space of Hermitian operators on m qubits is the m-qubit Pauli group $\mathcal{P}_m$, which has size $O(\text{poly}(n))$. The expansion of M' in that basis can be computed in polynomial time by basic linear algebra. Embedding the first m qubits into all n qubits gives the obvious embedding of $\mathcal{P}_m$ into $\mathcal{P}_n$, and this also gives (the same, still polynomial-size) expansion of M' into n-qubit Paulis. We write

$$M' = \sum_{\sigma \in \mathcal{P}_n \cap \mathcal{P}_m} \alpha_\sigma \sigma \,.$$

We emphasize that this is a sum over polynomially many terms, and that each coefficient can be calculated from knowledge of M and Q in polynomial time. Moreover, since V is a Clifford circuit, its conjugation action $\sigma \mapsto \sigma^V := V^\dagger \sigma V$ on a Pauli group element $\sigma \in \mathcal{P}_n$ is easily computed by direct gate-by-gate matrix multiplication (see, e.g., [11]).

We now return to the main calculation, to see that

$$\begin{aligned}
\langle\psi|U^\dagger(M \otimes I)U|\phi\rangle &= \langle\psi|Q^{\otimes n}V^\dagger(M' \otimes I)VQ^{\dagger\otimes n}|\phi\rangle \\
&= \sum_{\sigma \in \mathcal{P}_n \cap \mathcal{P}_m} \alpha_\sigma \langle\psi|Q^{\otimes n}V^\dagger \sigma V Q^{\dagger\otimes n}|\phi\rangle \\
&= \sum_{\sigma \in \mathcal{P}_n \cap \mathcal{P}_m} \alpha_\sigma \langle\psi|Q^{\otimes n}\sigma^V Q^{\dagger\otimes n}|\phi\rangle \\
&= \sum_{\sigma \in \mathcal{P}_n \cap \mathcal{P}_m} \alpha_\sigma \prod_{j=1}^{n} \langle\psi_j|Q\sigma_j^V Q^\dagger|\phi_j\rangle \,.
\end{aligned}$$

The sum and product in the final expression are both of polynomial size, and each term in the product can be computed in constant time. ◀

4 Qubit solutions to Yang-Baxter

4.1 The four solution families

Hietarinta classified all solutions to the Yang-Baxter equation in the qubit (i.e., 4 × 4) case [12]. The qubit solutions which are also unitary operators were identified by Dye [5]. All of these are of the form

$$R = k(Q \otimes Q)ST(Q \otimes Q)^{-1} \tag{5}$$

where k is a unit-norm scalar, T is the swap gate, and

$$Q = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is an invertible matrix. The trivial solution is $S = T$ which implies $R = kI$. There are four nontrivial solution families, depending on the possible values taken by S, which are listed below, along with the required conditions on the matrix entries.

$$S_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & q & 0 \\ 0 & 0 & 0 & r \end{pmatrix} \qquad 1 = |p| = |q| = |r| \;;\; c = -a\bar{b}/\bar{d}$$

$$S_2 = \begin{pmatrix} 0 & 0 & 0 & p \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ q & 0 & 0 & 0 \end{pmatrix} \qquad p = \frac{(b\bar{b} + d\bar{d})(\bar{a}b + \bar{c}d)}{(a\bar{a} + c\bar{c})(a\bar{b} + c\bar{d})} \;;\; q = 1/p \;;\; c \neq -a\bar{b}/\bar{d}$$

$$S_3 = \begin{pmatrix} 0 & 0 & 0 & p \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ q & 0 & 0 & 0 \end{pmatrix} \qquad p\bar{p} = \frac{(d\bar{d})^2}{(a\bar{a})^2} \ ; \ q\bar{q} = \frac{(a\bar{a})^2}{(d\bar{d})^2} \ ; \ |pq| = 1 \ ; \ c = -a\bar{b}/\bar{d}$$

$$S_4 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \qquad |a| = |d| \ ; \ c = -a\bar{b}/\bar{d}\,.$$

For $j = 1, 2, 3, 4$, let R_j be the Yang-baxter operator (5) resulting from choosing $S = S_j$.

4.2 Families one, two and three are unlikely to be universal

We will show that Theorem 6 applies to the single-element gate sets $\{R_1\}$, $\{R_2\}$, and $\{R_3\}$. We assume that all of the above matrix entries are exactly computable in constant time via a classical algorithm.

The gate R_1 has the form (3) where $C_i = I$, $P_i = T$, and $D_i = kS_1$. It remains to check that Q satisfies property (G) where G is the trivial group consisting only of the identity; this is confirmed by Lemma 8 below.

For the gate R_2, we set

$$M = \begin{pmatrix} 0 & \sqrt{p} \\ 1/\sqrt{p} & 0 \end{pmatrix}$$

and check that $M \otimes M = S_2$. It follows that $R_2 = kT(QMQ^{-1} \otimes QMQ^{-1})$ is not an entangling gate. Since R_2 is unitary, so is QMQ^{-1}. By the spectral theorem, there exist diagonal V and unitary U such that $UVU^{-1} = QMQ^{-1}$. Observe that $R_2 = (U \otimes U)k(V \otimes V)T(U \otimes U)^{-1}$ satisfies the conditions of Theorem 6.

For the gate R_3, we first rewrite the matrices as follows. Set

$$N = \begin{pmatrix} p^{-1/4} & 0 \\ 0 & p^{1/4} \end{pmatrix}$$

and $Q' = QN^{-1}$ and $S_3' = (N \otimes N)S_3(N \otimes N)^{-1}$. It's not hard to check that

$$R_3 = k(Q \otimes Q)S_3T(Q \otimes Q)^{-1} = k(Q' \otimes Q')S_3'T(Q' \otimes Q')^{-1} \ ,$$

and that Q' and S_3' satisfy the conditions of the third YBE solution family, with the additional property that $p = 1$ and $|q| = 1$. Note further that $S_3'(X \otimes X)$ is a diagonal unitary operator, where

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} .$$

We now see that R_3 is of the form (3) from Theorem 6, where $C_i = X$, $D_i = kS_3'(X \otimes X)$, and $P = T$. It remains to check that Q' satisfies property $(\langle X \rangle)$, which is done in Lemma 9 below.

▶ **Lemma 8.** *Let Q be an invertible 2×2 matrix defined by*

$$Q = \begin{pmatrix} a & b \\ c & d \end{pmatrix} ,$$

such that $c = -a\bar{b}/\bar{d}$. Then Q satisfies property (I).

Proof. Define the relevant matrices

$$A = \begin{pmatrix} |a| & |b| \\ |c| & |d| \end{pmatrix} \qquad \text{and} \qquad B = \frac{1}{|ad - bc|} \begin{pmatrix} |d| & |b| \\ |c| & |a| \end{pmatrix} .$$

Note that $a = 0$ implies $c = 0$, which would make Q non-invertible.

We compute each case separately. First let $k = l = 1$.

$$\begin{aligned} A_{11}B_{11} + A_{12}B_{21} &= \frac{|a||d| + |b||c|}{|ad - bc|} = \frac{|a||d| + |b||a\bar{b}/\bar{d}|}{|ad + ba\bar{b}/\bar{d}|} \\ &= \frac{|\bar{d}|(|a||d|^2 + |a||b|^2)}{|\bar{d}|(|ad\bar{d} + ab\bar{b}|)} = \frac{|a|(|d|^2 + |b|^2)}{|a||d\bar{d} + b\bar{b}|} = 1 . \end{aligned}$$

Next, let $k = l = 2$, and we again get

$$A_{21}B_{12} + A_{22}B_{22} = \frac{|c||b| + |d||a|}{|ad - bc|} = 1 .$$

Now suppose $k = 1$ and $l = 2$.

$$\begin{aligned} A_{11}B_{12} + A_{12}B_{22} &= \frac{|a||b| + |b||a|}{|ad - bc|} = \frac{2|a||b|}{|ad + ab\bar{b}/\bar{d}|} \\ &= \frac{2|a||b||d|}{|ad\bar{d} + ab\bar{b}|} = \frac{2|b||d|}{|d|^2 + |b|^2} . \end{aligned}$$

It remains to note that

$$|b|^2 + |d|^2 - 2|b||d| = (|b| - |d|)^2 > 0 .$$

Finally, we choose $k = 2$ and $l = 1$.

$$A_{21}B_{11} + A_{22}B_{21} = \frac{|c||d| + |d||c|}{|ad - bc|} = \frac{2|a||b|}{|ad - bc|} \leq 1 ,$$

by two applications of $c = -a\bar{b}/\bar{d}$ and the previous case. ◀

▶ **Lemma 9.** *Let Q be an invertible 2×2 matrix defined by*

$$Q = \begin{pmatrix} a & b \\ c & d \end{pmatrix} ,$$

such that $c = -a\bar{b}/\bar{d}$ and $|a|^2 = |d|^2$. Then Q satisfies property $(\mathcal{S}_2)$.

Proof. Define the matrices A and B as in Lemma 8. The case of π equal to the trivial permutation is handled by Lemma 8. We compute the remaining cases. Set $\pi = (12)$ and $k = l = 1$. Then

$$\begin{aligned} A_{12}B_{11} + A_{11}B_{21} &= \frac{|a||c| + |b||d|}{|ad - bc|} = \frac{|aa\bar{b}/\bar{d}| + |bd|}{|ad - ab\bar{b}/\bar{d}|} \\ &= \frac{|aa\bar{b}| + |bd\bar{d}|}{|ad\bar{d} + ab\bar{b}|} = \frac{|aa\bar{b}| + |ba\bar{a}|}{|aa\bar{a} + ab\bar{b}|} = \frac{|a\bar{b}| + |b\bar{a}|}{|a|^2 + |b|^2} , \end{aligned}$$

where we have applied the facts $c = -a\bar{b}/\bar{d}$ and $a\bar{a} = d\bar{d}$ and $a \neq 0$. Now note that

$$|a|^2 + |b|^2 - (|a\bar{b}| + |b\bar{a}|) = |a|^2 + |b|^2 - 2|a||b| = (|a| - |b|)^2 \geq 0 .$$

Hence $(|a\bar{b}| + |b\bar{a}|)/(|a|^2 + |b|^2) \leq 1$. For $k = l = 2$, we again get

$$A_{22}B_{12} + A_{21}B_{22} = \frac{|a||c| + |b||d|}{|ad - bc|} \leq 1\,.$$

Now set $k = 1$ and $l = 2$. Then

$$\begin{aligned} A_{12}B_{12} + A_{11}B_{22} &= \frac{|a|^2 + |b|^2}{|ad - bc|} = \frac{|a|^2 + |b|^2}{|ad + ab\bar{b}/\bar{d}|} \\ &= \frac{|\bar{d}|(|a|^2 + |b|^2)}{|ad\bar{d} + ab\bar{b}|} = \frac{|d|(|a|^2 + |b|^2)}{|a|(|d|^2 + |b|^2)} = 1\,. \end{aligned}$$

Finally, for $k = 2$ and $l = 1$, write $b = -\bar{c}d/\bar{a}$ and calculate

$$\begin{aligned} A_{22}B_{11} + A_{21}B_{21} &= \frac{|c|^2 + |d|^2}{|ad - bc|} = \frac{|c|^2 + |d|^2}{|ad + dc\bar{c}/\bar{a}|} \\ &= \frac{|\bar{a}|(|c|^2 + |d|^2)}{|da\bar{a} + dc\bar{c}|} = \frac{|a|(|c|^2 + |d|^2)}{|d|(|c|^2 + |a|^2)} = 1\,. \end{aligned}$$

◀

To conclude, we have shown the following.

▶ **Theorem 10.** *Let $R \in \{R_1, R_2, R_3\}$ be a unitary solution to the Yang-Baxter equation on qubits. Then $\mathcal{I}(\{R\})$ is in PromiseBPP.*

In particular, if one could perform (perhaps encoded) universal quantum computation with these circuits then PromiseBQP = PromiseBPP. We can also formulate the lack of universality for these solutions in the following terms.

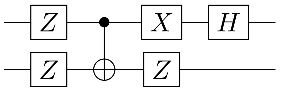
▶ **Theorem 11.** *Let $R \in \{R_1, R_2, R_3\}$ be a unitary solution to the Yang-Baxter equation on qubits, and let $\rho_n : B_n \to SU(2^n)$ be the corresponding unitary representation of the braid group. Then the image of ρ_n is not dense in $SU(2^n)$ for any $n \geq 2$, unless* PromiseBQP = PromiseBPP*.*

Proof. (Sketch) For a contradiction, suppose there exists an $n \geq 2$ such that the image of ρ_n is dense. Let C be an arbitrary m-qubit quantum circuit. We can assume without loss of generality that C only consists of 2-qubit gates acting on adjacent qubits, and that n is even. For each of the m qubits, assign $n/2$ qubits from the space of ρ_n. By the density of the image of ρ_n, we can then simulate C inside $\rho_{mn/2}$ gate-by-gate via the Solovay-Kitaev theorem. Then we can use the classical algorithm from Theorem 6 to approximate the relevant matrix entry of the resulting R-circuit, thus solving the PromiseBQP-hard problem of approximating the corresponding entry of C. ◀

4.3 Family four is unlikely to be universal

Recall that the fourth solution family is of the form $R_4 = k(Q \otimes Q)S_4T(Q \otimes Q)^{-1}$. We begin by demonstrating a Clifford circuit which is equal to the gate S_4T.

$$S_4T = \tfrac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} =$$



We also note that, in this solution family, Q is a scaled unitary operator. To see this, note that

$$Q^\dagger Q = \begin{pmatrix} |a|^2 + |c|^2 & \bar{a}b + \bar{c}d \\ a\bar{b} + c\bar{d} & |b|^2 + |d|^2 \end{pmatrix} = \begin{pmatrix} |a|^2 + |c|^2 & 0 \\ 0 & |b|^2 + |a|^2 \end{pmatrix} = \left(|a|^2 + |b|^2\right) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

where we first applied the condition $c = -a\bar{b}/\bar{d}$ to the off-diagonal elements and the condition $|a|^2 = |d|^2$ to the diagonal ones; the last equality follows from combining these two conditions to get $|c|^2 = |b|^2$. Now set $\alpha = (|a|^2 + |b|^2)^{1/2}$ and $Q_1 = \alpha^{-1}Q$. Using the above, one easily checks that Q_1 is unitary and that $Q_1^\dagger = \alpha Q^{-1}$. It follows that

$$(Q \otimes Q)A(Q \otimes Q)^{-1} = (\alpha Q_1 \otimes \alpha Q_1)A(\alpha^{-1}Q_1^\dagger \otimes \alpha^{-1}Q_1^\dagger) = (Q_1 \otimes Q_1)A(Q_1 \otimes Q_1)^\dagger$$

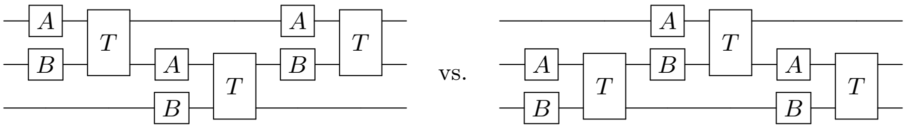
for any A. For us it will thus suffice to assume that Q is in fact unitary. This allows us to apply Theorem 7 and get the following result.

▶ **Theorem 12.** *Let U be a $\{R_4\}$-circuit on n qubits, M a Hermitian operator on $O(\log(n))$ qubits, and $|\psi\rangle, |\phi\rangle$ arbitrary n-qubit product states. Then $\langle\psi|U^\dagger(M \otimes I)U|\phi\rangle$ can be computed exactly in $O(poly(n))$ classical time.*

5 Some simple high-dimensional solutions

Finally, we list some simple unitary solution families to the Yang-Baxter equation that exist in every dimension, and to which Theorem 6 applies. We begin by observing that, whenever a 2-qudit gate S is a solution, then by (4) so is $(Q \otimes Q)S(Q \otimes Q)^{-1}$ for any 1-qudit gate Q.

For $A, B \in \mathrm{U}(V)$, the operator $T(A \otimes B)$ is a solution to the Yang-Baxter equation if and only if A and B commute. This is easily seen by following the wires in the circuits below.



If A and B do commute, then there's a unitary change of basis Q on V such that $Q^{-1}AQ$ and $Q^{-1}BQ$ are both diagonal. Therefore, Theorem 6 applies to $T(A \otimes B)$, so any circuits using this gate are classically simulable. Of course, this is not surprising, as they do not even entangle the qudits.

More generally, suppose $S \in \mathrm{U}(V \otimes V)$ is diagonal in the computational basis, and set

$$\lambda_{ij} = \langle ij|S|ij\rangle \qquad \text{for} \qquad i, j \in [d]\,,$$

where $d = \dim V$. Note that

$$S_{12} = S \otimes I = \bigoplus_{k\in[d]} P_k\,, \qquad S_{23} = I \otimes S = \bigoplus_{k\in[d]} S\,, \qquad I \otimes T = \bigoplus_{k\in[d]} T\,.$$

where $P_k = \oplus_{l\in[d]}\lambda_{kl}I$. We also have

$$S_{13} = (I \otimes T)(S \otimes I)(I \otimes T) = \bigoplus_{k\in[d]} TP_kT\,.$$

Substituting the above into the two sides of the algebraic Yang-Baxter equation (2), we get

$$\bigoplus_{k\in[d]} P_kTP_kTS \qquad \text{and} \qquad \bigoplus_{k\in[d]} STP_kTP_k$$

Clearly, P_k and S are symmetric. Since $\langle ab|T|cd\rangle = \delta_{ad}\delta_{bc} = \langle cd|T|ab\rangle$, so is T. By applying the transpose to one of the two sides above, we see that S satisfies algebraic Yang-Baxter. Thus ST is a solution to the YBE, one to which Theorem 6 clearly applies.

Acknowledgments. We acknowledge funding provided by the Institute for Quantum Information and Matter, an NSF Physics Frontiers Center with support of the Gordon and Betty Moore Foundation through Grant GBMF1250. We would also like to acknowledge the support from the Summer Undergraduate Research Fellowship (SURF) program, as well as the David L. Goodstein SURF endowment. Portions of this paper are a contribution of NIST, an agency of the US government, and are not subject to US copyright.

References

1 Emil Artin. Theorie der Zopfe. *Abh. Math. Sem. Univ. Hamburg*, 4:47–72, 1925.

2 John C. Baez. Braids and quantization (online lecture notes), May 1992.

3 M. Bremner, R. Jozsa, and D. J. Sheperd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proc. R. Soc. A*, 467:459, 2011.

4 J. L. Brylinski and R. Brylinski. Universal quantum gates. In *Mathematics of Quantum Computation*. Chapman & Hall, 2002.

5 H. A. Dye. Unitary solutions to the Yang-Baxter equation in dimension four. *Quantum Information Processing*, 2(1/2):117–152, April 2003.

6 Jennifer M. Franko. Braid group representations arising from the Yang-Baxter equation. *Journal of Knot Theory and Its Ramifications*, 19(04):525–538, 2010.

7 Jennifer M. Franko, Eric C. Rowell, and Zhenghan Wang. Extraspecial 2-groups and images of braid group representations. *Journal of Knot Theory and Its Ramifications*, 15(04):413–427, 2006.

8 Michael H. Freedman, Alexei Kitaev, and Zhenghan Wang. Simulation of topological field theories by quantum computers. *Communications in Mathematical Physics*, 227:587–603, 2002.

9 Michael H. Freedman, Michael J. Larsen, and Zhenghan Wang. A modular functor which is universal for quantum computation. *Communications in Mathematical Physics*, 227:605–622, 2002.

10 Michael H. Freedman, Michael J. Larsen, and Zhenghan Wang. The two-eigenvalue problem and density of Jones representation of braid groups. *Communications in Mathematical Physics*, 228:177–199, 2002.

11 Daniel Gottesman. The Heisenberg representation of quantum computers. *Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics*, 22:32–43, 1999.

12 Jarmo Hietarinta. Solving the two-dimensional constant quantum Yang-Baxter equation. *Journal of Mathematical Physics*, 34(5):1725–1756, 1993.

13 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

14 Dominik Janzing and Pawel Wocjan. A simple PromiseBQP-complete matrix problem. *Theory of Computing*, 3(4):61–79, 2007.

15 Vaughan F. R. Jones. A polynomial invariant for knots via von Neumann algebras. *Bull. Amer. Math. Soc. (N.S.)*, 12(1):103–111, 1985.

16 R. Jozsa and A. Miyake. Matchgates and classical simulation of quantum circuits. *Royal Society of London Proceedings Series A*, 464:3089–3106, December 2008.

17 Richard Jozsa and Maarten Van den Nest. Classical simulation complexity of extended clifford circuits. *arXiv*, 2013.

18 Louis H. Kauffman and Samuel J. Lomonaco Jr. Braiding operators are universal quantum gates. *New Journal of Physics*, 6(1):134, 2004.

19 Gabriele Nebe, E. M. Rains, and N. J. A. Sloane. The invariants of the Clifford groups. *Designs, Codes and Cryptography*, 24(1):99–122, 2001.

20 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

21 J. H. H. Perk and H. Au-Yang. Yang-Baxter equation. In *Encyclopedia of Mathematical Physics*, pages 465–473. Oxford: Elsevier, 2006.

22 John Preskill. Topological quantum computation (online lecture notes), 2004.

23 Eric C. Rowell and Zhenghan Wang. Localization of unitary braid group representations. *Communications in Mathematical Physics*, 311(3):595–615, 2012.

24 V. G. Turaev. The Yang-Baxter equation and invariants of links. *Invent. Math.*, 92:527–553, 1988.

A Appendix

We will now prove Lemmas 3 and 4.

Proof of Lemma 3. To sample from P_j, flip m unbiased coins to get an integer $0 \leq l \leq 2^m$. Subdivide 2^m into intervals according to

$$2^m = P_j(0)2^m + P_j(1)2^m + \cdots + P_j(d-1)2^m$$

and output k if l falls into the kth interval. Then the probability $D_j(k)$ with which you output k satisfies $|D_j(k) - P_j(k)| \leq 1/2^m$. Now do this for two indices, say 1 and 2 and note that

$$\begin{aligned} |P_1(k)P_2(l) - D_1(k)D_2(l)| &= |P_1(k)P_2(l) - D_1(k)D_2(l) + D_1(k)P_2(l) - D_1(k)P_2(l)| \\ &\leq |P_2(l)(P_1(k) - D_1(k))| + |D_1(k)(P_2(l) - D_2(l))| \\ &\leq 2/2^m \end{aligned}$$

Extending this to the case of multiplying all n distributions together, we get $|P(y) - D(y)| \leq n/2^m$ for all $y \in [d]^n$. The total variation distance then satisfies

$$|P - D| = \frac{1}{2} \sum_{x \in [d]^n} |P(x) - D(x)| \leq \frac{nd^n}{2^m} < 2^{-n}$$

so long as $m \geq 3n \log d$. ◀

Proof of Lemma 4. We expand the X_j into real and imaginary parts and apply the standard bound. Set $S_r = \sum_j \mathrm{Re}[X_j]/n$ and $S_i = \sum_j \mathrm{Im}[X_j]/n$ and $\mu_r = \mathbb{E}[\mathrm{Re}[X_j]]$ and $\mu_i = \mathbb{E}[\mathrm{Im}[X_j]]$. Note that $|\mathrm{Re}[X_j]| \leq b$ and $|\mathrm{Im}[X_j]| \leq b$. By the Chernoff-Hoeffding bound for real-valued random variables [13], we have

$$\Pr\left[|S_r - \mu_r| \geq \epsilon/2\right] \leq 2\exp\left(-n\epsilon^2/8b^2\right) ,$$

and likewise for the imaginary part. Taking the union bound, we have that

$$|S - \mu| = |S_r - \mu_r + i(S_i - \mu_i)| \leq |S_r - \mu_r| + |S_i - \mu_i| \leq \epsilon/2 + \epsilon/2 = \epsilon$$

except with probability $4\exp\left(-n\epsilon^2/8b^2\right)$. ◀

Blindness and Verification of Quantum Computation with One Pure Qubit

Theodoros Kapourniotis[1], Elham Kashefi[1], and Animesh Datta[2]

1 School of Informatics, University of Edinburgh
10 Crichton Street, Edinburgh EH8 9AB, UK
T.Kapourniotis@sms.ed.ac.uk, ekashefi@inf.ed.ac.uk

2 Clarendon Laboratory, Department of Physics,
University of Oxford, OX1 3PU, United Kingdom
animesh.datta@physics.ox.ac.uk

Abstract

While building a universal quantum computer remains challenging, devices of restricted power such as the so-called *one pure qubit* model have attracted considerable attention. An important step in the construction of these limited quantum computational devices is the understanding of whether the verification of the computation within these models could be also performed in the restricted scheme. Encoding via blindness (a cryptographic protocol for delegated computing) has proven successful for the verification of universal quantum computation with a restricted verifier. In this paper, we present the adaptation of this approach to the one pure qubit model, and present the first feasible scheme for the verification of delegated one pure qubit model of quantum computing.

1998 ACM Subject Classification Quantum computation theory, Cryptography, Model verification and validation

Keywords and phrases Delegated Computing, Verification, Measurement-based Model

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.176

1 Introduction

The physical realisation of quantum information processing requires the fulfilment of the five criteria collated by DiVincenzo [13]. While enormous progress had been made in realising them since, we are still some way from constructing a universal quantum computer. This raises the question whether quantum advantages in computation are possible without fulfilling one or more of DiVincenzo's criteria. From a more foundational perspective, the computational power of the intermediate models of computation are of great value and interest in understanding the computational complexity of physical systems. Several such models are known, including fermionic quantum computation [6], instantaneous quantum computation [7], permutational quantum computation [21], and boson sampling [1].

Deeply entwined with the construction of a quantum information processor is the issue of its verification. How do we convince ourselves that the output of a certain computation is correct and obtained using quantum-enhanced means. Depending on a given computation, one or both may be non-trivial. For instance, the correctness of the output of Shor's factoring algorithm [33] can be checked efficiently on a classical machine, but in general this is not known to be possible for all problems solvable by a quantum computer. On the other hand, by allowing a small degree of quantumness to the verifier [2, 18], or considering entangled non-commuting provers [17], the verification problem has been solved for universal quantum

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 176–204
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

computation. However, not much attention has been given to verifying restricted models of quantum-enhanced computation. It is in this direction that we endeavour to embark.

One of the earliest restricted models of quantum computation was proposed by Knill and Laflamme, named 'Deterministic Quantum Computation with One quantum bit (DQC1)', also referred to as the one pure qubit model [22]. It addresses the challenge of DiVincenzo's first criterion, that of preparing a pure quantum input state, usually the state of n separate qubits in the computational basis state zero. Instead, in the DQC1 model, only one qubit is prepared in a pure state (computational basis zero state) and the rest of the input qubits exist in the maximally mixed state. This model corresponds to a noisier, more feasible experimental setting and was initially motivated by liquid-state NMR proposals for quantum computing. The DQC1 model was shown to be capable of estimating the coefficients of the Pauli operator expansion efficiently. Following this, Shepherd defined the complexity class 'Bounded-error Quantum 1-pure-qubit Polynomial-time (BQ1P)', to capture the power of the DQC1 model [32], and proved that a special case of Pauli operator expansion, the problem of estimating the normalised trace of a unitary matrix to be complete for this class. This problem, and others that can be reduced to it, such as the estimation of the value of the Jones polynomial (see Ref. [12] for more such connections), is interesting from a complexity theoretical point of view since it has no known efficient classical algorithm. Moreover they are not known to belong to the class NP, therefore the problem of verifying the correctness of the result is non-trivial. More recently, it was shown that an ability to simulate classically efficiently a slightly modified version of this model would lead to the collapse of the polynomial hierarchy to the third level [29].

The approach of the Verifiable Universal Blind Quantum Computing (VUBQC) [18] is based on the intermediate step of blind computing, a cryptographic protocol where a restricted client runs the desired computation on a powerful server, such that the server does not learn anything about the delegated computation. A protocol for universal blind quantum computation with a client able to prepare only single qubits, based on Measurement-based Quantum Computing (MBQC) [31] model was introduced in [8]. Here, we take the same approach towards verification by first adapting this existing protocol for blind computing to the DQC1 model. Thus, the first goal is to define what it means to have a DQC1 computation in the MBQC setting. Fixing the input state to almost maximally mixed as it is done in the circuit picture of the DQC1 model does not suffice since the required auxiliary qubits for MBQC could potentially increase the number of pure qubits in the system by more than a logarithmic amount [1]. This adaptation is also necessary as almost all the optimal schemes [2, 8, 15, 25, 4, 27, 28, 34, 23, 19] for the blind computation exploit the possibility of adaptive computation based on the measurement, a freedom not allowed in the original DQC1 model [2]. The main results presented in this paper are the following.

- We introduce a new definition of DQC1 computation within the MBQC framework, called the DQC1-MBQC model [3], which captures the essential property of its original definition in the circuit model. Moreover, we show that the original definition of complexity class BQ1P is contained in DQC1-MBQC, where the latter is able to capture the process where new qubits are introduced or traced out during the execution of the computation.

[1] Increasing the number of pure qubits in the input to the order of logarithmic in the size of the computation is shown not to add extra power to the one pure qubit complexity class [32].

[2] Ref. [26] does not require the server to use measurement-based quantum computing.

[3] We use a different acronym than DQC1 to emphasis the structural distinction with the standard DQC1 model.

- We provide a sufficient condition for a graph state (underlying resource for an MBQC computation [20]) to be usable within DQC1-MBQC. A direct consequence of this is that the universal blind protocol, which satisfies this condition, can be directly adapted to the setting where the server is a DQC1-MBQC machine and the client is able to send one single qubit at a time.
- Building on the blind protocol and adapting the methods presented in [18], a verification protocol for the class DQC1-MBQC with a server restricted to DQC1-MBQC is given, where the probability of the client being forced to accept an incorrect result can be adjusted by setting the security parameter of the model. Since the protocol of [18] does not satisfy the sufficient condition and hence not runnable in the DQC1-MBQC, an alternative method is presented which also leads to different complexity results.

1.1 Preliminaries

We first introduce the notation necessary to describe a computation in MBQC [31, 11]. A generic pattern, consists of a sequence of commands acting on qubits:

- $N_i(|q\rangle)$: Prepare the single auxiliary qubit i in the state $|q\rangle$;
- $E_{i,j}$: Apply entangling operator controlled-Z to qubits i and j;
- M_i^α: Measure qubit i in the basis $\{\frac{1}{\sqrt{2}}(|0\rangle + e^{i\alpha}|1\rangle), \frac{1}{\sqrt{2}}(|0\rangle - e^{i\alpha}|1\rangle)\}$ followed by trace out the measured qubit. The outcome of measuring qubit i is called result and is denoted by s_i;
- $X_i^{s_j}, Z_i^{s_j}$: Apply a Pauli X or Z correction on qubit i depending on the result s_j of the measurement on the j-th qubit.

The corrections could be combined with measurements to perform 'adaptive measurements' denoted as ${}^{s_z}[M_i^\alpha]^{s_x} = M_i^{(-1)^{s_x}\alpha+s_z\pi}$. A pattern is formally defined by the choice of a finite set V of qubits, two not necessarily disjoint sets the input and the output, $I \subset V$ and $O \subset V$ determining the pattern inputs and outputs, and a finite sequence of commands acting on V.

▶ **Definition 1** ([10]). A pattern is said to be runnable if

(R0) no command depends on an outcome not yet measured;

(R1) no command (except the preparation) acts on a measured or not yet prepared qubit;

(R2) a qubit is measured (prepared) if and only if it is not an output (input).

The entangling commands $E_{i,j}$ define an undirected graph over V referred to as (G, I, O). Along with the pattern we define a partial order of measurements and a dependency function d which is a partial function from O^C to $\mathcal{P}^{I^C}$, where $\mathcal{P}$ denotes the power set. Then, $j \in d(i)$ if j gets a correction depending on the measurement outcome of i. In what follows, we will focus on patterns that realise (strongly) deterministic computation, which means that the pattern implements a unitary on the input up to a global phase. A sufficient condition on the geometry of the graph state to allow unitary computation is given in [10, 9] and will be used later in this paper. In what follows, $x \sim y$ denotes that x is adjacent to y in G.

▶ **Definition 2** ([10]). A *flow* $(f, \preceq)$ for a geometry (G, I, O) consists of a map $f : O^c \mapsto I^c$ and a partial order $\preceq$ over V such that for all $x \in O^c$

(F0) $x \sim f(x)$;

(F1) $x \preceq f(x)$;

(F2) for all $x, y : y \neq x, y \sim f(x)$ we have $x \preceq y$.

1.2 Main results

1.2.1 DQC1-MBQC

We define the class BQ1P formally as introduced by Shepherd [32], and then recast it into the MBQC framework.

▶ **Definition 3** (Bounded-error Quantum 1-pure-qubit Polynomial-time complexity class, [32]). BQ1P is defined using a bounded-error uniform family of quantum circuits – DQC1. A DQC1 circuit takes as input a classical string $\boldsymbol{x}$, of size n, which encodes a fixed choice of unitary operators applied on a standard input state $|0\rangle\langle 0| \otimes I_{w-1}/2^{w-1}$. The width of the circuit w is polynomially bounded in n. Let $Q_n(x)$ be the result of measuring the first qubit of the final state of a DQC1 circuit. A language in BQ1P is defined by the following rule:

$$\forall a \in L : Pr(Q_n(a) = 1) \geq \frac{1}{2} + \frac{1}{2q(n)} \tag{1}$$

$$\forall a \notin L : Pr(Q_n(a) = 1) \leq \frac{1}{2} - \frac{1}{2q(n)} \tag{2}$$

for some polynomially bounded q(n).

An essential physical property of DQC1 that we mean to preserve in DQC1-MBQC is its limited purity. To capture this we introduce the *purity parameter*:

$$\pi(\rho) = \log_2 (\mathrm{Tr}(\rho^2)) + d, \tag{3}$$

where d is the logarithm of the dimension of the state ρ. For a DQC1 circuit with k pure qubits, at each state of the computation the value of purity parameter π for that state remains constant equal to k. In fact, Shepherd showed that the class BQ1P is not extended by increasing the number of pure input qubits logarithmically. Thus, a purity that does not scale too rapidly with the problem size still remains in the same complexity class.

A characterisation of MBQC patterns compatible with the idea of the DQC1 model as introduced above is presented next. Any MBQC pattern is called DQC1-MBQC when there exists a runnable rewriting of this pattern such that after every elementary operation (for any possible branching of the pattern) the purity parameter π does not increase over a fixed constant. We assume that the system at the beginning has only the input state and at the end has only the output state.

We define a new complexity class that captures the idea of one pure qubit computation in the MBQC model. This complexity class, that we name DQC1-MBQC, can be based on any universal DQC1-MBQC resource pattern, which is defined analogously to the DQC1 circuits [32] as a pattern that can be adapted to execute any DQC1-MBQC pattern of polynomial size. A particular example of such a resource, as we will present later, can be built using the the brickwork state of [8] designed for the purpose of universal blind quantum computing. The input to a universal pattern is the description of a computation as a measurement angle vector and is used to classically control the measurements of the MBQC pattern. The quantum input of the open graph is always fixed to a mostly maximally mixed state, in correspondence to the DQC1 model.

▶ **Definition 4.** A language in DQC1-MBQC complexity class is defined based on a universal DQC1-MBQC resource pattern P_α that takes as input an angle vector $\boldsymbol{\alpha}$ of size n and is applied on the quantum state $|+\rangle\langle +| \otimes I_{w-1}/2^{w-1}$, $w \in O(n)$. A word $\boldsymbol{\alpha}$ belongs to the language depending of the probabilities of the measurement outcome ($R_n(\boldsymbol{\alpha})$) of the first output qubit of pattern P_α which are defined identically to Definition 3:

$$\forall a \in L : Pr(R_n(\alpha) = 1) \geq \frac{1}{2} + \frac{1}{2r(n)} \tag{4}$$

$$\forall a \notin L : Pr(R_n(\alpha) = 1) \leq \frac{1}{2} - \frac{1}{2r(n)} \tag{5}$$

for some polynomially bounded r(n).

▶ **Corollary 5.** *BQ1P ⊆ DQC1-MQBC.*

Proof. Any circuit description using a fixed set of gates can be efficiently translated into a measurement pattern applicable on the brickwork state. A specific example of translating each gate from the universal set {Hadamard, $\pi/8$, c-NOT} to a 'brick' element of the brickwork state is given in [8]. The quantum input state in the resulting measurement pattern is in the almost-maximally-mixed state, therefore the pattern is a valid DQC1-MBQC pattern. ◀

▶ **Definition 6.** An MBQC pattern is a DQC1-MBQC pattern if there is a runnable sequence of commands where for every elementary command and measurement outcome, there exists a fixed constant value c such that the overall quantum state of the system (ρ_i with dimension d_i) after the i^{th} operation satisfies the following relation

$$\pi(\rho_i) < \pi(\rho_{in}) + c, \tag{6}$$

where ρ_{in} is the quantum input of the pattern with dimension d_{in}, which is fixed to be the product of c_{in} (constant) pure qubits and a maximally mixed state of $d_{in} - c_{in}$ qubits.

The above definition captures the essence of DQC1 in that it maintains a low purity, high entropy state in MBQC, in contrast to DiVincenzo's first criterion. We derive a sufficient condition (that is also constructive) for the open graph state leading to DQC1-MBQC, capturing the universal blind quantum computing protocol as a special case. However, a general characterisation and further structural link with determinism in MBQC [10, 9, 24] is left as an open question for future work.

▶ **Theorem 7.** *Any measurement pattern on an open graph state (G, I, O) with flow $(f, \preceq)$ (as defined in Definition 2) and measurement angles $\boldsymbol{\alpha}$ where either $|I| = |O|$ or the flow function is surjective and all auxiliary preparations are on the $(X - Y)$ plane represents a DQC1-MBQC pattern.*

The full details and the proof of this theorem is provided in Section 2.

1.2.2 Blindness

A direct consequence of Theorem 7 is that the Universal Blind Computing Protocol (UBQC) introduced in [8] can be easily adapted to fit within the DQC1-MBQC class, since it is based on an MBQC pattern on a graph state with surjective flow.

In the blind cryptographic setting a client (Alice) wants to delegate the execution of an MBQC pattern to a more powerful server (Bob) and hide the information at the same time. The UBQC protocol is based on the separation of the classical and quantum operations when running an MBQC pattern. The client prepares some randomly rotated quantum states and sends them to the server and from this point on the server executes the quantum operations on them (entangling according to the graph and measuring) and the client calculates the measurement angles for the server and corrects the measurement outcomes she receives (to undo the randomness and get the correct result).

To define blindness formally we allow Bob to deviate from the normal execution in any possible way, and this is captured by modelling his behaviour during the protocol by an arbitrary CPTP map. The main requirement for blindness is that for any input and averaged over all possible choices of parameters by Alice, Bob's final state can always be written as a fixed CPTP map applied on his initial state, thus not offering any new knowledge to him. This definition of stand-alone blindness was presented first in [14] and takes into account the issue of prior knowledge.

▶ **Definition 8** (Blindness)**.** Let P be a protocol for delegated computation: Alice's input is a description of a computation on a quantum input, which she needs to perform with the aid of Bob and return the correct quantum output. Let ρ_{AB} express the joint initial state of Alice and Bob and σ_{AB} their joint final state, when Bob is allowed to do any deviation from the correct operation during the execution of P, averaged over all possible choices of random parameters by Alice. The protocol P is blind iff

$$\forall \rho_{AB} \in \mathcal{L}(\mathcal{H}_{AB}), \exists\, \mathcal{E} : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_B), \text{ s.t.} \quad \mathrm{Tr}_A(\sigma_{AB}) = \mathcal{E}(\mathrm{Tr}_A(\rho_{AB})) \tag{7}$$

To adapt the original UBQC protocol into the DQC1-MBQC setting we change the order of the operations so that the client does not send all the qubits to the server at the beginning, but during the execution of the pattern, following a rewriting of the pattern that is consistent with the purity requirement. The details are described in Section 2.

▶ **Theorem 9.** *There exists a blind protocol for any DQC1-MBQC computation where the client is restricted to BPP and the ability to prepare single qubits and the server is within DQC1-MBQC.*

1.2.3 Verification

In the verification cryptographic setting a client (Alice) wants to delegate a quantum computation to a more powerful server (Bob) and accept if the result is correct or reject if the result is incorrect (server is behaving dishonestly). The main idea of the original protocol of [18] is to test Bob's honesty by hiding a trap qubit among the others in the resource state sent to him by Alice. Blindness means that Bob cannot learn the position of the trap, nor its state. During the execution of the pattern Bob is asked to measure this trap qubit and report the result to Alice. If Bob is honest this measurement gives a deterministic result, which can be verified by Alice. Bob being dishonest means that Alice will receive the wrong result with no-zero probability. Depending on that result, Alice accepts or rejects the final output received by Bob.

To define verifiability formally we need to establish an important difference with the original protocol [18]: In a DQC1-MBQC pattern the quantum input is in a mixed state as opposed to a pure state. Reverting to the original definition that derives from the quantum authentication schemes in [3] we need to add an extra reference system R, that is used to purify the mixed input that exists in Alice's private system A. The assumption is that Bob does not learn anything about the reference system (ex. Alice is provided with the quantum input from a third trusted party which also holds the purification). Bob is allowed to choose any possible cheating strategy and our goal is to minimise the probability of Alice accepting the incorrect output of the computation at the end of the protocol.

▶ **Definition 10.** A protocol for delegated computation is ϵ-verifiable ($0 \leq \epsilon < 1$) if for any choice of Bob's strategy j, it holds that for any input of Alice:

$$\mathrm{Tr}(\sum_{\nu} p(\nu) P^{\nu}_{\text{incorrect}} B_j(\nu)) \leq \epsilon \tag{8}$$

where $B_j(\nu)$ is the state of Alice's system A together with the purification system R at the end of the run of the protocol, for choice of Alice's random parameters ν and Bob's strategy j. If Bob is honest we denote this state by $B_0(\nu)$. Let $P_\perp$ be the projection onto the orthogonal complement of the the correct (purified) quantum output. Then,

$$P^{\nu}_{\text{incorrect}} = P_\perp \otimes |\eta^{\nu_c}_t\rangle\langle\eta^{\nu_c}_t| \tag{9}$$

where $|\eta^{\nu_c}_t\rangle$ is a state that indicates if Alice accept or reject the result (see Section 3).

A verification protocol should also be correct, which means that in case Bob is honest Alice's state at the end of the run of the protocol is the correct output of the computation and an extra qubit set in the accept state (this property is also referred to as completeness).

In VUBQC, in order to adjust the parameter ϵ to any arbitrary value between 0 and 1 (a technique called probability amplification), one needs to add polynomially many trap qubits within the MBQC pattern. Specifically, adding polynomially many traps and incorporating the pattern into a fault tolerance scheme that corrects d errors, gives parameter ϵ exponentially small on d. As we explain in Section 3, adding polynomially many traps, following the same scheme as VUBQC, creates a pattern that is not a DQC1-MBQC pattern. Therefore to achieve an amplification of the error probability we need to develop a modified trapping scheme.

In Section 3 we give a verification protocol for DQC1-MBQC problems where, instead of running the pattern once, s computations of the same size are run in series, one being the actual computation and the others being trap computations. A similar approach is also considered for the restricted setting of the photonic implementation of VUBQC [5] and a verification protocol of the entanglement states [30]. In our setting each trap computation contains an isolated trap injected in a random position between the qubits of the pattern. We prove that in this verification protocol the server is within DQC1-MBQC complexity class, while the client is within BPP together with single qubit preparations (as in the original VUBQC). Moreover in this verification protocol we achieve the goal of probability amplification by choosing the appropriate value for parameter s.

▶ **Theorem 11.** *There exists a correct ϵ-verifiable protocol where the client is restricted to BPP and the ability to prepare single qubits and the server is within DQC1-MBQC. Using $O(sm)$ qubits and $O(sm)$ time steps, where m is the size of the input computation, we have:*

$$\epsilon = \frac{2m}{s} \tag{10}$$

2 DQC1-MBQC and Blindness

In this section we give a constructive proof of our main theorem for DQC1-MBQC and show how to construct a blind protocol as a consequence. The first step for proving Theorem 7 is the following rewriting scheme for patterns with flow.

▶ **Lemma 12.** *Any measurement pattern on an open graph state (G, I, O) with flow $(f, \preceq)$ (as defined in Definition 2) and measurement angles $\boldsymbol{a}$ where either $|I| = |O|$ or the flow*

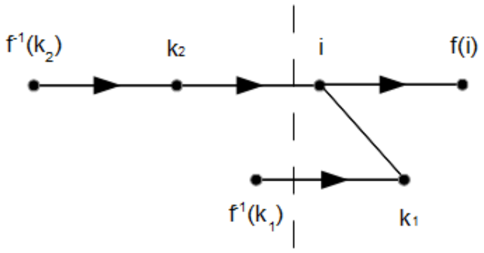


Figure 1 Qubit i gets an X correction from k_2 and Z corrections from $f^{-1}(k_2)$ and $f^{-1}(k_1)$. Qubits on the left of the dashed line are in the past of i. Qubit k_1 is created at timestep $f^{-1}(k_1)$ which is before timestep i from flow condition (F2).

function is surjective can be rewritten as

$$P_{\boldsymbol{a}} = \prod_{i\in O} X_i^{S_i^x} Z_i^{S_i^z} \prod_{i\in O^c}^{\preceq} \left({}^{S_i^z}[M_i^{a_i}]^{S_i^x} \left(\prod_{\{k:k\sim i, k\succeq i\}} E_{i,k} \right) N_{f(i)}(|+\rangle) \right) \tag{11}$$

where $S_i^x = s_{f^{-1}(i)}$ *for* $i \in I^c$*, else* $S_i^x = 0$ *and* $S_i^z = \sum_{\{k:k\in I^c, k\sim i, i\neq f^{-1}(k)\}} s_{f^{-1}(k)} \mod 2$. *The above pattern is runnable and implements the following unitary*

$$U_{G,I,O,\boldsymbol{a}} = 2^{|O^c|/2} \left(\prod_{i\in O^c} \langle +_{a_i}|_i \right) E_G N_{I^c} \tag{12}$$

where E_G *and* N_{I^c} *represent the global entangling operator and global preparation respectively.*

Proof. First we need to prove that $P_{\boldsymbol{a}}$ is runnable (cf. Definition 1). For condition (R0) we make the following observations: At step i, for $i \in I^c$, we need the result $s_{f^{-1}(i)}$ which is generated at step $f^{-1}(i)$, where $f^{-1}(i) \prec i$ from flow condition (F1). We also need the results $s_{f^{-1}(k)}$, for $\{k : k \in I^c, k \sim i, i \neq f^{-1}(k)\}$, which are generated at step $f^{-1}(k)$, where $f^{-1}(k) \prec i$ from flow condition (F2). Thus, condition (R0) is satisfied (see Figure 1 for a particular example). For condition (R1) we make the following observations: At step i, for $i \in O^c$, the entangling operator and measurement operator act on qubit i which either belongs in the set of inputs I or is created at step $f^{-1}(i)$, where $f^{-1}(i) \prec i$ from flow condition (F1). Entangling operator acts also on qubits $\{k : k \sim i, k \succeq i\}$. If $k = f(i)$ then qubit k is created at the same step (i) by operator $N_{f(i)}$. If $k \neq f(i)$ then qubit k is either an input or it is created at step $f^{-1}(k)$, and we have by flow condition (F2): i is a neighbour of k and $i \neq f^{-1}(k)$, thus $f^{-1}(k) \prec i$ (Figure 1). Final correction operators act on qubits that belong to the set of outputs O, which either belong also to the set of inputs I or are created at steps $\{f^{-1}(i) : i \in O\}$, where $\forall i \in O \setminus I, f^{-1}(i) \prec i$ from flow condition (F1). Moreover they have not yet been measured since $i \notin O^C$. Thus, condition (R1) is satisfied. It is easy to see that condition (R2) is satisfied.

Next we prove that the pattern of Equation 11 is implementing the unitary operation of Equation 12 when applied on an open graph with the properties described above. Since condition (R1) is satisfied, all preparation operators trivially commute with all previous operators

$$P_{\boldsymbol{a}} = \prod_{i\in O} X_i^{S_i^x} Z_i^{S_i^z} \prod_{i\in O^c}^{\preceq} \left({}^{S_i^z}[M_i^{a_i}]^{S_i^x} \left(\prod_{\{k:k\sim i, k\succeq i\}} E_{i,k} \right) \right) N_{I^c}\,.$$

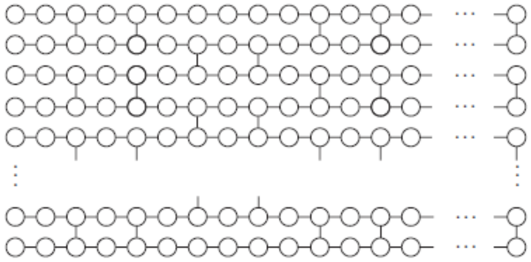


Figure 2 Brickwork state.

Each entangling operator commutes with all previous measurements since it is applied on qubits with indices $\succeq i$.

$$P_{\boldsymbol{a}} = \prod_{i\in O} X_i^{S_i^x} Z_i^{S_i^z} \prod_{i\in O^c}^{\preceq} \left({}^{S_i^z}[M_i^{a_i}]^{S_i^x}\right) E_G N_{I^c}\,.$$

We can decompose the conditional measurements into simple measurements and corrections

$$P_{\boldsymbol{a}} = \prod_{i\in O} X_i^{S_i^x} Z_i^{S_i^z} \prod_{i\in O^c}^{\preceq} \left(M_i^{a_i} X_i^{S_i^x} Z_i^{S_i^z}\right) E_G N_{I^c}\,.$$

By rearranging the order of correction operators we take

$$P_{\boldsymbol{a}} = \prod_{i\in O^c}^{\preceq} \left(X_{f(i)}^{s_i} \prod_{\{k:k\sim f(i),k\neq i\}} Z_k^{s_i} M_i^{a_i}\right) E_G N_{I^c}\,.$$

The above equation implements the unitary operation presented in the lemma (Equation 12) as proved in [10]. ◀

Next, we notice that there exist many universal families of open graph states satisfying the requirements of the above lemma. One such example is the brickwork graph state originally defined in [8]. In this graph state (Figure 2), the subset of vertices of the first column correspond to the input qubits I and the subset of vertices of the final column correspond to the output qubits O. This graph state has flow function $f((i,j)) = (i, j+1)$ and the following partial order for measuring the qubits: $\{(1,1),(2,1),\ldots,(w,1)\} \prec \{(1,2),(2,2),\ldots,(w,2)\} \prec \ldots \prec \{(1,d-1),(2,d-1),\ldots,(w,d-1)\}$, where w is the width and d is the depth of the graph and hence from Lemma 12 we obtain the following corollary.

▶ **Corollary 13.** *Any computation over the brickwork open graph state G with qubit index $(i \leq w, j \leq d)$ can be rewritten as follows.*

$$P_{\boldsymbol{a}} = \prod_{i=1}^{w} X_{(i,d)}^{S^x_{(i,d)}} Z_{(i,d)}^{S^z_{(i,d)}} \prod_{j=1}^{d-1}\prod_{i=1}^{w} {}^{S^z_{(i,j)}}\left[M_{(i,j)}^{a_{(i,j)}}\right]^{S^x_{(i,j)}} \left(\prod_{\substack{\{k,l:(k,l)\sim(i,j),\\ k\geq i,l\geq j\}}} E_{(i,j),(k,l)} \right) N_{(i,j+1)} \quad (13)$$

where

$S^x_{(i,j)} = s_{(i,j-1)}$ *for* $j > 1$, *else* $S^x_{(i,1)} = 0$, *and*

$S^z_{(i,j)} = \sum_{\{k,l:(k,l)\sim(i,j),l\leq j\}} s_{(k,l-1)} \mod 2$ *for* $j > 2$, *else* $S^z_{(i,j)} = 0$.

We show that patterns defined in Lemma 12 are within the framework of Definition 6 hence obtaining a sufficient condition for DQC1-MBQC.

▶ **Theorem 1.** *Any measurement pattern that can be rewritten in the form of Equation 11 represents a DQC1-MBQC pattern.*

Proof. A first general observation about the purity parameter π is that adding a new pure qubit σ to state ρ means that π increases by unity

$$\pi_{\rho\otimes\sigma} = \log_2 \mathrm{Tr}((\rho\otimes\sigma)^2) + d + 1 = \log_2 \mathrm{Tr}(\rho^2)\mathrm{Tr}(\sigma^2) + d + 1 = \pi_\rho + 1.$$

Additionally, applying any unitary U does not change the purity parameter π of the system since $\mathrm{Tr}((U\rho U^\dagger)^2) = \mathrm{Tr}(\rho^2)$ and dimension remains the same.

Returning to Equation 11, we notice that for every step $i \in O^c$ of the product the total computation performed corresponds mathematically to the following: On the qubit tagged with position i, a $J(a'_i)$ unitary gate is applied (where a'_i is an angle that depends on a_i and previous measurement results) up to a specific Pauli correction (depending on the known measurement result) and some specific Pauli corrections on the its entangled neighbours (again depending on the measurement result). At the end the qubit is tagged with position $f(i)$ (where f is the flow function). Since this mathematically equivalent computation is a unitary and the dimension of the system remains the same (there is only a change of position tags) we conclude that each step $i \in O^c$ does not increase the purity parameter of the system. To finish the proof we need to ensure that the individual operations within each step $i \in O^c$ and for $i \in O$ do not increase the purity parameter by more than a constant (and since there is only a constant number of operations within each step this does not increase the purity at any point more than constant). This is true since all these operations apply on (or add or trace over) a constant number of qubits. ◀

Building on this result, we can translate the UBQC protocol of [8] (and in fact many other existing protocols) to allow the blind execution of any DQC1-MBQC computation, where the server is restricted to DQC1-MBQC complexity class. The UBQC protocol is based on the brickwork graph state described above. Alice prepares all the qubits of the graph state, adding a random rotation around the (X, Y) plane to each one of them: $|+_{\theta_i}\rangle$, where θ_i is chosen at random from the set $A = \{0, \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/5, 3\pi/2, 7\pi/4\}$ and sends them to Bob, who entangles them according to the graph. The protocol then follows the partial order given by the flow: Alice calculates the corrected measurement angle α'_i for each qubit using previous measurement results according to the flow dependences. She sends to Bob measurement angle $\delta_i = \alpha'_i + \theta_i + r_i\pi$, using an extra random bit r_i. Bob measures according to δ_i, reports the result back to Alice who corrects it by XOR-ing with r_i. In the case of quantum output, the final layer is sent to Alice and is also corrected according to the flow dependences by applying the corresponding Pauli operators.

Since the brickwork graph state satisfies the requirements of Theorem 7 we can adapt the Universal Blind Quantum Computing protocol by making Alice and Bob follow the order of Equation 13 and operate on input $|+\rangle\langle+| \otimes I_{w-1}/2^{w-1}$. A detailed description is given in Protocol 1.

▶ **Theorem 4.** *Protocol 1 is correct.*

Proof. Correctness comes from the fact that what Alice and Bob jointly compute is mathematically equivalent to performing the pattern of Equation 13 on input $|+\rangle\langle+| \otimes I_{w-1}/2^{w-1}$. The argument is the same as in the original universal blind quantum computing protocol [8]

Protocol 1 Blind BQ1P protocol

Alice's input:

- A vector of angles $\boldsymbol{a} = (a_{1,1}, \ldots, a_{w,d})$, where $a_{i,j}$ comes from the set $A = \{0, \pi/4, 2\pi/4, \ldots, 7\pi/4\}$, that when plugged in the measurement pattern $P_{\boldsymbol{a}}$ of Equation 13 applied on the brickwork state, implements the desired computation. This computation is applied on a fixed input state $|+\rangle\langle+| \otimes I_{w-1}/2^{w-1}$.

Alice's output:

- The top output qubit (qubit in position $(1, d)$).

The protocol

1. Alice picks a random angle $\theta_{1,1} \in A$, prepares one pure qubit in state $R_z(\theta_{1,1})|+\rangle$ and sends it to Bob who tags it as qubit $(1, 1)$.
2. Bob prepares the rest of input state (qubits $(2, 1), \ldots, (w, 1)$) in the maximally mixed state $I_{w-1}/2^{w-1}$.
3. Alice and Bob execute the rest of the computation in rounds. For $j = 1$ to $d - 1$ and for $i = 1$ to w
   a. **Alice's preparation**
      i. Alice picks a random angle $\theta_{i,j+1} \in A$.
      ii. Alice prepares one pure qubit in state $R_z(\theta_{i,j+1})|+\rangle$.
      iii. Alice sends it to Bob. Bob tags it as qubit $(i, j + 1)$.
   b. **Entanglement and measurement**
      i. Bob performs the entangling operator(s):

$$\prod_{\{k,l:(k,l)\sim(i,j),k\geq i,l\geq j\}} E_{(i,j),(k,l)}$$

      ii. Bob performs the rest of the computation using classical help from Alice:
         A. Alice computes the corrected measurement angle $a'_{i,j} = (-1)^{S^x_{i,j}} a_{i,j} + S^z_{i,j}\pi$.
         B. Alice chooses a random bit $r_{i,j}$ and computes $\delta_{i,j} = a'_{i,j} + \theta_{i,j} + r_{i,j}\pi$.
         C. Alice transmits $\delta_{i,j}$ to Bob.
         D. Bob performs operation $M^{\delta_{i,j}}_{i,j}$ which measures and traces over the qubit (i, j) and retrieves result $b_{i,j}$.
         E. Bob transmits $b_{i,j}$ to Alice.
         F. Alice updates the result to $s_{i,j} = b_{i,j} + r_{i,j} \mod 2$.
4. Bob sends to Alice the final layer of qubits, Alice performs the required corrections and outputs the result.

repeated here for completeness. Firstly, since entangling operators commute with R_z operators, preparing the pure qubits in a rotated state does not change the underlying graph state; only the phase of each qubit is locally changed, and it is as if Bob had performed the R_z rotation after the entanglement. Secondly, since a measurement in the $|+_a\rangle, |-_a\rangle$ basis on a state $|\phi\rangle$ is the same as a measurement in the $|+_{a+\theta}\rangle, |-_{a+\theta}\rangle$ basis on $R_z(\theta)|\phi\rangle$, and since $\delta = a' + \theta + \pi r$, if $r = 0$, Bob's measurement has the same effect as Alice's target measurement; if $r = 1$, all Alice needs to do is flip the outcome. ◀

Note that Protocol 1 can be trivially simplified by omitting all the measurements that are applied on maximally mixed states (i.e. all measurements applied on qubits in rows 2 to w from the beginning of the computation until each one is entangled with a non-maximally mixed qubit). However this does not give any substantial improvement in the complexity of the protocol.

▶ **Theorem 5.** *Protocol 1 is blind.*

(Proof Sketch). A detailed proof is provided in Appendix A. Intuitively, rotation by angle $\theta_{i,j}$ serves the purpose of hiding the actual measurement angle, while rotation by $r_{i,j}\pi$ hides the result of measuring the quantum state. This proof is consistent with definition of blindness based on the relation of Bob's system to Alice's system which takes into account prior knowledge of the secret and is a good indicator that blindness can be composable [14]. ◀

Regarding the complexity of the protocol, Alice needs to pick a polynomially large number of random bits and perform polynomially large number of modulo additions that is to say Alice classical computation is restricted to the class BPP. However Alice's quantum requirement is only to prepare single qubits, she has access to no quantum memory or quantum operation. Therefore assuming $BQ1P \not\subset BPP$ suggests Alice's quantum power is more restricted than $BQ1P$ and hence DQC1-MBQC. On the other hand, Bob performs a pattern of the form given in Equation 13, with the difference that instead of preparing the pure qubits himself, he receives the pure qubits through the quantum channel that connects him with Alice. Also, the qubits are not prepared in state $|+\rangle$, but in some state on the (X, Y) plane, but this doesn't alter the reasoning in the complexity proofs. Thus, Bob has computational power that is within the DQC1-MBQC complexity class according to the Corollary 13 and Theorem 7.

3 Verification

VBQC protocol is based on the ability to hide a trap qubit inside the graph state while not affecting the correct execution of the pattern. Both the trap qubit and the qubits which participate in the actual computation are prepared in the (X, Y) plane of the Bloch sphere. To keep them disentangled, some qubits (called dummy) prepared in the computational basis $\{|0\rangle, |1\rangle\}$, are injected between them. Being able to choose between the two states is essential for blindness (Theorem 4 in [18]). In particular, if a dummy qubit is in state $|0\rangle$, applying the entangling operator cZ between this qubit and a qubit prepared on the (X, Y) plane has no effect. If a dummy qubit is in state $|1\rangle$ then applying cZ will introduce a Pauli Z rotation on the qubit prepared on the (X, Y) plane. This effect can be cancelled by Alice in advance, by introducing a Pauli Z rotation on all the neighbours of $|1\rangle$'s when preparing the initial state.

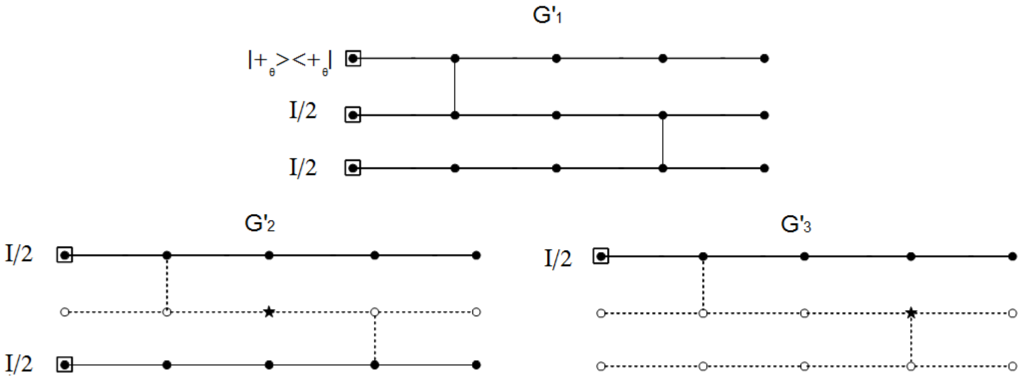


Figure 3 Let G' be the graph which consists of s isolated brickwork graphs (each denoted as G'_i), each of the same dimensions required for the desired computation. An example construction with $s = 3$ and one trap per graph together with a small brickwork state for computation is given above. Black vertices correspond to auxiliary qubits prepared on the $(X - Y)$ plane or mixed state when they are inputs (inside square), star vertices correspond to trap qubits and white vertices to auxiliary qubits prepared in the computational basis. Edges represent entangling operators, dashed where entangling has no effect (except of local rotations).

In the simplest version of VUBQC, a single trap, prepared in state $|+_{\theta_t}\rangle$, where θ is chosen at random from the angles set A (defined above) and placed at position t, chosen at random between all the vertices of the open graph state (G, I, O). During the execution of the pattern, if $t \notin O$, Bob is asked to measure qubit t with angle $\theta_t + r\pi$ and return the classical result b_t to Alice. If $b_t = r_t$ Alice sets an indicator bit to state acc (which means that this computation is accepted), otherwise she sets it to rej (computation is rejected). If $t \in O$, Alice herself measures the trap qubit and sets the indicator qubit accordingly. This version of the protocol is proven to be correct and ϵ-verifiable, with $\epsilon = (m - 1)/m$, where m is the size of the computation.

A generalisation of this technique which allows for arbitrary selection of parameter ϵ is also presented in [18]. By allowing for a polynomial number of traps to be injected in the graph state and adapting the computation inside a fault tolerant scheme with parameter d one can have ϵ inversely exponential to d. The question is whether this amplification method can also be used to design a verification protocol for DQC1-MBQC with arbitrary small ϵ. Unfortunately the underlying graph state used by this protocol does not have flow and not all qubits are prepared in the (X, Y) plane, so that one can not apply Theorem 7 to get a compatible rewriting of the pattern. Moreover, having the requirement that we should be able to place every trap qubit (which is a pure qubit) at any position in the graph, means that there exist patterns that will never be possible to be rewritten to satisfy the purity requirement. This leads us to seek a different approach for probability amplification for verification in the DQC1-MBQC model.

Instead of placing a polynomial number of isolated traps within the same graph, which is also used to perform the actual computation, we utilise s isolated brickwork subgraphs, one used for the computation and the rest being trap subgraphs (see Figure 3). Therefore at the beginning of the protocol, Alice chooses random parameter t_g, which denotes which graph will be the computational subgraph, and for each of the remaining trap subgraphs i, she chooses a random position t_i to hide one isolated trap. The rest of each trap subgraph will be a trivial computation (all measurement angles set to 0) on a totally mixed state, and a selected set of dummy qubits are placed to isolate this computation from the trap. Computation subgraph and trap subgraphs are of the same size, and by taking advantage of the blindness of the protocol, Bob cannot distinguish between them. Therefore, to be able to cheat, he needs to deviate from the correct operation only during the execution on the computational subgraph and never deviate while operating on any of the traps. This gives the desirable ϵ parameter that will be proved later. The full description of protocol is given in Protocol 2. Each isolated pattern k is executed separately and according to the DCQ1-MBQC rewriting on the brickwork state given in Equation 13 in the blind setting. Pre-rotations on the neighbours of dummy qubits guarantee that the computation is not affected by the choice of dummies as described before. To prove the complexity of the protocol we need to notice that although the graph used satisfies the conditions of Theorem 7, the existence of the dummy qubits prepared in the computational basis creates the need of a new proof.

▶ **Theorem 6.** *The computational power of Bob in Protocol 2 is within DQC1-MBQC.*

Proof. Note that the s patterns are executed in series and Bob does not keep any qubits between executions. The inputs to these patterns are almost maximally mixed, in accordance with the purity requirement and this 'mixedness' propagates through both computational and trap subgraphs. For the computational subgraph (which is not entangled with the rest) the reasoning of the proof of Theorem 7 applies, since this subgraph satisfies the sufficient conditions and no dummy qubits are used. In the case of a trap subgraph k consider first those operations that apply on the isolated trap and dummy subgraph only. Then for each step $(i, j)_k$ of the main iteration of the protocol (where $(i, j)_k$ is a trap or a dummy) a new pure qubit is sent to Bob, which increases the purity parameter by 1. Entangling will not have any effect on the purity parameter. While the measurement does not increase the purity of the qubit since it was already pure (dummy or trap remain always pure through the computation), and tracing out the resulting qubit will decrease the purity by 1. Thus, the whole step will not change the purity. On the other hand, for the remaining operations the reasoning of the proof of Theorem 7 goes through, since this subgraph satisfies the sufficient conditions. Also operations that apply on both subgraphs are all unitaries therefore they do not affect purity. ◀

Using the definition of verifiability given in Definition 10 we prove the main theorem for the existence of a correct and verifiable DQC1-MBQC protocol (Theorem 11). The full proof is given in Appendix B, while here we describe the main steps.

Proof of Theorem 11 (Sketch). Correctness of Protocol 2 comes from the fact that the computational subgraph is disentangled from the rest of the computation and if Bob performs the predefined operations, from the correctness of the blind protocol Alice will receive the correct output. Also, in this case, (and since the traps are corrected to cancel the effect of their entanglement with their neighbouring dummies) the measurement of the traps will give the expected result and Alice will accept the computation.

The proof of verifiability follows the same general methodology of the proof of the original VUBQC protocol [18], except the last part which contains the counting arguments. For the rest we use single indexing for the qubits, where subgraph G'_i consists of m qubits indexed $(i-1)+1$ to im. Therefore the total number of qubits in the protocol is sm. Parameter n represents the size of the input of each subgraph (parameter w in the protocol).

Based on Definition 10 we need to bound the probability of the (purified) output collapsing onto the wrong subspace and accepting that result. To explicitly write the final state $B_j(\nu)$

Protocol 2 Verifiable DQC1-MBQC protocol with $s-1$ trap computations

Alice's input:

- An angle vector $\boldsymbol{a} = (a_{1,1}, \dots, a_{w,d-1})$, where $a_{i,j}$ comes from the set $A = \{0, \pi/4, 2\pi/4, \dots, 7\pi/4\}$, that, when plugged in the measurement pattern $P_{\boldsymbol{a}}$ of Equation 13 on the brickwork open graph state G of dimension (w, d) and flow $(f, \preceq)$, it implements the desired computation on fixed input $|+\rangle\langle+| \otimes I_{w-1}/2^{w-1}$.

Alice's output:

- The top output qubit of G (qubit in position $(1, d)$ in G) together with a 1-bit, named *acc*, that indicates if the result is accepted or not.

The protocol

- **Preparation steps.** Alice picks t_g at random from $\{1, \dots, s\}$. Let G' be the graph which consists of s isolated brickwork graphs, each of the dimension the same as G. Then the t_g-th isolated graph (named G'_{t_g}) will be the computational subgraph for this run of the protocol.
- Alice maps the measurement angles of the computational subgraph G'_{t_g} to angles of graph G: $\boldsymbol{a}'_{G_{t_g} \setminus O_{t_g}} = \boldsymbol{a}$ and appropriately set the dependency sets S^x and S^z for all the vertices of G'_{t_g} (according to the standard flow), while for the rest of the vertices (graph $G' \setminus G'_{t_g}$) the sets S^x and S^z are empty.
- For $k = 1$ to s except t_g:
  1. Alice chooses one random vertex $\boldsymbol{t}_k = (t_x, t_y)_k$ among all vertices of G'_k for placing the trap.
  2. By G'_k's geometry, vertex (t_x, t_y) may be connected by a vertical edge to vertex (t'_x, t_y), where t'_x represents either $t_x + 1$ or $t_x - 1$. We add in D (set of dummies) all vertices of rows t_x, t'_x (if it exists) of G'_k, except the trap itself.
  3. All elements of $\boldsymbol{a}'_{G_k}$ are mapped to 0.
- Alice chooses random variables $\boldsymbol{\theta}_{G' \setminus D}$, each uniformly at random from A.
- Alice chooses random variables $\boldsymbol{r}_{G'}$ and $\boldsymbol{d}_D$, each uniformly at random from $\{0, 1\}$.
- For $k = 1$ to s:
  1. **Initial step.** If $k = t_g$ then: Let $(1, 1)_k$ be the position of the top input qubit in G'_k. Alice prepares the following states and sends them to Bob:
$$\begin{array}{rl} \{(1,1)_k\} & \left|+_{\theta_{(1,1)_k}}\right\rangle \\ \forall (i,1)_k \notin \{(1,1)_k\} & I/2 \end{array}$$
     Otherwise: Alice prepares the following states and sends them to Bob:
$$\begin{array}{rc} \forall (i,1)_k \in D & \left|d_{(i,1)_k}\right\rangle \\ (i,1)_k = \boldsymbol{t}_k & \prod_{\{m,l:(m,l)_k \sim (i,1)_k, (m,l)_k \in D\}} Z^{d_{(m,l)_k}} \left|+_{\theta_{(i,1)_k}}\right\rangle \\ \forall (i,1)_k \notin \{D, \boldsymbol{t}_k\} & I/2 \end{array}$$

Protocol 2 (cont'd)

2. **Main Iteration.** For $j = 1$ to $d-1$, for $i = 1$ to w:
   a. **Alice's preparation**
      i. Alice prepares one pure qubit in one of the following states, depending on $(i, j+1)_k$:
         $(i, j+1)_k \in D$ $\quad\left|d_{(i,j+1)_k}\right\rangle$
         $(i, j+1)_k \notin D$ $\quad\prod_{\{m,l:(m,l)_k \sim (i,j+1)_k,(m,l)_k \in D\}} Z^{d_{(m,l)_k}}\left|+_{\theta_{(i,j+1)_k}}\right\rangle$
      ii. Alice sends it to Bob. Bob labels it as qubit $(i, j+1)_k$.
   b. **Entanglement and measurement**
      i. Bob performs the entangling operator(s):

$$\prod_{\{m,l:(m,l)_k \sim (i,j)_k, m \geq i, l \geq j\}} E_{(i,j)_k,(m,l)_k}$$

      ii. Bob performs the rest of the computation using classical help from Alice:
         A. Alice computes the corrected measurement angle $a''_{(i,j)_k} = (-1)^{S^x_{(i,j)_k}} a'_{(i,j)_k} + S^z_{(i,j)_k}\pi$.
         B. Alice computes actual measurement angle $\delta_{(i,j)_k} = a''_{(i,j)_k} + \theta_{(i,j)_k} + r_{(i,j)_k}\pi$.
         C. Alice transmits $\delta_{(i,j)_k}$ to Bob.
         D. Bob performs operation $M^{\delta_{(i,j)_k}}_{(i,j)_k}$ which measures and traces over the qubit $(i,j)_k$ and retrieves result $b_{(i,j)_k}$.
         E. Bob transmits $b_{(i,j)_k}$ to Alice.
         F. Alice updates the result to $s_{(i,j)_k} = b_{(i,j)_k} + r_{(i,j)_k} \mod 2$.
3. Bob sends the final layer to Alice and Alice applies the final corrections if needed (only in round t_g).
4. If the trap qubit is within the qubits received, Alice measures it with angle $\delta_{\mathbf{t}_k} = \theta_{\mathbf{t}_k} + r_{\mathbf{t}_k}\pi$ to obtain $b_{\mathbf{t}_k}$. Also, Alice discards all qubits received by Bob in this round except qubit $(1,d)_{t_g}$.

- Alice outputs qubit in position $(1,d)_{t_g}$ and sets bit *acc* to 1 if $b_{\mathbf{t}_k} = r_{\mathbf{t}_k}$ for all k.

we need to define the following notations. Alice's chosen random parameters are denoted collectively by ν, a subset of those are related to the traps: ν_T including t_g, t_k's and θ_{t_k}'s for $k \in \{1, \dots, s\} \setminus t_g$. Also $\nu_C = \{\nu \setminus \nu_T\}$. The projection onto the correct state for each trap t_k is denoted by $\left|\eta_{t_k}^{\nu_T}\right\rangle$, where $\left|\eta_{t_k}^{\nu_T}\right\rangle = \left|+_{\theta_{t_k}}\right\rangle$ when $t_k \in O_k$ and $\left|\eta_{t_k}^{\nu_T}\right\rangle = |r_{t_k}\rangle$ otherwise (since the trap has been already measured). $C_{\boldsymbol{r}}$ denotes the Pauli operators that map the output state of the computational subgraph to the correct one. $c_{\boldsymbol{r}}$ is used to compactly deal with the fact that in the protocol each measured qubit i is decrypted by XOR-ing them with r_i, except for the trap qubits which remain uncorrected: $\forall k : (c_{\boldsymbol{r}})_{t_k} = 0$. $\rho_{M_k^\nu}$ denotes the density matrix representing the total quantum state received by Bob from Alice for each round k of the protocol. A special case is the t_kth round where $\rho_{M_k^\nu}$ represents the total state received by Bob together with its purification (not known to Bob). The classical information received by Bob at each elementary step i (measurement angles) are represented by $|\delta_i\rangle$'s.

We allow Bob to have an arbitrary deviation strategy j, at each elementary step i which is represented as CPTP map $\mathcal{E}_i^j$, followed by a Pauli Z measurement of qubit i (since Bob has to produce a classical bit at each step and return it to Alice), which is represented by taking the sum over projectors on the computational basis $|b_i\rangle$, for $b_i \in \{0, 1\}$. All measurement operators can be commuted to the end of the computation and all CPTP maps can be gathered to a single map $\mathcal{E}^j$ after Bob has received everything from Alice, so that the failure probability can be written as:

$$p_{incorrect} = \sum_{\boldsymbol{b'},\nu} p(\nu)\mathrm{Tr}(P_\perp \bigotimes_{k=1}^{s} \left|\eta_{t_k}^{\nu_T}\right\rangle\left\langle\eta_{t_k}^{\nu_T}\right|$$

$$C^{\boldsymbol{b'},\nu_C}|\boldsymbol{b'}+\boldsymbol{c^r}\rangle\langle\boldsymbol{b'}|\mathcal{E}^j \left(\bigotimes_{k=1}^{s}\bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes \rho_{M_k^\nu}\right)|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|C^{\boldsymbol{b'},\nu_C\dagger})$$

Our strategy will be to rewrite this probability by introducing the correct execution of the protocol before the attack, on each subgraph k: $\mathcal{P}_k = \bigotimes_{i=1}^{m-n}(H_{(k-1)m+i}Z_{(k-1)m+i}(\delta_{(k-1)m+i}))E_{G'_k}$ and at the same time decomposing the attack to the Pauli basis, using general Paulis $\sigma_{i,k}$ applying on qubits $(k-1)m+1 \leq \gamma \leq km$ for each k.

$$p_{incorrect} = \sum_{\boldsymbol{b'},\nu,v,i,j} \alpha_{vi}\alpha_{vj}^{*}p(\nu)\mathrm{Tr}(P_\perp \bigotimes_{k=1}^{s} \left|\eta_{t_k}^{\nu_T}\right\rangle\left\langle\eta_{t_k}^{\nu_T}\right|C^{\boldsymbol{b'},\nu_C}|\boldsymbol{b'}+\boldsymbol{c^r}\rangle\langle\boldsymbol{b'}|$$

$$\bigotimes_{k=1}^{s}(\sigma_{i,k}\left(\mathcal{P}_k\bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes \rho_{M_k^\nu}\mathcal{P}_k^\dagger\right)\sigma_{j,k})|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|C^{\boldsymbol{b'},\nu_C\dagger}$$

This way we can characterise which Pauli attacks give non-zero failure probability when the final state is projected on the correct one. For convenience we introduce the following sets for an arbitrary Pauli $\sigma_{i,k}$:

$$\begin{aligned}
A_{i,k} &= \{\gamma \text{ s.t. } \sigma_{i|\gamma} = I \text{ and } (k-1)m+1 \leq \gamma \leq km\} \\
B_{i,k} &= \{\gamma \text{ s.t. } \sigma_{i|\gamma} = X \text{ and } (k-1)m+1 \leq \gamma \leq km\} \\
C_{i,k} &= \{\gamma \text{ s.t. } \sigma_{i|\gamma} = Y \text{ and } (k-1)m+1 \leq \gamma \leq km\} \\
D_{i,k} &= \{\gamma \text{ s.t. } \sigma_{i|\gamma} = Z \text{ and } (k-1)m+1 \leq \gamma \leq km\}
\end{aligned}$$

We use the superscript O to denote subsets subject to the constraint $km \geq \gamma \geq km - n + 1$. For an arbitrary t_g, the only attacks that give the corresponding term of the sum not equal to zero: are those that (i) produce an incorrect measurement result for qubits $(t_g-1)m+1 \leq \gamma \leq$

$t_g m - n$ or (ii) operate non-trivially on qubits $t_g m - n < \gamma \leq t_g m$. We denote this condition by $i \in E_{i,t_g}$ and $j \in E_{j,t_g}$: $|B_{i,t_g}| + |C_{i,t_g}| + |D^O_{i,t_g}| \geq 1$ and $|B_{j,t_g}| + |C_{j,t_g}| + |D^O_{j,t_g}| \geq 1$.

The next step will be to characterise which attacks of these subsets remain undetected by the trap mechanism and try to find an upper bound on their contribution to the failure probability. By applying blindness and observing that only the terms where $\sigma_{i,k} = \sigma_{j,k}$ contribute we obtain the following upper bound (details in Appendix B):

$$p_{incorrect} \leq \sum_{t_g} \sum_{v,i \in E_{i,t_g}} |\alpha_{vi}|^2 p(t_g) \prod_{k=\{1,\ldots,s\} \setminus t_g} \left(\sum_{\substack{km-n<t_k \leq km,\\ \theta_{t_k}}} p(t_k, \theta_{t_k}) (\left\langle +_{\theta_{t_k}} \middle| \sigma_{i|t_k} \middle| +_{\theta_{t_k}} \right\rangle \right)^2$$
$$+ \sum_{\substack{(k-1)m<t_k \leq km-n,\\ r_{t_k}}} p(t_k, r_{t_k}) (\langle r_{t_k} | \sigma_{i|t_k} | r_{t_k} \rangle)^2)$$

The rest is based on a counting argument using $\forall k$, $|A_{i,k}| + |B_{i,k}| + |C_{i,k}| + |D_{i,k}| = m$.

$$p_{incorrect} \leq \sum_{t_g} \sum_{v,i \in E_{i,t_g}} |\alpha_{vi}|^2 \frac{1}{s} \prod_{k=\{1,\ldots,s\} \setminus t_g} \frac{1}{2m} (2|A_{i,k}| + |B^O_{i,k}| + |C^O_{i,k}| + 2|D_{i,k} \setminus D^O_{i,k}|)$$
$$\leq \sum_{t_g} \sum_{v,i \in E_{i,t_g}} |\alpha_{vi}|^2 \frac{1}{s} \prod_{k=\{1,\ldots,s\} \setminus t_g} \frac{1}{2m} (2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$$

We denote the product term $\prod_{k=\{1,2,3,\ldots,s\} \setminus z} \frac{1}{2m} (2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$ as $P_{i,z}$. We also denote each set $\{E^*_{i,1} \cap E^*_{i,2} \cap \ldots \cap E^*_{i,s}\}$, where each term $E^*_{i,w}$ is either $E_{i,w}$ or its complement, $E^C_{i,w}$, depending on whether the w-th value of a binary vector $\boldsymbol{y}$ (size s) is 1 or 0 respectively, as $W_{i,\boldsymbol{y}}$. Let the function $\#\boldsymbol{y}$ give the number of positions i such that y_i=1.

$$= \frac{1}{s} (\sum_{k=1}^{s} \sum_{\{\boldsymbol{y}:\#\boldsymbol{y}=k\}} \sum_{i \in W_{i,\boldsymbol{y}},v} (|\alpha_{vi}|^2 \sum_{\{z:y_z=1\}} P_{i,z}))$$

The condition $i \in W_{i,\boldsymbol{y}}$ means that the following conditions hold together: $\{|B_{i,w}| + |C_{i,w}| + |D^O_{i,w}| \geq 1 : y_w = 1\}$,$\{|B_{i,w}| + |C_{i,w}| + |D^O_{i,w}| = 0 : y_w = 0\}$.

$$\leq \frac{1}{s} (\sum_{k=1}^{s} \sum_{\{\boldsymbol{y}:\#\boldsymbol{y}=k\}} \sum_{i \in W_{i,\boldsymbol{y}},v} |\alpha_{vi}|^2 k \left(\frac{2m-1}{2m} \right)^{k-1}) = \frac{1}{s} (\sum_{k=1}^{s} c_k k \left(\frac{2m-1}{2m} \right)^{k-1})$$

where $c_k = \sum_{\{\boldsymbol{y}:\#\boldsymbol{y}=k\}} \sum_{i \in W_{i,\boldsymbol{y}},v} |\alpha_{vi}|^2$.

An upper bound on the above expression is:

$$p_{incorrect} < \frac{2m}{s} \tag{14}$$

◀

4 Conclusion

In this paper we present the first study of the delegation of quantum computing in a restricted model of computing and show that the general framework of the verification via blindness could be adapted to the setting of one-pure qubit model. In order to improve the obtained bound on the security parameter two open questions has to be addressed. The first one aims to expand the class of resource states for DQC1 model so that several techniques from the MBQC domain could be applicable here. The second question will complement the first by searching for fault-tolerant schemes based on any new resource state for DQC1 model. More concretely we propose the study of following questions:

- A sufficient condition for compatibility with DQC1 based on the step-wise determinism criteria is presented in Theorem 7. Is this approach extendable to weaker notions of determinism such as information preserving maps as defined in [24]? Which is a necessary condition for a family of MBQC resource states to be universal for the DQC1 computation?
- Theorem 11 presents a scheme for verification where by adjusting the number of rounds one could obtain an ϵ-verifiable delegated DQC1-MBQC computing with ϵ being polynomially small on computation size. How can we efficiently amplify this bound to any desired exponentially small one? Is there a way to adapt the proposed probability amplification method of [18] based on a quantum error correcting code, into the DQC1-MBQC model?

Acknowledgements. We would like to thank Joe Fitzsimons, Vedran Dunjko and Alistair Stewart for useful discussions. AD was supported in part by EPSRC (Grant No. EP/H03031X/1), U.S. EOARD (Grant No. 093020), and the EU Integrated Project SIQS. TK was supported by Mary and Armeane Choksi Postgraduate Scholarship and School of Informatics Graduate School.

References

1 S. Aaronson and A. Arkhipov. The computational complexity of linear optics. In *STOC*, 2011.

2 D. Aharonov, M. Ben-Or, and E. Eban. Interactive proofs for quantum computations. In *Proceedings of Innovations in Computer Science 2010*, page 453, 2010.

3 H. Barnum, C. Crépeau, D. Gottesman, A. Smith, and A. Tapp. Authentication of quantum messages. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science (FOCS 2002)*, page 449, 2002.

4 S. Barz, E. Kashefi, A. Broadbent, J. F. Fitzsimons, A. Zeilinger, and P. Walther. Demonstration of blind quantum computing. *Science*, 335(6066):303–308, 2012.

5 Stefanie Barz, Joseph F Fitzsimons, Elham Kashefi, and Philip Walther. Experimental verification of quantum computation. *Nature Physics*, 2013.

6 Sergey B. Bravyi and Alexei Yu. Kitaev. Fermionic quantum computation. *Annals of Physics*, 298:210, 2002.

7 M. Bremner, R. Jozsa, and D. Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proc. Roy. Soc. A*, 467:459, 2011.

8 A. Broadbent, J. Fitzsimons, and E. Kashefi. Universal blind quantum computing. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009)*, page 517, 2009.

9 D. Browne, E. Kashefi, M. Mhalla, and S. Perdrix. Generalized flow and determinism in measurement-based quantum computation. *New Journal of Physics*, 9:250, 2007.

10 V. Danos and E. Kashefi. Determinism in the one-way model. *Physical Review A*, 74:052310, 2006.

11 V. Danos, E. Kashefi, and P. Panangaden. The measurement calculus. *Journal of ACM*, 54:8, 2007.

12 A. Datta and A. Shaji. Quantum discord and quantum computing – an appraisal. *International Journal of Quantum Information*, 9:1787, 2011.

13 D. DiVincenzo. The physical implementation of quantum computation. *Fortschr. Phys.*, 48:771, 2000.

14 V. Dunjko, J. Fitzsimons, C. Portmann R., and Renner. Composable security of delegated quantum computation. *arXiv preprint arXiv:1301.3662*, 2013.

15 V. Dunjko, E. Kashefi, and A. Leverrier. Universal blind quantum computing with coherent states. *arXiv preprint arXiv:1108.5571*, 2011.

16 Vedran Dunjko. Ideal quantum protocols in the non-ideal physical world. *PhD Thesis, Heriot-Watt University*, 2012.

17 B. Reichardt F., Unger, and U. Vazirani. Classical command of quantum systems. *Nature*, 496, 2013.

18 Joseph F Fitzsimons and Elham Kashefi. Unconditionally verifiable blind computation. *arXiv preprint arXiv:1203.5217*, 2012.

19 V. Giovannetti, L. Maccone, T. Morimae, and T. Rudolph. Efficient universal blind computation. *arXiv preprint arXiv:1306.2724*, 2013.

20 M. Hein, J. Eisert, and H. J. Briegel. Multi-party entanglement in graph states. *Physical Review A*, 69, 2004. quant-ph/0307130.

21 Stephen P. Jordan. Permutational quantum computing. *Quantum Info. Comput.*, 10(5):470–497, May 2010.

22 E. Knill and R. Laflamme. Power of one bit of quantum information. *Phys. Rev. Lett.*, 81:5672, 1998.

23 A. Mantri, C. Perez-Delgado, and J. Fitzsimons. Optimal blind quantum computation. *arXiv preprint arXiv:1306.3677*, 2013.

24 M. Mhalla, M. Murao, S. Perdrix, M. Someya, and P. Turner. Which graph states are useful for quantum information processing? In *TQC Theory of Quantum Computation, Communication and Cryptography 2011*, 2010.

25 T. Morimae, V. Dunjko, and E. Kashefi. Ground state blind quantum computation on aklt state. *arXiv preprint arXiv:1009.3486*, 2011.

26 Tomoyuki Morimae. Continuous-variable blind quantum computation. *Phys. Rev. Lett.*, 109:230502, Dec 2012.

27 Tomoyuki Morimae and Keisuke Fujii. Blind topological measurement-based quantum computation. *Nature Communications*, 3:1036, 2012.

28 Tomoyuki Morimae and Keisuke Fujii. Blind quantum computation protocol in which alice only makes measurements. *Phys. Rev. A*, 87:050301, May 2013.

29 Tomoyuki Morimae, Keisuke Fujii, and Joseph F Fitzsimons. On the hardness of classically simulating the one clean qubit model. *arXiv preprint arXiv:1312.2496*, 2013.

30 Anna Pappa, André Chailloux, Stephanie Wehner, Eleni Diamanti, and Iordanis Kerenidis. Multipartite entanglement verification resistant against dishonest parties. *Physical Review Letters*, 108(26), 2012.

31 R. Raussendorf and H. J. Briegel. A one-way quantum computer. *Physical Review Letters*, 86:5188–5191, 2001.

32 D. Shepherd. Computing with unitaries and one pure qubit. *arXiv:quant-ph/0608132*, 2006.

33 P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26:1484–1509, 1997. First published in 1995.

34 T. Sueki, T. Koshiba, and T. Morimae. Ancilla-driven universal blind quantum computation. *Physical Review A*, 87, 2013.

A Proof of Theorem 5

Proof. In this proof of blindness for Protocol 1 we use techniques developed in [16]. The basic difference from the proof of [16] arises from the different order in which Bob receives the states from Alice. Nevertheless, after commuting all CPTP maps into a single operator at the end, the methodology for proving blindness is the same as in the original proof. We give the full proof here for the sake of clarity.

To prove blindness we do not separate Alice's system into a classical and a quantum part but we consider the whole of Alice's system as quantum. This is a reasonable assumption since a classical system can be viewed as a special case of a quantum system. Therefore, by proving blindness for the more general case we also prove blindness for the special case.

For the sake of clarity we use single indexing for all the qubits of the resource state. The total number of qubits is denoted by m and the number of qubits in a single column of the brickwork state is denoted by n.

Our goal will be to explicitly write the state $\sigma_B = \mathrm{Tr}_A(\sigma_{AB})$ that Bob holds at the end of the execution of the protocol. To achieve this we express Bob's behaviour at each step i of the protocol as a collection of completely-positive trace-preserving (CPTP) maps $\mathcal{E}_i^{b_i}$, each for every possible classical response b_i from Bob to Alice.

At step 1 of the main loop of the protocol Bob has already been given the top input qubit at position 1 (position $(1,1)$ in the protocol notation) and the qubit at position $f(1) = 1+n$ (position $(1,2)$ in the protocol notation) together with the angle for measuring qubit 1 (angle can be represented as a quantum state composed of 3 qubits). State $\mathrm{Tr}_A(\rho_{AB})$ represents Bob's state before the protocol begins and can, in general, be dependent on Alice's secret measurement angles. The state of Bob averaged over all possible choices of Alice and possible classical responses from Bob, after step 1 is:

$$\sum_{b_1,r_1,\theta_1,\theta_{1+n}} \mathcal{E}_1^{b_1}\left(\left|\delta_1^{\theta_1,r_1}\right\rangle\left\langle\delta_1^{\theta_1,r_1}\right| \otimes |+_{\theta_{1+n}}\rangle\langle+_{\theta_{1+n}}| \otimes |+_{\theta_1}\rangle\langle+_{\theta_1}| \otimes \mathrm{Tr}_A(\rho_{AB})\right)$$

Note the all binary parameters in sums range over 0 and 1, ex. $\sum_{b_1}$ stands for $\sum_{b_1=0}^{1}$ and all angles range over the 8 possible values in A.

We can write the state of Bob after step 2 of the main iteration as:

$$\sum_{b_2,b_1,r_2,r_1,\theta_{2+n},\theta_{1+n},\theta_2,\theta_1} \mathcal{E}_2^{b_2}\left(\left|\delta_2^{\theta_2,r_2}\right\rangle\left\langle\delta_2^{\theta_2,r_2}\right| \otimes |+_{\theta_{2+n}}\rangle\langle+_{\theta_{2+n}}| \right.$$
$$\left. \otimes\, \mathcal{E}_1^{b_1}\left(\left|\delta_1^{\theta_1,r_1}\right\rangle\left\langle\delta_1^{\theta_1,r_1}\right| \otimes |+_{\theta_{1+n}}\rangle\langle+_{\theta_{1+n}}| \otimes |+_{\theta_1}\rangle\langle+_{\theta_1}| \otimes \mathrm{Tr}_A(\rho_{AB})\right)\right)$$

Following this analysis, after the last step of the iteration Bob's state will be:

$$\sigma_B = \sum_{\substack{\boldsymbol{b}_{\leq m-n},\\ \boldsymbol{r}_{\leq m-n},\boldsymbol{\theta}_{\leq m}}} \mathcal{E}_{m-n}^{b_{m-n}}\left(\left|\delta_{m-n}^{\boldsymbol{b}_{<m-n},\boldsymbol{r}_{\leq m-n},\theta_{m-n}}\right\rangle\left\langle\delta_{m-n}^{\boldsymbol{b}_{<m-n},\boldsymbol{r}_{\leq m-n},\theta_{m-n}}\right| \otimes |+_{\theta_m}\rangle\langle+_{\theta_m}|\right.$$
$$\otimes \ldots \otimes \mathcal{E}_2^{b_2}\left(\left|\delta_2^{\theta_2,r_2}\right\rangle\left\langle\delta_2^{\theta_2,r_2}\right| \otimes |+_{\theta_{2+n}}\rangle\langle+_{\theta_{2+n}}|\right.$$
$$\left.\left. \otimes\, \mathcal{E}_1^{b_1}\left(\left|\delta_1^{\theta_1,r_1}\right\rangle\left\langle\delta_1^{\theta_1,r_1}\right| \otimes |+_{\theta_{1+n}}\rangle\langle+_{\theta_{1+n}}| \otimes |+_{\theta_1}\rangle\langle+_{\theta_1}| \otimes \mathrm{Tr}_A(\rho_{AB})\right)\right)\ldots\right)$$

Notation $\boldsymbol{b}_{<m-n}$ stands for all the elements of $\boldsymbol{b}$ with index less than $m-n$.

Collecting all CPTP maps by commuting them with systems which they do not apply on into a single operator $\mathcal{E}$ and rearranging the terms of the tensor product inside gives:

$$= \sum_{\substack{\boldsymbol{b}_{\leq m-n},\\ \boldsymbol{r}_{\leq m-n},\boldsymbol{\theta}_{\leq m}}} \mathcal{E}^{\boldsymbol{b}_{\leq m-n}}\Big(\bigotimes_{i=m-n}^{m} |+_{\theta_i}\rangle\langle+_{\theta_i}| \bigotimes_{i=n+1}^{m-n-1} (\left|\delta_i^{\boldsymbol{b}_{<i},\boldsymbol{r}_{\leq i},\theta_i}\right\rangle\left\langle\delta_i^{\boldsymbol{b}_{<i},\boldsymbol{r}_{\leq i},\theta_i}\right| \otimes |+_{\theta_i}\rangle\langle+_{\theta_i}|)$$

$$\bigotimes_{i=2}^{n}(\left|\delta_i^{\theta_i,r_i}\right\rangle\left\langle\delta_i^{\theta_i,r_i}\right|) \otimes \left|\delta_1^{\theta_1,r_1}\right\rangle\left\langle\delta_1^{\theta_1,r_1}\right| \otimes |+_{\theta_1}\rangle\langle+_{\theta_1}| \otimes \mathrm{Tr}_A(\rho_{AB})\Big)$$

We introduce the controlled unitary:

$$U = \prod_{n+1\leq i\leq m-n-1, i=1} Z_i(-\delta_i)$$

and rewrite the state as:

$$\sum_{\substack{\boldsymbol{b}_{\leq m-n},\\ \boldsymbol{r}_{\leq m-n},\boldsymbol{\theta}_{\leq m}}} \mathcal{E}^{\boldsymbol{b}_{\leq m-n}}\Big(U^{\dagger}U \bigotimes_{i=m-n}^{m} |+_{\theta_i}\rangle\langle+_{\theta_i}| \bigotimes_{i=n+1}^{m-n-1} (\left|\delta_i^{\boldsymbol{b}_{<i},\boldsymbol{r}_{\leq i},\theta_i}\right\rangle\left\langle\delta_i^{\boldsymbol{b}_{<i},\boldsymbol{r}_{\leq i},\theta_i}\right| \otimes |+_{\theta_i}\rangle\langle+_{\theta_i}|)$$

$$\bigotimes_{i=2}^{n}(\left|\delta_i^{\theta_i,r_i}\right\rangle\left\langle\delta_i^{\theta_i,r_i}\right|) \otimes \left|\delta_1^{\theta_1,r_1}\right\rangle\left\langle\delta_1^{\theta_1,r_1}\right| \otimes |+_{\theta_1}\rangle\langle+_{\theta_1}| U^{\dagger}U \otimes \mathrm{Tr}_A(\rho_{AB})\Big)$$

After applying the innermost unitary and absorbing the outermost into the CPTP-map we have:

$$\sum_{\substack{\boldsymbol{b}_{\leq m-n},\\ \boldsymbol{r}_{\leq m-n},\boldsymbol{\theta}_{\leq m}}} \mathcal{E'}^{\boldsymbol{b}_{\leq m-n}}\Big(\bigotimes_{i=m-n}^{m} |+_{\theta_i}\rangle\langle+_{\theta_i}|$$

$$\bigotimes_{i=n+1}^{m-n-1} \left(\left|\delta_i^{\boldsymbol{b}_{<i},\boldsymbol{r}_{\leq i},\theta_i}\right\rangle\left\langle\delta_i^{\boldsymbol{b}_{<i},\boldsymbol{r}_{\leq i},\theta_i}\right| \otimes \left|+_{-a_i^{'}{}^{\boldsymbol{b}_{<i},\boldsymbol{r}_{<i}}-r_i\pi}\right\rangle\left\langle+_{-a_i^{'}{}^{\boldsymbol{b}_{<i},\boldsymbol{r}_{<i}}-r_i\pi}\right|\right)$$

$$\bigotimes_{i=2}^{n}\left(\left|\delta_i^{\theta_i,r_i}\right\rangle\left\langle\delta_i^{\theta_i,r_i}\right|\right) \otimes \left|\delta_1^{\theta_1,r_1}\right\rangle\left\langle\delta_1^{\theta_1,r_1}\right| \otimes |+_{-a'_1-r_1\pi}\rangle\langle+_{-a'_1-r_1\pi}| \otimes \mathrm{Tr}_A(\rho_{AB})\Big)$$

It is essential for the proof that each term with index i in the tensor products depends only on parameters with index $\leq i$. This allows to break the summations over $\boldsymbol{r}_{\leq m-n}$ and $\boldsymbol{\theta}_{\leq m}$ and calculate them iteratively from left to right, given the following:

$$\sum_{\theta_i} |+_{\theta_i}\rangle\langle+_{\theta_i}| = \frac{I_1}{2}$$

where $I_n = \bigotimes_n I$. Also,

$$\sum_{r_i,\theta_i} \left|\delta_i^{\boldsymbol{r}_{\leq i},\theta_i}\right\rangle\left\langle\delta_i^{\boldsymbol{r}_{\leq i},\theta_i}\right| \otimes \left|+_{-a_i^{'}{}^{\boldsymbol{r}_{<i}}-r_i\pi}\right\rangle\left\langle+_{-a_i^{'}{}^{\boldsymbol{r}_{<i}}-r_i\pi}\right|$$

$$= \sum_{r_i}\Big(\sum_{\theta_i}\Big(\left|a_i^{'}{}^{\boldsymbol{r}_{<i}} + \theta_i + r_i\pi\right\rangle\left\langle a_i^{'}{}^{\boldsymbol{r}_{<i}} + \theta_i + r_i\pi\right|\Big) \otimes \left|+_{-a_i^{'}{}^{\boldsymbol{r}_{<i}}-r_i\pi}\right\rangle\left\langle+_{-a_i^{'}{}^{\boldsymbol{r}_{<i}}-r_i\pi}\right|\Big)$$

$$= \sum_{r_i} \frac{I_3}{2^3} \otimes \left|+_{-a_i^{'}{}^{\boldsymbol{r}_{<i}}-r_i\pi}\right\rangle\left\langle+_{-a_i^{'}{}^{\boldsymbol{r}_{<i}}-r_i\pi}\right|$$

$$= \frac{I_4}{2^4}$$

and

$$\sum_{r_i,\theta_i} \left|\delta_i^{\theta_i,r_i}\right\rangle\left\langle\delta_i^{\theta_i,r_i}\right| = \frac{I_3}{2^3}$$

This procedure will produce the state:

$$\sigma_B = \mathcal{E}'\left(\frac{I_{4m-4n+1}}{2^{4m-4n+1}} \otimes \mathrm{Tr}_A(\rho_{AB})\right) = \mathcal{E}''(\mathrm{Tr}_A(\rho_{AB}))$$

where $\mathcal{E}''$ is some CPTP map. Therefore Definition 8 is satisfied. ◀

B Proof of Theorem 11

Proof. The same notation is used as in Section 3. The first step is to write the state of Alice's system at the end of the execution of the protocol for fixed Bob's behaviour j and choices of Alice ν. We have utilised the fact that all measurements can be moved to the end. Also, we have commuted all Bob's operations to the end (before the measurements) merging them to a single CPTP map. The state of Alice is:

$$B_j(\nu) = \sum_{\boldsymbol{b}} \otimes_{i=k}^{s} \left|+_{\theta_{t_k}+b_{t_k}\pi}\right\rangle\left\langle+_{\theta_{t_k}+b_{t_k}\pi}\right| C^{\boldsymbol{b},\nu_C} |\boldsymbol{b}+\boldsymbol{c^r}\rangle\langle\boldsymbol{b}|$$

$$\mathcal{E}^j\left(\bigotimes_{k=1}^{s}\bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b},\nu}\right| \otimes \rho_{M_k^\nu}\right) |\boldsymbol{b}\rangle\langle\boldsymbol{b}+\boldsymbol{c^r}| C^{\boldsymbol{b},\nu_C\dagger} \otimes_{i=k}^{s} \left|+_{\theta_{t_k}+b_{t_k}\pi}\right\rangle\left\langle+_{\theta_{t_k}+b_{t_k}\pi}\right|$$

where $\left|+_{\theta_{t_k}+b_{t_k}\pi}\right\rangle\left\langle+_{\theta_{t_k}+b_{t_k}\pi}\right|$ are used to define Alice's measurement of the traps which are part of the output state of each round k (if they exist).

To bound the failure probability, observe that projectors orthogonal to $\left|\eta_{t_k}^{\nu_T}\right\rangle$'s vanish, thus we have (where $b' = \{b_i\}_{i\neq t_1\ldots t_s}$):

$$p_{incorrect} = \sum_{\boldsymbol{b'},\nu} p(\nu)\mathrm{Tr}(P_\perp \bigotimes_{k=1}^{s} \left|\eta_{t_k}^{\nu_T}\right\rangle\left\langle\eta_{t_k}^{\nu_T}\right|$$

$$C^{\boldsymbol{b'},\nu_C}|\boldsymbol{b'}+\boldsymbol{c^r}\rangle\langle\boldsymbol{b'}|\mathcal{E}^j\left(\bigotimes_{k=1}^{s}\bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes \rho_{M_k^\nu}\right)|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|C^{\boldsymbol{b'},\nu_C\dagger})$$

We introduce the following unitary, which characterises the correct operation on each subgraph k: $\mathcal{P}_k = \bigotimes_{i=1}^{m-n}(H_{(k-1)m+i}Z_{(k-1)m+i}(\delta_{(k-1)m+i}))E_{G'_k}$.

We can rewrite the failure probability, introducing $\mathcal{P}_k^\dagger\mathcal{P}_k$'s on both sides of the quantum state of the system before the attack, and absorbing the outermost unitaries into the updated CPTP map $\mathcal{E}'^j$:

$$p_{incorrect} = \sum_{\boldsymbol{b'},\nu} p(\nu)\mathrm{Tr}(P_\perp \bigotimes_{k=1}^{s} \left|\eta_{t_k}^{\nu_T}\right\rangle\left\langle\eta_{t_k}^{\nu_T}\right| C^{\boldsymbol{b'},\nu_C}$$

$$|\boldsymbol{b'}+\boldsymbol{c^r}\rangle\langle\boldsymbol{b'}|\mathcal{E}'^j\left(\bigotimes_{k=1}^{s}(\mathcal{P}_k\bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes \rho_{M_k^\nu}\mathcal{P}_k^\dagger)\right)|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|C^{\boldsymbol{b'},\nu_C\dagger})$$

We decompose $\mathcal{E}'^j$ using the following facts: There exist some matrices $\{\chi_v\}$ of dimension $s(4m-3n) \times s(4m-3n)$, with $\sum_v \chi_v\chi_v^\dagger = I$ such that for every density operator ρ:

$\mathcal{E}'^{j}(\rho) = \sum_v \chi_v \rho \chi_v^\dagger$. Also, each χ_v can be decomposed to the Pauli basis: $\chi_v = \sum_i \alpha_{vi}\sigma_i$, with $\sum_{v,i} \alpha_{vi}\alpha_{vi}^* = 1$. Setting $\sigma_{i,k}$ to be the part of σ_i that applies on the qubits $(k-1)m+1 \leq \gamma \leq km$.

$$p_{incorrect} = \sum_{\boldsymbol{b'},\nu,v,i,j} \alpha_{vi}\alpha_{vj}^* p(\nu)\mathrm{Tr}(P_\perp \bigotimes_{k=1}^{s} |\eta_{t_k}^{\nu_T}\rangle\langle\eta_{t_k}^{\nu_T}| C^{\boldsymbol{b'},\nu_C}$$

$$|\boldsymbol{b'}+\boldsymbol{c^r}\rangle\langle\boldsymbol{b'}| \bigotimes_{k=1}^{s}(\sigma_{i,k}\left(\mathcal{P}_k \bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes \rho_{M_k^\nu}\mathcal{P}_k^\dagger\right)\sigma_{j,k})|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|C^{\boldsymbol{b'},\nu_C\dagger}$$

Without loss of generality we can assume that σ_i, σ_j do not change the δ's.

For an arbitrary t_g, the only attacks that give the corresponding term of the sum not equal to zero:

$$P_\perp(C^{\boldsymbol{b'},\nu_C}|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|\sigma_{i,t_g}$$

$$(\mathcal{P}_{t_g}\bigotimes_{i=1}^{m-n}\left|\delta_{(t_g-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(t_g-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes \rho_{M_{t_g}^\nu}\mathcal{P}_{t_g}^\dagger)\sigma_{j,t_g}|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|C^{\boldsymbol{b'},\nu_C\dagger}) \neq 0$$

are those that (i) produce an incorrect measurement result for qubits $(t_g-1)m+1 \leq \gamma \leq t_g m - n$ or (ii) operate non-trivially on qubits $t_g m - n < \gamma \leq t_g m$. We denote this condition by $i \in E_{i,t_g}$ and $j \in E_{j,t_g}$.

We can rewrite the probability by eliminating $P_\perp$ (observing that it applies to a positive operator) and $C^{\boldsymbol{b'},\nu_C}$ (by the cyclical property of the trace):

$$p_{incorrect} \leq \sum_{\nu,v,i\in E_{i,t_g},j\in E_{j,t_g}} \alpha_{vi}\alpha_{vj}^* p(\nu)\prod_{k=1}^{s}\mathrm{Tr}(\left|\eta_{t_k}^{\nu_T}\right\rangle\left\langle\eta_{t_k}^{\nu_T}\right|$$

$$|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|\sigma_{i,k}\left(\mathcal{P}_k\bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes \rho_{M_k^\nu}\mathcal{P}_k^\dagger\right)\sigma_{j,k})$$

We extract a trace over R from $\rho_{M_{t_g}^\nu}$. And extract the sums over $\nu_{C,k}$'s from the general sum, where $\nu_{C,k}$ is the subset of random parameters ν_C that are used for the computation of round r:

$$= \sum_{\nu_T,v,i\in E_{i,t_g},j\in E_{j,t_g}} \alpha_{vi}\alpha_{vj}^* p(\nu_T)\prod_{k=1}^{s}\mathrm{Tr}(\left|\eta_{t_k}^{\nu_T}\right\rangle\left\langle\eta_{t_k}^{\nu_T}\right|$$

$$|\boldsymbol{b'}\rangle\langle\boldsymbol{b'}+\boldsymbol{c^r}|\sigma_{i,k}\left(\mathcal{P}_k\sum_{\nu_{C,k}}(p(\nu_{C,k})\bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes \mathrm{Tr}_R(\rho_{M_k^\nu}))\mathcal{P}_k^\dagger\right)\sigma_{j,k})$$

To take advantage of the blindness property we use the following lemma where the proof is given later.

▶ **Lemma 7** (Blindness (excluding the traps)).

$$\forall k, \sum_{\nu_{C,k}} p(\nu_{C,k})\bigotimes_{i=1}^{m-n}\left|\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{(k-1)m+i}^{\boldsymbol{b'},\nu}\right| \otimes Tr_R(\rho_{M_k^\nu})$$

$$= \frac{I_k^{t_k}}{Tr(I_k^{t_k})} \otimes \left|\delta_{t_k}^{\theta_{t_k},r_{t_k}}\right\rangle\left\langle\delta_{t_k}^{\theta_{t_k},r_{t_k}}\right| \otimes \left|+_{\theta_{t_k}}\right\rangle\left\langle+_{\theta_{t_k}}\right|$$

If $k \neq t_g$, $I_k^{t_k} = \bigotimes_{4m-3n-1} I$ when $km - n < t_k \leq km$ and $I_k^{t_k} = \bigotimes_{4m-3n-4} I$ when $(k-1)m < t_k \leq km - n$. And if $k = t_g$, $I_k^{t_k} = \bigotimes_{4m-3n} I$.

Lemma 7 allows us to simplify the big sum above based on the position of the traps. We also sum over $\boldsymbol{b'}$ since there are no longer any dependencies on it in the sum, obtaining:

$$= \sum_{t_g, v, i\in E_{i,t_g}, j\in E_{j,t_g}} \alpha_{vi}\alpha^*_{vj}\, p(t_g) \prod_{k=1}^{s} \mathrm{Tr}(
\sum_{\substack{km-n<t_k\le km,\\ \theta_{t_k}}} p(t_k,\theta_{t_k}) \left|+_{\theta_{t_k}}\right\rangle\left\langle +_{\theta_{t_k}}\right| \sigma_{i,k}(\frac{\mathcal{I}}{\mathrm{Tr}(\mathcal{I})} \otimes \left|+_{\theta_{t_k}}\right\rangle\left\langle+_{\theta_{t_k}}\right|)\sigma_{j,k}
+ \sum_{\substack{(k-1)m<t_k\le km-n,\\ r_{t_k}}} p(t_k,r_{t_k})|r_{t_k}\rangle\langle r_{t_k}|\sigma_{i,k}(\frac{\mathcal{I}}{\mathrm{Tr}(\mathcal{I})}\otimes|r_{t_k}\rangle\langle r_{t_k}|)\sigma_{j,k})$$

where $\mathcal{I} = \bigotimes_{4m-3n-1} I$ when $k \neq t_g$. And $\mathcal{I} = \bigotimes_{4m-3n} I$ when $k = t_g$.

Note that $\sum_{\theta_{t_k}} \mathrm{Tr}(\left|+_{\theta_{t_k}}\right\rangle\left\langle +_{\theta_{t_k}}\right|\sigma_{i,k}(\frac{\mathcal{I}}{\mathrm{Tr}(\mathcal{I})}\otimes\left|+_{\theta_{t_k}}\right\rangle\left\langle +_{\theta_{t_k}}\right|)\sigma_{j,k})$ is zero if $\sigma_{i,k} \neq \sigma_{j,k}$. The same is true for $\sum_{r_{t_k}} \mathrm{Tr}(|r_{t_k}\rangle\langle r_{t_k}|\sigma_{i,k}(\frac{\mathcal{I}}{\mathrm{Tr}(\mathcal{I})}\otimes|r_{t_k}\rangle\langle r_{t_k}|)\sigma_{j,k})$. Therefore we can only keep those terms where $\sigma_{i,k} = \sigma_{j,k}$ and the failure probability becomes:

$$= \sum_{t_g} \sum_{v, i\in E_{i,t_g}} |\alpha_{vi}|^2 p(t_g) \prod_{k=\{1,\ldots,s\}\setminus t_g} (\sum_{\substack{km-n<t_k\le km,\\ \theta_{t_k}}} p(t_k,\theta_{t_k})(\left\langle +_{\theta_{t_k}}\right|\sigma_{i|t_k}\left|+_{\theta_{t_k}}\right\rangle)^2
+ \sum_{\substack{(k-1)m<t_k\le km-n,\\ r_{t_k}}} p(t_k,r_{t_k})(\langle r_{t_k}|\sigma_{i|t_k}|r_{t_k}\rangle)^2)$$

The rest of the proof is based on a counting argument. For convenience we introduce the following sets for an arbitrary Pauli $\sigma_{i,k}$:

$$\begin{aligned}
A_{i,k} &= \{\gamma \text{ s.t. } \sigma_{i|\gamma} = I \text{ and } (k-1)m+1 \le \gamma \le km\}\\
B_{i,k} &= \{\gamma \text{ s.t. } \sigma_{i|\gamma} = X \text{ and } (k-1)m+1 \le \gamma \le km\}\\
C_{i,k} &= \{\gamma \text{ s.t. } \sigma_{i|\gamma} = Y \text{ and } (k-1)m+1 \le \gamma \le km\}\\
D_{i,k} &= \{\gamma \text{ s.t. } \sigma_{i|\gamma} = Z \text{ and } (k-1)m+1 \le \gamma \le km\}
\end{aligned}$$

and use the superscript O to denote subsets subject to the constraint $km \ge \gamma \ge km - n + 1$.

The failure probability is then:

$$= \sum_{t_g} \sum_{v, i\in E_{i,t_g}} |\alpha_{vi}|^2 \frac{1}{s} \prod_{k=\{1,\ldots,s\}\setminus t_g} ((\frac{1}{8m}(8|A^O_{i,k}| + 4|B^O_{i,k}| + 4|C^O_{i,k}|) +
\frac{1}{2m}(2|A_{i,k} \setminus A^O_{i,k}| + 2|D_{i,k} \setminus D^O_{i,k}|))$$

Merging the terms:

$$= \sum_{t_g} \sum_{v, i\in E_{i,t_g}} |\alpha_{vi}|^2 \frac{1}{s} \prod_{k=\{1,\ldots,s\}\setminus t_g} \frac{1}{2m}(2|A_{i,k}| + |B^O_{i,k}| + |C^O_{i,k}| + 2|D_{i,k} \setminus D^O_{i,k}|)$$

Using the fact that for every k, $|A_{i,k}| + |B_{i,k}| + |C_{i,k}| + |D_{i,k}| = m$:

$$\le \sum_{t_g} \sum_{v, i\in E_{i,t_g}} |\alpha_{vi}|^2 \frac{1}{s} \prod_{k=\{1,\ldots,s\}\setminus t_g} \frac{1}{2m}(2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$$

The conditions $i \in E_{i,t_g}$ that we obtained at the first part of the proof are translated to $|B_{i,t_g}| + |C_{i,t_g}| + |D^O_{i,t_g}| \geq 1$. In order to be able to use these conditions we need to rewrite the formula. First we expand it:

$$= \frac{1}{s}(\sum_{v,i \in E_{i,1}} |\alpha_{vi}|^2 \prod_{k=\{2,3,\ldots,s\}} \frac{1}{2m}(2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$$

$$+ \sum_{v,i \in E_{i,2}} |\alpha_{vi}|^2 \prod_{k=\{1,3,4,\ldots,s\}} \frac{1}{2m}(2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$$

$$\ldots + \sum_{v,i \in E_{i,d}} |\alpha_{vi}|^2 \prod_{k=\{1,2,\ldots,s-1\}} \frac{1}{2m}(2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|))$$

We denote the product term $\prod_{k=\{1,2,3,\ldots,s\}\setminus z} \frac{1}{2m}(2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$ as $P_{i,z}$. We also denote each set $\{E^*_{i,1} \cap E^*_{i,2} \cap \ldots \cap E^*_{i,s}\}$, where each term $E^*_{i,w}$ is either $E_{i,w}$ or its complement, $E^C_{i,w}$, depending on whether the w-th value of a binary vector $\boldsymbol{y}$ (size s) is 1 or 0 respectively, as $W_{i,\boldsymbol{y}}$. Then we have:

$$= \frac{1}{s}(\sum_{\boldsymbol{y} \setminus (0\ldots0)} \sum_{i \in W_{i,\boldsymbol{y}}, v} (|\alpha_{vi}|^2 \sum_{\{z : y_z = 1\}} P_{i,z}))$$

Let the function $\#\boldsymbol{y}$ give the number of positions i such that y_i=1.

$$= \frac{1}{s}(\sum_{k=1}^{s} \sum_{\{\boldsymbol{y} : \#\boldsymbol{y} = k\}} \sum_{i \in W_{i,\boldsymbol{y}}, v} (|\alpha_{vi}|^2 \sum_{\{z : y_z = 1\}} P_{i,z}))$$

We separately consider the following term for any arbitrary $\boldsymbol{y}$ with $\#\boldsymbol{y} = r$.

$$\sum_{i \in W_{i,\boldsymbol{y}}} (|\alpha_{vi}|^2 \sum_{\{z : y_z = 1\}} P_{i,z})$$

The condition $i \in W_{i,\boldsymbol{y}}$ means that the following conditions hold together: $\{|B_{i,w}| + |C_{i,w}| + |D^O_{i,w}| \geq 1 : y_w = 1\}$,$\{|B_{i,w}| + |C_{i,w}| + |D^O_{i,w}| = 0 : y_w = 0\}$. We expand:

$$= \sum_{i \in W_{i,\boldsymbol{y}}} (|\alpha_{vi}|^2 \sum_{\{z : y_z = 1\}} \prod_{k=\{1,2,3,\ldots,s\}\setminus z} \frac{1}{2m}(2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$$

$$= \sum_{i \in W_{i,\boldsymbol{y}}} (|\alpha_{vi}|^2 \sum_{\{z : y_z = 1\}} \prod_{\{k : y_k = 1, k \neq z\}} \frac{1}{2m}(2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$$

$$\prod_{\{k : y_k = 0\}} \frac{1}{2m}(2m - |B_{i,k}| - |C_{i,k}| - |D^O_{i,k}|)$$

And by using the above conditions:

$$\leq \sum_{i \in W_{i,\boldsymbol{y}}} (|\alpha_{vi}|^2 \sum_{\{z : y_z = 1\}} \prod_{\{k : y_k = 1, k \neq z\}} \frac{1}{2m}(2m - 1) \prod_{\{k : y_k = 0\}} \frac{1}{2m}(2m)$$

$$= \sum_{i \in W_{i,\boldsymbol{y}}} (|\alpha_{vi}|^2 \sum_{\{z : y_z = 1\}} \left(\frac{2m - 1}{2m}\right)^{r-1}$$

$$= \sum_{i \in W_{i,\boldsymbol{y}}} |\alpha_{vi}|^2 r \left(\frac{2m - 1}{2m}\right)^{r-1}$$

Therefore the bound of our failure probability will be:

$$\begin{aligned}
p_{\text{incorrect}} &\leq \frac{1}{s}(\sum_{k=1}^{s} \sum_{\{\boldsymbol{y}:\#\boldsymbol{y}=k\}} \sum_{i\in W_{i,\boldsymbol{y}},v} |\alpha_{vi}|^2 k \left(\frac{2m-1}{2m}\right)^{k-1}) \\
&= \frac{1}{s}(\sum_{k=1}^{s} k \left(\frac{2m-1}{2m}\right)^{k-1} \sum_{\{\boldsymbol{y}:\#\boldsymbol{y}=k\}} \sum_{i\in W_{i,\boldsymbol{y}},v} |\alpha_{vi}|^2) \\
&= \frac{1}{s}(\sum_{k=1}^{s} c_k k \left(\frac{2m-1}{2m}\right)^{k-1})
\end{aligned}$$

where $c_k = \sum_{\{\boldsymbol{y}:\#\boldsymbol{y}=k\}} \sum_{i\in W_{i,\boldsymbol{y}},v} |\alpha_{vi}|^2$
subject to conditions:

$$\sum_{k=1}^{s} c_k \leq 1 \tag{15}$$

and

$$\forall k : c_k \geq 0 \tag{16}$$

◀

Proof of Lemma 7. First we define state $|q_i\rangle$ as:

$$\begin{aligned}
&i \in D && |q_i\rangle \equiv |d_i\rangle \\
&i \notin D && |q_i\rangle \equiv (\prod_{\{j:j\sim i, j\in D\}} Z^{d_j})|+_{\theta_i}\rangle
\end{aligned}$$

By substituting $\rho_{M_k^{\nu}}$'s and taking the trace over R:
If $k \neq t_g$ the state becomes:

$$\begin{aligned}
\sum_{\nu_{C,k}} p(\nu_{C,k})(\bigotimes_{i=km-n+1}^{km} |q_i\rangle\langle q_i| \bigotimes_{i=(k-1)m+n+1}^{km-n} &\left(\left|\delta_i^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_i^{\boldsymbol{b'},\nu}\right| \otimes |q_i^{\nu}\rangle\langle q_i^{\nu}|\right) \\
&\bigotimes_{i=1}^{2} \left(\left|\delta_{p_{i,k}}^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_{p_{i,k}}^{\boldsymbol{b'},\nu}\right| \otimes \left|q_{p_{i,k}}^{\nu}\right\rangle\left\langle q_{p_{i,k}}^{\nu}\right|\right) \otimes I_{4(n-2)}/2^{4(n-2)})
\end{aligned}$$

where $\left|q_{p_{i,k}}^{\nu}\right\rangle$ denote the first layer pure qubits (a maximum of two) of the k-th graph state, used as padding (dummies) or trap and their positions are defined as: $1 + (k-1)m \leq \{p_{1,k}, p_{2,k}\} \leq n + (k-1)m$.

Otherwise, if $k = t_g$ the state becomes:

$$\begin{aligned}
\sum_{\nu_{C,k}} p(\nu_{C,k})(\bigotimes_{i=t_g m-n+1}^{t_g m} |q_i\rangle\langle q_i| \bigotimes_{i=(t_g-1)m+n+1}^{t_g m-n} &\left(\left|\delta_i^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_i^{\boldsymbol{b'},\nu}\right| \otimes |q_i^{\nu}\rangle\langle q_i^{\nu}|\right) \\
&\otimes \left|\delta_u^{\theta_u,r_u}\right\rangle\left\langle\delta_u^{\theta_u,r_u}\right| \otimes \left|q_u^{\theta_u}\right\rangle\left\langle q_u^{\theta_u}\right| \otimes I_{4(w-1)}/2^{4(w-1)})
\end{aligned}$$

where $u = (t_g-1)m+1$ is the position of the single pure qubit of the input to the DQC1-MBQC computation.

An implicit assumption was that all δ's that are used to implement the measurements of maximally mixed inputs are maximally mixed states themselves, without any loss of generality.

We define a new controlled unitary:

$$\mathcal{P}'_k = \left(\prod_{\{i: i \notin D, (k-1)m+1 \leq i \leq km-n\}} Z_i(-\delta_i) \right) \prod_{\{i: i \notin D_k\}} \prod_{\{j: j \sim i, j \in D_k\}} Z_i(d_j) \qquad (17)$$

where D_k denotes the set of dummies of subgraph G'_k.

Using this unitary we rewrite the state. If $k \neq t_g$ it becomes:

$$\sum_{\nu_{C_k}} p(\nu_{C,k}) \mathcal{P}'^{\dagger} \mathcal{P}' (\bigotimes_{i=km-n+1}^{km} |q_i\rangle\langle q_i| \bigotimes_{i=(k-1)m+n+1}^{km-n} \left(\left|\delta_i^{\boldsymbol{b}',\nu}\right\rangle \left\langle \delta_i^{\boldsymbol{b}',\nu}\right| \otimes |q_i^{\nu}\rangle\langle q_i^{\nu}| \right)$$
$$\bigotimes_{i=1}^{2} \left(\left|\delta_{p_{i,k}}^{\boldsymbol{b}',\nu}\right\rangle \left\langle \delta_{p_{i,k}}^{\boldsymbol{b}',\nu}\right| \otimes \left|q_{p_{i,k}}^{\nu}\right\rangle \left\langle q_{p_{i,k}}^{\nu}\right| \right) \otimes I_{4(n-2)}/2^{4(n-2)}) \mathcal{P}'^{\dagger} \mathcal{P}'$$

Otherwise:

$$\sum_{\nu_{C,k}} p(\nu_{C,k}) \mathcal{P}'^{\dagger} \mathcal{P}' \bigotimes_{i=t_g m-n+1}^{t_g m} |q_i\rangle\langle q_i| \bigotimes_{i=(t_g-1)m+n+1}^{t_g m-n} \left(\left|\delta_i^{\boldsymbol{b}',\nu}\right\rangle \left\langle \delta_i^{\boldsymbol{b}',\nu}\right| \otimes |q_i^{\nu}\rangle\langle q_i^{\nu}| \right)$$
$$\otimes \left|\delta_u^{\theta_u, r_u}\right\rangle\left\langle \delta_u^{\theta_u, r_u}\right| \otimes \left|q_u^{\theta_u}\right\rangle \left\langle q_u^{\theta_u}\right| \otimes I_{4(w-1)}/2^{4(w-1)}) \mathcal{P}'^{\dagger} \mathcal{P}'$$

After applying the innermost unitary, if $k \neq t_g$:

$$\sum_{\nu_{C,k}} p(\nu_{C,k}) \mathcal{P}'^{\dagger} (\bigotimes_{i=km-n+1}^{km} |q'_i\rangle\langle q'_i| \bigotimes_{i=(k-1)m+n+1}^{km-n} \left(\left|\delta_i^{\boldsymbol{b}',\nu}\right\rangle \left\langle \delta_i^{\boldsymbol{b}',\nu}\right| \otimes \left|q_i^{'\nu}\right\rangle \left\langle q_i^{'\nu}\right| \right)$$
$$\bigotimes_{i=1}^{2} \left(\left|\delta_{p_{i,k}}^{\boldsymbol{b}',\nu}\right\rangle \left\langle \delta_{p_{i,k}}^{\boldsymbol{b}',\nu}\right| \otimes \left|q_{p_{i,k}}^{'\nu}\right\rangle \left\langle q_{p_{i,k}}^{'\nu}\right| \right) \otimes I_{4(n-2)}/2^{4(n-2)}) \mathcal{P}'$$

where state $|q'_i\rangle$ is defined as:

$$\begin{array}{rl} i \in D & |q'_i\rangle \equiv |d_i\rangle \\ i \notin D, \forall k : km \geq i \geq km-n+1 & |q'_i\rangle \equiv |+_{\theta_i}\rangle \\ i \notin D, \forall k : km-n \geq i \geq (k-1)m+1 & |q'_i\rangle \equiv \left|+_{-a_i^{''\,\boldsymbol{b}',\boldsymbol{r}<i}-r_i\pi}\right\rangle \end{array}$$

Otherwise, if $k = t_g$:

$$\sum_{\nu_{C,k}} p(\nu_{C,k}) \mathcal{P}'^{\dagger} (\bigotimes_{i=t_g m-n+1}^{t_g m} |q'_i\rangle\langle q'_i| \bigotimes_{i=(t_g-1)m+n+1}^{t_g m-n} \left(\left|\delta_i^{\boldsymbol{b}',\nu}\right\rangle \left\langle \delta_i^{\boldsymbol{b}',\nu}\right| \otimes \left|q_i^{'\nu}\right\rangle \left\langle q_i^{'\nu}\right| \right)$$
$$\otimes \left|\delta_u^{\theta_u, r_u}\right\rangle\left\langle \delta_u^{\theta_u, r_u}\right| \otimes \left|q_u^{'\theta_u}\right\rangle \left\langle q_u^{'\theta_u}\right| \otimes I_{4(w-1)}/2^{4(w-1)}) \mathcal{P}'$$

It is essential for the proof that each term with index i in the tensor product depends only on parameters with index $\leq i$ and the term with index $(t_g-1)m+1$ (input qubit) and the trap qubit and its measurement angle (if it is not an output) depend only on their own parameters. This allows to break the summations and calculate them iteratively from left to right, given the following:

$$\sum_{d_i} p(d_i)|d_i\rangle\langle d_i| = \frac{I}{2}$$

$$\sum_{\theta_i} p(\theta_i)|+_{\theta_i}\rangle\langle +_{\theta_i}| = \frac{I}{2}$$

$$\sum_{\theta_i,r_i,d_i} p(\theta_i, r_i, d_i)\left|\delta_i^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_i^{\boldsymbol{b'},\nu}\right| \otimes |d_i\rangle\langle d_i| = \frac{I_4}{2^4}$$

$$\begin{aligned}
&\sum_{\theta_i,r_i} p(\theta_i, r_i)\left|\delta_i^{\boldsymbol{b'},\nu}\right\rangle\left\langle\delta_i^{\boldsymbol{b'},\nu}\right| \otimes \left|+_{-a_i^{''\ \boldsymbol{b'},\boldsymbol{r}_{<i}}-r_i\pi}\right\rangle\left\langle+_{-a_i^{''\ \boldsymbol{b'},\boldsymbol{r}_{<i}}-r_i\pi}\right| \\
&= \sum_{r_i} p(r_i)\left(\sum_{\theta_i} p(\theta_i)\left|a_i^{''\ \boldsymbol{b'},\boldsymbol{r}_{<i}} + \theta_i + r_i\pi\right\rangle\left\langle a_i^{''\ \boldsymbol{b'},\boldsymbol{r}_{<i}} + \theta_i + r_i\pi\right|\right) \\
&\qquad\qquad \otimes\left|+_{-a_i^{''\ \boldsymbol{b'},\boldsymbol{r}_{<i}}-r_i\pi}\right\rangle\left\langle+_{-a_i^{''\ \boldsymbol{b'},\boldsymbol{r}_{<i}}-r_i\pi}\right| \\
&= \sum_{r_i} p(r_i)\frac{I_3}{2^3} \otimes \left|+_{-a_i^{''\ \boldsymbol{b'},\boldsymbol{r}_{<i}}-r_i\pi}\right\rangle\left\langle+_{-a_i^{''\ \boldsymbol{b'},\boldsymbol{r}_{<i}}-r_i\pi}\right| \\
&= \frac{I_4}{2^4}
\end{aligned}$$

where $I_n = \bigotimes_n I$. The last step was possible because each corrected computation angle a_i'' depends only on past r's.

And finally (for $u = (t_g - 1)m + 1$),

$$\begin{aligned}
&\sum_{\theta_u,r_u} p(\theta_u, r_u)|\delta_u^{\theta_u,r_u}\rangle\langle\delta_u^{\theta_u,r_u}| \otimes \left|+_{-a_u'-r_u\pi}\right\rangle\left\langle+_{-a_u'-r_u\pi}\right| \\
&= \sum_{r_u} p(r_u)\left(\sum_{\theta_u} p(\theta_u)\left|a_u^{'} + \theta_u + r_u\pi\right\rangle\left\langle a_i^{'} + \theta_u + r_u\pi\right|\right) \\
&\qquad\qquad \otimes\left|+_{-a_u'-r_u\pi}\right\rangle\left\langle+_{-a_u'-r_u\pi}\right| \\
&= \frac{I_4}{2^4}
\end{aligned}$$

For $k \neq t_g$, if $km \geq t_k \geq km - n + 1$ the above procedure will eventually give:

$$\begin{aligned}
&\mathcal{P}'^{\dagger}(\frac{I_{4m-3n-1}}{2^{4m-3n-1}} \otimes \left|+_{\theta_{t_k}}\right\rangle\left\langle+_{\theta_{t_k}}\right|)\mathcal{P}' \\
&= \frac{I_{4m-3n-1}}{2^{4m-3n-1}} \otimes \left|+_{\theta_{t_k}}\right\rangle\left\langle+_{\theta_{t_k}}\right|
\end{aligned}$$

If $km - n \geq t_k \geq (k-1)m + 1$ the above procedure will eventually give:

$$\begin{aligned}
&\mathcal{P}'^{\dagger}(\frac{I_{4m-3n-4}}{2^{4m-3n-4}} \otimes |\delta_{t_k}^{\nu_T}\rangle\langle\delta_{t_k}^{\nu_T}| \otimes \left|+_{r_{t_k}\pi}\right\rangle\left\langle+_{r_{t_k}\pi}\right|)\mathcal{P}' \\
&= \frac{I_{4m-3n-4}}{2^{4m-3n-4}} \otimes |\delta_{t_k}^{\nu_T}\rangle\langle\delta_{t_k}^{\nu_T}| \otimes \left|+_{\theta_{t_k}}\right\rangle\left\langle+_{\theta_{t_k}}\right|
\end{aligned}$$

And for $k = t_g$ the result will be: $\bigotimes_{4m-3n} I$, which concludes the proof. ◀

Device-independent Randomness Extraction for Arbitrarily Weak Min-entropy Source

Jan Bouda[1,2,3], Marcin Pawłowski[4,5], Matej Pivoluska[1], and Martin Plesch[1,6]

1 Faculty of Informatics, Masaryk University
Botanická 68a, 602 00 Brno, Czech Republic
bouda@fi.muni.cz

2 Física Teórica: Informació i Fenómens Quántics Universitat Autónoma de Barcelona
08193 Bellaterra (Barcelona), Spain

3 LIQUID: Lepanto Institute for Quantum Information and Decoherence
Carrer de Lepant 307, 08025 Barcelona, Spain

4 Instytut Fizyki Teoretycznej i Astrofizyki, Uniwersytet Gdański
PL-80-952 Gdańsk, Poland
dokmpa@univ.gda.pl

5 School of Mathematics, University of Bristol
Bristol BS8 1TW, United Kingdom

6 Institute of Physics, Slovak Academy of Sciences
Bratislava, Slovakia
martin.plesch@savba.sk

Abstract

In this paper we design a protocol to extract random bits with an arbitrarily low bias from a single arbitrarily weak min-entropy block source in a device independent setting. The protocol employs Mermin devices that exhibit super-classical correlations. Number of devices used scales polynomially in the length of the block n, containing entropy of at least two bits. Our protocol is robust, it can tolerate devices that malfunction with a probability dropping polynomially in n at the cost of constant increase of the number of devices used.

1998 ACM Subject Classification G.3 Random Number Generation

Keywords and phrases Randomness Extraction, Device Independence

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.205

1 Introduction

High quality randomness is a very useful resource in many computation and cryptographic tasks. In fact it has been shown that many protocols, including quantum ones, vitally require perfect randomness for their security [1, 2].

Unfortunately, even though we cannot fully predict certain processes it is very difficult to argue that they produce *perfect randomness* – independent and unbiased bits. The problem of imperfect randomness has a long history in classical computer science and long line of research was devoted to randomness extraction – algorithms to transform imperfect sources of randomness into (close to) perfect ones [3].

The drawback of randomness extractors are twofold. Firstly, extractors typically require at least two independent sources of (imperfect) randomness. Worse still, even imperfect randomness of classical processes has to be assumed, because in principle classical physics is

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 205–211
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

deterministic. Quantum physics, with its intrinsic randomness allows us, in theory, to drop the second assumption. Preparation of a pure state and measurement in its complementary basis will yield a perfectly random result. In practice, however, we are replacing the assumption of randomness by yet another assumption – perfect control of quantum devices. This assumption is also very problematic, as we have learned in case of quantum key distribution [4].

Luckily enough, thanks to Bell-type experiments, it is possible to certify by classical procedures that quantum processes are being observed and therefore intrinsic randomness is being produced. This is the basic idea behind device independent randomness extraction. Effectively, we are exchanging the assumption of independent randomness of the second source by a much weaker assumption – validity of quantum mechanics. Alternatively, one can view device-independent randomness extraction as quantum protocol for extracting randomness from a single weak source – a task that is classically impossible.

In this paper we work with (n,k)block min-entropy random sources. These are sources with n-bit blocks of output with guaranteed min-entropy k. Such a source can be modeled as a sequence of n-bit random variables $X_1, X_2, \ldots,$ such that

$$\begin{aligned}&\forall x_1,\ldots,x_{i-1} \in \{0,1\}^n, \forall e \in \mathcal{I}(E), \\ &H_\infty(X_i|X_{i-1}=x_{i-1},\ldots,X_1=x_1,E=e) \geq k,\end{aligned} \tag{1}$$

where E is a random variable describing all adversary's information about the source and $\mathcal{I}(E)$ is it's image. Therefore, each new block has high min-entropy, even conditioned on the previous ones and any information of the adversary. This is a generalization of Santha-Vazirani sources [5], which can be viewed as block sources with $n=1$.

Note that the task of transforming a single block source into a fully random bit is known to be impossible [3]. Furthermore, it is impossible to turn a block source with $n>1$ into Santha-Vazirani source, therefore we cannot use existing randomness extraction protocols [6, 7, 8, 9].

It is also worth to note that similar results were independently obtained by Chung, Shi and Wu [10]. The main difference between the two results is that we work with min-entropy block sources, while their results hold also for general min-entropy sources.

2 Device Independent Concept and Mermin Inequality

In this paper we use the three-partite Mermin inequality. Let's consider three non–communicating boxes, each of them having a single bit input and a single bit output. Let us denote the input bits of the respective boxes by X, Y and Z and the corresponding output bits A, B and C. Input bits are correlated and it holds that $XYZ \in \{111, 100, 010, 001\}$. The inputs are simultaneously passed to all boxes, so each box only knows it's input. The value v of the Mermin term is a function of the 4 conditional probabilities defined by the behavior of the device and of the probability distribution on inputs

$$\begin{aligned}v =& P(A\oplus B\oplus C=1|XYZ=111)P(XYZ=111)+\\ &+P(A\oplus B\oplus C=0|XYZ=100)P(XYZ=100)+\\ &+P(A\oplus B\oplus C=0|XYZ=010)P(XYZ=010)+\\ &+P(A\oplus B\oplus C=0|XYZ=001)P(XYZ=001).\end{aligned} \tag{2}$$

In particular, for the uniform input distribution we set $P(XYZ=111) = P(XYZ=010) = P(XYZ=001) = P(XYZ=100) = \frac{1}{4}$ and denote the Mermin term by v_u.

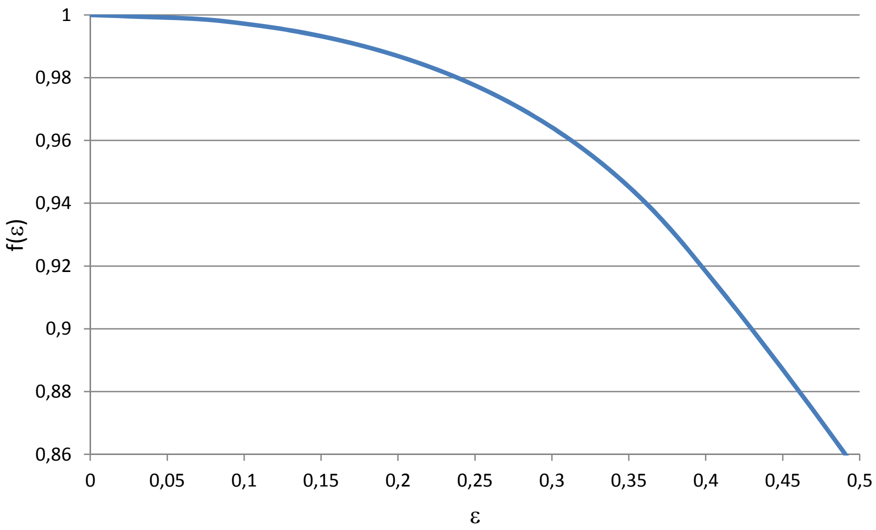


Figure 1 Depicted is the value of Mermin variable $v = f(\varepsilon)$ needed to certify the bias of the output bit to be at most ε.

Assuming the uniform distribution on all four inputs, the maximal value of v_u achievable by a classical device [11] is $\frac{3}{4}$ (thus the Mermin inequality reads $v_u \leq \frac{3}{4}$) and there exists a classical device that can make any 3 conditional probabilities simultaneously equal to 1. With the use of quantum mechanics we can achieve $v_u = 1$ and satisfy perfectly all 4 conditional probabilities using the tripartite GHZ state $\frac{1}{\sqrt{2}}(|000\rangle + |111\rangle)$ and measuring σ_X (σ_Y) when receiving 0 (1) on input.

The beautiful property of the Mermin inequality is that the violation v gives us directly the probability that the device passes a specific test

$$A + B + C = X \cdot Y \cdot Z, \tag{3}$$

where addition and product are both taken modulo 2. The probability of failing the test is therefore $1 - v$.

Mironowicz and Pawlowski [9] showed the following result: Take a linearly ordered sequence of ℓ Mermin devices $D_1 \dots D_\ell$ (ℓ being arbitrary) that have uniform distribution on inputs, and each device knows inputs and outputs of its predecessors, but devices cannot signal to its predecessors. Let us assume that the inputs of devices are described by random variables $XYZ_1, \dots, XYZ_\ell$, and the outputs by $ABC_1, \dots, ABC_\ell$. Then there exists a function $f(\varepsilon)$ such that if the value of the Mermin term (2) using uniform inputs is at least $v_u \geq f(\varepsilon)$, then the output bit A_ℓ has a bias at most ε conditioned on the input and output of all its predecessors and the adversarial knowledge. This function can be lower bounded by a Semi-Definite Program (SDP) using any level of the hierarchy introduced in [12]. By using the second level of the hierarchy one can obtain the bound on $f(\varepsilon)$ as a function of ε shown in Fig. 1.

We can set $\ell = 1$ (having just a single device) and get the lower bound on the detection probability of producing a bit biased by more than ε, which is greater than $1 - f(\varepsilon)$. Our protocol uses many devices, which are forbidden to communicate at all, therefore they can be ordered arbitrarily and thus this limit holds for all of these devices simultaneously.

3 Single-round protocol

In the rest of our analysis we will be working with (n,k) block sources for an arbitrary n and $k \geq 2$. This is to simplify the explanation, since by taking $\lceil \frac{2}{k'} \rceil$ blocks of an arbitrary (n',k') source with $k' > 0$ we get a (n,k) source with $n = \lceil \frac{2}{k'} \rceil n'$ and $k = \lceil \frac{2}{k'} \rceil k' \geq 2$.

Let us start with a min-entropy $(n,2)$ source (recall that (n,k) source with $k > 2$ is also an $(n,2)$ source) and define $N = 2^n$. Let $H = \{h_1,$
$dots, h_m\}$ be a family of hash functions s.t. $h_i : \{0,\dots,N-1\} \to \{0,1,2,3\}$. Each hash-function h_i is used to provide input for a Mermin-type device D_i, where outputs of the function $0,1,2,3$ identify $111,100,010,001$ inputs for the device.

We want to construct H with the property that for every 4-element set $S \subseteq \{0,\dots,N-1\}$ there exist at least one hash function $h \in H$ such that $h(S) = \{0,1,2,3\}$. This is trivially satisfied for the set of all possible hashing functions $H_{full} = \{0,1,2,3\}^N$, however, such a class of functions with its 4^N elements is impractically large. There exists a construction of such class of hash functions with logarithmic number of functions in N (see [13]), thus the number of devices needed scales polynomially with the length of the sequence n. We also stress that for large n one hash function covers as many as 9% of all four-tuples, independently on n. So the size of an optimal set of hash functions might not depend on n at all. Let us denote $m = |H|$. The protocol works as follows:

1. Obtain a (weakly) random n bit string r from the (n,k) block source.
2. Into each device D_i input the 3 bit string r_i chosen from set $\{111,100,010,001\}$ – each one corresponding to one of the possible outputs of $h_i(r)$ – and obtain the outputs A_i, B_i and C_i.
3. Verify whether for each device D_i the condition $X_i + Y_i + Z_i = A_i \cdot B_i \cdot C_i$ holds. If this is not true, abort the protocol.
4. Output $b = \bigoplus_{i=1}^m A_i$.

The protocol is depicted in the Fig. 2.

Let us now examine the properties of the bit b_i. First consider only flat $(n,2)$ distributions. Recall that these are exactly distributions that are uniform on 4-element subsets of the sample space. Our construction of the class H of hash functions assures that for any flat probability distribution there is a function $h_j \in H$ and the corresponding device D_j such that inputs of D_j (hashed by h_j) are uniform. Although output bits A_i are not independent in general, as most of them can be produced by fully deterministic strategies, (1) together with the arbitrary ordering we can impose on devices $\{D_i\}_{i=1}^m$ we have that if the adversary wants to bias (conditioned on the inputs and outputs of other devices) A_j by amount greater than ε, she must risk getting caught with probability at least $1 - f(\varepsilon)$. Therefore A_j is partially independent of other $\{A_i | i \neq j\}$, and the output of the round b is biased by at most ε with probability at least $1 - f(\varepsilon)$.

The set of all $(n,2)$ distributions is convex and flat distributions are exactly all extremal points of this convex set [14]. Thus any $(n,2)$ distribution d can be expressed as a convex combination of at most N $(n,2)$ flat distributions d_i (Caratheodory theorem) as $d = \sum_{i=1}^N p_i d_i$ for some $p_i \geq 0$, $\sum_{i=1}^N p_i = 1$. The lower bound on probability that the adversary is not detected is given by the successful cheating probabilities when using flat distribution $d_i \in \{d_i\}_{i=1}^N$ averaged through the probability distribution on these flat distributions

$$v_u \leq \sum_{i=1}^N p_i P(\text{not detected} | d_i) \leq f(\varepsilon) \textstyle\sum_{i=1}^N p_i \,. \tag{4}$$

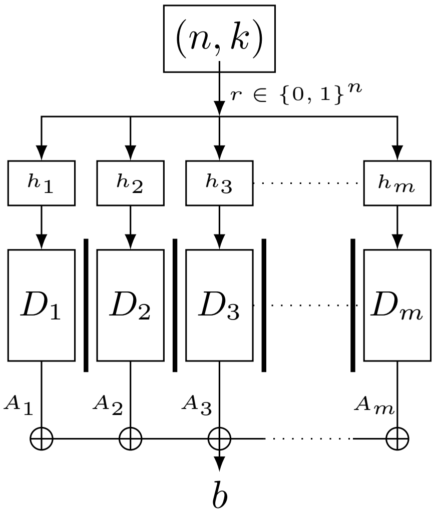


Figure 2 Depiction of a single round protocol. Bit string drawn from the flat random source is hashed into m inputs into Mermin devices so that at least one device receives all four inputs with non-zero probability. This guarantees at least one result almost perfectly random with high probability, which holds also for the product of individual results.

Thus the upper bound $v_u \leq f(\varepsilon)$ holds for non-flat distributions as well.

To summarize this part, having an (n,k) source with $k \geq 2$, with a single round of a protocol, we can produce a single bit that is biased at most by ε with a certainty of $1 - f(\varepsilon)$.

4 Multiple-round protocol

Let us state the most general task: we have an (n,k) block source with arbitrary n and $k \geq 2$. We would like to produce a bit that is biased by no more than ε with certainty of at least $1 - \delta$.

If the one-round version does not meet these parameters, we will repeat the whole protocol l times. By using new devices and new outputs of the block source, each of the runs j will produce a bit b_j that is biased by ε from perfectly random bit conditioned on all the previous bits $\{b_i | i < j\}$ up to a probability $f(\varepsilon)$. Thus, in order to achieve the bias of the output bit

$$b = \bigoplus_{j=1}^{l} b_j \tag{5}$$

of at least ε, all bits b_i has to have at least this bias. Therefore, after l rounds, the probability of the adversary not being detected will be upper bounded by $f(\varepsilon)^l$. Note that the product form does not come from the fact that the detection probabilities are independent (they are not). This is a product of a chain of conditional probabilities. Recall that the bound $f(\varepsilon)$ holds conditioned on any inputs and outputs of the previous devices (in an arbitrarily ordering that respects the causality). Thus choosing

$$l > \frac{\log \delta}{\log f(\varepsilon)} \tag{6}$$

will guarantee the fulfillment of the conditions for the parameters ε and δ.

Summing up, with an (n, k) block source and

$$O\left(\frac{\log \delta}{\log f(\varepsilon)} Poly\left[n\left\lceil\frac{2}{k}\right\rceil\right]\right) \tag{7}$$

Mermin devices we can produce a single random bit with bias smaller than ε with probability larger than $1-\delta$. For producing more bits we simply repeat the whole procedure: all the bits produced will have bias smaller than ε conditioned on the bits produced so far, with linear scaling of the resources.

5 Robustness

Aborting the protocol after even a single mistake of the devices is certainly highly impractical from the implementation point of view. Therefore we expand our analysis to a situation where we tolerate certain noise on the devices, which would manifest itself by occasional failing of the test condition even for honest devices. More specifically, we shall tolerate a certain fraction of the devices to malfunction without aborting the protocol.

In more technical version of this work [13] we show, that if we tolerate

$$\frac{(1-f(\varepsilon))}{2} l \tag{8}$$

devices to fail in the whole protocol and want to achieve security parameters ε, δ we can do so by increasing

$$l > \frac{8 \ln \delta}{f(\varepsilon) - 1} . \tag{9}$$

This translates into increasing the number of rounds of the protocol comparing to the case of ideal devices by a factor of $\frac{8 \ln(f(\varepsilon))}{f(\varepsilon)-1}$. For small ε the parameter $f(\varepsilon)$ approaches 1 and the multiplication factor saturates by 8.

On the other hand we also show that for honest but faulty devices with individual failure probability bounded by

$$\frac{(1-f(\varepsilon))}{4m}, \tag{10}$$

the probability of aborting the protocol decreases exponentially with the number of protocol rounds l.

6 Conclusion

In this paper we have introduced a protocol that extracts weak randomness obtained from a min-entropy source in the device independent setting. The protocol works for arbitrarily weak block min-entropy sources with a reasonable scaling of the number of devices. Our protocol is also robust, as it allows tolerating some fraction of malfunctioning devices at the cost of a constant increase of the number of devices used.

Acknowledgements. Authors thank P. Horodecki, A. Winter, and S. Massar for insightful and stimulating discussions and Piotr Mironowicz for supplying the raw data for Fig. 1. JB, MP2 and MP3 acknowledge the support of the Czech Science Foundation GA CR project P202/12/1142 and support of the EU FP7 under grant agreement no 323970 (RAQUEL). MP3 acknowledges VEGA 2/0072/12. JB acknowledges support by the European Research Council through Advanced Grant "IRQUAT". MP1 acknowledges FNP TEAM, NCN grant 2013/08/M/ST2/00626 and ERC QOLAPS.

References

1 Jan Bouda, Matej Pivoluska, Martin Plesch, and Colin Wilmott. Weak randomness seriously limits the security of quantum key distribution. *Phys. Rev. A*, 86:062308, 2012.

2 Marcus Huber and Marcin Pawłowski. Weak randomness in device-independent quantum key distribution and the advantage of using high-dimensional entanglement. *Phys. Rev. A*, 88:032309, 2013.

3 Ronen Shaltiel. An introduction to randomness extractors. In *Automata, Languages and Programming*, volume 6756 of *Lecture Notes in Computer Science*, pages 21–41. 2011.

4 L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov. Hacking commercial quantum cryptography systems by tailored bright illumination. *Nature Photonics*, 4:686–689, 2010.

5 Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33(1):75 – 87, 1986.

6 F. G. S. L. Brandão, R. Ramanathan, A. Grudka, K. Horodecki, M. Horodecki, and P. Horodecki. Robust Device-Independent Randomness Amplification with Few Devices. 2013.

7 R. Colbeck and R. Renner. Free randomness can be amplified. *Nature Physics*, 8:450–454, 2012.

8 R. Gallego, L. Masanes, G. de la Torre, C. Dhara, L. Aolita, and A. Acín. Full randomness from arbitrarily deterministic events. *Nature Communications*, 4, 2013.

9 P. Mironowicz and M. Pawlowski. Amplification of arbitrarily weak randomness, arXiv: 1301.7722, 2013.

10 K.-M. Chung,Y. Shi, X. Wu, Physical Randomness Extractors, arXiv: 1402.4797 2014.

11 N. David Mermin. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.*, 65:1838–1840, 1990.

12 M. Navascués, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

13 J. Bouda, M. Pawlowski, M. Pivoluska, and M. Plesch. Device-independent randomness extraction for arbitrarily weak min-entropy source, arXiv: 1402.0974, 2014.

14 Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988.

Graph Homomorphisms for Quantum Players

Laura Mančinska[1] and David Roberson[2]

1 Centre for Quantum Technologies, National University of Singapore
2 School of Physical and Mathematical Sciences, Nanyang Technological University

Abstract

A homomorphism from a graph X to a graph Y is an adjacency preserving mapping $f : V(X) \rightarrow V(Y)$. We consider a nonlocal game in which Alice and Bob are trying to convince a verifier with certainty that a graph X admits a homomorphism to Y. This is a generalization of the well-studied graph coloring game. Via systematic study of quantum homomorphisms we prove new results for graph coloring. Most importantly, we show that the Lovász theta number of the complement lower bounds the quantum chromatic number, which itself is not known to be computable. We also show that other quantum graph parameters, such as quantum independence number, can differ from their classical counterparts. Finally, we show that quantum homomorphisms closely relate to zero-error channel capacity. In particular, we use quantum homomorphisms to construct graphs for which entanglement-assistance increases their one-shot zero-error capacity.

1998 ACM Subject Classification G.2.2 Graph Theory

Keywords and phrases graph homomorphism, nonlocal game, Lovász theta, quantum chromatic number, entanglement

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.212

1 Graph homomorphism game as a generalization of coloring game

In the (X, c)-coloring game, Alice and Bob are trying to convince a verifier with certainty that the graph $X = (V, E)$ is c-colorable [10, 6]. The verifier sends Alice and Bob vertices $a, b \in V$ respectively and they respond with colors $\alpha, \beta \in [c]$ accordingly. To win Alice an Bob need to respond with $\alpha = \beta$ for $a = b$ and with $\alpha \neq \beta$ for $ab \in E$. Classical Alice and Bob can win with probability 1 if and only if X is c-colorable. In contrast, quantum Alice and Bob using shared entanglement can sometimes win the (X, c)-coloring game even when X is not c-colorable [6, 1, 5, 16].

We introduce a natural generalization of the graph coloring game: the graph homomorphism game. A graph homomorphism is a function $\varphi : V(X) \rightarrow V(Y)$ such that $\varphi(x)$ and $\varphi(x')$ are adjacent whenever x and x' are adjacent. When such a map exists we say that X has a homomorphism to Y and write $X \rightarrow Y$. A c-coloring of X can be viewed as a homomorphism $\varphi : X \rightarrow K_c$, where K_c is the complete graph on c vertices. Graph homomorphisms have been used to prove results about different types of chromatic numbers, graph products etc.; they have applications in areas like complexity theory, statistical physics and others (see [12, 13] for a general reference).

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 212–216
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Our motivation for this work is that a systematic study of quantum homomorphisms can yield

- better understanding of and new results concerning quantum graph coloring (see Section 4);
- new examples of nonlocal games with perfect quantum but not classical strategies (see Section 2);
- new results for zero-error capacity via the connections that we establish in Section 3.

In the (X,Y)-*homomorphism game* the verifier sends Alice and Bob vertices $x,x' \in V(X)$ respectively and they respond with vertices $y,y' \in V(Y)$ accordingly. To win players need to respond with $y = y'$ to questions $x = x'$ and with $yy' \in E(Y)$ to questions $xx' \in E(X)$. Like the coloring game, the (X,Y)-homomorphism game can be won with certainty by classical players if and only if $X \rightarrow Y$. If quantum players using shared entanglement can win the (X,Y)-homomorphism game with certainty we say that X has a *quantum homomorphism* to Y and write $X \xrightarrow{q} Y$. As we know from the case of coloring and will see from new examples in the next section, sometimes $X \xrightarrow{q} Y$ even though $X \not\rightarrow Y$ (i.e., X does not admit a homomorphism to Y).

It is known that whenever X is quantum c-colorable, the (X,c)-coloring game can be won using projective measurements on maximally entangled state [5]. Moreover, Bob's projectors are the complex conjugates of Alice's. We have verified that the proof of [5] extends to the case of the (X,Y)-homomorphism game. This allows the following combinatorial reformulation:

▶ **Lemma 1.** *We have $X \xrightarrow{q} Y$ if and only if there exists an assignment of projectors P_{xy} to pairs of vertices $(x,y) \in V(X) \times V(Y)$ such that $\sum_y P_{xy} = I$ for all $x \in V(X)$ and*

$$P_{xy}P_{x'y'} = 0 \text{ whenever } (x = x' \ \& \ y \neq y') \text{ or } (x \sim x' \ \& \ y \not\sim y').$$

This reformulation is instrumental in proving many of the results in the coming sections. The other proof technique that we employ only uses the players' ability to win certain homomorphism games to conclude that they can also win some other homomorphism game. For example, this kind of reasoning easily shows that quantum homomorphisms are transitive, i.e., $X \xrightarrow{q} Y$ and $Y \xrightarrow{q} Z$ implies that $X \xrightarrow{q} Z$.

Curiosly, if instead of entanglement Alice and Bob are given access to non-signalling correlations, they can win the (X,K_2)-homomorphism game with certainty for any graph X. This implies that they can win any (X,Y)-homomorphism game for arbitrary graphs X,Y as long as $E(Y) \neq \emptyset$.

2 Quantum parameters

The quantum chromatic number, $\chi_q(X)$, is defined as the smallest c for which quantum players can win the (X,c)-coloring game with certainty [10, 6]. This parameter has been relatively well-studied [1, 5, 9, 17, 16]. In particular, it is known that for the family of graphs Ω_{4n} there is an exponential separation between $\chi(\Omega_{4n})$ and $\chi_q(\Omega_{4n})$. Here, the so-called Hadamard graph Ω_n is the graph with vertex set $\{\pm 1\}^n$ and edge set $\{(v,w) : v^T w = 0\}$. Also, a complete characterization of graphs with $\chi_q(X) < \chi(X)$ has been given [16]. However, many questions remain open. For example, it is not known whether $\chi_q(X)$ is computable, or whether there exists a family of graphs X_n such that $\lim_{n\to\infty} \chi(X_n) = \infty$ but $\lim_{n\to\infty} \chi_q(X_n) < \infty$. A systematic study of quantum homomorphisms could aid in answering these and other questions

Using the framework of quantum homomorphisms, we can introduce a quantum analogue for any graph parameter defined in terms of graph homomorphisms (e.g., clique number, independence number, odd girth, etc.). Here we only consider the following:

- quantum clique number, $\omega_q(X) = \max\{n : K_n \xrightarrow{q} X\}$;
- quantum independence number, $\alpha_q(X) = \omega_q(\overline{X})$ where $\overline{X}$ denotes the complement of X.

Let us remark that by now, the quantum independence number has been further used by many other authors exploring parallel repetition, zero-error communication, binary constraint system games etc.

We are about to see that quantum clique and independence number can be different from their classical counterparts. Moreover, we show how to construct a graph with such a separation using any two graphs X and Y such that $X \xrightarrow{q} Y$ but $X \not\to Y$.

For graphs X and Y, their homomorphic product, $X \ltimes Y$, is the graph with vertex set $V(X) \times V(Y)$, and vertex (x, y) is adjacent to (x', y') if either $(x = x'$ and $y \neq y')$ or $(xx' \in E(X)$ and $yy' \notin E(Y))$. This definition is motivated by the fact that $X \to Y$ if and only if $\alpha(X \ltimes Y) = |V(X)|$. We have proved the quantum version of this fact, i.e., $X \xrightarrow{q} Y$ if and only if $\alpha_q(X \ltimes Y) = |V(X)|$. Combining these two facts gives:

▶ **Theorem 2.** *Let X, Y be graphs such that $X \xrightarrow{q} Y$ but $X \not\to Y$. Then we have that $\alpha(X \ltimes Y) < \alpha_q(X \ltimes Y)$ and $\omega(\overline{X \ltimes Y}) < \omega_q(\overline{X \ltimes Y})$.*

This theorem allows to obtain separations for clique and independence numbers starting from any graph X with $\chi_q(X) < \chi(X)$. For example, the fact that $\Omega_n \xrightarrow{q} K_n$ [1] but $\Omega_{4n} \not\to K_{4n}$ for $n > 2$ [11] implies that $\alpha(\Omega_{4n} \ltimes K_{4n}) < \alpha_q(\Omega_{4n} \ltimes K_{4n})$ for all $n > 2$.

3 Relationship to entanglement-assisted zero-error capacity

The *one-shot zero-error capacity*, $c_0(X)$, of a graph X is the maximum number of different messages that can be sent without error by one use of any classical noisy channel $\mathcal{N}$ with confusability graph X [18, 15]. In the scenario where the communicating parties can use shared entanglement, we speak about *entanglement-assisted zero-error capacity*, $c_0^*(X)$ [7].

The separations between $c_0^*(X)$ and $c_0(X)$ and their asymptotic analogues have been investigated in [7, 14, 16, 3]. It is an open question how large these separations can be. As [16] shows, a separation between the one-shot zero-error capacities can be obtained starting from any graph X with $\chi_q(X) < \chi(X)$.

A somewhat analogous relationship can be shown to hold for quantum homomorphisms in general:

▶ **Theorem 3.** *Let X, Y be graphs such that $X \xrightarrow{q} Y$ but $X \not\to Y$. Then we have that*

$$c_0(X \ltimes Y) < c_0^*(X \ltimes Y).$$

It turns out that the quantum independence number, $\alpha_q(X)$, is closely related to and might equal the one-shot entanglement-assisted zero-error capacity:

▶ **Theorem 4.** *For any graph X we have $\alpha_q(X) \leq c_0^*(X)$ with equality if and only if $c_0^*(X)$ can be achieved using a strategy in which all of Alice's measurements are projective and the shared state is maximally entangled.*

By the above theorem, proving that $\alpha_q(X) = c_0^*(X)$ for all graphs X would settle the open question of whether projective measurements on maximally entangled state suffice to achieve $c_0^*(X)$. If this was the case, the results from [16] would imply a complete characterization of graphs for which $c_0(X) < c_0^*(X)$.

Finally, we show that quantum homomorphisms respect the order of both the one-shot and asymptotic entanglement-assisted zero-error capacities.

▶ **Theorem 5.** *Let Θ^* denote the entanglement-assisted Shannon capacity. For any graphs X, Y we have that $X \xrightarrow{q} Y$ implies both*

$$c_0^*(\overline{X}) \leq c_0^*(\overline{Y}) \text{ and } \Theta^*(\overline{X}) \leq \Theta^*(\overline{Y}).$$

The above theorem can be used to lower bound $\Theta^*(Y)$ in the case when $\overline{X} \xrightarrow{q} \overline{Y}$ and $\Theta^*(X)$ is known for some graph X.

4 Relationship to Lovász ϑ

The Lovász theta number of X, denoted $\vartheta(X)$, was introduced in [15] as an efficiently computable upper bound for the Shannon capacity $\Theta(X)$. It has been shown that $\vartheta(X)$ upper bounds even the entaglement-assisted Shannon capacity $\Theta^*(X)$ [2, 8]. We have established that quantum homomorphisms respect the order of Lovász theta:

▶ **Theorem 6.** *For any graphs X, Y we have that $X \xrightarrow{q} Y$ implies $\vartheta(\overline{X}) \leq \vartheta(\overline{Y})$.*

Applying the above theorem with Y being the complete graph on $\chi_q(X)$ vertices gives the following:

▶ **Corollary 7.** *For any graph X we have $\vartheta(\overline{X}) \leq \chi_q(X)$.*

Corollary 7 gives us an efficiently computable lower bound on the quantum chromatic number $\chi_q(X)$, which itself is not even known to be computable (By now our lower bound on $\chi_q(X)$ has been strengthened by replacing ϑ with ϑ^+ [4]). The lower bound from Corollary 7 can also be used to conclude that the previously established [1] upper bound $\chi_q(\Omega_n) \leq n$ is actually tight for all Hadamard graphs Ω_n with $4|n$. (The other cases are not interesting since Ω_n is either empty or bipartite.)

References

1 David Avis, Jun Hasegawa, Yosuke Kikuchi, and Yuuya Sasaki. A quantum protocol to win the graph colouring game on all hadamard graphs. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, E89-A(5):1378–1381, 2006.

2 Salman Beigi. Entanglement-assisted zero-error capacity is upper-bounded by the Lovász theta function. *Phys. Rev. A*, 82:10303–10306, 2010.

3 Jop Briët, Harry Buhrman, and Dion Gijswijt. Violating the shannon capacity of metric graphs with entanglement. *Proceedings of the National Academy of Sciences*, 110(48):19227–19232, 2013.

4 Jop Briët, Harry Buhrman, Monique Laurent, and Giannicola Scarpa. Zero-error souce-channel coding with entanglement. arXiv:1308.4283, 2013.

5 Peter J. Cameron, Ashley Montanaro, Michael W. Newman, Simone Severini, and Andreas Winter. On the quantum chromatic number of a graph. *Electr. J. Comb.*, 14(1), 2007.

6 Richard Cleve, Peter Hoyer, Ben Toner, and John Watrous. Consequences and limits of nonlocal strategies. In *19th IEEE Annual Conference on Computational Complexity*, pages 236–249, 2004.

7 Toby S. Cubitt, Debbie Leung, William Matthews, and Andreas Winter. Improving zero-error classical communication with entanglement. *Phys. Rev. Lett.*, 104:230503, Jun 2010.

8 Runyao Duan, Simone Severini, and Andreas Winter. Zero-error communication via quantum channels and a quantum Lovász ϑ function. In *IEEE International Symposium on Information Theory*, pages 64–68, 2011.

9 Junya Fukawa, Hiroshi Imai, and François Le Gall. Quantum coloring games via symmetric SAT games. In *Asian Conference on Quantum Information Science (AQIS'11)*, 2011.

10 Viktor Galliard and Stefan Wolf. Pseudo-telepathy, entanglement, and graph colorings. In *IEEE International Symposium on Information Theory*, page 101, 2002.

11 Chris Godsil and Michael W. Newman. Coloring an orthogonality graph. *SIAM J. Discret. Math.*, 22(2):683–692, 2008.

12 Geňa Hahn and Claude Tardif. Graph homomorphisms: structure and symmetry. In *Graph symmetry*, volume 497 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 107–166. Kluwer Acad. Publ., 1997.

13 Pavol Hell and Jaroslav Nešetřil. *Graphs and homomorphisms.* Oxford University Press, 2004.

14 Debbie Leung, Laura Mančinska, William Matthews, Maris Ozols, and Aidan Roy. Entanglement can increase asymptotic rates of zero-error classical communication over classical channels. *Commun. Math. Phys.*, 311:97–111, 2012.

15 László Lovász. On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory*, 25(1):1–7, 1979.

16 Laura Mančinska, Giannicola Scarpa, and Simone Severini. New separations in zero-error channel capacity through projective Kochen-Specker sets and quantum coloring. *IEEE Transactions on Information Theory*, 59(6):4025–4032, 2013.

17 Giannicola Scarpa and Simone Severini. Kochen-Specker sets and the rank-1 quantum chromatic number. *IEEE Trans. Inform. Theory*, 58(4):2524–2529, 2012.

18 Claude E. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, 2(3):8–19, 1956.

Quantum Linear Network Coding as One-way Quantum Computation*

Niel de Beaudrap[1] and Martin Roetteler[2]

1 Centrum Wiskunde & Informatica, Science Park 123, Amsterdam, The Netherlands
beaudrap@cwi.org

2 Microsoft Research, One Microsoft Way, Redmond, WA, USA
martinro@microsoft.com

Abstract

Network coding [1] is a technique to maximize communication rates within a network, in communication protocols for simultaneous multi-party transmission of information. Linear network codes are examples of such protocols in which the local computations performed at the nodes in the network are limited to linear transformations of their input data (represented as elements of a ring, such as the integers modulo 2). The quantum linear network coding protocols of Kobayashi *et al.* [17, 18] coherently simulate classical linear network codes, using supplemental classical communication. We demonstrate that these protocols correspond in a natural way to measurement-based quantum computations with graph states over qudits [21, 4, 8] having a structure directly related to the network.

1998 ACM Subject Classification E.4 Coding and Information Theory, F.1.1 Models of computation

Keywords and phrases Network coding, quantum computing, measurement-based computation, simulation

Digital Object Identifier 10.4230/LIPIcs.TQC.2014.217

1 Introduction

Network coding [1] is a technique to maximize the rate at which a set of *source nodes* can simultaneously transmit a set of independent messages to certain *target nodes* through a fixed network. For this purpose, it is sufficient to give each communication link enough bandwidth to accommodate multiple messages to be transmitted at once: however, less bandwidth may be required at each link if one allows nodes to distribute information about the messages across the network. A classic example is the *two-pair problem* on the "butterfly network" (illustrated in Figure 1): rather than halve the bandwidth between two messages at an apparent bottleneck in the network, the internal nodes may perform simple local computations on the messages, to allow the input data to be reconstructed at the targets. *Linear network coding* is the special case in which the protocol only requires each node to compute a linear transformation of its inputs to achieve this goal.

We consider *quantum network coding*, in which we perform similar tasks with quantum states transmitted through noiseless quantum channels. It is immediately apparent that some problems which can be sensibly posed for "classical" network coding are impossible in general

* Supported by a Vidi grant from the Netherlands Organisation for Scientific Research (NWO) and the European Commission project QALGO.

9th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC'14).
Editors: Steven T. Flammia and Aram W. Harrow; pp. 217–233
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

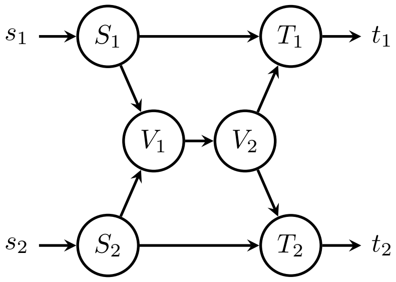


Figure 1 The *butterfly network*, with source nodes S_1 and S_2 and target nodes T_1 and T_2. The two-pair problem on this network is for S_1 to communicate their input to the target T_2, and simultaneously for S_2 to communicate their input to the target T_1, assuming that each edge can carry at most one message (represented *e.g.* by a single bit, 0 or 1). The classic solution is for S_1, S_2, and V_2 to duplicate their inputs, and for V_1, T_1, and T_2 to compute the parity of their inputs, in which case $(t_1, t_2) = (s_2, s_1)$.

for quantum network coding. For instance, while a classical network code allows for the each of the source nodes to each send a copy of their inputs to *both* targets in the butterfly network (see page 220), this is clearly not possible for quantum states due to the no-cloning theorem [24]. Other problems which do not require multiple copies of the input states to be re-created at the output (such as the two-pairs problem above) are still potentially unsolvable with fixed-capacity quantum channels alone, even when the corresponding classical problem is solvable [15, 19]. However, some of these problems become feasible for quantum states when the network nodes share prior entanglement [14], or if the capacities of the communication links scale as the logarithm of the number of target nodes [22].

Because classical information is easier to faithfully transmit and transform than quantum information, it is common to consider quantum protocols which also allow classical communication, and where fewer restrictions are imposed on the classical than the quantum communication (see Ref. [20]). In a setting where *no* restrictions are imposed on classical communication, Kobayashi *et al.* [17] describe a quantum protocol for the *k-pairs problem*: the problem in which each of k source nodes wish to communicate their input message to one of k distinct target nodes. Their protocol is in effect a coherent simulation of a classical linear network code. More generally, for any classical linear network code which performs some injective linear transformation $\mathbf{t} = M\mathbf{s}$ of the input data, Ref. [17] yields a corresponding quantum procedure to coherently simulate that network over for arbitrary superpositions of input data. We call such a protocol a (classically assisted) *quantum linear network code*. For the k-pairs problem, the protocols of Ref. [17] were subsequently extended in two different ways by Ref. [18]: to restrict the classical communication to the same network as the quantum communication (albeit with multiple rounds of communication, and sending a single message backwards as well as forwards along each communication link) and to accommodate non-linear protocols as well.

In this article we show that classically assisted quantum linear network codes in the style of Ref. [18] are in effect an instance of *one-way measurement based quantum computation* (MBQC) [21, 4, 8, 9]: a model of quantum computation in which one may entangle an arbitrary input state $|\psi\rangle$ with a graph state, which is then subjected to a sequence of measurements, leaving a final residual state which contains a transformed state $U|\psi\rangle$ for some unitary transformation[1] U. Furthermore, the graph state used as a resource is closely related

[1] In general, the transformation which is performed on an input state $|\psi\rangle$ is not necessarily a unitary

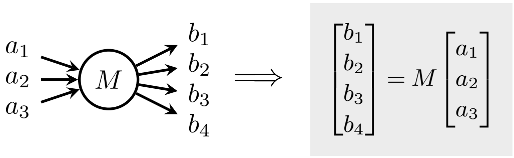

a_1
a_2
a_3
M
b_1
b_2
b_3
b_4
$\Longrightarrow$
$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = M \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$

Figure 2 An illustration of the transformation of messages performed by a single network node in a linear coding protocol.

structurally to the network used in the coding protocol. This demonstrates a link between MBQC and linear network coding, construed as distributed models of computation, and suggests novel ways of interpreting measurement-based procedures. At the same time, this suggests MBQC as a unifying framework in which to consider multi-party quantum networking protocols, including cryptographic applications formulated in the one-way model [3, 16] as well as standard security proofs of BB84 [23].

2 Preliminaries

In this section, we present introductory remarks on classical linear network coding, and summarize the development of Refs. [17, 18]. We assume familiarity with standard models of quantum computation on qubits, as well as measurement-based quantum computation (see *e.g.* Refs. [21, 4, 8, 9] for introductory references). We introduce the notation and the definitions for the operators used over qudits of dimension d below.

2.1 Classical network coding

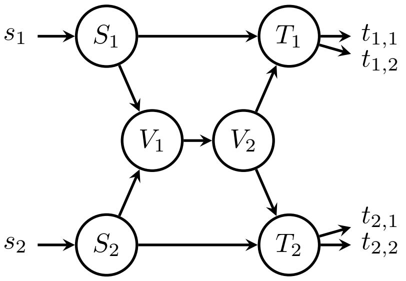
We model a communications network by a directed graph of communications links, each of which can be used to transmit a single message from some message set M. In this article we suppose that M consists of a cyclic ring[2] $\mathbb{Z}_d = \mathbb{Z}/d\mathbb{Z}$. The messages are sent between co-operative agents (represented by nodes of the digraph) who may perform some non-trivial transformation of the data they receive from ingoing links. In the context of linear network codes, the transformations performed by each node are linear transformations, as represented in Figure 2.

The result of this computation is then sent as output messages to other nodes. We restrict ourselves to directed acyclic networks, and assume that each node waits for all inputs to arrive before computing its outputs.

The canonical network coding problems involve distributing information from a collection of *source* nodes $\mathsf{S} = \{S_1, S_2, \ldots\}$ to a collection of *target* nodes $\mathsf{T} = \{T_1, T_2, \ldots\}$, such as the *multicast problem* (in which each source S_h must transmit their data to every one of the targets T_j), and the k*-pairs problem* (in which each source S_h tries to send their message to a single target $T_{\pi(h)}$, for some permutation $\pi \in \mathfrak{S}_k$ of the indices). The source nodes

transformation, but rather some completely positive trace preserving map Φ acting on $\rho_0 = |\psi\rangle\langle\psi|$. However, standard treatments of the one-way model describe how measurements on graph states may be used to simulate the transformations performed by unitary circuits, which by construction would transform the input state $|\psi\rangle$ unitarily.

[2] In the setting where messages represent elements of a finite field $\mathrm{GF}(p^r)$ (see *e.g.* Ref. [13]), we may replace each communication link with r parallel communications links, representing elements of $\mathrm{GF}(p^r)$ as r-dimensional vectors over $\mathrm{GF}(p) \cong \mathbb{Z}_p$. In the case of linear network codes, this leads to no loss of generality, as every $\mathrm{GF}(p^r)$-linear transformation of messages is also a $\mathrm{GF}(p)$-linear transformation.



Multicast problem on the butterfly network:

$$\begin{bmatrix} t_{1,1} \\ t_{1,2} \\ t_{2,1} \\ t_{2,2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

Decompose:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} T_1 & 0 \\ 0 & T_2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & V_2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & V_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S_1 & 0 \\ 0 & S_2 \end{bmatrix}$$

$$\begin{aligned} \text{where } S_1, S_2 &: \mathbb{Z}_d \to \mathbb{Z}_d^2 \\ V_1 &: \mathbb{Z}_d^2 \to \mathbb{Z}_d \\ V_2 &: \mathbb{Z}_d \to \mathbb{Z}_d^2 \\ T_1, T_2 &: \mathbb{Z}_d^2 \to \mathbb{Z}_d^2 \end{aligned}$$

Figure 3 The multicast problem on the butterfly network, formulated as a linear transformation over the ring $\mathbb{Z}_d$. A solution by linear network coding decomposes this transformation as a product of block matrices according to the network structure. A typical solution to this problem is presented in Eqn. (1).

S_j each have some piece of information, usually represented as a single element $s_j \in \mathbb{Z}_d$ or vector $\mathbf{s}_j \in \mathbb{Z}_d^{n_j}$. To put the source and target nodes on an equal footing to the other network nodes, we suppose that the inputs s_j of the sources S_j are messages received from elsewhere (*e.g.* storage devices owned by the source nodes), and the outputs t_j to be computed by the targets T_j are also transmitted to somewhere, as depicted in Figure 1. A solution via linear network codes simply assigns linear transformations to each node, in such a way that the composite transformation performs the correct redistribution of input messages.

We regard linear network coding as a distributed model of computation, in which linear transformations are decomposed into block matrices, where each non-trivial block is represented by a single node. For *any* linear function f – of which the k-pairs and multicast problems are special cases – we consider which transformations the nodes may perform (if any) to compute f. Figure 3 presents the multicast problem on the butterfly network in this form, to which one solution is the following assignment of matrices to each node in the network:

$$S_1 = S_2 = V_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \qquad V_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}, \qquad T_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \qquad T_2 = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}. \tag{1}$$

2.2 Classically assisted quantum network coding

We now outline the constructions of Ref. [17], and also of Ref. [18] in the special case of linear coding protocols over the ring $\mathbb{Z}_d$ of integers modulo d, for protocols using message qudits of dimension d.

Consider a node V performing some coding operation $\mathbf{y} = V\mathbf{x}$ for $\mathbf{x} \in \mathbb{Z}_d^\ell$ and $\mathbf{y} \in \mathbb{Z}_d^m$ in a classical coding network. We may simulate this node by initializing an output register $\mathbf{y} = \mathbf{0} \in \mathbb{Z}_d^m$, performing a bijective mapping $(\mathbf{x}, \mathbf{y}) \mapsto (\mathbf{x},\, \mathbf{y} + V\mathbf{x})$ in the larger space $\mathbb{Z}_d^{\ell+m}$, and then discarding the input $\mathbf{x}$. The bijective mapping can be performed by elementary row transformations on $\mathbf{x}$, which in the quantum setting may be performed by controlled-X operations,

$$\Lambda X_{j,k} = \sum_{c=0}^{d-1} |c\rangle\langle c|_j \otimes X_k^c, \tag{2}$$

where $X\,|q\rangle = |q+1 \bmod d\rangle$ is an analogue of the unitary Pauli operator σ_x on qubits. Consider a generic node V which accepts a collection of input qudits $a_1, \ldots, a_\ell$ as input and produces output qudits $b_1, \ldots, b_m$, coherently simulating the transformation $|\mathbf{x}\rangle_{a_1\cdots a_\ell} \longmapsto |T\mathbf{x}\rangle_{b_1\cdots b_m}$. In the construction of Ref. [17] for quantum linear codes, V simulates this transformation by preparing the qudits $b_1, \ldots, b_k$ in the $|0\rangle$ state, and performing the transformations

$$\Lambda X^{V_{j,k}}\Big(|x_k\rangle \otimes |0\rangle\Big) = |x_k\rangle \otimes |V_{j,k}x_k\rangle \tag{3}$$

on the qudits a_k and b_j, for every index $1 \leqslant j \leqslant \ell$ and $1 \leqslant k \leqslant m$ in any order. For standard basis states, the result is to transform $|\mathbf{x}\rangle\,|\mathbf{0}\rangle \mapsto |\mathbf{x}\rangle\,|V\mathbf{x}\rangle$. This characterizes a linear transformation

$$\tilde{U}_V \;=\; \left(\prod_{j=1}^{m}\prod_{k=1}^{\ell} \Lambda X^{V_{j,k}}_{a_k,b_j}\right)\left(\mathbb{1}_{\mathbf{a}} \otimes |\mathbf{0}\rangle_{\mathbf{b}}\right), \tag{4}$$

which is a unitary embedding for any transformation V. (An example of such a circuit is illustrated in Figure 4.) If the qudits $a_1, \ldots, a_\ell$ where originally in standard basis states, we could simply discard them; but if they are initially not in standard basis states, they will become entangled with $b_1, \ldots, b_m$. To decouple them, we attempt to project each of the qudits a_j to the $|+\rangle$ state by measurement,

$$|+\rangle \;=\; \tfrac{1}{\sqrt{d}}\Big(|0\rangle + |1\rangle + \cdots + |d-1\rangle\Big). \tag{5}$$

Successfully doing so on a generic input state $|\psi\rangle = \sum_{\mathbf{x}} u_{\mathbf{x}}\,|\mathbf{x}\rangle$ would lead to the sequence of transformations

$$\begin{aligned} |\psi\rangle \;\longmapsto\; \sum_{\mathbf{x}} u_{\mathbf{x}}\,|\mathbf{x}\rangle_{\mathbf{a}}\,|\mathbf{0}\rangle_{\mathbf{b}} \;\longmapsto\; & \sum_{\mathbf{x}} u_{\mathbf{x}}\,|\mathbf{x}\rangle_{\mathbf{a}}\,|V\mathbf{x}\rangle_{\mathbf{b}} \\ \longmapsto\; & \frac{1}{\sqrt{d^\ell}}\left(\bigotimes_{k=1}^{\ell} |+\rangle_{a_k}\right) \otimes \sum_{\mathbf{x}} u_{\mathbf{x}}\,|V\mathbf{x}\rangle_{\mathbf{b}}\,. \end{aligned} \tag{6}$$

This mapping is of course non-unitary: projection onto $|+\rangle$ must be performed as part of a measurement onto some basis. Ref. [17] considers a measurement of the qudits a_j in the Fourier basis,

$$|\omega_r\rangle \;=\; \frac{1}{\sqrt{d}}\sum_{x=0}^{d-1} \mathrm{e}^{2\pi i x r/d}\,|x\rangle \;=\; F\,|r\rangle, \qquad\qquad \text{where } F = \frac{1}{\sqrt{d}}\sum_{x,r=0}^{d-1} \mathrm{e}^{2\pi i k x/d}\,|x\rangle\langle r|\,. \tag{7}$$

The operator F is the *quantum Fourier transform over* $\mathbb{Z}_d$. We may attempt to simulate projection of each qudit a_j onto $|+\rangle$ by Fourier basis measurements, where a result of $|\omega_0\rangle$ is a success, as $|\omega_0\rangle = |+\rangle$. If we obtain results $\big|\omega_{r_j}\big\rangle$ for $r_j \neq 0$ instead of $|+\rangle$, the post-measurement state is

$$\left(\bigotimes_{k=1}^{\ell} |\omega_r\rangle_{a_k}\right) \otimes \sum_{\mathbf{x}} u_{\mathbf{x}} \mathrm{e}^{-2\pi i (\mathbf{r}\cdot\mathbf{x})/d}\,|V\mathbf{x}\rangle_{\mathbf{b}} \tag{8}$$

up to normalization. If V is injective, the relative phase $\mathrm{e}^{-2\pi i (\mathbf{r}\cdot\mathbf{x})/d}$ can be undone by a suitable application of Z operations on the qudits $b_1, \ldots, b_m$, where Z is the unitary generalization of σ_z:

$$Z \;=\; \sum_{q=0}^{d-1} \mathrm{e}^{2\pi i q/d}\,|q\rangle\langle q|\,. \tag{9}$$

If V is not injective, then only certain vectors $\mathbf{r}$ of measurement outcomes can be immediately corrected, resulting in a non-unitary CP map. However, regardless of whether some nodes in coding network perform non-invertible operations, the relative phases which accumulate on the entire state are linear functions. Then if the transformation performed by the whole network is injective, the phases which have accumulated due to the measurements can be undone if the target nodes have sufficient information about the measurement outcomes.

The protocol of Ref. [17] solves the k-pairs problem: thus the transformation it performs is indeed injective. Each node simply transmits their measurement outcomes to each target node, which performs a suitable combination of Z operations to correct the relative phases. Ref. [18] presents an alternative protocol in which the measurements are deferred until after all quantum messages have been sent, and in which the internal nodes of the network do the majority of the phase corrections, as follows. Consider a node which attempts to coherently simulate a transformation $L : \mathbb{Z}_d^\ell \to \mathbb{Z}_d^m$ in the middle of a coding network which attempts to coherently simulate a transformation $M : \mathbb{Z}_d^{\mathsf{S}} \to \mathbb{Z}_d^{\mathsf{T}}$ on an input state $|\psi\rangle = \sum_{\mathbf{x}} u_{\mathbf{x}} |\mathbf{x}\rangle$. Suppose that we perform the simulation procedure above, but omitting the Fourier basis measurements. For some linear maps H and K, the state after the final quantum messages is in general an entangled state of the form[3]

$$|\Psi\rangle = \sum_{\mathbf{x}} u_{\mathbf{x}} |\mathbf{x}\rangle_{\mathsf{S}} \otimes |M\mathbf{x}\rangle_{\mathsf{T}} \otimes \Big(|K\mathbf{x}\rangle_{a_1,\ldots,a_\ell} \otimes |LK\mathbf{x}\rangle_{b_1,\ldots,b_m} \Big) \otimes |H\mathbf{x}\rangle_{\mathrm{rest}} \,, \tag{10}$$

where the factors in parentheses are the input and output qudits to the node L. If the qudits $b_1, \ldots, b_m$ are measured in the Fourier basis by the nodes to which they are sent, they yield some outcomes $r_1, \ldots, r_m$, and the remaining qudits are transformed to

$$|\Psi'\rangle = \sum_{\mathbf{x}} u_{\mathbf{x}} |\mathbf{x}\rangle_{\mathsf{S}} \otimes |M\mathbf{x}\rangle_{\mathsf{T}} \otimes \Big(\mathrm{e}^{-2\pi i (\mathbf{r}\cdot LK\mathbf{x})/d} |K\mathbf{x}\rangle_{a_1,\ldots,a_\ell} \Big) \otimes |H\mathbf{x}\rangle_{\mathrm{rest}} \,, \tag{11}$$

where $\mathbf{r}$ is the vector of the outcomes. Let $\boldsymbol{\tau} = L^\top \mathbf{r}$: we have $\boldsymbol{\tau} \cdot K\mathbf{x} = \mathbf{r} \cdot LK\mathbf{x}$ by construction. If the nodes which perform these measurements send the outcomes to the node L, then L can undo the phases induced by measurement of the qudits b_k by performing the operation $Z^{\boldsymbol{\tau}} := Z_{a_1}^{\tau_1} Z_{a_2}^{\tau_2} \cdots Z_{a_\ell}^{\tau_\ell}$, which performs the mapping

$$\begin{aligned} Z_{a_1}^{\tau_1} Z_{a_2}^{\tau_2} \cdots Z_{a_\ell}^{\tau_\ell} \Big| (K\mathbf{x})_1 (K\mathbf{x})_2 \cdots (K\mathbf{x})_\ell \Big\rangle &= \exp\!\Big(\tfrac{2\pi i}{d} \big[\tau_1 (K\mathbf{x})_1 + \cdots + \tau_\ell (K\mathbf{x})_\ell \big] \Big) |K\mathbf{x}\rangle \\ &= \mathrm{e}^{2\pi i (\boldsymbol{\tau}\cdot K\mathbf{x})/d} |K\mathbf{x}\rangle \,. \end{aligned} \tag{12}$$

Performing these corrections on $|\Psi'\rangle$ then yields the state

$$|\Psi''\rangle = \sum_{\mathbf{x}} u_{\mathbf{x}} |\mathbf{x}\rangle_{\mathsf{S}} \otimes |M\mathbf{x}\rangle_{\mathsf{T}} \otimes |K\mathbf{x}\rangle_{a_1,\ldots,a_\ell} \otimes |H\mathbf{x}\rangle_{\mathrm{rest}} \,, \tag{13}$$

which has fewer unmeasured qudits than $|\Psi\rangle$, and no relative phases. This simulates projecting the qudits $b_1, \ldots, b_m$ to the $|+\rangle$ state. By induction, if each node aside from the source nodes (but including the target nodes) measures their input qudits in the Fourier basis, and communicates the outcomes backwards along their incoming links to the nodes which

[3] The final tensor factor is on the remaining nodes entangled with the sources, whose components in the standard basis are again some linear transformations of the standard basis on the source nodes' inputs; by induction on the depth of the coding network, one may show that H and K are indeed linear transformations.

provided those qudits, those nodes can correct for the effect of the measurements. Eventually one obtains the state

$$\left|\Psi^{(n)}\right\rangle = \sum_{\mathbf{x}} u_{\mathbf{x}} \left|\mathbf{x}\right\rangle_{\mathsf{S}} \otimes \left|M\mathbf{x}\right\rangle_{\mathsf{T}}, \tag{14}$$

which is an entangled state of the (collective) inputs to the source nodes and the outputs of the target nodes. If the source nodes measure their qudits in the Fourier basis, it suffices for them to communicate the outcomes to target nodes in such a way that the outcomes can be corrected.

For arbitrary linear transformations M, direct communication among target nodes or between the source and the target nodes may be required to undo the relative phases induced by measurement. If the source nodes measure their qudits and collectively obtain a vector $\mathbf{s}$ of outcomes, the resulting state on the remaining qudits is

$$\left|\Psi^{(n+1)}\right\rangle = \sum_{\mathbf{x}} u_{\mathbf{x}} \mathrm{e}^{-2\pi i(\mathbf{s}\cdot\mathbf{x})/d} \left|M\mathbf{x}\right\rangle_{\mathsf{T}}. \tag{15}$$

If M has a left-inverse A, and we let $B = A^{\top}$, it suffices for the sources to somehow communicate $\sigma_j := \sum_k B_{jk} s_k$ to the target node T which is responsible for producing the message t_j. This would allow T to perform a Z^{σ_j} correction and undo the relative phase on the j^{th} output qudit. Specifically, if the sources collectively communicate $\boldsymbol{\sigma} = B\mathbf{s}$ to the targets, who collectively perform the phase operations $Z^{\boldsymbol{\sigma}} = Z_{t_1}^{\sigma_1} Z_{t_2}^{\sigma_2} \cdots$ on the target qudits, the resulting state is

$$\begin{aligned}\left|\Psi^{(n+2)}\right\rangle &= \sum_{\mathbf{x}} u_{\mathbf{x}} \mathrm{e}^{2\pi i\left[\boldsymbol{\sigma}\cdot(M\mathbf{x})-\mathbf{s}\cdot\mathbf{x}\right]/d} \left|M\mathbf{x}\right\rangle_{\mathsf{T}} = \sum_{\mathbf{x}} u_{\mathbf{x}} \mathrm{e}^{2\pi i[\mathbf{s}^{\top}(AM-\mathbb{1})\mathbf{x}]/d} \left|M\mathbf{x}\right\rangle_{\mathsf{T}} \\ &= \sum_{\mathbf{x}} u_{\mathbf{x}} \left|M\mathbf{x}\right\rangle_{\mathsf{T}}; \end{aligned} \tag{16}$$

There are special cases where the amount of communication required outside of the network can be bounded. In particular, for the k-pairs problem where M is a permutation matrix (so that $(M^{-1})^{\top} = M$), it suffices to perform the classical linear coding protocol on the vector $\mathbf{s}$ to transmit $\boldsymbol{\sigma} = M\mathbf{s}$ to the target nodes. In this case, all classical communications may be restricted to the same network as the quantum communications – albeit using each communication link once in reverse, for the measurements of the qudits involved in the intermediate messages. More generally, if M is injective and there is a block-diagonal matrix B (where the blocks act on collections of messages held by individual target nodes) such that $M^{\top}BM = \mathbb{1}$, the sources may communicate $M\mathbf{s}$ to the targets, allowing the target nodes to compute $\boldsymbol{\sigma} = B^{\top}M\mathbf{s}$ and use this to govern phase corrections.

3 Classically assisted quantum linear coding is one-way MBQC

We now show how any coherent linear coding protocol, as described in Section 2.2, is in essence a measurement computation in the one-way model. The graph states of the MBQC procedures constructed in this way are easily derived from the coding network itself: allocate two entangled qudits at either end of each communications link in the network (one for the node on either side of the link), with further entangling operations between the qudits corresponding to the incoming links and the outgoing links. The corrections are the same as for the coherent coding network, albeit with some supplemental corrections arising from the way that the ΛX operations are simulated. If we follow the protocol of Ref. [17],

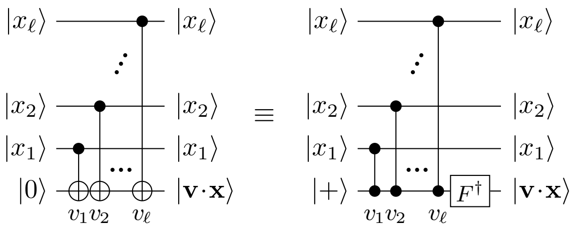


Figure 4 Equivalent ways to decompose a unitary transformation $\tilde{U}_V$ which prepares a single message qudit, for a single-row matrix $V = \mathbf{v}^\top$. The left-hand circuit represents the decomposition of Eqn. (4). Variables v_j below operations denote the power to which the circuit operation is raised. Multi-row coding transformations V may be simulated by several such circuits, acting on different target qudits.

the corrections are all deferred to the end of the procedure, as in standard treatments of measurement-based computation.

Again, we assume familiarity with the measurement based model: see Refs. [21, 7, 4, 9] for references applicable to qubits (similar results and constructions apply over arbitrary qudits).

3.1 MBQC simulation of a single coding node

The main element of the correspondence between quantum linear network coding and MBQC is the observation that ΛX operations differ by only a Fourier transform from a controlled-phase operation,

$$\Lambda Z \;=\; (\mathbb{1}\otimes F)\Lambda X(\mathbb{1}\otimes F^\dagger) \;=\; \sum_{c=0}^{d-1} |c\rangle\langle c| \otimes Z^c, \tag{17}$$

which are the diagonal operations used to construct the entanglement structures in measurement-based computation. This means that the injective maps $\tilde{U}_V$ used to perform the coding at each node may be straightforwardly represented in terms of preparing the state $|+\rangle = F\,|0\rangle$ for each output qudit b_j to be sent, performing the entangling operation $\Lambda Z^{V_{j,k}}$ between each input qudit a_k and each output qudit b_j, and then acting on b_j with a Fourier transform, as represented in Figure 4.

Note that the inverse Fourier transform acting on the output-message qudit may be simulated by a Fourier basis measurement by introducing another auxiliary qudit, using a standard MBQC construction. Consider a qudit v in an arbitrary pure state $|\psi\rangle = \sum_{x=0}^{d-1} u_x\,|x\rangle$. We may introduce a qudit w prepared in the state $|+\rangle$, and entangle them using a $\Lambda Z^\dagger$ operation, obtaining the state

$$|\Psi\rangle_{vw} = \Lambda Z^\dagger_{vw}\,|\psi\rangle_v\,|+\rangle_w\,. \tag{18}$$

We then measure v in the Fourier basis, obtaining a state $|\omega_r\rangle$, and perform the operation X^{-r} on w. We may use the stabilizer formalism (see *e.g.* Ref. [10]) to succinctly verify how this sequence of transformations, considered as CP maps, transform X and Z: as these generate an operator basis for single-qudit states, this will suffice to show how it transforms $|\psi\rangle_v$ to $F^\dagger\,|\psi\rangle_w$. Specifically, we wish to see how the group of Pauli operators which *stabilize* the state (*i.e.*, at each point in time, those Pauli operators for which the state is a +1-eigenvector) transforms, for states on v and/or w. We use the following facts:

- We write $\omega = \exp(\frac{2\pi i}{d}) \in \mathbb{C}$ as a minor abuse of notation: it is easy to verify that $X\,|\omega_r\rangle = \omega^r\,|\omega_r\rangle$. In particular, $|+\rangle$ is the unique $+1$-eigenvector of X up to scalar factors.
- Measuring v in the Fourier basis is equivalent to measuring the eigenstates of X_v, obtaining some state $|\omega_r\rangle$: the post-measurement state is then stabilized by $\omega^{-r}X_v$, as well as by operators (but only those operators) which commute with X_v and stabilized the pre-measurement state.
- Conjugating X_v by $\Lambda Z^\dagger_{vw}$ yields $X_v Z^\dagger_w$, and similarly conjugating X_w by ΛZ_{vw} yields $Z^\dagger_v X_w$. As they are diagonal, conjugating Z_v or Z_w by ΛZ_{vw} has no effect. Conjugating by X_w^{-r} transforms $Z^\dagger_w$ to $\omega^{-r}Z^\dagger$, and leaves X_w unchanged.

We may then describe the sequence of transformations on $|\psi\rangle_v$ as follows: for any scalar $\phi \in \mathbb{C}$, the operator ϕX_v transforms as follows:

$$\begin{aligned}\big\langle \phi X_v \big\rangle \xmapsto{\text{prep. } |+\rangle_w} \big\langle \phi X_v\,,\, X_w \big\rangle \xmapsto{\Lambda Z^\dagger_{vw}} & \big\langle \phi X_v Z^\dagger_w\,,\, Z^\dagger_v X_w \big\rangle \\ \xmapsto{X_v \text{ meas.}} & \big\langle \phi X_v Z^\dagger_w\,,\, \omega^{-r} X_v \big\rangle \;=\; \big\langle \omega^{-r} X_v \big\rangle \otimes \big\langle \phi \omega^r Z^\dagger_w \big\rangle \\ \xmapsto{X_w^{-r} \text{ corr.}} & \big\langle \omega^{-r} X_v \big\rangle \otimes \big\langle \phi Z^\dagger_w \big\rangle \;, \end{aligned} \tag{19a}$$

so that these operations transform $\phi X_v \mapsto \phi Z^\dagger_w$; and similarly,

$$\begin{aligned}\big\langle \phi Z_v \big\rangle \xmapsto{\text{prep. } |+\rangle_w} \big\langle \phi Z_v\,,\, X_w \big\rangle \xmapsto{\Lambda Z^\dagger_{vw}} & \big\langle \phi Z_v\,,\, Z^\dagger_v X_w \big\rangle \;=\; \big\langle \phi Z_v\,,\, \phi X_w \big\rangle \\ \xmapsto{X_v \text{ meas.}} & \big\langle \omega^{-r} X_v\,,\, \phi X_w \big\rangle \\ \xmapsto{X_w^{-r} \text{ corr.}} & \big\langle \omega^{-r} X_v \big\rangle \otimes \big\langle \phi X_w \big\rangle \;, \end{aligned} \tag{19b}$$

so that we obtain $\phi Z_v \mapsto \phi X_w$. Similarly, for any Weyl operator $W_{a,b}$ [10, Definition II], the operator $\phi W_{a,b}$ acting on v will be transformed to a Weyl operator $\phi W_{-a,b}$ on w; the calculation is straightforward. This implies (*c.f.* [10, Eqn. 17]) that aside from the teleportation from v to w, the effect is an inverse Fourier transform of the state.

Thus, we may simulate the coding procedure of a node V as described in Section 2.2 as follows. Provided a collection of incoming qudits $a_1, \ldots, a_\ell$, we may prepare output qudits $b_1, \ldots, b_m$ by:

1. preparing output message qudits $b_1, \ldots, b_m$ and auxiliary qudits $b'_1, \ldots, b'_m$ in the state $|+\rangle$;
2. entangling the qudits b_j and b'_j by a $\Lambda Z^\dagger$ operation, and performing $\Lambda Z^{V_{jk}}$ operations between each pair of qudits a_k and b'_j;
3. measuring each qudit b'_j in the Fourier basis, obtaining some outcome r_j, and performing an X^{-r_j} operation on the corresponding output qudit b_j.

This describes a MBQC procedure with inputs and outputs which we may illustrate by a *geometry* (in the terminology of Ref. [9, 7]) specifying the input and output qubits.

Figure 5 presents geometries for the partial coding operation performed by $\tilde{U}_V$ as in Figure 4, and for the entire operation of a single coding node (including the eventual measurement of the input qubits): input qudits have arrows pointing inwards, and output qudits have arrows pointing outwards.

3.2 MBQC geometries to simulate entire network coding protocols

In the diagrammatic convention of this article, composition of MBQC procedures may be represented by contracting the arrows between the outputs of earlier procedures and the inputs of later ones. For MBQC procedures to simulate the linear network codes, composing

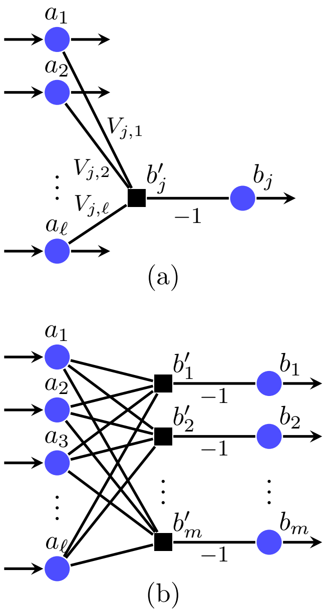


Figure 5 Geometries of MBQC procedures for a single node performing a transformation $V : \mathbb{Z}_d^\ell \to \mathbb{Z}_d^m$ of the standard basis. Incoming/outgoing message qudits are represented by blue circles; auxiliary qudits by black squares. **(a)** The geometry associated to coding a single message qudit, simulating the right-hand circuit of Figure 4. Edges are labeled by their "weights", *i.e.* the necessary power of ΛZ in the procedure. As the qudits a_k remain unmeasured, these are depicted as being outputs as well as inputs of this procedure. **(b)** The geometry associated to the entire operation of a coding node, including measurement of the incoming message qudits. Edge weights between the qudits a_k and α_j depend on the coding operation being simulated: if the coding operation being performed is sparse, many of these edge weights will be zero (corresponding to edges which should be omitted entirely). Only the qudits b_j form the output of this procedure.

the geometries associated to each node yields a bipartite graph with a structure closely related to that of the coding network itself. Specifically, one associates a qudit for the output qudits of the coding network, as well as for each incoming and outgoing message qudit at each node (with qudits at the outgoing links being the "auxiliary" qudits described above), and connecting them by a bipartite graph corresponding to the non-zero coefficients V_{jk} of the coding node. The edges of the coding network are replaced by *undirected* edges with weights -1, corresponding to the entangling operations between the outgoing message qudits (which are either the inputs for some other node, or the outputs of the entire network). The directionality of the communication links are represented by the order of the measurement and correction operations, as well as the classical communication involved in the correction subroutine.

As an example, we illustrate this construction in Figure 6 for procedure for the two-pair problem performing a SWAP operation on two qudits (*e.g.* in which we use the coding operations $S_1 = S_2 = V_2 = [\,1\;\;1\,]^\top$ and $V_1 = T_1 = T_2 = [\,-1\;-1\,]$).

As every measurement involved is performed in the Fourier basis (equivalently: the eigenbasis of the X operator), the only information which this graphical representation omits are the order in which the measurements occur, and the correction procedures, which we consider next.

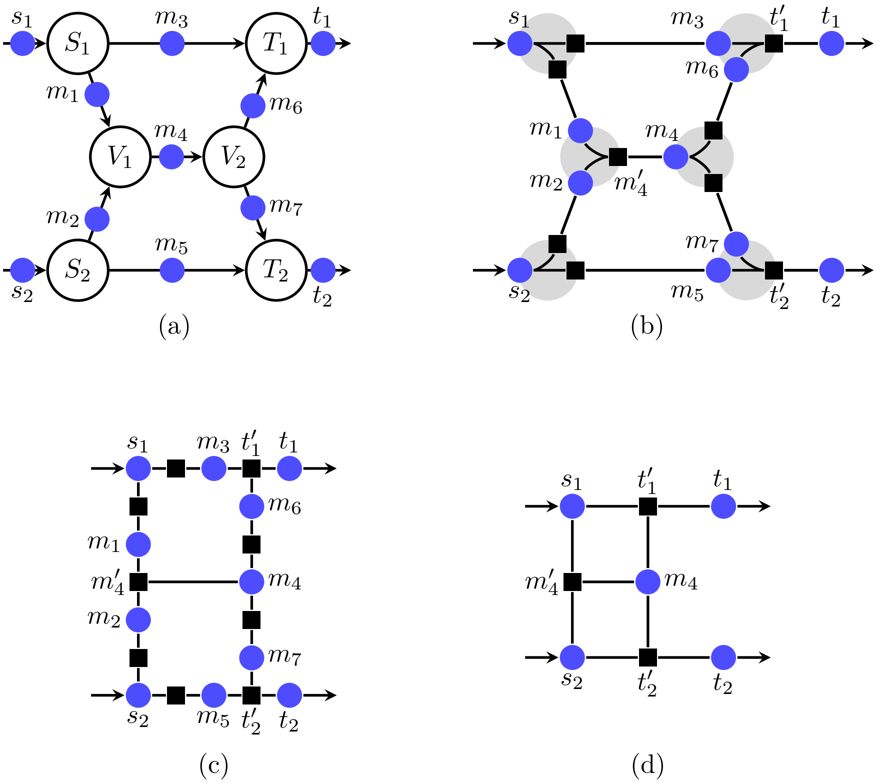


Figure 6 Construction of a MBQC geometry for a procedure simulating a coding protocol for the 2-pair problem on **(a)** the butterfly network, shown with message qudits for each communication link. **(b)** The graph obtained by substituting each coding node, with the geometry for the corresponding MBQC procedure. This is derived by adding vertices for "auxiliary" qudits (black squares) for each output message qudit, and associating each "auxiliary–output" pair to an outbound network link. Edges represent powers of ΛZ operations, which are used for single-qudit teleportation along the network links. The input and output message qudits of the linear code become the source and target subsystems of the MBQC procedure. **(c)** The same geometry, presented in grid formation. **(d)** The geometry of a MBQC procedure (*c.f.* Ref. [5, Figure 7]) for the SWAP operation.

3.3 Measurement and communication of outcomes

The corrections required to use X measurements to simulate projection onto $|+\rangle$ may be performed in two natural ways, corresponding to the protocols of Refs. [17] and [18] respectively.

3.3.1 Free classical communication

In a setting as in Ref. [17] where classical communication is free, all corrections may be deferred to the target nodes of the coding network, which prepare the output qudits. This is a natural approach for simulating the network code as a MBQC procedure: in measurement-based computation, it is conventional to simulate CP maps in such a way that the output qudits are the only qudits on which unitary correction operations are performed. As in Ref. [17], successful projection onto the $|+\rangle$ state (or a "0" outcome of a X measurement) is the ideal case; it then suffices to determine how the errors (or *byproduct operations* in the terminology of Ref. [21]) propagate to the output qudits, in order to correct them. We describe this in terms of communication directly to the targets, as well as some amount of communication within the coding network.

When simulating the coding procedure at each node using auxiliary qudits, measuring those auxiliary qudits introduces an additional source of error: if the correction is not immediately performed on the outgoing message qudits, this induces additional phase errors. Commuting an $X^{-r}_{b_j}$ operation past an entangling operation $\Lambda Z^{U_{ij}}_{b_j c'_i}$, where c'_i is an auxiliary qudit for a subsequent node performing a coding operation U, yields an error operation $X^{-r}_{b_j} Z^{-rU_{ij}}_{c'_i}$. The operation $X^{-r}_{b_j}$ does not affect the outcome of the measurement on b_j, as the states $|\omega_r\rangle$ are eigenvectors of X. The Z error on c'_i induced by postponing the correction on b_j is significant, but we may account for this error by classical post-processing of the measurement result r' on c'_i itself. Let $\tilde{r} = rU_{ij}$ for the sake of brevity: because $XZ^{-\tilde{r}} \propto= \omega^r Z^{-\tilde{r}} X$, we may account for an uncorrected $Z^{-\tilde{r}}$ operation on c'_i by performing an X measurement, obtaining some outcome r'_0, and then subtracting $\tilde{r}$ from that outcome to obtain an adjusted outcome $r' = r'_0 - \tilde{r}$ for future corrections.

More generally, c'_i will accumulate uncorrected Z errors arising from the uncorrected X errors on each of the input messages on which it depends. If those input qubits b_j have errors X^{-r_j} associated with them, these collectively induce an error

$$Z^{-(r_1 U_{i1} + r_2 U_{i2} + \cdots)} = Z^{-\hat{\mathbf{e}}_i \cdot U\mathbf{r}} \tag{20}$$

on c'_i. We may simulate this correction after the Z measurement by subtracting $\tilde{r} = \hat{\mathbf{e}}_i \cdot U\mathbf{r}$ from the measurement outcome r'_0, yielding $r' = r'_0 - \hat{\mathbf{e}}_i \cdot U\mathbf{r}$. By propagating the results of the auxiliary qudit measurements forward through the coding network, subsequent coding nodes may locally adapt the measurement outcomes in order to simulate the correction of errors on their own auxiliary qudits, allowing the target nodes to perform the necessary X corrections on the output qudits of the network. Alternatively, all of the results may be transmitted directly to the target nodes, which can simulate this sequential adaptation of measurement outcomes themselves.

For a coding network performing an injective transformation $M : \mathbb{Z}_d^{\mathcal{S}} \to \mathbb{Z}_d^{\mathcal{T}}$, the phase errors induced by measurement of the message qudits may be corrected in the manner described in Ref. [17]. Without loss of generality, we may suppose that the agents at each network coding node prepare their auxiliary and message qudits, and all nodes except the target nodes communicate their outgoing messages to their recipients. Afterwards, they measure their auxiliary nodes in some order consistent with the topological ordering of the network, and similarly communicate the outcomes forward, allowing subsequent nodes to adjust their auxiliary measurement outcomes, and allowing target nodes to perform what X corrections are necessary on the output qudits. The remaining measurement operations and classical messages are identical to those of Ref. [17], in which it does not matter if nodes transmit outgoing message qudits before they measure incoming message qudits.

For the sake of completeness, we sketch an inductive approach to the Z correction protocol of the target nodes in this setting. Let A be a left-inverse of M, and consider an input state $|\psi\rangle$ to the coding network, expressed as

$$|\psi\rangle = \sum_{\mathbf{x} \in \mathbb{Z}_d^{\mathcal{S}}} u_{\mathbf{x}} \, |\mathbf{x}\rangle = \sum_{\mathbf{y} \in \mathrm{img}(M)} u_{A\mathbf{y}} \, |A\mathbf{y}\rangle . \tag{21}$$

The state obtained after performing the preparation and entanglement phases of the MBQC procedure, and after performing the auxiliary qudit measurements and X corrections on the output qudits, is exactly a state of the form in Eqn. (10), of the form

$$|\Psi\rangle \;=\; \sum_{\mathbf{y}\in\mathrm{img}(M)} u_{A\mathbf{y}}\,|A\mathbf{y}\rangle_{\mathsf{S}} \otimes |MA\mathbf{y}\rangle_{\mathsf{T}} \otimes |HA\mathbf{y}\rangle_{\mathrm{rest}} \;=\; \sum_{\mathbf{y}\in\mathrm{img}(M)} u_{A\mathbf{y}}\,|A\mathbf{y}\rangle_{\mathsf{S}} \otimes |\mathbf{y}\rangle_{\mathsf{T}} \otimes |HA\mathbf{y}\rangle_{\mathrm{rest}} \tag{22}$$

for some linear map H. (The latter equality holds because for any $\mathbf{y} = M\mathbf{x}$, we have $MA\mathbf{y} = MAM\mathbf{x} = \mathbf{y}$.) Indeed, the distinction between the input qudits S and the other non-target qudits is unimportant: we may subsume the linear map A on the standard basis of $\mathcal{S}$ and the map HA on the standard basis of the other qudits into a map

$$K \;=\; \left[\begin{array}{c} A \\ \hline HA \end{array}\right] \tag{23}$$

where the upper rows correspond to indices in S, and the lower rows to the other non-output qudits. We may then write

$$|\Psi\rangle \;=\; \sum_{\mathbf{y}\in\mathrm{img}(M)} u_{A\mathbf{y}}\,|\mathbf{y}\rangle_{\mathsf{T}} \otimes |K\mathbf{y}\rangle_{\Omega\smallsetminus\mathsf{T}}\;. \tag{24}$$

We may isolate any non-output qudit $u \in \Omega \smallsetminus T$. Let $\Omega' = \Omega \smallsetminus \{u\}$, and consider another decomposition

$$K \;=\; \left[\begin{array}{c} \boldsymbol{\kappa}_u^{\top} \\ \hline K' \end{array}\right] \tag{25}$$

where the upper row corresponds to the index for the qudit u and contains a row-vector $\boldsymbol{\kappa}_u^{\top}$, and K' corresponds to all of the other non-output qudits; we may then once more re-write

$$|\Psi\rangle \;=\; \sum_{\mathbf{y}\in\mathrm{img}(M)} u_{A\mathbf{y}}\,|\mathbf{y}\rangle_{\mathsf{T}}\;\;|\boldsymbol{\kappa}_u\cdot\mathbf{y}\rangle_u \otimes |K'\mathbf{y}\rangle_{\Omega'\smallsetminus\mathsf{T}}. \tag{26}$$

Measuring u in the Fourier basis and obtaining the outcome r, the resulting state on the remaining qudits is

$$|\Psi'\rangle \;=\; \sum_{\mathbf{y}\in\mathrm{img}(M)} u_{A\mathbf{y}}\,\omega^{-r(\boldsymbol{\kappa}_u\cdot\mathbf{y})}\,|\mathbf{y}\rangle_{\mathsf{T}}\;\;|K'\mathbf{y}\rangle_{\Omega'\smallsetminus\mathsf{T}}, \tag{27}$$

following Eqn. (11). If the outcome r is transmitted to the target nodes, and who know the value of $\boldsymbol{\kappa}_u$, they may simply compute $\boldsymbol{\sigma} := r\boldsymbol{\kappa}_u$ and collectively perform $Z^{\boldsymbol{\sigma}} = Z_{t_1}^{\sigma_1} Z_{t_2}^{\sigma_2}\cdots$ on the qudits of T, thereby obtaining

$$|\Psi''\rangle \;=\; \sum_{\mathbf{y}\in\mathrm{img}(M)} u_{A\mathbf{y}}\,|\mathbf{y}\rangle_{\mathsf{T}}\;\;|K'\mathbf{y}\rangle_{\Omega'\smallsetminus\mathsf{T}}, \tag{28}$$

which is again a state of the same form as in Eqn. (10), on one fewer qudits. By induction, we may measure each of the qudits of $\Omega \smallsetminus \mathsf{T}$ in any order (or simultaneously), and transmit them to the target nodes, which then make the appropriate Z corrections to obtain the state

$$\left|\Psi^{(n)}\right\rangle \;=\; \sum_{\mathbf{y}\in\mathrm{img}(M)} u_{A\mathbf{y}}\,|\mathbf{y}\rangle_{\mathsf{T}} \;=\; \sum_{\mathbf{x}\in\mathbb{Z}_d^{\mathcal{S}}} u_{\mathbf{x}}\,|M\mathbf{x}\rangle_{\mathsf{T}}\;. \tag{29}$$

In summary, provided free classical communication to the targets and within the coding network, all measurements may be performed simultaneously, with the results of the measurement of incoming messages being transmitted directly to the targets to perform Z corrections on the output qudits. Measurement results of the auxiliary qudits may be communicated along the coding network, and used to adapt the outcomes of subsequent measurements, culminating in measurement information useful to the target nodes to perform X corrections on the output qudits.

3.3.2 Constrained classical communication

In the setting of Ref. [18], we attempt to reduce the amount of classical communication which takes place outside of the network (but allowing messages to pass in either direction). To this end, we allow the source nodes and the intermediate nodes of the network to perform Z corrections. The way in which these corrections are performed follows from **(a)** the description of how X corrections may be simulated in the setting of "free" classical communication, as this already can be performed only with communication within the coding network; and **(b)** the phase correction procedure of Ref. [18] which was outlined in Section 2.2. These corrections may be performed as follows:

- All auxiliary qudits may be measured simultaneously, and their outcomes propagated forward through the network, as in the previous section. Alternatively, one may instead perform X correction operations for the auxiliary qudits at each node: this imposes an order on the measurement of the auxiliary qudits which is consistent with the topological order of the network, so that each node may use the measurement outcomes for preceding auxiliary qudits when correcting its own auxiliary qudits.
- The measurement of each node's incoming message qudits must be performed in an order opposite to the topological order of the coding network, in order to allow the node which sent each message qudit to perform the necessary corrections involving its own incoming message qudits.

From this, one may derive schedules for measuring each qudit in the network, and for communicating classical messages forward or backward through the network to allow the necessary X or Z corrections.

For the correction of phases induced by measurement of the input qubits of the source, following As in Section 2.2, whether the corrections arising from the measurement of the input qudits managed by the source nodes can be corrected without communicating outside of the network, may depend on the transformation which the network performs. For any linear transformation M for which $M^{\top}BM = \mathbb{1}$ for some block-diagonal B acting on blocks of qudits held by target nodes – *e.g* for permutation matrices M – classical network coding of of the outcomes of measuring the inputs of the source nodes will suffice.

3.4 Overview of the MBQC construction

The above construction rests on the fact that the protocol of Ref. [17] is unaffected if the measurements are deferred until each node sends its messages. (The protocol of Ref. [18] in fact requires this modification.) The result of doing so causes these protocols to give rise to large distributed entangled states, on which local measurements are performed to simulate projection onto the $|+\rangle$ state. In this sense, these protocols are literally quantum computation by measurements; the modifications described in this Section – namely, replacement of ΛX operations by ΛZ operations, introduction and measurement of auxiliary qudits in order to make the previous modification possible, and communication of the results of measuring auxiliary qudits – are straightforward modifications which demonstrate that they are effectively computations in the one-way MBQC model of Refs. [21, 7].

The MBQC procedures which result from these transformations have comparable complexity to the original protocols of Refs. [17, 18], differing essentially only in the various operations performed on the auxiliary qudits, as well as the communication and transformation of their measurement outcomes. For a coding network with k input messages, ℓ output messages, and m internal communication links, the total number of qudits involved in the MBQC procedure is easily verified to be $k+2\ell+2m$, following Section 3.2. The number of entangling operations involved for each node (disregarding exponents) is simply the same as the number of ΛX operations involved in simulating $\tilde{U}_V$, plus twice the out-degree (involved in entangling the auxiliary and outgoing message qudits for the node). Thus there are exactly $2(m+\ell)$ more entangling operations, in the form of ΛZ operations, in the MBQC protocol than there are ΛX operations in the original presentation of the protocols in Refs. [17, 18]. There are also exactly $2(m+\ell)$ additional classical messages sent in the MBQC protocol, either directly to the targets or entirely within the network, again as a result of measuring the auxiliary qudits.

4 Open questions

In this article, we have illustrated the way in which classically-assisted quantum linear network coding over $\mathbb{Z}_d$ as described by Kobayashi *et al.* [17, 18] is in effect an instance of measurement-based computation in the one-way model [21, 7], in particular using measurements only in the Fourier basis (the eigenbasis of the X cyclic shift operator on d-dimensional qudits). While not explicitly presented as an instance of MBQC, the differences between the protocols of Refs. [17, 18] and one-way measurement-based procedures are straightforward, and involve no substantial differences in *e.g.* the amount of classical communication required. We may ask to what extent these results (particularly the bounds on classical communication outside of the network) hold for classically assisted *non-linear* quantum codes as well.

While the MBQC model is sometimes described as a distributed model of computation, little emphasis has been placed on the communication cost of MBQC computation. A common presentation (*e.g.* as in Refs. [3, 2]) is that measurement results are recorded by an effectively delocalized classical control, which receives messages containing measurement outcomes from one or more agents which manage individual qudits, and which responds with instructions of how to perform subsequent measurements. Bounding the communication requirements of a MBQC procedure, to eliminate the need of a delocalised control center, may be necessary to realize the reduction in the computational depth of a MBQC procedure (one of the theoretical selling points of the MBQC model [21]).

As network coding subsumes constant-depth distributed computation, we may interpret these results as recommending measurement-based computation as a framework for analyzing multiparty communication protocols, as we have suggested in the introduction. We may also consider this as an alternative means of approaching the problem of assigning semantics to measurement-based computations, a problem of some interest in models of quantum computation [7, 9, 12, 6]. Specifically: rather than interpreting a measurement-based procedure as a quantum circuit with some potentially exotic features (such as closed time-like curves [6]), we may interpret pieces of measurement-based computations as coherently simulating transformations of the standard basis on several qudits at once. Such simple semantics is likely to prove useful to any programme to find novel ways of using measurement-based computation as a medium in which to develop algorithms (see Ref. [11]).

As a final open question, we ask whether a converse to our results hold, the form of a classical simulation algorithm for certain measurement-based computations by linear network codes. This article shows that (a coherent quantum simulation of) a classical linear network code is in effect a measurement-based procedure which performs only X-eigenbasis measurements, on a graph state with similar structure to the coding network. This is a special case of an efficiently simulatable class of computations: the unitary transformations realized by MBQC procedures performing only Pauli-eigenbasis measurements are *Clifford group* *operations*,[4] which can be simulated *e.g.* on standard basis states by linear transformations on a cyclic ring [10]. This raises the question: is there a sense in which a MBQC procedure on a graph G, which implements unitary a transformation using only measurements in a Pauli eigenbasis (or only the X-eigenbasis) and Pauli corrections, can be "locally" simulated by a classical linear code – in such a way that the expectation value of any observable on a single given qudit can be evaluated from information available at a corresponding target node – on a network similar to G?

Acknowledgements. This work was done in part while MR was with NEC Laboratories America, and NdB was at the University of Cambridge with support from the EC project QCS. NdB would like to thank Peter Høyer for helpful comments at the beginning of this research.

References

1 R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung. Network information flow. *IEEE Transactions on Information Theory*, 46:1204–1216, 2000.

2 J. Anders and D. E. Browne. Computational power of correlations. *Phys. Rev. Lett.*, 2009. arXiv:0805.1002.

3 A. Broadbent, J. Fitzsimons, and E. Kashefi. Universal blind quantum computation. In *Proc. 50th IEEE FOCS*, pages 517–526, 2009. arXiv:0807.4154.

4 D. E. Browne and H. J. Briegel. One-way quantum computation – a tutorial introduction. arXiv:quant-ph/0603226, 2006.

5 D. E. Browne, E. Kashefi, M. Mhalla, and S. Perdrix. Generalized flow and determinism in measurement-based quantum computation. *New J. Phys.*, 9, 2007. arXiv:quant-ph/0702212.

6 R. D. da Silva, E. F. Galvão, and E. Kashefi. Closed timelike curves in measurement-based quantum computation. *Phys. Rev A*, 83, 2011. arXiv:1003.4971.

7 V. Danos, E. Kashefi, and P. Panangaden. Robust and parsimonious realisations of unitaries in the one-way model. *Phys. Rev. A.*, 72, 2006. arXiv:quant-ph/0411071.

8 V. Danos, E. Kashefi, and P. Panangaden. The measurement calculus. *J. ACM*, 54, 2007. arXiv:0704.1263.

9 N. de Beaudrap. Unitary-circuit semantics for measurement-based computations. *Int'l J. Quant. Info.*, 8:1–91, 2010. arXiv:0906.4261.

10 N. de Beaudrap. A linearized stabilizer formalism for systems of finite dimension. *Quant. Info. & Comp.*, pages 73–115, 2013. arXiv:1102.3354.

11 N. de Beaudrap, V. Danos, E. Kashefi, and M. Roetteler. Quadratic form expansions for unitaries. In *Proc. TQC 2008*, pages 29–46, 2008. arXiv:0801.2461.

12 R. Duncan. A graphical approach to measurement-based quantum computing. arXiv:1203.6242, 2012.

13 M. Grassl, M. Roetteler, and T. Beth. Efficient quantum circuits for non-qubit quantum error-correcting codes. *Int'l J. Found. Comp. Sci.*, 14:757–775, 2003. arXiv:quant-ph/0211014.

14 M. Hayashi. Prior entanglement between senders enables perfect quantum network coding with modification. *Phys. Rev. A*, 76, 2007. arXiv:0706.0197.

15 M. Hayashi, K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita. Quantum network coding. In *Proc. 24th Annual STACS*, pages 610–621, 2007. arXiv:quant-ph/0601088.

[4] This is well-known for qubits [4]; on qudits it follows from how stabilizer states are transformed under measurements, see Ref. [10].

16 E. Kashefi, D. Markham, M. Mhalla, and S. Perdrix. Information flow in secret sharing protocols. *EPTCS*, 9:87–97, 2009. arXiv:0909.4479.

17 H. Kobayashi, F. Le Gall, H. Nishimura, and M. Roetteler. General scheme for perfect quantum network coding with free classical communication. In *Proc. 36th ICALP*, pages 622–633, 2009. arXiv:0908.1457.

18 H. Kobayashi, F. Le Gall, H. Nishimura, and M. Roetteler. Constructing quantum network coding schemes from classical nonlinear protocols. In *Proc. 2011 IEEE Int'l Symp. Info. Theory*, pages 109–113, 2011. arXiv:1012.4583.

19 D. Leung, J. Oppenheim, and A. Winter. Quantum network communication – the butterfly and beyond. *IEEE Trans. Inf. Theory*, 56:3478–3490, 2010. arXiv:quant-ph/0608223.

20 M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

21 R. Raussendorf, D. E. Browne, and H. J. Briegel. Measurement-based quantum computation on cluster states. *Phys. Rev. A*, 68, 2003. arXiv:quant-ph/0301052.

22 Y. Shi and E. Soljanin. On multicast in quantum network. In *Proc. 40th Annual Conf. Info. Sci. and Systems*, pages 871–876, 2006.

23 P. W. Shor and J. Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85:441–444, 2000. arXiv:quant-ph/0003004.

24 W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.